TWO PRELIMINARY MODELS FOR THE ANALYSIS OF CHANGES IN A SOCIAL INDICATOR OF JOB SATISFACTION

Kenneth C. Land, University of Illinois at Urbana-Champaign

1. Introduction

Research in social indicators has been public now for nearly a decade, since the release of the Bauer (1966) volume which was the first to use the label. Roughly, as Land (1974b) has noted, it is now generally agreed that social indicators are statistics which measure social conditions or activities and changes therein over time for various segments of a population. By social conditions, it is meant both the external (social and physical) and the internal (subjective and perceptional) contexts of human existence in a given society. This definition is a highly general characterization of social indicators, for it allows almost any index of social activity to be classified as a social indicator provided that the index can be construed as reflecting a social condition of some population. For example, if an index such as the "sex ratio" of a population is construed as reflecting some condition of life in that population, then it could be regarded as a social indicator by this definition. Since social indicators have been considered to be generalizations of economic indicators and since economic indicators are just indexes of economic conditions or activities, this characterization is appropriate. It gives a general and somewhat permeable cognitive orientation to the notion of a social indicator.

Nevertheless, as Land (1974b) observes, it is useful to have a more stringent external validity criterion for classifying a social statistic as a social indicator in order that the definition may be made more precise. Two such criteria have been proposed in the social indicators literature.

First, social planners and policymakers (see, e.g., Executive Office of the President, Office of Management and Budget, 1974; Organization for Economic Cooperation and Development, 1973; Terleckyj, 1972; U.S. Department of Health, Education and Welfare, 1969) generally require that social indicators be arguments in a social welfare function, where this function is taken to be some combination of the utility (satisfaction) functions of the members of the society which, in turn, depend upon the conditions of life indexed by the social indicators. This criterion reduces social indicators to goal output indicators, as they can be taken to measure the degree of achievement of a goal objective in a social welfare function.

A second criterion, usually associated with social scientists rather than social planners (see, e.g., Campbell and Converse, 1972; Land and Spilerman, 1974; Sheldon and Moore, 1968), requires that social indicators demonstrate a consistent historical pattern of timing and covariation with social change. This criterion leads to the class of indicators of social change. Certainly, it is not necessary for the list of goal output indicators to be identical to the list of indicators of social change. In fact, it is very difficult for a meaningful social welfare function to take into account more than a few aspects of social life. But the contents and relative weights of social welfare functions will undoubtedly change from time to time as the interests of members of a society and of policymakers change. As this occurs, social scientists will be called upon to provide different indicators as components of the welfare function. Ideally, these indicators will be derivable from a general list of "indicators of social change," a list which has been verified to satisfy the external validity criterion of social indicators which is distinct from the "design" or "planning" function associated with social policy.

It is readily seen that these two criteria and the corresponding types of indicators are generalizations of the two main criteria of external validity which have been employed in the standard economic indicators. First, with respect to economic "analysis," economic indicators are just defined as time series that have shown a historical, consistent pattern of timing and of conformity to business cycles (G. Moore, 1961; 1967). Second, from the perspective of economic "design" or "policy," economic indicators are components of economic welfare (see, e.g., Fox, Sengupta, and Throbecke, 1973). Thus, an economic indicator such as the unemployment rate possesses external validity not only as the objective of economic policy but also through its past history of covariation with business cycles. Indeed, it would be rather irresponsible to make economic policy with respect to some indicator if nothing more were known about it than that it is the goal of a policy. Fortunately, we do not live in such a state of abysmal ignorance with respect to the standard economic indicators.

In the case of most of the generalized social indicators which have been proposed thus far, our knowledge is less well endowed, and Land (1974c) has observed that this is true for "quality of employment indicators." The latter indicators are usually conceived of as "outcomes" of employment which contribute to job satisfaction or utility (see Seashore, 1973 and Taylor, 1973 for lists of possible outcomes of employment).

In order to interpret the variations of a social indicator, Land (1971; 1974a; 1974b; 1974c) has emphasized the indispensability of a social indicator model, that is, a model which determines the value and variations of the social indicator in question as a function of policy instrument indicators and non-manipulable indicators (data). Moreover, this assistance in interpretation occurs at the conceptual level as well as in the quantitative sense. For example, some index of "schooling" is usually included among lists of social indicators. In order to interpret this index, one must specify fairly precisely the nature of the system and process under consideration, because, while schooling is an "output" of a school system, it is an "input" in terms of the characteristics of an individual with respect to the job market. As a second example, we note that a health index such as "life expectancy" is only partially determined as an "output" of a narrowly defined health care system. In many cases, in fact, this index may be more strongly determined by such factors as genetics, nutrition, life style, etc. than by "delivery of health care." These examples illustrate how the careful definition of the system and process determining an indicator is essential to its interpretation.

The purpose of this paper is to describe two preliminary sociological models for an aggregate social indicator of job satisfaction. These models are preliminary in the sense that they are only first-order approximations to the limits of existing data and estimation methods. Both models link the job satisfaction indicator in a consistent historical pattern to referent indicators of social change. Thus, they lend external validity directly to this index as an indicator of social change and, hence, indirectly to any social indicator which covaries with job satisfaction. Moreover, this is of particular importance in the case of job satisfaction because of the role which satisfaction plays in the determination of social welfare and, hence, in planning, as noted above.

Land (1974b) has remarked that social indicator models typically take one of two forms. The first is a macro-sociological time-series model in which the purpose is to study the covariations of an "aggregate level of well-being" indicator with other aggregate indexes. The second type of model usually addresses equity values in terms of measures of dispersion of skewness and takes the form of a life-cycle and cohort analysis. In terms of these types, the first model to be described below is a macro model, in which aggregate job satisfaction is written as a function of aggregate unemployment, where the latter can be taken as a referent index of cycle social change. The second model complements the first by describing the effects of population structure and cohort replacement on job satisfaction.

2. A Macro Time-Series Model of Job Satisfaction

As a first approach to a macro-sociological time-series model for job satisfaction, consider an analysis of the data on trends in job satisfaction summarized in Table 1. These data from fifteen national sample surveys were reported in a recent Manpower Administration monograph by Quinn, Staines, and McCullough (1974) on the trend in job satisfaction, in which the authors (p. 1) conclude:

Table 1. Percentage of "Satisfied" Workers (Men Only, Ages 21 through 65)*, 1958-1973**

Date	Percentage Estimate (Sample Size)	Source
1958	81% (852)	Survey Research Center, U. of Michigan
1962	84% (1,028)	National Opinion Research Center
July 16, 1963	89% (1,469)	Gallup Poll
1964	90% (1,025)	Survey Research Center, U. of California
1964	92% (3,086)	National Opinion Research Center
August 6, 1965	87% (1,338)	Gallup Poll
September 6, 1966	92% (1,361)	Gallup Poll
September 29, 1966	89% (1,340)	Gallup Poll
March 25, 1969	92% (585)	Gallup Poll
November-December, 1969	88% (1,528)	Survey Research Center, U. of Michigan
1971	91% (1,025)	Survey Research Center, U. of Michigan
August 17, 1971	88% (516)	Gallup Poll
December 7, 1971	86% (558)	Gallup Poll
January 23, 1973	88% (526)	Gallup Poll
February-March, 1973	91% (1,493)	Survey Research Center, U. of Michigan

* Except for the 1958 survey which was based on men 21 or older

** "Don't know" answers were excluded from the percentage bases

Source: Quinn, Staines, and McCullough (1974)

"In spite of public speculation to the contrary, there is no conclusive evidence of a widespread, dramatic decline in job satisfaction. Reanalysis of 15 national surveys conducted since 1958 indicates that there has not been any significant decrease in overall levels of job satisfaction over the last decade."

Let us analyze the fluctuations in these data in more detail from a social indicator model perspective.

Before taking the analysis task seriously, however, we should determine to what extent the data are to be taken as an accurate representation of national trends in job satisfaction. Certainly, the series is affected by both methods variance and sampling variance. For example, over the years, the Gallup Poll Organization has asked the following question of a national sample of adults: "On the whole, would you say you are satisfied or dissatisfied with the work you do?" On the other hand, the other surveys tend to ask us a question of the form: "Are you very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied with your job?" But, unlike the Gallup question, the wording of the latter has varied somewhat from survey to survey. Moreover, the Gallup polls do not restrict respondents to employed persons, whereas the other surveys are "surveys of working conditions" and hence are so restricted. But, for the sub-samples of men, ages 21 through 65, reported in Table 1, the civilian labor force participation rate is usually on the order of 90 percent or more (Quinn, Staines, and McCullough, 1974: p. 3). Also, each of the estimates in Table 1 is subject to sampling error. At the 95-percent confidence interval, the sampling error for percentages near 90 percent is about 2 percent for random sample sizes near 1500 and about 3 percent for random sample sizes near 500. Of course, since the surveys of Table 1 are not based on simple random samples, these sampling errors are best used only as approximate guidelines.

These error sources are substantial and cannot be ignored. Taken together, they make the data series in Table 1 less meaningful than we would like as a basis for building a social indicator model. Nevertheless, it is the only available national time-series on job satisfaction and, hence, the best. Therefore, proceeding on the assumption that some data is better than no data, we shall attempt an interpretation relative to other trends during the last decade and a half.

In our discussion of indicators of social change, we noted that one procedure for giving external social change validity to a social indicator is to determine its degree of covariation with referent time-series on social change. Moreover, the economy is a major source of cyclical social change. Since job satisfaction is a function of one's experience with a job, and since jobs are created in the economy, it follows that these strategies imply that we should explore the degree of covariation of job satisfaction with an indicator of economic activity as a first approximation to a social indicator model of the macro variety. In particular, we note that, among the various stochastic components of the series in Table 1, there seems to be a significant covariation with the trend in unemployment rate for the 1958-1973 period. Table 2 shows that, from a peak of nearly 7 percent in 1958, the unemployment rate shows a general decline during the next several years, with significant peaks in 1961, 1963, and 1971.

Table 2. Total Unemployment Rates, 1958-1972

Year	Rate	Year	Rate
1958	6.8	1966	3.8
1959	5.5	1967	3.8
1960	5.5	1968	3.6
1961	6.7	1969	3.5
1962	5.5	1970	4.9
1963	5.7	1971	5.9
1964	5.2	1972	5.6
1965	4.5	1973	4.9

Source: Social Indicators: 1973, p. 136, and Manpower Report of the President, 1974, p. 23.

Regressing the data in Table 1 on those in Table 2 yields the following estimated equation:

$$(1) \qquad S_t = 99.3 - \underset{(0.89)}{2.2}\, U_t + \varepsilon_t; \quad (R = -0.68)$$

where S_t = job satisfaction at time t, U_t = unemployment rate at time t, ε_t = a stochastic disturbance term at time t, the number in parentheses under the regression coefficient is its standard error, and the correlation coefficient is given at the right in parentheses. This equation is estimated for the nine time points for which the data can be matched in Tables 1 and 2. Therefore, it is only a first-order estimate and the correlation could probably be improved by dating the job satisfaction surveys by the month they were in the field and using monthly unemployment data. Nevertheless, the results in equation (1) are statistically significant and meaningful. Briefly, equation (1) implies that a 1-percent increase in the unemployment rate produces more than a 2-percent decrease in job satisfaction. It also implies that, if this linear relationship of unemployment and job satisfaction holds throughout the range of the job satisfaction variable, then job satisfaction could reach approximately 99 percent if the unemployment rate approximates zero. But the assumption on which this inference is based is probably not tenable; that is, equation (1) would probably exhibit some non-linearities at the upper end of the job satisfaction scale. Moreover, the regression coefficient of U in (1) may be biased, for we have not adjusted for the possibility of autocorrelation of the disturbances in our estimation procedure. However, for this analysis, we shall be content with the least-squares estimates of equation (1).

For our present purposes, it is more interesting to tease out another implication of equation (1) for the job satisfaction data. Because the Gallup polls do not exclude unemployed men from the job satisfaction question, one implication of this model is that the Gallup series should show more sensitivity to the unemployment variable than the other series. That is, if unemployed men are more likely to be "dissatisfied" than are employed men, and if the unemployment rate increases by, say 1 or 2 percent, one anticipates at least 5 to 10 more "dissatisfied" responses in a sample of about 500 men. This is sufficient to register a 2- to 3-percent decrease in aggregate job satisfaction rates. Making the appropriate comparisons of the data in Tables 1 and 2, we do indeed find that the correlation of the Gallup series with the unemployment series is somewhat higher (-0.74, $p < 0.05$) than that for the total series. On the other hand, equation (1) shows that even the inclusion of the job satisfaction rates based on the "working conditions surveys," which restrict respondents to employed persons, does not make the correlation vanish. It is tempting to conjecture why there should be any relationship between job satisfaction and unemployment rates for employed persons. One fairly direct hypothesis is that a rise in unemployment signals a decrease in opportunities for job promotion or for changing jobs and thus decreases job satisfaction. A more complicated theory would posit a relationship between "general satisfaction" and job satisfaction and argue that the former is decreased when the economy turns down and hence depresses the latter. Such conjectures cannot be subjected to evaluation by the present analysis.

3. A Life-Cycle-Cohort Model of Job Satisfaction

As a second example of a model constructed to make sense of the data in Table 1, consider a life-cycle model constructed on the basis of two consistently reported relationships in the job satisfaction literature. First, as Quinn, Staines, and McCullough (1974: p. 9) note, professional-technical workers and managers, officials, and proprietors register the highest levels of job satisfaction. In fact, Kahn (1972: pp. 181-184) observes that a strong positive correlation of occupational status and job satisfaction has been consistently reported in the literature. Explicitly, for occupational status scored in Duncan units (Blau and Duncan, 1967), the occupational status-job satisfaction correlation for the grouped data in Quinn, Staines, and McCullough (1974: p. 12) report a strong positive correlation between the age of a worker and his job satisfaction (approximately 0.95 for their grouped data) which has remained fairly consistent for the timespan for which data are available. We quote their interpretation of this relationship:

> "The tenuous nature of generalizations about 'generation gaps' or related longitudinal trends seems all the more apparent when a far simpler explanation is considered--that older workers, especially in the case of men, are more satisfied with their jobs than younger workers simply because they have better jobs. In an achievement-oriented society, the 'best' jobs are reserved for those who can perform them best. Generally such performance depends on a worker's job experience, accrued skills, and demonstrated competence in related jobs. While this may not be true in all cases, certainly a job candidate's previous background and experience weigh heavily in the deliberations of those who will promote him or her to a 'better' job. Younger workers lack sufficient background to qualify them for the best jobs around. In addition, the fact that our society, like most others, places a high value on seniority increases the probability that better jobs will go to workers over 30. 'Beginners' in every sense of the word, younger workers are confined in consequence to positions that are often less than wholly satisfying."

In brief, the much talked-about decline in the job satisfaction of younger workers over the last decade has not been substantiated in the survey data. Rather, the positive age-job satisfaction relationship has been quite consistent over time. We shall use it and the occupational status-job satisfaction relationship to construct some life-cycle models to interpret trends in the data of Table 1.

Using first the positive age-job satisfaction relationship, we can construct a very simple life-cycle model for aggregate job satisfaction on the basis of the male age distribution data shown in Table 3. Assuming this relationship is constant and not cancelled out by other effects, we immediately conjecture that aggregate job satisfaction rates should be consistently higher in the 1960's than in the 1950's. Put simply, the cohorts of men in the low job satisfaction ages (21-29) in the 1960's were relatively much smaller than in the prior decade. Consequently, merely as an effect of the relative weights of the respective cohorts, we would anticipate that aggregate job satisfaction rates in the 1960's would be on a higher plateau than those of the 1950's. Indeed, the data in Table 1 do show that the job satisfaction rates for 1958 and 1962 are significantly lower than those for the remainder of the decade. Given the above observations about the effects of the unemployment rate and the fact that 1958 was a year of high unemployment, it is likely that the 1958 percentage is depressed even lower than it would be otherwise. Nevertheless, it is clear that the job satisfaction rate does not again reach the low levels of 1958 and 1962 even though it dips several times in the succeeding observations.

Examining the age distribution for 1970, we would project, on the basis of this simple model, a stabilizing or declining aggregate job satisfaction rate for the 1970's as relatively large cohorts of male workers enter the low job satisfaction years of the life-cycle, thus weighting the aggregate average toward lower rates. Indeed, in Table 1, we find that the Michigan Survey Research Center rates are stable for 1971 and 1973, whereas the recent Gallup polls show a slight average decline from their 92-percent peaks of 1966 and 1969. But we can construct a more sophisticated model which allows for real social forces as opposed to the simple weighting effects of this model. The general nature of such a model is based on the following considerations.

Table 3. Male Population, by Age: 1960 and 1970

Age	Percent of Male Population	
	1960	1970
15 to 19 years	7.5	9.7
20 to 24 years	6.0	8.0
25 to 29 years	6.0	6.7
30 to 34 years	6.6	5.6
35 to 39 years	6.9	5.5
40 to 44 years	6.4	5.9
45 to 49 years	6.1	5.9
50 to 54 years	5.4	5.4
55 to 59 years	4.7	4.8
60 to 64 years	3.9	4.1

Source: Social Indicators: 1973, p. 255

Relative to the positive occupational prestige-job satisfaction relationship, we note first that recent research (Hauser and Featherman, 1973) has found that there was more upward mobility during 1962-1970 than during 1952-1962. In addition, on the basis of White's (1970) research, it is now clear that one of the principal ways in which upward mobility occurs is through the creation of "vacant" positions at the top of the job structure. Such a vacancy reverberates throughout the occupational structure by creating a "chain of vacancies" as men are successively pulled up through the structure. On a society-wide scale, there are essentially two ways in which vacancies at the top of the occupational prestige scale can be created. One is through the expansion of those sectors of the economy in which higher prestige occupations are dominant, and the other is through the replacement of older cohorts as they retire or otherwise exit from the occupational structure. Both of these forces were very strong during the 1960's. That is, the 1960's were years of great expansion of the professional-technical, managerial-proprietor, and sales-occupational categories (Hauser and Featherman, 1973) and a decade in which replacement vacancies for men aged 21-29 in 1960 were high because of the larger cohorts ahead of this cohort, as can be seen in the age distribution for 1960 shown in Table 3.

Now, if the cohort of men in the low job satisfaction years (ages 21 to 29) in the early 1960's did experience higher rates of mobility into upper status occupations during the 1960's due to higher rates of job expansion in these occupations and higher vacancy replacement rates due to the smaller size of their cohorts, and if the occupational status-job satisfaction relationship is relatively constant, then, other things equal, we would expect the cohort of men aged 21 to 29 in the early 1960's to show a large increase in job satisfaction over this decade than the age 30 to 39 and 40 to 49 cohorts. Some data for making a rough check on this hypothesis are shown in Table 4. These data correspond to four of the fifteen surveys reported in Table 1 which give age tabulations with equal age intervals. We find that the 21-29 category in 1962 shows an increase of 16-18 percent in job satisfaction by the 1971 and 1973 surveys, whereas the 30-39 and 40-49 categories of 1962 show an increase of only 11-12 percent over this interval. Thus, the increase of the satisfaction percentage in the younger cohort is approximately one-half greater than that of the older cohorts. But to verify that this change is not due only to the life-cycle positive aging-job satisfaction relationship reported above, we should also find that there is an increase in job satisfaction for the same age intervals between 1962 and 1973. Examining Table 4, we find that they are about 10 percent on the average for the different age intervals. Thus, the observed 16-18 percent increase in job satisfaction of the 1962 21-29 cohort cannot be fully accounted for by either the life-cycle or the period effects alone. We conclude that the observed increase in satisfaction for this cohort is due both to the effect of the positive aging-job satisfaction relationship and to the effect of the period changes of the 1960's. It is worth noting that these conclusions are somewhat different from those of Quinn, Staines, and McCullough (1974: p. 7) who adduce that jobs have improved in terms of: (1) real wages and fringe benefits; (2) Federal and State legislation on working conditions; (3) computerization of menial jobs; and (4) a more "employee-centered" management. Of these four factors, our model does not explicitly incorporate the second and fourth. Although these job improvements may have had an impact in the movement of job satisfaction rates to the relatively higher plateaus of the 1960's, their effects are very difficult to document and the foregoing analysis shows that they may not be necessary to account for the observed changes. Moreover, they do not lead to any predictions for the 1970's as does the model sketched above.

Table 4. Percentage of "Satisfied" Workers, 1962-1973, by Age for Four Surveys with Equal Age Intervals

Age Intervals	Year and Source			
	1962	1969	1971	1973
	NORC	SRC, U of M	SRC, U of M	SRC, U of M
21 to 29 years	74	75	85	84
30 to 39 years	82	76	90	92
40 to 49 years	84	88	93	94
50 years and older	88	87	95	96

Source: Quinn, Staines, and McCullough (1974: p. 54)

Examining the age distribution for 1970 in Table 3 relative to this more complex life-cycle model, we again project a stabilizing or declining aggregate job satisfaction rate for the 1970's. In brief, the cohorts of young men coming onto the job market in the 1970's are no longer smaller than the ones preceding them. Thus, the upward mobility of these men will not be assisted by this replacement aspect of the vacancy chain effect to such an extent as were their peers in the 1960's. Moreover, even if the economy expands at its traditional 3 to 4 percent per year rate throughout the 1970's, it will be hard pressed to give upward mobility rates to the oncoming cohorts of young men comparable to those which were experienced by their peers in the previous decade. The lower rates of upward mobility will thus give these men lower job satisfaction rates, which, together with their larger relative cohort sizes, will produce a lower aggregate job satisfaction rate. More precisely, we expect the aggregate rate to be stable at first and then to decline in the middle and later 1970's.

To summarize this brief excursion into the analysis of job satisfaction, we have presented rough arguments for two types of social indicator models for the data of Table 1. First, our macro time-series model argues that the aggregate job satisfaction rate does go down when the unemployment rate goes up, particularly if unemployed persons are allowed to answer a job satisfaction question. Second, our life-cycle model argues that the "high plateau" of job satisfaction percentages of the 1960's was due to the interaction of the normal life-cycle relationship of age to job satisfaction with the unique size and mobility experience of the low satisfaction-prone cohorts of this period.

In terms of our macro model as expressed in equation (1), this implies that the constant term shifted upwards in the mid-1960's from a lower value for the earlier period. Unfortunately, this implication cannot be tested directly, due to a paucity of time points on job satisfaction prior to 1962. These models are only very tenuously formulated and are meant to be illustrative rather than definitive. Nevertheless, they do permit a rough forecast of expected job satisfaction rates for the 1970's, something which no other analysis has yet allowed. That is, our first model forecasts a fluctuation of this rate with the unemployment rate, whereas our second model forecasts a tapering off of the rate of job satisfaction improvement and a slow movement to a plateau (constant term) of aggregate rates which will probably be lower than that of the 1960's. Of course, these forecasts assume that the effects in the models remain constant and are not cancelled out by other changes which have not been taken into consideration.

4. Conclusion

Just as the great social issues for the late 1960's centered around problems associated with teenagers such as drugs and crime, one can project that the 1970's will see considerable attention given to issues associated with quality of employment and job satisfaction. In both cases, the fundamental driving force is the age structure of the population. On the basis of our distinction between macro time-series and life-cycle social indicator models, we were led to explore the effects of the population age composition and of the economy on a time series of aggregate job satisfaction rates. By demonstrating that job satisfaction has a consistent relationship with these referent indicators of social change, we have given it (and, indirectly, any quality of employment indicator that is correlated with job satisfaction) some degree of validation as an indicator of social change in the sense described in Sections 3 and 4, and this holds regardless of whether job satisfaction is treated as a goal output indicator in a social policy.

Relative to social policy with respect to job satisfaction, it is worth noting that our models indicate that aggregate job satisfaction could be maximized in a number of ways: e.g., (1) by decreasing the unemployment rate to near zero; (2) by not allowing anyone under age 29 to be employed; and (3) by abolishing blue-collar jobs. None of these alternatives is practicable in an extreme sense, although it is national economic policy to keep the unemployment rate as low as feasible and the shift in the labor force has been away from operative and nonfarm labor jobs for some time.

References

[1] Bauer, Raymond A.
1966 "Detection and anticipation of impact: the nature of the task," Chapter 1 in Raymond A. Bauer (ed.), Social Indicators. Cambridge, Massachusetts: The Massachusetts Institute of Technology Press.

[2] Blau, Peter and Otis Dudley Duncan
1967 The American Occupational Structure. New York: Wiley.

[3] Campbell, Angus and Philip E. Converse (eds.)
1972 The Human Meaning of Social Change. New York: Russell Sage Foundation.

[4] Executive Office of the President, Office of Management and Budget
1974 Social Indicators 1973. Washington, D.C.: U.S. Government Printing Office.

[5] Fox, Karl A., Jati K. Sengupta, and Erik Throbecke
1973 The Theory of Quantitative Economic Policy: With Applications to Economic Growth, Stabilization, and Planning. Second Revised Edition. New York: American Elsevier.

[6] Hauser, Robert M. and David L. Featherman
1973 "Trends in the occupational mobility of U.S. men, 1962-1970." American Sociological Review 38 (June):302-310.

[7] Kahn, Robert L.
1972 "The meaning of work: interpretation and proposals for measurement," Chapter 5 in A. Campbell and P.E. Converse (eds.), The Human Meaning of Social Change. New York: Russell Sage Foundation.

[8] Land, Kenneth C.
1971 "On the definition of social indicators." The American Sociologist 6 (November): 322-325.

[9] Land, Kenneth C.
1974a "Social indicator models: an overview," Chapter 2 in K.C. Land and S. Spilerman (eds.), Social Indicator Models. New York: Russell Sage Foundation.

[10] Land, Kenneth C.
1974b "Theories, models and indicators of social change." Paper prepared for presentation to a Conference on Indicators of Social and Economic Change at Unesco, Paris, France, May 20-22, 1974.

[11] Land, Kenneth C.
1974c "The role of quality of employment indicators in general social indicator systems." Paper presented to a Conference on Indicators of Quality of Employment, Bureau of Social Science Research, Inc., Washington, D.C., April 26-28, 1974.

[12] Land, Kenneth C. and Seymour Spilerman (eds.)
1974 Social Indicator Models. New York: Russell Sage Foundation.

[13] Moore, Geoffrey H.
1961 Business Cycle Indicators. Princeton, N.J.: Princeton University Press.

[14] Moore, Geoffrey H.
1967 Indicators of Business Expansion and Contractions. New York: Columbia University Press.

[15] Organization for Economic Co-operation and Development
1973 List of Social Concerns Common to Most OECD Countries. Paris.

[16] Quinn, Robert P., Graham L. Staines, and Margaret R. McCullough
1974 Job Satisfaction: Is There a Trend? Manpower Research Monograph No. 30, U.S. Department of Labor.

[17] Seashore, Stanley
1973 "Job satisfaction as an indicator of quality of employment." Paper presented to a Symposium on Social Indicators of the Quality of Working Life, Canada Department of Labor, Ottawa.

[18] Sheldon, Eleanor Bernert and Wilbert E. Moore (eds.)
1968 Indicators of Social Change: Concepts and Measurements. New York: Russell Sage Foundation.

[19] Taylor, James C.
1973 "Concepts and problems in studies of the quality of working life." Report to the Manpower Administration, U.S. Department of Labor.

[20] Terleckyj, N.E.
1972 National Goals Accounting: A Framework for Evaluating Opportunities for the Achievement of National Goals. Washington, D.C.: National Planning Association.

[21] United States Department of Health, Education and Welfare
1969 Toward a Social Report. Washington, D.C.: U.S. Government Printing Office.

[22] White, Harrison C.
1970 Chains of Opportunity. Cambridge: Harvard University Press.

STRATEGY ISSUES IN THE DEVELOPMENT OF QUALITY OF EMPLOYMENT INDICATORS

Robert P. Quinn, Survey Research Center, The University of Michigan

Problems confronted in the development and application of systems of social indicators are not qualitatively different from those of any other scientific endeavor. Several characteristics of research on social indicators tend, however, to highlight a number of strategy issues that under more customary circumstances can either be easily resolved or comfortably ignored. Among these characteristics are: the magnitude and continuity of the activity; the contribution to it of multiple investigators; its more-than-usual social relevance; and its multiple audiences.

This paper discusses several general strategy issues that must be confronted in the development of social indicators systems in general and systems for monitoring quality of employment in particular. Each of these issues confronted my colleagues and I in the development of two national surveys of the American workforce (Quinn, et al., 1971; Quinn & Mangione, 1973; Quinn & Shepard, 1974). Hopefully, others engaged in similar activities in the future will be able to profit from our efforts in this area.

Defining the Domain of Investigation

Recent discussions of employment and working conditions have been characterized by a boomlet of new terms that generally begin with the phrase "quality of" and are rounded off by such words as "employment," "work," or "working conditions." None of these terms has really been adequately defined, perhaps for good reason. The phrases have often been used almost as passwords to signify to the listener that something about work is being considered that is somehow new and different from the more traditional concerns of management, labor, and government--a strictly exclusionary definition. Where non-exclusionary definitions have been attempted, they have for the most part been so broad as to include "traditional" areas of concern along with the "newer" ones.

Given this broadness, the selection of measures to be used in an indicator system often becomes highly arbitrary. The system based on such a definition may as a result become a catch-all and vulnerable to the inclusion in perpetuity of measures that happened to be relevant to research problems that were "hot" only at the time when the system was developed.

My colleagues and I faced this problem of selection at a practical level when we developed the second of our national surveys in 1973. The first survey's measures of working conditions had covered a wide variety of content areas as suited the heterogeneity of those contributing to the survey and their particular research needs--the domain had indeed been broadly and loosely defined. The crunch came three years later when a decision to repeat the national survey was made. We were faced at that time with the problem of identifying those materials that would be repeated, those that would be added, and those that would be scrapped.

The "quality-effectiveness strategy" was developed in order to solve this problem and to capitalize on the survey's analytic potentials. This strategy defines three general concepts: working conditions, effectiveness, and quality of employment.

The term "working conditions" refers to descriptions of characteristics of a worker's job obtained from any informed source. These descriptions may focus on any characteristic of the job from the cleanliness of the physical work environment to the degree of time pressure for performance, or from the degree of challenge the job provides to the income a worker receives. Since working conditions characterize a job, they are independent of the individual who does the job. This means that they do not include the worker's evaluation of the conditions measured and that different people doing the same job should describe it similarly.

Effectiveness. The term "effectiveness" refers to states or events that have a positive or negative value from the perspective of some person or set of people. Three such perspectives are distinguished: those of employees, their employers, and society as a whole.

Quality of employment. The term "quality of employment" refers to a judgment about working conditions based on the impact that the working conditions have on effectiveness. Thus, good quality of employment from the perspective of a worker would be some combination of working conditions that produce health (one criterion of effectiveness as judged by the worker). Good quality of employment from the perspective of an employer would be working conditions that lead to a productive, profitable organization.

These definitions provided a standard for selecting measures of quality of employment to be carried forward into the second national survey: select only those measures of working conditions that had a demonstrable association with some criterion of effectiveness.

In the 1969 survey, job satisfaction served as the only criterion measure of effectiveness, a criterion governing the selection of the working conditions to be treated subsequently as quality of employment indicators. The 1973 national survey uses four major criteria of effectiveness other than job satisfaction: physical health; depressed mood; drinking behavior; and life satisfaction. A companion study conducted in five employing establishments focuses upon the criteria of performance, absenteeism, turnover, and participation in activities outside of work. In each of those studies the measures of working conditions (i.e., potential quality of employment indicators) have been expanded to include conditions likely to be associated with the particular criteria of effectiveness under investigation. To the extent that they are in fact able to predict these criteria, they will be included as quality of employment measures in later monitoring efforts. What is unusual about this strategy is that measures of satisfaction, health, and so forth, although relevant to the system of indicators, are not themselves indicators. Once they have served their purpose, they are shelved until some later time when it may be necessary to use them again to determine whether the quality of employment indicators have retained their predictive powers.

Making Value Assumptions Explicit

Identifying and measuring the "quality" of anything requires standards against which quality can be evaluated. As much as we would like to avoid doing so, the identification of such standards may require that social scientists take normative stands. Unfortunately, social scientists may not be without their biases in their selection of standards.

While it is impossible to rule out entirely the intrusion of arbitrary standards in the measurement of quality of employment, it may nevertheless be possible to circumscribe their effects and to make them more obvious. Two types of normative assertions are often invoked and confused in discussions of quality of employment. The first type prescribes various desirable working conditions, such as having a job that provides a great deal of autonomy. The second type of assertion prescribes only desirable outcomes or criteria of effectiveness, such as good health or satisfaction with one's life. The quality-effectiveness strategy, while permitting normative statements, restricts such statements to outcome criteria. Whether a particular working condition is "good" or "bad" thereby becomes an empirical question rather than a value issue: it is "good" to the extent that it is associated with any or all of the criteria of effectiveness thus specified. This certainly does not avoid value questions. It simply makes them a little more amenable to discussion by limiting the range of variables to which they apply and by confining them to the assignment of priorities among criteria of effectiveness.

To keep reminding ourselves of precisely whose values were being given priority in the indicators of quality of employment we have used in our two surveys, we have found it useful to distinguish three different perspectives for evaluating criteria of effectiveness. The first perspective, that of employers, is a familiar one and includes such matters as productivity, withdrawal from work, counter-productive behavior, and adaptability to changing work procedures. Among the outcomes desired from a second perspective, that of employees, are the equally familiar ones of job satisfaction, mental health, physical health, and so forth.

A third perspective can also be invoked: that of the community or the society. Some of the costs and benefits associated with working do not enter into the personal accounting of either employers or employees. For example:

- --Workers whose jobs undermine their health place an additional demand on the nation's already overburdened system of health-care delivery.
- --Workers whose skills and education are underutilized by their jobs represent an obvious social waste.
- --A worker whose expression of dissatisfaction takes the form of reactions that result in termination may become a candidate for subsequent collection of unemployment compensation, an obvious drain on local resources.
- --The income-deficient worker may burden society with a family prone to illness, future welfare costs, and substandard economic contribution.

The assignment of any criterion to a particular perspective may at times be somewhat arbitrary and perhaps even uncharitable. The assignment does not mean to imply, for example, that from the point of view of employers the physical or mental disorders of their workers are of no importance, only that from the perspectives of most employers there are other more important outcomes. Conversely, the assignment does not imply that employees are necessarily indifferent to productivity. Indeed, the harder it is intellectually to assign a criterion to a particular perspective, the more important that criterion is likely to be. According to this rule of thumb, monitoring and action priorities might profitably be assigned to those working conditions that affect outcomes that are patently relevant to all three perspectives. Work-related illnesses and injuries are one example. They are obviously important to the ill or injured worker, represent a cost to his or her employer (in terms, for example, of sick-pay and filling the worker's position while he or she is away from work), and are costly to society as well (e.g., in terms of the illnesses' or injuries' drain on the nation's already scarce medical resources).

The categorization of criteria of effectiveness in terms of perspective serves principally to clarify the value priorities of the investigator. For example, when we first compared the priorities of our two national surveys against our own list of criteria, we discovered that we had been overwhelmingly concerned with criteria of effectiveness that were valuable from an employee's perspective. Using the quality-effectiveness strategy, our resulting indicators of quality of employment would therefore be similarly biased. Our five-establishment study was designed to compensate for this by attempting to identify those working conditions that were associated with criteria of effectiveness that were important principally from an employer's and a societal perspective. Periodic reviews of the investment of our research energy with reference to this list have generally been very sobering by providing constant reminders of when we have gone overboard in our pursuit of matters that involve only limited perspectives.

Differentiating the Domain of Investigation

Given some agreement upon the domain of investigation and the value perspectives involved, there remains yet a question that must be answered before a monitoring system can be instituted: What are the basic dimensions underlying the domain? Simply talking about "good" and "bad" aspects of jobs is not sufficient. Further differentiation is obviously necessary. But how is such differentiation to be achieved, and what should constitute the basic vocabulary of the system? There are two obvious ways of answering this question. One of these casts the basic dimensions to be investigated in terms of existing theories of behavior. The second defines such dimensions empirically with only minimal reference to theory.

There are certainly plentiful a priori schemes for categorizing working conditions or workers' motivational orientations toward work, and each of these has its well-reasoned, if not always well-documented, theoretical foundations. At the simplest end are those categorizations that comfortably divide the world into two classes: jobs that are "motivating" versus those that are simply "satisfying." At the more complex extreme are those categorizations that define a variety of "needs" for working and assume that each achieves prominence as lower-level needs are progressively satisfied. Other classification schemes proliferate between these two extremes. All of them, however, present a common problem in that adopting them demands an implicit subscription to their assumptions and their logic. To adopt any one of them in a system of social indicators involves the simultaneous adoption of all its attendent theoretical trappings. While this may provide some short-run theoretical continuity to the monitoring effort, it may reduce the usefulness of the data thus collected for subsequent "secondary" analyses by other investigators who do not subscribe to the same theory. A theoretically "tight" monitoring system with an idiosyncratic theoretical orientation may therefore be self-defeating in that it may provide rich data for those subscribing to the theory involved but may toss only scraps to the non-believers.

On the other hand, the basic dimensions of the monitoring effort can be determined on a wholly empirical basis. This would involve subjecting the system's measures to some kind of dimensional analysis--factor analysis, cluster analysis, small-space analysis, or what-have-you. Many such dimensional analyses of jobs are already available, and they concur in their identification of from about five to eight "basic" dimensions of work. Such agreement is, however, based exclusively upon dimensional analyses of satisfaction ratings of job characteristics. Unfortunately, the same facets do not emerge when dimensional analyses are made of the importance ratings that workers assign to job facets. Nor do they routinely emerge from dimensional analyses of working conditions (i.e., job descriptors).

How, therefore, are facets of jobs to be compartmentalized and differentiated? If this is to be a wholly theoretical procedure, whose theory should be subscribed to and at what cost in terms of making the indicator system unpalatable to others? If it is to be strictly an empirical matter, what are to be the inputs to the relevant dimensional analyses: importance ratings of job characteristics? working conditions with regard to these job facets? satisfaction with regard to these facets? or something else entirely?

Using "Standards" versus "Fit" Models

An important issue in understanding responses to work is whether people are better off by maximizing all the "good" qualities of their jobs or whether they are better off by each person obtaining personally-attuned optimum levels of the particular outcomes that he or she desires most. Measures of the "fit" between the worker and his or her work environment have often been hypothesized to be better predictors of effectiveness than are simply the amounts of the working conditions in question. Indicators of quality of employment that do not accomodate differences in what workers want from their jobs rankle. They are superficially contrary to many theories of human behavior, not to mention common sense. They seem at odds with the advocacy of individually-oriented programs or worker training or job change. And they run counter to the ideology that every person should be treated as an individual.

But indicators of quality of employment that ignore individual differences thrive. Moreover, they constitute the vast majority of those indicators that are publicly available. That such indicators have persisted and remain useful is due in large part to the selectivity of their foci and their assumptions about workers' needs or desires. They focus principally upon those areas of working life that are of concern to most workers and areas where what is desirable to the vast majority of them can be safely assumed (e.g., health and safety, income adequacy).

While "fit" models relate more meaningfully to theories of human behavior, how feasible is their application in a system monitoring quality of employment? A classic example of attempts to translate "fit" models into measuring instruments are those that weight satisfaction ratings of particular job facets (e.g., fringe benefits, security, competent supervision) by the importance that different individuals assign to these facets. While such a weighting procedure is theoretically persuasive and seems easily performable, attempts to improve job satisfaction measures by this procedure have without exception failed to justify it. Whether to include measures of individual preferences in systems of quality of employment indicators seems, therefore, less a theoretical issue than a practical one. It is difficult to argue against their inclusion on theoretical grounds. The only remaining problem is to determine when they constitute a necessity and when they are only a nicety.

r b ms f C mm ca i[1]

Because of its presumed usefulness to society, many results of research involving social indicators are likely to reach a broader audience than social scientists are accustomed to addressing. This, naturally, creates uncommon problems of data presentation and interpretation, only some of which can be overcome by careful use of language. A deeper-rooted problem involves the use of complex measures, particularly psychological measures based on continuous scales. Those who are not social scientists expect social scientific data to be quantifiable in the same terms that the data of other sciences are. That is, absolute quantities of things can be tallied, or the percentages of people of various types can be specified. Pollsters tend to perpetuate this expectation by declaring, for example, that such-and-such percentage of Americans disapprove of what the President is doing--as if approval and disapproval had been measured as absolutes.

The question we are most frequently asked about our national surveys is a very straightforward one: "How satisfied with work is the average American worker?" Our answer, characterized by the customary lucidity of the social scientist is: "minus two." This answer, while perfectly accurate, is not very helpful to the lay questioner. Our inability to answer otherwise has two sources. First, all the questions in our job satisfaction measure use four- or five-point, fixed-alternative scales, the interval properties of which can certainly be questioned. The second reason involves the complexity of the construction of the satisfaction measure. The five "facet-free" questions using five-point scales are averaged, as are the 34 "facet-specific" questions that use four-point scales. The distributions of each of these means are then converted to standardized _z_ scores. Next, a mean of these two _z_ scores is obtained for each individual, and the resulting mean is finally multiplied by 100 to remove decimals.

The latter arithmetic whoop-de-do, which is not really very complicated as psychological measures go, has certainly not made the initial set of job satisfaction questions lose sight of the psychological reality under investigation. It nevertheless leaves non-scientists completely in the dark, even those accustomed to weather reports that present such equally (or more esoteric) numbers as wind-chill factors and MURC indices. None of this helps the image of social science as a publicly useful discipline. While we can answer some complex questions about job satisfaction based on this measure, superficially simple ones, like "How satisfied with work is the average American worker?" are embarrassingly difficult.

There seems no obvious way around this difficulty. Talk of "educating the public"--or even educating relevant sectors of the public--to the complexities of psychological measurement is largely wishful thinking. This will continue to be true so long as the pollsters persist in leaving the impression that matters are really very simple and so long as posters in the New York subways announce the absolute percentages of Americans with "mental health problems." Occasionally, awkward, but workable compromises can be reached. One such successful, apparent compromise is the widely-used measure of drinking behavior developed by Cahalan, Cisin, and Crossley (1969) for use in population surveys.[2] This measure, far more complex than our own job satisfaction measure, involves very complicated combinations of the frequency of alcohol consumption, the amount consumed at each sitting, and individual variability in amount consumed. These combinations produce a continuous scale of values that the measure's developers break at arbitrary, but reasonable points. To each part of the distribution thus differentiated a publicly meaningful label is assigned: "heavy drinkers," "infrequent drinkers," etc. No one suffers greatly from this assignment. The non-professional users of the measure feel comfortable in having available exact percentages of "heavy drinkers." The professionals use the categories simply as a set of ordered classes, ignore the labels assigned them, and wink knowingly at each other.

But segmenting the continuous distribution of a scale into ordered classes and assigning a label to each can at times result in misunderstandings. To simplify analysis and data presentation, a continuous distribution is often dichotomized at its median and the upper and lower halves of the distribution designated "high" and "low." One of our reports from the 1969 national survey did so. A median split on job satisfaction was performed and a table distinguished "Satisfied" from "Dissatisfied" in order to show the association between job satisfaction and some other variable. Naturally, 50 percent of the workers fell into each category. An early reader of the table, unsophisticated in such "conventional" presentations of data, found the table's marginals very interesting. The marginals "obviously" indicated that half of the work force were dissatisfied with their jobs. The incident is instructive in that it confounded two problems--the problem of communicating data to those unfamiliar with certain conventions and the problem of balancing descriptive and analytic goals in monitoring systems. The user of the data in this example was expecting the survey to provide only descriptive statistics in readily interpretable terms; the marginal distributions were therefore very interesting. Our staff, on the other hand, was concerned with the association between a particular working condition and job satisfaction; the marginal statistics and their verbal designations were only following journal conventions.

Any measure used in a population survey or in a system of social indicators is likely to achieve a premature legitimacy. This is especially likely to be true where the sample size is large and there is an impressive array of institutions or people directing, conducting, or financing the effort. This situation can produce yet another problem of communication. Even a good measure can usually be improved, and its inclusion in a system of social indicators provides an excellent opportunity for such improvement. This raises the question of how much alteration a monitoring system can undergo during its history without losing touch with its users, particularly those who would borrow some of its measures for use in more limited studies. The latter group present an especially difficult problem, since it is particularly tempting to them to appropriate the system's measures. Not only are the measures legitimized, albeit often prematurely, but national norms are available. Given writing and publication lags, borrowers of a system's measures may find themselves using early versions of measures that have been long since improved upon by the system's designers. For example, some of the difficulty in comparing surveys using the Cahalan, Cisin, and Crossley measure of drinking behavior stems from the survey's uses of various "developing" versions of the final instrument. Clearly, freezing measures for eternity does little for the quality of a monitoring system. Good judgment must be exercised, however, in the timing of changes and improvements, as well as their magnitude. Frequent changes of only minor importance may do a disservice to the system's users.

FOOTNOTES

[1] Problems of Communication

[2] Another example is Belloc, Breslow, and Hochstim's (1971) survey measure of overall physical health.

REFERENCES

Belloc, N., Breslow, L., & Hochstim, J. Measurement of physical health in a general population survey. American Journal of Epidemiology, 1971, 93(5), 328-329.

Cahalan, D., Cisin, I., & Crossley, H. American drinking practices. Rutgers, N.J.: Rutgers Center of Alcohol Studies, 1969.

Quinn, R., Seashore, S., Kahn, R., Mangione, T., Campbell, D. Staines, G., & McCullough, M. Survey of working conditions. Washington, D.C.: U.S. Government Printing Office, 1971. Document 2916-0001.

Quinn, R.P., & Mangione, T.W. The 1969-1970 survey of working conditions: chronicles of an unfinished enterprise. Ann Arbor, Michigan: Survey Research Center, 1974.

Quinn, R.P., & Shepard, L.J. The 1972-73 quality of employment survey: descriptive statistics, with comparison data from the 1969-70 survey of working conditions. Ann Arbor, Michigan: Survey Research Center, 1974.

EQUITY CONCEPTS AND THE WORLD OF WORK

Lester C. Thurow, Massachusetts Institute of Technology

For a decade or more governmental manpower efforts have focused on equalizing the distribution of earnings. Blacks are to catch up with whites; women are to be paid the same as men; the poor are to escape from poverty. While there has been some limited progress toward all of these goals, the distribution of earnings has been remarkedly resistent to changes. Explicit manpower programs and enormous changes in educational attainment of the labor force have made almost no dent in the pattern of earnings. In the post-war period the distribution of earnings among adult white males--the group not subject to discrimination--has become slightly more unequal.

If you look at the theories lying behind social efforts to reduce inequalities in the distribution of earnings in the post-World War II period, they: (1) follow neo-classical economics and (2) ignore a wide variety of psychological, sociological, and labor economics research.

Almost all government manpower and education programs are based on the neo-classical economic view of the world. Neo-classical economics assumes that workers come into the labor market with a definite, preexisting, set of skills (or lack of skills), and that they then compete against one another on the basis of wages. According to this theory education and formal training are crucial since they create the skills which people bring into the market. As a result it is possible to governmental authorities to impose a more equal distribution of earnings on the labor market by injecting a more equal distribution of skills into the labor market.

Injecting a more equal distribution of skills into the labor market has a powerful three-pronged impact on the distribution of earnings. First, a training program that transforms a low-skill person into a high-skill person raises his productivity and therefore his earnings. Second, it reduces the total supply of low-skill workers, which leads in turn to an increase in their market wages. Third, it increases the supply of high-skills workers, and this lowers their wages. The net result is that total output rises (because of the increase in productivity among formerly unskilled workers), and the distribution of earnings becomes more equal. What could be more ideal?

From this point of view there is no need to worry about, or even to know, anything about the equity judgements of the work force. If society decides to equalize the distribution of earnings, it can do so indirectly through equalizing investments in human capital. The equity judgements of the work force play no role in the economy although they may play a role in the political arena.

The rest of this paper will attempt to argue that the equity judgements of the work force are not irrelevant. They may, in fact, be at the heart of the current distribution of earnings. Any attempt to alter the distribution of earnings may have to take them into account and may have to change them. The first step in any such effort would be more explicit knowledge about the equity judgements of the work force. Some indicator of the equity judgements of the work force, and how they change, might be of crucial importance in government policies to change the distribution of earnings. But this is to get ahead of the argument. The first step is to understand the importance of equity judgements to government manpower policy making.

I. The Importance of Interdependent Preferences

Neo-classical economics and the marginal productivity theory of distribution implicitly assumes that individuals look only at their own wages and productivity to determine whether they are fairly or unfairly paid, or that they are unable to do anything about it if they do in fact look at their relative wages and become unhappy with what they see. Both of these assumptions are open to question.

The history of individuals looking at their relative wages and becoming unhappy with what they see is at least as old as the history of mankind. Man's interests have existed for at least the last 2000 years.

> "For the kingdom of heaven is like a household who went out early in the morning to hire laborers for his vineyard. After agreeing with the laborers for a denarius a day, he sent them into his vineyard. And going out about the third hour he saw others standing idle in the market place: and to them he said, 'You go into the vineyard too, and whatever is right I will give you.' Going out again about the sixth and ninth hour, he did the same. And about the eleventh hour he went out and found others standing; and he said to them, 'Why do you stand here idle all day?' They said to him, 'Because no one has hired us.' He said to them, 'You go into the vineyard too.'
>
> And when the evening came the owner of the vineyard said to his steward, 'Call the laborers and pay them their wages beginning with the last, up to the first.' And when those hired about the eleventh hour came, each of them received a denarius. Now when the first came, they thought they would receive more; but each of them also received a denarius. And on receiving it they grumbled at the householder, saying, 'These last worked only one hour, and you have made them equal to us who have borne the burden of the day and the scorching heat.' But he replied to one of them, 'Friend, I am doing you no wrong,

did you not agree with me for a
denarius? Take what belongs to
you and go."[1]

While an interest in my fellow worker's earnings and the resulting unhappiness and disruptions when norms of social justice are broken may be contrary to the norms of both biblical man and economic man, they seem to be an endemic part of human man. Imagine what would have happened if the parable of Lord of the Vineyard had extended over to another day. Hiring labor and operating the vineyard on the second day would have been a real headache.

A wide variety of more recent evidence points to the existence of the same type of interdependent preferences. Over the past three decades the Gallup poll has asked, "What is the smallest amount of money a family of four needs to get along in this community?" The 17 answers to this question have all fallen between 53 percent and 59 percent of the average income of the year in which the question was asked.[2] The responses are consistent with respect to the average income in the year in which the question was asked but grew in absolute terms as average incomes grow. A Harvard sociologist, Lee Rainwater, has shown that when people are asked to categorize others as "poor, getting along, comfortable, prosperous, or rich," they do so rather consistently relative to average incomes.[3] A University of Pennsylvania economist, Richard Esterlin, has reviewed the evidence as to how happiness is related to income in different countries of the world.[4] He finds that happiness (utility?) is almost completely dependent upon one's relative income position within his own country and almost not at all as to whether the individual is located in a high income country or a low income country.

The same phenomena are reported in labor negotiations. Bargaining about relative wages is at least as pervasive as bargaining about absolute wages. Perhaps the best recent example occurred in Sweden where college workers struck to increase their pay relative to non-college workers. Their demand was not for more income, but for a wider wage differential. Conversely it is difficult, or impossible, to find any employer who systematically attempts to measure worker's marginal products and then sets pay scales in accordance.

Actual utility functions seem to be heavily, if not completely, determined by relative incomes and interdependent preferences rather than absolute incomes and independent preferences. Sociologists call interdependent preferences 'relative deprivation'; psychologists would label the same phenomenon 'envy'; labor economists refer to wage contours. Whatever the name, interdependent preferences seemed to be a widespread phenomenon.

To say that utility functions are highly interdependent, however, is not to say that men are going to be able to implement their interdependent preferences in the labor market. What allows individuals to exercise their interdependent preferences in the labor market? My utility may depend upon the income of my neighbor, but this would not influence my own wages or productivity in the standard wage competition model. Like it or not, each individual would be paid his marginal product.

The assumed irrelevance of interdependent preferences flows from two counterfactual assumptions implicitly contained in the wage competition model. First, individuals are assumed to have fixed marginal products--skills--that they sell in the labor market. In fact individuals have a vector of possible marginal products depending upon their motivation. An unhappy worker can lower his productivity. Often the reduction can occur in such a manner that it is difficult and expensive to determine whether a worker has or has not reduced his productivity. While a worker's happiness or utility is irrelevant if he has a fixed marginal product, it is highly relevant if he has a variable marginal product. Employers need to set a structure of wages that illicites voluntary cooperation and motivates their work force. The net result is an avenue whereby interdependent preferences can influence the wage structure.

Second, individuals are assumed to be interchangeable parts in the production process. In fact most production processes require a degree of teamwork that can only be acquired through on-the-job experience and a high degree of internal harmony. A team with a revolving membership or a team that is unhappy with its wage structure has a lower productivity than a team with a stable membership and satisfied with its wage structure. There is a high degree of truth in the old aphorism, "there is no institution that cannot be brought to its knees by working to rule." Efficient economic production does not occur if everyone does just what is required or what is compelled. The net result is an avenue whereby group preferences about a "just" wage structure can have a major impact on production.

While economists have ignored the problem of getting individuals and groups of individuals up to their maximum productivity, industrial psychologists have made this a key problem. They ask how wages and other incentive systems can be used to promote maximum productivity. Economists see the work decisions as a zero-one decision where the individual either does or does not sell his time and a fixed productivity for the offered bribe. Industrial psychologists see the work decision as a more continuous decision. A person decides to work, but he also decides how much effort and cooperation to provide. Economists might respond that workers can always be fired if they are not producing at the agreed upon level, but this ignored the costs of hiring and firing, the costs of determining whose productivity is below the norm, and the costs of disrupting the production team. While there is a limited role for inspection and punishment, productivity basically depends upon voluntary cooperation and this requires a wage structure that is in harmony with the interdependent preferences of the work force.

Team wage structures lead to different wages for the same skill (a major puzzle for neo-classical economics). Some workers with a particular occupational skill play on high productivity teams while others play on low productivity teams. Raw unskilled labor makes a very different wage depending upon whether it plays for General Motors or for a Mississippi plantation. The two workers have exactly the same skill, but they are effectively segregated from each other. The low wage Mississippi farm worker is not allowed to make a bid for the job of the unskilled auto worker. One's employer becomes an important element in determining one's wages in a way that could not occur in a simple wage competition. The net result is a structure of wages that is often more homogenous within firms or industries than it is within occupations.

The variability of individual and team production functions creates problems for a marginal productivity theory of distribution since there is not a distribution of marginal products but many potential distributions of marginal products. If an employer attempts to pay a group its marginal products and these ran counter to the interdependent preferences of the group, the employer may find a completely different set of marginal products from what he originally found. What is worse, an employer that attempts to impose a marginal productivity distribution of earnings on a contrary set of interdependent preferences finds that productivity substantially decreases in the process. Interdependent preferences lead to a situation where group and individual performances depend upon having a set of relative wages that the group itself regards as fair and equitable.

Since their profits depend upon it, employers are anxious to establish a wage structure that their employers regard as equitable.. There is a profit maximizing wage structure, but it need not be a marginal productivity wage structure. Individual marginal products may have little to do with the structure of wages even if average wages are governed by average productivity. The structure of wages within this average depends upon the structure of interdependent preferences rather than upon the structure of individual marginal products.

Interdependent preferences combined with self-controlled individual and team production functions leads to the rigidity that is prevelent in the wage structure. Wages are not flexible in the manner that neo-classical economics would hypothesize since rapid wage flexibility becomes counter-productive in production environment where wage increases for one worker show up as a real wage (utility) reductions for others. This loss in utility causes them to lower their own productivity and to disrupt team activities. Given the need for production teamwork and the existence of interdependent preferences, wages are negotiated and set on a team rather than an individual basis. Unions formalize and perhaps strengthen the process, but they do not cause it. Non-union profit maximizing employers have the same interests. The same wage rigidity is noted in the non-union sectors of the economy.

II. The Analytical Problem

Analytically the problem is to know what factors produce and alter interdependent preferences and group norms of industrial justice. Sociologists have extensively studies this process under the title of "relative deprivation."[5] Sociological studies indicate that individuals have strong feelings that economic benefits should be proportional to costs (i.e. effort, hardships, talents, and the like), but that equals should be treated equally. Since there are various "costs" and rewards (income, esteem, status, power, etc.) in any situation, the problem immediately arises as to how equals are defined and how proportionality is to be determined.

This leads to the difficult problem of "reference group" determination. To what group do you belong and to what groups do you compare yourself when trying to determine whether you are being treated equally and proportionally. In any historical situation it is relatively easy to describe the different reference groups that exist, but it has proven difficult, or impossible, to find general principles that govern reference group determination.

Reference groups seem to be both stable and restricted. People look at groups that are economically close to themselves and require great social shocks, such as wars and economic depressions, to change specifications of relative deprivation. Conceptions of what constitutes proportionality and equality tends to be heavily determined by history and culture. Distributions of the past are fair until proven unfair.

This explains why inequalities in the distribution of economic rewards that are much larger than inequalities in the distribution of personal characteristics seem to cause little dissatisfaction, and why people tend to ask for rather modest amounts if they are asked how much additional income they would like to be making. The happiest people seem to be those that do relatively well within their own reference group rather than those that do relatively well across the entire population.

The importance of social shocks can be seen in the income changes caused by the Great Depression and World War II.[6] In the Great Depression, an economic collapse was the mechanism for change. Large incomes simply had further to fall than small incomes. In World War II there was a concensus that the economic burdens of the war should be relatively equal ("equal sacrifice") shared, so the federal government used its economic controls over wages to achieve more equality. Wage policies during World War II were a manifestation of a change in the sociology of what constitutes "fair" wage differentials or relative deprivation. As a consequence of the widespread consensus that wage differentials should be reduced, it was possible to reduce wage differentials deliberately.

After the Great Depression and World War II wage differentials had become embedded in the labor market for a number of years, these differentials became the new standard of relative deprivation and were regarded as the "just" wage differentials, even after the egalitarian pressures of World War II had disappeared. Basically the same differentials exist to this day--30 years later.

It is important to note, however, that the new standards were not imposed by government or a reluctant population but were imposed on the labor market by population beliefs as to what constituted equity in wartime. No one knows how to engineer such changes in less extreme situations.

The labor economics literature has the concept of relative deprivation under a different name--wage contours.[7] Different groups of workers expect to be treated relatively equally and to have a fixed structure of wages with respect to other groups. Historical differentials are to be observed. In peace as opposed to war, wage controls are used to reestablish historical wage differentials in an effort to control inflation. It is thought that one of the major elements leading to wage inflation is the leapfrogging that occurs when wage structures start go get out of line with historical wage contours. One group gets ahead of its historical position and other groups attempt to reestablish their historical position, or even to get ahead so as to "get even" for the initial violation of "equity." As with relative deprivation, the wage contour theory runs into problems in that it seems to be impossible to find general principles for why specific wage contours exist. They are easy to describe but hard to explain. The inability to find analytical explanations of reference groups or wage contours makes it difficult to know how to alter reference groups or wage contours, but it in no way diminished their importance to the structure of wages.

If utility functions are interdependent and conditioned by experience and history, relative wages may be rigid regardless of changes in the underlying distribution of marginal products. The historical wage differentials have the sanction of time. They are assumed to be just until proven unjust. Even more importantly, the longer they exist, the more they condition worker's beliefs about what constitutes justice and injustice.

To say that relative earnings are conditioned by interdependent preferences is not to say that relative earnings are immutable. Slow changes in relative earnings might be accepted since they never seem to challenge the accepted norms. Relative deprivation does, however, stop short-run wage changes from being used as a market clearing mechanism. The static benefits to be gained by clearing markets with wage changes simply are not large enough to offset the losses from the labor disruptions that would follow.

As a consequence to understand the structure of earnings and the factors that produce changes in it, it is necessary to have a sociology or psychology of interdependent preferences. Lacking a consistent theory of reference group determination, the sociology of wage determination is in a rudimentary form, but this does not diminish its importance. Worker's views about what constitutes an "equitable" wage structure have an important role to play in the determination of wages. Relative deprivation, wage contours, interdependent preferences, envy--they all mean that economic stratification is man-made, but that it is to a large extent self-perpetuating and autonomous.

III. The Need for Indicators of Equity Judgements

This brings us back to the subject of labor force statistics. If it granted that interdependent preferences both exist and can influence the actual structure of wages, interdependent preferences become important in any efforts to alter the structure of earnings. To know how interdependent preferences affect the structure of earnings, it is necessary to know something about the structure of interdependent preferences. What is in fact regarded as a fair differential between two jobs? Do norms of social justice differ across geographic, racial, occupational, and industrial groups or are they general widely-shared American norms. Answers to all of these questions might be important in efforts to alter the structure of wages.

Knowledge about the structure of interdependent preferences and changes in the structure of interdependent preferences is also necessary to study what, if any, factors cause changes in interdependent preferences. It may be true that only major wars can bring about changes, but this is only a hypothesis until some efforts are made empirically to measure interdependent preferences and until some efforts are made to see if any "controllable" variables have an influence upon them.

I do not pretent to be an expert on attitudinal studies and the appropriate questions to ask in constructing valid statistical measures of equity judgements, but let me briefly indicate some of the kinds of things that might be useful. The basic need is to go beyond satisfaction measures to determine norms of social justice. What does the average worker perceive as the "fair" wage differential between the highest and lowest skilled worker in his place of work? Does this differential differ from the actual wage differential? Similar results would be interesting for white collar workers and for managers. How do these differentials differ from what would be perceived as a just distribution of wages across the entire economy? Should the highest paid worker make 100 times as much as the lowest worker or should he make 10 times as much? Obviously these are crude questions that need refinement, but they indicate the general type of information that might be useful. While it is interesting to know whether a worker is happy or unhappy (satisfied or unsatisfied), it would also be interesting to have some quantitative measures of what he perceives as fair or unfair. Satisfaction and fairness are not necessarily the same thing.

FOOTNOTES

1. Bible, Matthew, Chapter 20, Verse 1-14.

2. Lee Rainwater, "Poverty, Living Standards and Family Well-Being," Joint Center for Urban Studies of MIT and Harvard, Working Paper No. 10, Page 45.

3. IBID, Page 49.

4. Richard Esterline, "Does Money Buy Happiness?" The Public Interest, Winter 1973.

5. For an example see: Walter Garrison Runcimen, Relative Deprivation and Social Justice, Routledge & Paul, London, 1966.

6. World War II is the most recent period where the distribution of earnings became noticeably more equal.

7. John Dunlop, Wage Determination Under Trade Unions. Kelly, New York, 1950.

DISCUSSION

Richard P. Shore, U. S. Department of Labor

Being neither a statistician nor an expert on social indicators, I feel somewhat constrained in expressing many definitive judgments about the three papers which have been presented--other than commenting on their uniform excellence. What I believe I can do well, and perhaps what should have been done in a prefatory statement, is to provide something of a historical backdrop and a context for the papers which might enhance their meaning. Having accomplished this, I will then give some reactions to each.

Six years ago, in the annual Manpower Report of the President, the then Secretary of Labor expressed the following conviction: "What a man's or a woman's work is like and what employment means are crucial to the quality of American life. There is a danger of forgetting that the ultimate purpose of the economy--and of employment as part of it--is to satisfy the needs of individuals, instead of the other way around. We must begin to consider and examine the meaning of employment--in terms of human satisfaction--going beyond the earnings it provides. The full significance of work can be identified only through examination of all the varied gratifications--and deprivations--to which it leads. We are undertaking that examination."

In a subsequent section of the Report, prefaced by the heading "Quality of Employment," this statement can be found: "If meaningful and generally acceptable indexes of the quality of employment are to be developed...the current limited efforts to refine concepts and measures, and to expand research on the complex interrelationships among the characteristics of the individual, his job, and his environment must be greatly intensified. Efforts to date have served the more limited objectives of employers and academic scholars better than the much broader and more stringent requirements of national planning."

Regretably, these requirements had not been met when, less than five years later, the Department of Health, Education, and Welfare released its now well-known examination of Work In America in the form of a report which, at least, sought "to lay the groundwork for changes in policy" regarding work and employment. As you may well know, many of the report's judgments about the condition of American workers, constituting much of the raison d'etre for its policy recommendations, proved to be highly controversial. I think that it would be useful for us to re-examine a few of them.

According to Work In America, things have been changing, but "work has not changed fast enough to keep up with the rapid and widescale changes in worker attitudes, aspirations and values." As a consequence, "significant numbers of American workers are dissatisfied with the quality of their working lives...creating an increasingly intolerable situation /with/ severe repercussions...likely to be experienced in other parts of the social system." More specifically, we are told that "as work problems increase, there may be a consequent decline in physical and mental health, family stability, community participation and cohesiveness, and 'balanced' sociopolitical attitudes, while there is an increase in drug and alcohol addiction, aggression, and delinquency." Finally, as to implications for Government Policy, "we have sufficient information about the relationship between work and heart disease, longevity, mental illness, and other health problems to warrant governmental action. That jobs can be made more satisfying and that this will lead to healthier and more productive workers and citizens is no longer in doubt."

No one would dispute that these are pretty bold assertions, but many are likely to regard them as highly equivocal ones, perhaps more accurately characterized as hyptheses than as established principles. But let us continue to trace the path of recent history.

In 1973, the Office of Management and Budget completed its extensive collation of data culminating in the publication of Social Indicators 1973, an extension, in a sense, of the Federal Government effort that began with the publication in 1969 of Toward A Social Report. In addition to its fairly predictable treatment of Employment Opportunities, the OMB report also makes a very commendable effort to characterize the Quality of Employment Life. Chosen to represent this concept were a measure of job satisfaction (as defined and operationalized by the Michigan Survey Research Center) and a range of working conditions measures representing both contributory (benefit coverage) and outcome (work injuries) variables.

While Social Indicators 1973 certainly deserves high marks for its accomplishments in collating a range of data which hitherto had not been assembled in such fashion, I suspect that most readers,whose expectations were heightened by the volume's impressive title, could not help but be disappointed. I do not intend this so much as a criticism of what was done as an expression of my own disappointment of what was produced in relation to what was needed. In short, it was a reasonable beginning, but certainly more illustrative than definitive. "Quality of Employment," though still a useful referent term, was still in need of definition, conceptually and operationally.

This need was probably nowhere more acute than in the Department of Labor where a little known but intensive debate was taking place regarding the "true" condition of the American worker. When it was proposed, as a few had the temerity to do, that the Department do battle with the obscure, if not unseen, enemy called the quality of work or employment, the predictable, perhaps invariable, response was: "Show me the data which establish the existence of a problem that justifies or requires Government attention." Regretably, the data necessary to build a case for action did not seem to exist. There was no compelling evidence, for example, that job satisfaction had declined either precipitously or gradually, or that absenteeism was raging, or that productivity was disasterously sagging, or that we had been experiencing wholesale defection from the labor force.

On the other hand, the protagonists in this dispute could point to the notable strides taken since World War II to improve the well-being of workers, in terms of wage levels, income security, job protection, and the like. If there were in fact a case to be made, it obviously would have to be based on reliable data rather than catchwords, slogans, polemic and theory. It would require that we first establish just what it is that is important to measure and to improve.

So it was on to the Bureau of Social Science Research with this "simple" request for assistance: "There seems to be something of a campaign afoot to improve the quality of work (or employment, or working life). It sounds like a reasonable national goal, but what specifically is it that we are or should be striving to change and how can we gage the progress of our change efforts?"

The three conceptual papers which have been presented here today deal with a subset of the several issues which are being addressed as part of that exercise. Each of them, in my view, makes a notable contribution to the overall purpose of the project.

I must confess to having a particular affinity for the Quinn paper, not because of the fact that much of its substance grows out of research--very excellent research--supported by the Department of Labor, but because of what I have found to be a extraordinarily useful approach to conceptualizing and assessing the quality of employment. As Quinn points out, the quality-effectiveness strategy differs significantly from the OMB Social Indicators approach in that it deals with processes instead of outcomes alone.

It not only seeks to provide explanations of what causes what, but more important, it establishes a basis for selecting points of intervention in the social system which are likely to bring about desired changes in outcome measures.

The examination of the quality of employment simultaneously from three different perspectives has much to commend. This approach expands the more traditional conception of social indicators as criteria of individual well-being to include as well the many broader and equally legitimate organizational and societal interests in the character of employment. Realistically, policy decisions are not going to be based exclusively on the measurable impact of employment on workers alone, but will take into account as well the consequences for other concerned parties as well as society generally.

This three-way approach to assessment serves well both to highlight the value dimension of social indicators and to locate it more sensibly at the outcome or effectiveness end of the cause-effect equation than at the working conditions end as is commonly done. Too many contentious statements already have been made regarding the attributes which are supposed to characterize good and bad jobs. This kind of debate is better resolved empirically than theoretically or polemically, by establishing what the specific consequences are of any particular job attribute.

Of course the beauty of the tri-perspective approach to assessing the quality of employment is somewhat marred by an inevitable complication. Despite the optimistic assumption subscribed to by some that conditions of employment which best serve the needs of workers also further the interests of employers and others, there is no sound basis for concluding that all good things necessarily go together. Research now underway at the Survey Research Center, with Labor Department sponsorship, hopefully will shed badly needed light on what form is in fact taken by the "structure of effectiveness." It seems reasonable to anticipate, however, that there are likely to be conflicts--perhaps substantial ones--among the three basic perspectives with respect to at least some conditions of employment. The method for their resolution is by no means obvious. Notwithstanding Quinn's assertion that the selection of dominant perspectives in cases of conflict should be made by "those individuals, groups, and organizations whose positions confer on them the legitimacy (emphasis added) to make such decisions." If there are any universally accepted criteria of legitimacy, I am not aware what they may be or how they became established.

A somewhat related problem concerns the selection from among the broad set of working condition variables some subset that might best occupy our attention in monitoring the quality of work. I agree with Quinn that this choice should certainly include those conditions that impact significantly on effectiveness criteria that are important to all three perspectives. However, this is a necessary but not sufficient basis of choice. It stands to reason that there are conditions which significantly affect workers, for example, that have little impact on employer interests. I doubt the wisdom of assigning such conditions to a low(er) order of priority merely on the basis of their negligible implications for employer goal achievement.

Land, like Quinn, argues the importance of establishing cause-effect linkages between outcome indicators and antecedent conditions, although the relationships he examines are with broad social and economic conditions rather than specific features of employment. Using job satisfaction as an aggregate social indicator, Land describes two tentative models he has developed to illustrate how job satisfaction can be established as an indicator of social change.

In attempting to relate trends in aggregate job satisfaction to the basic economic indicator of the unemployment rate, Land is immediately confronted with a formidible methodological problem: the doubtful reliability and interpretability of available statistics on job satisfaction. It is no easy matter to plot trends in job satisfaction, notwithstanding the deceptively facile judgments that some observers have made in the last few years. In addition to suffering from the limitation of excessively high standard errors, existing data on job satisfaction reflect rather substantial variance in question wording. There is a reasonable basis for doubting the comparability, for example, of answers to questions about satisfaction with "your job" and satisfaction with "the work you do." This measurement problem is rather nicely illustrated in the appendix of a monograph written by Quinn for the Department of Labor (Job Satisfaction: Is There a Trend?, Manpower Administration Monograph No. 30). I urge you to read it.

In examining the linkage between satisfaction and unemployment rate, Land finds a significant inverse relationship between the two indexes, with the former increasing as the latter decreases. The hypothesis he offers up to account for this correlation seem plausbile enough, although I suspect that other equally plausuible ones could be generated. I do find particular favor, however, with the notion that the relationship may in part be a function of general changes in life satisfaction. I have long been puzzled by the neglect of the obverse of the popular "spillover" hypothesis; i.e., general life circumstances influencing work experience.

It is notable too that Land's findings and interpretations run counter to what I sense is the prevailing wisdom around the Labor Department. Although the popular view is that job satisfaction and unemployment are likely to be correlated, the relationship is assumed to be direct rather than inverse, with increases in unemployment accompanied by increases in job satisfaction. Presumably, during periods of rising unemployment, workers are so damned thankful to have any job at all that they overlook or play down those features of employment which might otherwise distress them. I think the lesson to be learned here is the importance of defining issues and developing policy on the basis of empirical data rather than gut feelings and highly personalized views of the world. A system of social indicators of the quality of employment obviously has much to offer in this respect.

Land's second model, establishing linkages between satisfaction and both occupational status and age, is also provocative. Although I find the logic of his analysis and the predictive power of the model fairly impressive, I think it best that we reserve final judgment until there is some opportunity to test the accuracy of his projections to the decade of the 1970's. It is well that we accept his acknowledgement that both these models are quite tentative and that they are merely illustrative of how job satisfaction can be given interpretive value as a social indicator.

Of course, even if we assume that both models are essentially valid, we are faced with no small challenge in deriving policy implications regarding the enhancement of job satisfaction. With his tongue in perhaps both cheeks, Land suggests three possibilities--eliminate unemployment, bar those under 29 from employment, and abolish blue-collar jobs. I trust that we can be equally imaginative in discovering some more viable options. It may be that the much maligned Work In America treatise deserves more careful consideration in this regard.

Finally, a few thoughts about Thurow and his treatment of equity in the world of work. Having been subjected these past few years to the tyranny of neo-classical economists, it is exhilarating to find that there are other economic perspectives that are alive and well, and to hear it conceded that there are disciplines other than economics which may have something to contribute to the definition and solution of social problems. I might add that I was forewarned that the Thurow brand of economics is both mushy soft and dangerous, and that I should be wary of giving any implicit endorsement to his efforts to subvert or mongrelize the discipline by conceding that sociologists and psychologists have something of value to contribute in attempting to understand and shape the world of work. For the while at least I will take my chances and do penance if necessary after I rejoin the proponents of the true faith in Washington.

The concept of interdependent preferences and the issue of equity are of obvious concern in the development of a social indicators scheme. Indeed, I would be inclined to go beyond their pertinence to the matter of wage structure and apply a criterion of equity to other, non-economic (or at least less economic) conditions of employment. Equity is a basis for differentiating among occupational groups with respect to job security, comfort and safety, and other amenities of employment as well as in terms of wage benefits.

Although I disagree with many of his interpretations, I do concur with the basic tenor of Robert Schrank's treatment of what he calls "schmoozing" in the May 1974 issue of Industrial Relations. This quite unique term is used to describe a range of opportunities available to some work groups that are not available to other groups--the freedom to make adjustments in standard working hours, the opportunity to use a telephone while at work, and the chance to engage in casual socializing in the work place. Government white-collar workers and assembly line workers present picture of striking contrasts in this regard. Schrank argues that the many amenities available to white-collar workers generally which are denied to blue-collar workers produce considerable resentment. I suspect that he is correct, although I know of no basis for gaging either its prevalence or its intensity.

AN INVESTIGATION OF INTERVIEW METHOD, THREAT AND RESPONSE DISTORTION

William Locander, University of Houston
Seymour Sudman, University of Illinois, Urbana-Champaign
Norman Bradburn, University of Chicago

Introduction

Response effects in surveys are a major concern to users and gatherers of survey data and have received wide attention in the literature. Sudman and Bradburn list more than 900 articles in their recent book Response Effects in Surveys. (35, 36) Two of the most important variables that have been studied are question threat (6, 12, 13, 14, 15, 17, 21, 26) and method of administration. (4, 5, 7, 8, 9, 10, 18, 20, 22, 23, 25, 28, 29, 31, 32, 33, 34, 38, 40, 42, 43, 44, 45) Usually these have not been studied jointly, nor has there been any validation of response.

The purpose of this study was to examine the joint effects of question threat and method of administration on response distortion. Another major objective was to study the randomized response model which has been described in several recent articles (2, 3, 11, 16, 19, 30, 41) as a technique to reduce or completely eliminate response distortion of threatening or personal questions. A brief discussion of this method is given below.

Random Response Model

The randomized response model was developed by Warner (41) as a technique to reduce response distortion of threatening or personal questions. By using a probability mechanism, the respondent answers one of two questions selected randomly, but the interviewer does not know what question was answered. Generally, the respondent would answer one of the following questions "Yes" or "No":

I am a member of Group A.
I am not a member of Group A.

By knowing the probability of answering each question, the sample size, and the total number of "yes" replies; the true proportion of the population that are members of Group A can be estimated. Warner feels that the potential advantages of the technique depend on the actual cooperation that is achieved by the model. Warner notes that the randomized response technique can be used to estimate distributions other than a simple dichotomous variable. For example, estimating the proportion of a population in particular income classes can be accomplished by asking the respondent to make five separate randomized responses of whether or not he is in each of the five classes.

Horvitz, Shah, and Simmons (19) suggest that the technique developed by Simmons of using unrelated questions in the randomized response model is a valuable modification. It is felt that this will help to overcome the respondent's suspicions and thus increase cooperation. With this modification, one question should be threatening and the other innocuous and unrelated. The questions should read as follows:

I am a member of Group A.
I am a member of Group B.

With this method, the unrelated question data are treated as a separate sample for estimating purposes. As long as the probabilities are not equal, the sample estimate can be obtained in the following manner:

Given:

π_A = true proportion with attribute A

P_1 = probability that the statement "I am a member of Group A" is answered by the respondent

P_2 = probability that the statement "I am a member of Group B" is answered by the respondent

π_y = true proportion with attribute y

λ_1 = proportion of "yes" answers

The sample estimate can be determined by:

$$\pi_y = \frac{\lambda_1 - \pi_y (1 - P_1)}{P_1}$$

Design of the Study

The study looks at four interview techniques: face-to-face, telephone, self-administered, and the random response model. Figure 1 shows the design including the four levels of threatening questions. It was planned that there would be fifty respondents per cell and therefore a total sample size of eight hundred.

The threat dimension includes questions about the ownership of a Chicago Public Library card, voter registration and voting behavior, involvement in bankruptcy, and being charged with drunken driving. These four topics were chosen because a priori we believed that the level of threat increases as one goes from a question on having a library card to one on being charged with drunken driving. In addition, for these questions it was possible to obtain validation information from public records. Thus, in the results it is possible not only to see what differences there are by method of administration and threat, but also the actual response error.

The respondents in the face-to-face bankruptcy cell had all declared bankruptcy in the recent past. The respondents in the drunken driving cells had all been charged with drunken driving in a time period not less than 6 months or more than 12 months from the date of the study. The respondents in the library card and voting behavior cells were drawn from a random sample and validated from Chicago Public Library and city voting records. This was done after the questionnaire was administered. We ignore in this experiment possible errors in lists of library card holders and voters such as mis-filing or misspelling of names. There is no reason to believe that such errors would be related to method of administration of questions.

It was recognized that our a priori judgement of threat might not be the same as the respondent's. Thus, while we thought admission of bankruptcy to be highly threatening because it is viewed as a personal failure, some people could see it as a shrewd business tactic to alleviate debt. After the main part of the interview was completed, respondents were asked about how threatening they found the questions. The responses to these questions were combined to form an acute anxiety scale. A measure of chronic anxiety was also obtained so that response effect could be related to chronic, acute and total anxiety. Chronic anxiety was measured by the Bendig Short Form (1) of the Taylor Manifest Anxiety Scale (37).

The interviewing was done in Chicago by interviewers trained and supervised by the National Opinion Research Center of the University of Chicago. Interviewer assignments were randomized over the different methods of administration, but interviewers were matched on race.

Interview Completion Rates

The rate of completed interviews varied by method and group. Table 1 shows the percent completed by method and sample. The overall completion rate was 72.2% of 941 interviews. This is about average for a sample in a large city. Use of the telephone achieved the highest interview completion rate. It did better than the other three methods across all sample types except in the bankruptcy sample when comparing it to personal interviewing.

Table 1

Percent of Completed Interviews by Method of Administration and Sample

Methods	Sample			
	General Sample	Bankrupts	Drunken Drivers	Total
Personal Interview	76.0 n=125	70.3 n=54	57.1 n=63	67.8 n=242
Telephone	89.9 n=109	68.3 n=60	77.8 n=63	76.6 n=232
Self Administered	75.4 n=114	59.3 n=59	47.5 n=61	60.7 n=234
Random Response	77.6 n=116	67.2 n=55	58.1 n=62	67.6 n=233
TOTAL	79.7 n=464	66.2 n=228	60.1 n=249	

It is interesting to note that the telephone was relatively more successful in getting completed interviews with drunken drivers. The self-administered technique where the interviewer left a questionnaire and picked it up later did about as well as personal interviewing and the random response model in the general sample. However, self administration did not do nearly as well for the bankruptcy and drunken driving samples where a large fraction of the respondents had not finished high school. The random response method achieved completion rates similar to personal interviewing.

It was much more difficult to locate the bankrupts and drunken drivers and this factor was the major source of not completing interviews with these groups. About 90 per cent of the non-interviews with bankrupts and 80 per cent of non-interviews with drunken drivers were due to the interviewers inability to locate the respondent.

RESULTS

The main findings of the study are shown in Table 2. The data are presented as proportions of distortion for each of the twenty cells representing different conditions. The results represent the proportion of respondents in each condition who gave distorted answers. The proportion of distortion in each cell is defined as the absolute value of the difference divided by the total sample size:

$$\text{Distortion} = \frac{\text{Response - Validated}}{\text{Total N}}$$

Table 2 is organized so that the threat dimension increases for most method conditions from low to high distortion.

In all cases, except the random response-bankruptcy cell (.00), the proportions of distortion increase as threat increases. The threat dimension has been reordered because the voter registration question was least distorting across all four methods. The library card question appeared to be more threatening than voter registration. Voting in the primary election had a higher rate of distortion than the bankruptcy question in all method conditions. (Some of this may have been due to errors in the list of primary voters.) The drunken driving question, with the exception of the random response cell, had the highest distortion rate.

Looking down the columns of Table 2, the method treatment findings are not as clear cut. At the low end of the threat dimension (voter registration and library card), the self-administered method does have a lower distortion rate than the two more personal methods (face-to-face and telephone interviewing). As question threat increases, self-administration does not continue to reduce distortion relative to the other two methods. In the drunken driving cell, self-administered forms had a higher bias than the more personal methods. The fact that the self-administered technique did no better in reducing bias for threatening questions than the other methods is in keeping with Cannell and Fowler's (5, p.254) findings. They found that comparisons of self-administered and personal interviewing data showed no method effect in the reporting of threatening material.

The relation between personal interviewing and telephoning is also interesting. At low threat levels, the distortion on the face-to-face interview is slightly lower than on phone interviews. As the questions become more threatening, face-to-face interviews have larger errors than phone interviews. This supports, although weakly, Hyman's theory that the degree of social

Low threat ⟷ High threat

Highly Personal ⟷ Impersonal

	Library Card and Voting Behavior	Bankruptcy	Drunken Driving	
Face to Face	100	50	50	n = 200
Random Response	100	50	50	n = 200
Telephone	100	50	50	n = 200
Self Administered	100	50	50	n = 200
				n = 800

Figure 1. Study design--question threat by method of administration

Table 2

Proportion of Distorted Responses

Methods	Threat: Voter Registration	Library Card	Bank-ruptcy	Vote Primary	Drunken Driving
Face-to-face	.15 n=92	.19 n=93	.32 n=38	.39 n=80	.47 n=30
Telephone	.17 n=89	.21 n=97	.29 n=41	.31 n=77	.46 n=46
Self-Administered	.12 n=80	.18 n=82	.32 n=31	.36 n=74	.54 n=28
Random Response	.11 n=61	.26 n=61	.00 n=26	.48 n=50	.35 n=23

involvement or "physical presence" of the interviewer can contribute to response distortion. As the interviewer is removed to a telephoning situation and the questions become more personal, the physical absence of the interviewer tends to reduce social involvement and telephoning becomes less distorting than personal interviewing.

The random response model tended to produce higher variances across threat treatments. The range of results ran from (.00) distortion in the bankruptcy cell to .48 in the voting cell. In the highest threat condition, random response yielded the lowest rate of distortion. However, in the March primary voting question the model yielded the greatest distortion. In a somewhat threatening bankruptcy question, the distortion was zero.

The bankruptcy variable is unique in that part of the sample might not have perceived the question as threatening but rather an opportunity to tell of a shrewd business maneuver. This point causes difficulty in evaluating the raw proportions. However, in the drunken driving condition, the model was the lowest in response bias. Initially as one looks across the treatments, it is difficult to evaluate the model's performance in total. The random response technique did not, however, remove all error from the responses.

Another way of looking at the data in Table 2 is to note that response errors may be due either to over-reporting a socially desirable act such as owning a library card, being registered to vote or voting in the primary election, or under-reporting a socially undesirable act such as being involved in a bankruptcy proceeding or being charged with drunken driving. Generally, admitting a socially undesirable act is more threatening, but Table 1 indicates that not voting in the primary election was considered more threatening than having declared bankruptcy.

It is evident that random response procedures are least effective in reducing over-reporting of socially desirable acts. If anything, they are even worse than face-to-face interviews. (The randomized response questions were, of course, asked in the context of a face-to-face interview.) Self-administered and telephone interviews work best on reducing over-statements. The same results were found by Weiss (43).

On the other hand, randomized response procedures are most effective in reducing under-reporting of socially undesirable acts while self-administered forms are least effective, with face-to-face and telephone methods in the middle.

In order to test the statistical significance of the results in Table 2, an analysis of variance was run on the raw data. Table 3 presents the results of these computations. It can be seen that the threat treatment was significant at the .01 level. The method effect, however, was not significant.

Table 3

Analysis of Variance of Proportional Data

Source	Sum of Squares	df	Mean Square	F Ratio
Treatment A (method)	.75	3	.25	.516
Treatment B (threat)	13.40	4	3.35	5.41*
Residual	5.808	12	.484	

*Significant at .01 level

With only one observation per cell it is impossible to determine within cell variance. Kirk (24 p.227) outlines a procedure for handling ANOVA with n=1. Tukey's (39) test procedure for one degree of freedom was modified by Kirk for a factoial design experiment. If Tukey's F-test for non-additivity is insignificant, the interaction term may be used to test the treatment effects. The interaction in the data for this study was not significant.

The data were also transformed using an arcsin transformation. (Winer (46), pp. 399-400) An analysis of variance was run on the transformed data to see if the results differed from those reported in Table 3. Again the method effect was not significant and the threat effect was significant at the .01 level.

Random Response Model

In addition to the results of Table 2, there are other observations that one might make about this relatively new technique. Generally the model was well received by both interviewer and respondent. During the course of the interview, only 5% of the sample who used the random response model said it was confusing, silly, or unnecessary.

The interviewers were asked to evaluate each respondent's reaction to the random response box. Table 4 shows the results of each question by sample type. The general sample and bankrupts appear to follow the same distribution. However, the percentages of "yes" responses for the drunken drivers drops off somewhat.

Chronic and Acute Anxiety

Chronic anxiety was significantly related to response distortion for personal interviewing and self-administration but not for telephoning. The telephone method appeared to be more stable than the other methods across anxiety groups.

One of the principal reasons for taking acute measures was to validate the threat dimension. Generally, the dimension was validated with library card and voting behavior at the low end and court and traffic questions at the high end. Response distortions were significantly different between the high and low acute groups. The acute effect was significant at the .05 level but method of administration proved to be insignificant. There were significant differences between high and medium acute groups with personal interviewing and between high and low acutes using self-administered forms. Telephoning did not show any significant differences between groups that differed in acute anxiety. The distortion rate for the telephone method was also the lowest across acute groups. The random response model produced acute anxiety scores that were generally higher than or close to personal interviewing. Further discussion of these relations is found in Locander (27).

DISCUSSION AND IMPLICATIONS

It is clear from the findings of this experiment that no data collection method is superior to all other methods for all types of threatening questions. If one looks for significant differences, then none of the methods differ among themselves and one is free to use whatever procedure is most convenient.

If one accepts the results at face value, then each of the data-gathering methods is best under certain conditions. The randomized response procedure gives the lowest distortion on threatening questions asking about the performance of socially undesirable acts. It is obvious, however, that one does not always obtain unbiased answers using random response models. The 35 per cent understatement of drunken driving is still a major response bias, although lower than for other methods. The use of randomized response procedures does not permit any multivariate analysis of the relation between the threatening question and independent variables, unless very large samples are screened. For many uses, the loss of information from using randomized response would not be compensated by a modest reduction in response bias.

Self-administered procedures are slightly better than other methods for reducing the overstatements on questions asking about performance of socially desirable acts, but are worst on questions that ask about undesirable acts. In addition, the cooperation rate is lowest for self-administered questionnaires.

There do not appear to be any meaningful differences in response bias between telephone and face-to-face interviews in this experiment except that, for this large city sample, the

Table 4

Interviewer's Answers to Questions About the Random Response Model

Question		Response: General Sample	Bankrupts	Drunken Drivers
1) Do you think the respondent understood the use of the random response box?	Yes	90 %	89.2%	78.4%
	No	8.9%	5.4%	18.9%
	DK	1.1%	5.4%	2.7%
2) Do you think the respondent accepted the explanation of the box and believed that his/her answers really were private?	Yes	92.2%	89.2%	78.4%
	No	6.7%	5.4%	5.4%
	DK	1.1%	5.4%	16.2%
		n=90	n=37	n=37

sample cooperation was highest by telephone. This study again indicates the usefulness of telephone procedures especially in metropolitan areas.

Ultimately the conclusion is that highly threatening questions have high response biases that are not greatly affected by the way in which the question is asked, even if privacy is preserved.

1. Bendig, A.W. "The Development of a Short Form of the Manifest Anxiety Scale." Journal of Consulting Psychology, 20, 1956, 384.

2. Boruch, Robert F. "Assuring Confidentiality of Responses in Social Research: A Note on Strategies." The American Sociologist, 6, November, 1971, 308-31.

3. Boruch, Robert F. "Relations Among Statistical Methods for Assuring Confidentiality of Social Research Data." Social Science Research, 1972, Vol. 1, 4, 403-413.

4. Cahalan, D. "Measuring Newspaper Readership by Telephone: Two Comparisons with Face-to-Face Interviews." Journal of Advertising Research 1, December 1960, 1-6.

5. Cannell, C.F. and Fowler, F.J. "A Comparison of a Self-Enumerative Procedure and a Personal Interview: A Validity Study." Public Opinion Quarterly, 27, 1963, 250-64.

6. Clark, J., and Tifft, L. "Polygraph and Interview Validation of Self-Reported Deviant Behavior," American Sociological Review, 31, 1966, 516-23.

7. Colombotos, J. "Personal versus Telephone Interviews: Effect on Responses." Public Health Reports, 84, 1969, 773-782.

8. Dohrenwend, B.S. "An Experimental Study of Directing Interviewing," Public Opinion Quarterly, 34, 1970, 117-125.

9. Ellis, A. "Questionnaire versus Interview Methods in the Study of Human Love Relationships." American Sociological Review, 12, 1947, 541-553.

10. Enterline, P.E., and Capt, K.G. "A Validation of Information Provided by Household Respondents in Health Surveys." American Journal of Public Health, 49, 1959, 205-212.

11. Eriksson, Sven A. "A New Model for Randomized Response," International Statistics Review, 1973, Vol. 41, 1, 101-113.

12. Ferber, R. "On the Reliability of Responses Secured in Sample Surveys." Journal of the American Statistical Association, 50, 1955, 788-810.

13. Ferber, R. The Reliability of Consumer Reports of Financial Assets and Debts. Studies in Consumer Savings, No. 6, Urbana, Ill.: Bureau of Economic and Business Research, University of Illinois, 1966.

14. Ferber, R., Forsythe, J., Guthrie, H.W., and Maynes, E.S. "Validation of a National Survey of Consumer Financial Characteristics: Saving Accounts." Review of Economics and Statistics, 51, 1969, 436-444.

15. Gray, P.G. "The Memory Factor in Social Surveys." Journal of the American Statistical Association, 50, 1955, 344-363.

16. Greenberg, B., Abul-Ela, A., Simmons, W., and Horwitz, D. "The Unrelated Question Randomized Response Model: Theoretical Framework." Journal of the American Statistical Association, 64, 1969, 520-539.

17. Hagburg, E.D. "Validity Questionnaire Data: Reported and Observed Attendance in an Adult Education Program." Public Opinion Quarterly, 32, 1968, 453-456.

18. Hochstim, J.R. "A Critical Comparison of Three Strategies of Collecting Data from Households." Journal of the American Statistical Association, 62, 1967, 976-989.

19. Horwitz, D.G., Shah, B.V. and Simmons, Walt R. "The Unrelated Question Randomized Response Model." 1967 Social Statistics Section Proceeding of the American Statistical Association, 67-72.

20. Hyman, H. Interviewing in Social Research. Chicago: University of Chicago Press, 1954.

21. Ito, R. "An Analysis of Response Error: A Case Study." Journal of Business, 36, 1963, 440-447.

22. Jackson, R.M., and Rothney, J.W.M. "A Comparative Study of the Mailed Questionnaire and the Interview in Follow-up Studies Personnel and Guidance Journal, 39, 1961, 569-571.

23. Kahn, R.L. "A Comparison of Two Methods of Collecting Data for Social Research: The Fixed-Alternative Questionnaire and the Open-ended Interview." Unpublished Ph.D. dissertation, University of Michigan, 1952.

24. Kirk, Roger E. Experimental Design: Procedures for the Behavioral Sciences, Belmont, California, Brooks/Cole Publishing Company, 1968, 227.

25. Knudsen, D.D., Pope, H., and Irish, D.P. "Response Differences to Questions on Sexual Standards: An Interview-Questionnaire Comparison." Public Opinion Quarterly, 31, 1967, 290-297.

26. Lansing, J.B., Ginsburg, G.P., and Braaten, K. An Investigation of Response Error. Studies in Consumer Savings, No. 2. Urbana, Ill.: Bureau of Economic and Business Research, University of Illinois, 1961.

27. Locander, William, An Investigation of Interview Method, Threat, and Response Distortion, Unpublished Doctoral Dissertation, Urbana, Illinois, University of Illinois, 1974.

28. McCord J., and McCord W. "Cultural Stereotypes and the Validity of Interviews for Research in Child Development." Child Development, 32, 1961, 171-185.

29. Mooney, H.W., Pollack, B.R., and Corsa, L. "The Use of Telephone Interviewing to Study Human Reproduction." Public Health Reports, 83, 1968, 1049-1060.

30. Moors, J.J.A. "Optimization of the Unrelated Question Randomized Response Model." Journal of American Statistical Association, September 1971, 66, 627-29.

31. Price, D.O., and Searles, R. "Some Effects of Interviewer-Respondent Interaction on Responses in a survey Situation." Proceedings of the American Statistical Association, Social Statistics Section, 1961, 211-21.

32. Schmiedeskamp, J.W., "Reinterviews by Telephone." Journal of Marketing, 26, January 1962, 28-34.

33. Suchman, E.A., and Phillips, B.S. "An Analysis of the Validity of Health Questionnaires." Social Forces, 36, 1958, 223-232.

34. Sudman, S., Greeley, A.M., and Pinto, L. "The Effectiveness of Self-Administered Questionnaires." Journal of Marketing Research, 2, 1965, 293-297.

35. Sudman, Seymour. "Toward a Theory of Response Effects in Survey Research." Proceedings of the American Statistical Association, Social Statistics Section, 1970, 14-16.

36. Sudman, Seymour and Bradburn, Norman, Response Effects in Surveys, Chicago: Aldine Press, 1974.

37. Taylor, Janet A. "A Personality Scale of Manifest Anxiety." Journal of Abnormal and Social Psychology, 48, 1953, 285-290.

38. Tittle, C.R., and Hill, R.J. "The Accuracy of Self-Reported Data and Prediction of Political Activity." Public Opinion Quarterly, 31, 1967, 103-106.

39. Tukey, John W. "One Degree of Freedom for Non-Additivity." Biometrics, 1949, 5, 232-242.

40. Walsh, W.B. "Validity of Self-Report." Journal of Counseling Psychology, 14, 1967, 18-23.

41. Warner, Stanley, L. "Randomized Response: A Survey Technique for Eliminating Error Answer Bias." Journal of the American Statistical Association, 60, 1965, 63-69.

42. Wedell, C., and Smith, K.V. "Consistency of Interview Methods in Appraisal of Attitudes." Journal of Applied Psychology, 35, 1951, 392-396.

43. Weiss, C. "Validity of Welfare Mothers' Interview Responses." Public Opinion Quarterly, 32, 1968-69, 622-633.

44. Williams, J.A., Jr. "Interviewer-Respondent Interaction." Sociometry 27, 1964, 338-352.

45. Williams, J.A., Jr. "Interviewer Role Performance: A Further Note on Bias in the Information Interview." Public Opinion Quarterly, 32, 1968, 287-294.

46. Winer, B.J. Statistical Principles in Experimental Design, McGraw-Hill, New York, 1971.

DISCUSSION

D. G. Horvitz, Research Triangle Institute

My evaluation of this paper is, in general, favorable. The authors are to be congratulated for carrying out a type of study which is badly needed by those of us interested in the design of surveys and in the overall quality of survey data. Despite the widely recognized need to be concerned with the magnitude of reporting biases in surveys, significant validation studies have been rather few in number.

The results of this study indicate, at least in part, the magnitude and range of the net bias for questions of a "threatening" nature. The authors prefer to use the term "response distortion" rather than "response bias" and I would appreciate a comment as to why they choose to deviate from usual terminology. Since the estimated "distortions" or "biases" for the random samples (i.e., library card, voter registration, and primary voter) range from .11 to .48 over the four methods of administration, I would be quite concerned about the validity of survey data on these items regardless of the method of administration. The proportions of distorted responses shown in Table 2 for "bankruptcy" and "drunken driving" do not have the same immediate interpretation (of consequences) since the samples for these items were all in the sensitive behavioral category. Thus, if only 2 percent of all drivers have been charged with driving under the influence then an estimate with a .35 reporting bias among "drunk drivers" would lead to an estimate of 1.3 percent in this category among all drivers or a net bias of .7 percent. The relative bias of .7/2 or 35 percent remains large, of course.

While the authors devote a considerable portion of their paper to comparisons of the four methods of administration, I am not certain that very much of their discussion is justified since, as the paper notes, there are few, if any, statistically significant differences. This does not imply that, to quote the paper, "one is free to use whatever procedure is most convenient" end quote. A more appropriate way to proceed, in my opinion, would be to take a total survey design approach in which a portion of the total resources available for data collection is allocated to validating the actual state or behavioral class of a sample of the respondents (provided such validation is possible). Thus, in the case of the item "voting in the most recent primary election," some of the survey budget would be spent on checking official voting records for some random portion of the total sample in order to estimate the net bias in this item. Thus, a two phase or double sampling scheme is envisaged in which a cost function (including the costs of collecting the data directly from a large sample together with the costs of error validation for a subsample) together with the variance function is used to determine the most efficient allocation of resources to the two phases.

This is but a simple example of the total survey design approach in which a balanced allocation of available resources among the various sources of error is sought. Recently, Judith Lessler, completed a thesis concerned with this very problem. She was able to determine, using total error models for both self-administered and interviewer administered instruments, conditions under which one should (a) allocate their entire resources to a single but faulty measurement method or (b) use a double sampling scheme with some resources allocated to "error-free" (or validation) measurements or (c) allocate all resources to "error-free" measurements. The error models used by Lessler included response variance components as well as sampling variance and net bias in the self-administered case and (in addition) correlated response variance terms in the interviewer administered case.

The optimum allocation (or sample sizes) depends, naturally, on costs as well as on the type and magnitude of the various components of error. Clearly, data on the costs of the various alternative data collection methods as well as the biases and error variances associated with each method are essential to achieving optimum or balanced survey designs. Studies such as just was reported here should emphasize estimating the cost components as well as measurement error components.

The comparison (in this paper) of the randomized response (RR) technique with direct questions apparently is based solely on the estimated bias (response distortion). Since the variance of the RR estimate is larger than for a direct question with the same sample size, it would have been better to compare root mean square errors, rather than just the net biases. Or better yet, root mean square errors for a given data collection investment.

The paper does not describe the RR devices that were used. It clearly should in any published version. The paper also attempts to minimize the potential value of RR procedures which I don't feel is warranted. RR as a device for reducing the bias in responses to threatening questions <u>is in the earliest stages of its development</u>. It has had rather varied success in tests by others, some providing very positive and heartening results; others were complete failures or at best not very good.

It is noted that RR provided an exact answer for the bankruptcy item in this study, but an underreport bias of .35 for "drunk driving." A recent validation test by Folsom with a sample of indicted drunk drivers (DUI's) had rather different results, with 15 percent underreporting with the direct question (self-administered) and 28 percent underreporting with RR. The results here today, taken at face value, seem to contradict Folsom's results. Folsom noted that Locander <u>et al</u>. also used the RR device on several less sensitive questions before the DUI trial which may have helped to convince respondents that the device was truly random and did in fact protect their privacy.

Folsom concluded that there is a definite need for more extensive research into alternate ways of asking sensitive questions and I agree. This means testing a wide range of questioning methods, including RR. Coin flip devices which have been tried by several researchers offer some promise. One coin flip version, for example, suggested by Dawes instructs the respondent to say "yes" if the coin comes up "heads" and "to answer the sensitive question" if the coin comes up "tails." There is a large variance penalty with this technique--the variance of the estimate of the proportion with the sensitive attribute is four times the variance for a direct question with no response bias. On the other hand, the respondents know that at least half the time (on the average) there will be a "yes" response which offers quite a bit of protection. [Folsom used a device with 35 red, 4 white and 11 blue beads. Thus the probability was only .22 that a "yes" response refers to the selection of a nonsensitive blue bead compared to .50 with the coin flip.]

Locander _et al_. conclude that "threatening questions have high response biases" and I agree. They also conclude that these response biases "are not greatly affected by the way in which the question is asked, even if privacy is preserved." I cannot agree that this conclusion is warranted from their study. Much more research and testing of alternatives is needed. Ultimately, as I have attempted to say earlier, we need survey designs which will permit the total error to be assessed by the survey data, just as probability samples permit objective assessment of the sampling error. Such survey designs could involve a combination of direct questioning and RR.

We live in a period where concerns about preserving privacy and confidentiality are paramount. As survey researchers we need to be prepared with techniques which will protect respondents to the fullest extent from invasions of privacy. I urge you all to undertake methodological studies--particularly validating studies--of the now very numerous alternative RR procedures for collecting data on personally threatening subjects. These studies are essential since, in my view, without them we may find ourselves in a position of being prevented by law from collecting any data on these subjects in any form or manner whatsoever. This may sound overly pessimistic to many of you. I don't think it is--there is already legislation before the Congress which suggests, if enacted, that the ability to collect important social and behavioral data considered to be of a private nature will be limited in the future.

References Cited

Dawes, Robyn M., "Guttman Scaling Randomized Responses: A Technique for Evaluating the Underlying Structures of Behaviors to Which People May Not Wish to Admit," Oregon Research Institute, University of Oregon, March 1974 (Mimeo.).

Folsom, Ralph E., "A Randomized Response Validation Study: Comparison of Direct and Randomized Reporting of DUI Arrests," Final Report, 255U-807, Research Triangle Institute, August 1974.

Lessler, Judith T., "A Double Sampling Scheme Model for Eliminating Measurement Process Bias and Estimating Measurement Errors in Surveys," Ph.D. Dissertation, University of North Carolina at Chapel Hill, 1974.

DISCUSSION

Roy D. Hickman, Iowa State University

The work by Locander, Sudman and Bradburn represents an interesting and provacative investigation in an area of survey methodology needing more intensive research. I recall a statement made by W. E. Deming many years ago which I paraphrase as closely as possible: "Much research time and effort has gone into the development of methods by which sampling error may be reduced or minimized; however, I am convinced that nonsampling errors (response and nonresponse errors) make up the larger part of the total error incurred in most sample surveys." This paper bears directly on the problem of nonsampling errors, since it considers the effects of question threat and method of administration on response distortion.

It was surprising that the telephone achieved a response rate so much higher than the personal interview. This may have been due to difficulty in "getting in the door" on the personal interviews. Regarding the self-administered technique, I am not sure how level of education may have influenced response rate. It is probable that this differential results from more basic psychological characteristics of the subgroups.

For the methods other than randomized response, it would be interesting to consider the empirical distribution of the response errors, that is, of the persons possessing the characteristic, the proportion who reported they did not, and alternatively, of the persons not possessing the characteristic, the proportion who reported that they did.

We note that the questions were reordered on the threat dimension on the basis of the distortion estimates. For the library and voting items, it is difficult to assess the degree to which errors in the validation procedure may have influenced these distortion estimates. Also, the authors seem to imply that response errors, as they relate to the threat dimension, are conscious and willful errors by the respondent. However, there is evidence that on questions regarding voting behavior, errors may be related more to recall and memory, that is, nonwillful errors. One might hypothesize that degree of recall is related to the threat dimension. For instance, a respondent might incorrectly recall and report voting behavior, but one is not likely to forget a drunken driving citation.

The randomized response technique seems to have been well received, as measured by the interviewer's evaluation of the respondent's reaction. It would have been interesting to have asked questions of the respondent using the randomized response technique itself; it is conceivable that these questions may be, in some cases, rather high on the threat dimension. It would also be informative to know the interviewer's reaction to the use of the technique.

I would like to make a few comments on our experiences with various interviewing methods at the Statistical Laboratory, Iowa State University. We have continued to rely upon face-to-face personal interview as the primary method for data collection. On occasion, we have used the telephone, primarily on reinterviews and on surveys where respondents have been screened on certain characteristics for further interview. As compared to the face-to-face method, the costs associated with such telephone interviews have generally been about half as large, and there has been little difference in response rates. I mention two problems associated with the telephone method: 1) the difficulty of obtaining an accurate and usable frame for sample selection, and 2) the restriction upon length and complexity of the questionnaire.

We have not utilized the methods of self-administration or of randomized response. Self-administration does eliminate potential bias due to interviewer-respondent interaction; however, for attitudinal questions relating to a specific person, there is a real hazard of confounding this individual's responses with those of other members of the household or family. We hope to experiment with the randomized response method in the future, although our reservations are 1) the possibility of an adverse respondent reaction, and 2) the difficulties encountered in performing multivariate data analysis, such as regression, with randomized responses. Nevertheless, we realize that, for certain very sensitive questions, one may be willing to (and should) forego such reservations in order to decrease response bias.

On a survey of farm operators in western Iowa, we are currently tape-recording face-to-face interviews. This study employs a somewhat difficult questionnaire which attempts to assess a farmer's reaction to alternative soil and water conservation plans for his farm. In the pretest, we found that respondents had no negative reaction to the method and displayed little reluctance to the use of the recorder. It remains to be seen how effective the method will be in this particular instance. We foresee some difficulty in the coding of the responses.

Regarding the subject of response errors, we have found that they may be sizable, depending upon the type of questions and the length of time reference. In a recent study of response error and interviewer effect with a survey of farm operators, utilizing a reinterview procedure, respondent response variances for certain questions were as high as 56 percent of the total variation. Of the 21 characteristics tested, eight had respondent response variances in excess of 25 percent of the total variance. Interviewer effects were a negligible portion of total variation.

For the more sensitive questions, we suspect that the response errors may be related to the facility of the interviewers. Less experienced interviewers may be reluctant to ask these types of questions, thus incurring a higher response error and a higher refusal rate.

COMMENTS ON AN INVESTIGATION OF INTERVIEW METHODS, THREAT AND RESPONSE DISTORTION BY LOCANDER, SUDMAN, AND BRADBURN DELIVERED AT THE 1974 ANNUAL JOINT STATISTICAL MEETING OF THE AMERICAN STATISTICAL ASSOCIATION

Charles D. Palit, University of Wisconsin-Extension (Madison)

First, I must congratulate Dr.'s Locander, Sudman, and Bradburn on their paper and its delivery. Their work represents a significant addition to a growing, but still inadequate, list of investigations on the reaction of respondents to survey questions. The results of such explorations help to formalize and clear up the thinking of the builders of interview schedules or questionnaires and the designers of studies in general. The basic findings for the study:

I. that distortion is relatively independent of the method used to deliver the question; and

II. that (high) levels of threat in a question brings about distortion (otherwise known as bias) in the respondent's replies

represent in the first case a valuable addition to the body of information we must draw upon when designing a study and in the second case confirmation of what I know to be part of the lore of survey research.

At the same time, this study confirms another bit of lore about threatening questions and that is that it is sometimes difficult to determine a priori what questions are threatening to the public and what questions are not. The authors' post-study reordering of their threat scale is just one more example of this difficulty. I agree with their reordering of the library card and voter regulation variables, even though this reordering might seem to some to be an adjustment of the data to fit the hypothesis. It is, after all, a very minor switch. I do, however, have some difficulty with the placement of the "voted in the primary" variable. The magnitude of the reordering seems too large. I would wonder if the authors' alternative hypothesis as to the cause of the large distortion levels, i.e., poor primary voting records is not a more viable one, and if this were indeed the case, whether it should not be dropped from the analysis. However, it is entirely possible that what made this variable (primary voting) more threatening to the respondents was the context in which it appeared. In the appropriate context, it might be a very threatening question indeed.

As an example of threat being modified by context, I can quote some informal pretest results for the cover sheet format on a telephone interview using what has come to be known as a random digit sample. Our first question on the cover sheet is..."Is this telephone number (GIVE TELEPHONE NUMBER LISTED ABOVE)?". A "yes" response is followed immediately by the question: "Is this a residential number?". Then in the City of Milwaukee at least the response rate will be somewhere between 10 and 15 percent worse than if the cover sheet format is...

1. Is this telephone number ________?

↓ Yes

Good, I am Mary Smith from the University of Wisconsin's Survey Research Laboratory and we are conducting a survey...

2. Is this a residential number?

The brief introduction apparently removes some of the threat associated with the question: "Is this a residential number?". The respondent then becomes willing to continue the interview.

Moving on to the method study, I find it surprising that the randomized response method does not perform better. I am equally surprised that the authors spend so much time discussing the method effects only to inform us in the end that there is no statistically significant method effect. My own feeling is that the randomized response method fails to show a statistically significant effect because of the small sample size involved, and because of the extra noise introduced by the "Vote Primary" variable. If this variable were dropped, I suspect we would have a significant method effect. It is worth noting that if we treat the distortion measure as a binomial variable (which is certainly a correct treatment except perhaps for the randomized response case) and analyze the bankruptcy and the drunken driving columns separately, we find "statistically significant results."

Turning to the response rate results, it is indeed gratifying to see that the telephone produces the highest response rate. This is in accord with our own and other informal observations and again it is worth noting that under random sampling assumptions, the difference is statistically significant. It is nice to have the hypothesis that telephone interviews are less anxiety-producing than face-to-face interviews confirmed.

Reports on mail questionnaires to the general population show in many cases poor response rates. I have always felt that the problem with such mail surveys has been the inability of the sampler to designate a specific respondent. These results help to reinforce this suspicion.

I am, however, dubious about the authors' claim that poor performance of self-administered questionnaires with bankrupts and drunken drivers was due to their lack of education and hence, reading and writing ability; especially in light of the authors' later claim that 80 to 90 percent of the non-response in these classifications was due to the inability of the interviewers to locate the respondent. Without a knowledge of the percent of non-response which is in the "cannot locate" category for the other methods, it is, of course, impossible to reach a hard conclusion. However, I wonder if the drop in the response rate is not an interviewer effect!

Finally, in closing, I would again like to congratulate the authors on their stimulating paper and would hope that their results would lead to a greater acceptance of mixed mode surveys. I am sure that the practictioners of the art here recognize the great utility of mixing face-to-face interviews with telephone interviews as a means of dealing with response problems in metropolitan areas; particularly in dealing with high-security apartment buildings. A method of operation which the results of this paper would indicate as practical.

THE RIGHT OF PRIVACY AND THE NEED TO KNOW

Vincent P. Barabba, Bureau of the Census

The term "rights" is used a lot these days. It is tossed around in newspaper headlines, by TV commentators, by protest groups, and from the pulpit. We talk about the rights of minorities ...of women...of unborn children...of future generations...and increasingly, of the rights of the individual in today's fast-paced, crowded world.

To me, the nostalgia craze, and the growing popularity of getting back to nature indicates that some of us long for a time when it was easier to be a self-sufficient individual.

Nostalgia is fun---and it's also refreshing to open your eyes to nature. But it's naive to think that we can turn back the clock to a simpler time. What we need to do is to come to grips with today and define the problem---which is how to maintain the dignity of the individual while keeping order, and, at the same time, develop and maintain programs to improve our quality of life.

That's what I want to talk about today: the rights of the individual in our society, specifically the right of privacy and its relationship to the legitimate need of government to have information about the individual.

The specter of government intrusion into the affairs of individual citizens has always been a highly emotional subject---and unfortunately, one mainly discussed in the framework of newspaper headlines. Events of recent years have brought the issue into sharp focus: military surveillance of civilians, wiretapping, the bugging of offices and industrial and political espionage.

My files are crammed with clippings on the subject, with headlines such as-----

"Reversing the Rush to 1984."

Or---"Big Brother Society Feared."

And---"Federal Data-Gathering Like Octopus."

I see two types of anxiety reflected in these headlines. One is on the part of the individual, the other by society in general.

Let's look briefly at society as a whole. Under this heading, there are two general concerns. First, there is the concern flowing from the myriad of advertising and mass media campaigns.

By the late 1950's, computers and television had become part of the everyday scene in America. Vance Packard's book The Hidden Persuaders crystalized a fear that had been forming for some years---the fear that increased knowledge about society in general would lead to the ability to manipulate the public mind. Words such as "brainwashing" and "subliminal advertising" were current on the cocktail party circuit.

This fear was put into perspective shortly afterward by Raymond Bauer. He said the fear of manipulation is an old one on the part of society, and pointed out that superstition has been replaced by fear of the unknown potential of new technological inventions.

Bauer's main point was that those who would desire to manipulate the public mind are always one jump behind the public mind. In other words, as the potential for manipulation increases, so does our sophistication, and with it, our resistance level to persuasion. An example of growing resistance is the consumer movement---who could have foretold such a powerful grassroots phenomenon just a few years ago?

The most realistic description I have seen of the capability of mass media to manipulate society came from Bernard Berelson. He said: "Some kinds of communication, on some kinds of issues, brought to the attention of some kinds of people, under some kinds of conditions, have some kinds of effects."

While thinking about this subject, it occured to me that in just ten years it will be 1984. Nineteen eighty-four----and George Orwell's powerful work about what totalitarianism can do to the human spirit. The very title has come to be a shorthand way of referring to anything which infringes on the rights of the individual. The relationship between government and society it describes is probably the ultimate example of manipulation.

While the manipulation Orwell envisioned was brutally direct, there is also a subtle variation of this fear. This second concern is based on the fact that dictators throughout history have provided diversions to keep the minds of the masses off their real troubles, so it follows that a government---utilizing statistical information---could cater to public opinion on certain emotional issues, leaving it free to pursue its real aims, which might be counter to public interest.

But this theory ignores one fundamental truth. In America, the thrust of the government's statistical programs has been to provide increased amounts of summary data to all groups within our society at a reasonable cost, whatever their political philosophy.

I think the other anxiety in the headlines I mentioned is more valid and of much greater concern to the average person. That's the fear of misuse of personal information----that information gathered for a legitimate purpose will be used later in a different context which could injure either the individual or his family.

This fear was summed up very well by the noted Russian author Alexander Solzhenitsyn in his novel, Cancer Ward. He wrote:

> "As every man goes through life he fills in a number of forms for the record, each containing a number of questions...there are thus hundreds of little threads radiating from every man, millions of threads in all. If these threads were suddenly to become visible, the whole sky would look like a spider's web....They are not visible, they are not material, but every man is constantly aware of their existence. Each man, permanently aware of his own invisible threads, naturally develops a respect for the people who manipulate the threads."

This fear of the misuse of personal information is exaggerated by the popular image of the computer. That image often casts the computer in the role of a villain. It becomes the tool of the all-pervasive, yet unidentified "they." Whenever we find fault with some action of government, business, school, or any other segment of society, it's always "they" who did it, and increasingly the computer is blamed for making it possible.

The facts, however, are clear: the computer has made a profound contribution to the public good---and done it so well we take it for granted. Just one example is the millions of checks Social Security recipients get each month. Those checks simply wouldn't arrive as fast or as accurately with manual processing.

In addition, the public generally over-estimates the abilities and the applications of computers---thanks in part to spy movies and television. Together with recent headlines, this image of the computer has led to vague fears of an ominous National Data Bank, which would store every facet of our personal lives for instant retrieval by any government agency which requested information.

For these two fears I have described to be realized in the United States---for our society to move that close to the nightmares described by Orwell and Solzhenitsyn---America would have to abrogate not only current law, but its entire democratic tradition.

Let me underline this thought by quoting to you a prophetic portion of a paper written by Otis Dudley Duncan which concerned plans for the 1970 census:

> "...in this country we have proved that a statistical system can incorporate rigid safeguards of confidentiality. The institutionalization of these safeguards has proceeded to the point where it is inconceivable that they would break down, except in the catastrophic event of a breakdown in our whole system of institutions protecting the rights of the individual. In the case of such a catastrophe, my guess is that much more direct ways of infringing these rights would be found than that of making inappropriate use of statistical records secured ostensibly in confidence.

Many of the headlines I mentioned before are followed by articles which list the Census Bureau in their catalog of information-gathering agencies. So far the Bureau has not been identified as a culprit, but some of the articles leave the impression that it has the potential to be one.

My concern is that as real culprits are identified, the Census Bureau will be tarred with the same brush---and that this will occur without an understanding of the vital role the Bureau plays in the Nation's decision making process, or of its standards in regard to privacy and confidentiality. I am also concerned that lawmakers, in trying to guarantee that personal information is safeguarded, will overreact, with the result that legitimate data-collection machinery will be hampered.

Because of these concerns, I appreciate the opportunity to share my thoughts with this audience---which represents such an influential group within the statistical community. I would like to review for you just how the Bureau looks on the rights of privacy of the citizens from whom it collects data, and how we keep confidential the information we do collect.

We at the Census Bureau believe our mission is to gather accurate, timely, and complete data from individuals, businesses, and governments, and to make available to the public general statistical summaries of that data.

But there is an inherent conflict in gathering data from individuals. That conflict is between the individual's right of privacy on the one hand, and, on the other, government's use of mandatory processes to obtain the information it needs for valid purposes.

Basic to this discussion is the question: what is the right of privacy? It is a very easy term to use, but a very difficult one to define.

American legal and academic scholars have wrestled with the problem for the better part of a century. Several West European nations have commissions reviewing present privacy safeguards, and what is needed for the future. Sweden recently enacted a comprehensive data act. And several West German states have ombudsmen whose job is to report to elected officials on problems of individual data privacy and confidentiality.

Also in Germany, psychologists have offered the definition of privacy as those areas of people's lives in which they can act without fearing that information may be passed on in a way disfunctional to themselves. Put another way, it is the concept of selective transmission of information, be it to family, friends, one's doctor, or a government organization.

In attempting a definition, let us assume that privacy does not mean a sacrosanct area where no questions can be asked. Rather, we have a situation in which questions may be asked on a voluntary or mandatory basis. And, the individual believes the answers he gives will not be used for purposes other than those described.

This "security" of belief on the part of the individual, when coupled with a promise from the receiver that the requested information will only be used in specified ways, creates a bond between the two. I would call this a "confidential relationship."

The Most Reverened Mark J. Hurley, writing for the Knights of Columbus, has defined confidentiality as "secrets"---two types of which are the "committed secret" and the "promised secret."

The committed secret is one kept by reason of a tacit agreement between the parties that the information will not be divulged. Examples would be secrets between lawyer and client, and doctor and patient.

The promised secret is one kept by virtue of a promise made prior to learning the secret. Such a promise might be our Bureau's pledge that responses to census questionnaires will be seen only by sworn employees of the Bureau. Another would be the presidential proclamations defining the nature of the census and stating the sort of activities for which individual data will not be used.

Here it is important to distinguish between personal information gathered for statistical purposes as opposed to that gathered for administrative purposes. The information may be the same in both cases, but administrative records are intended to affect the individual directly---for instance, those used by the Internal Revenue Service, or the Social Security system. Statistical records---such as those maintained by the Census Bureau---do not deal with the individual directly when used only for the compilation and analysis of summary data.

Obviously, privacy does not exist in an absolute sense, any more than freedom does. As Justice Oliver Wendell Holmes said in his celebrated opinion: "Freedom of speech does not include the freedom to yell 'fire' in a crowded theater."

Privacy, as freedom, has meaning only in the context of human society, and society changes as time passes. As our society becomes more complex, we need to know more about ourselves in order to establish priorities and properly allocate our human, financial, and natural resources.

In the rural life style of 1790, for example, it would have been hard to justify the government's interest in whether a household had its own bathroom facilities. Today, with society's commitment to eliminate slums and sub-standard housing, that information is needed to identify the number and location of such housing before millions of taxpayers' dollars are spent on such programs.

To contend otherwise is to say that the interests of the individual transcend the interests of society, and have priority over public efforts to eliminate sub-standard housing.

The question about complete plumbing is just one of the many well-publicized examples which critics of the census have used. Taking a question out of context to imply invasion of privacy is the most persistent technique for criticizing the Bureau, and the most intellectually dishonest.

The right of privacy is often expressed as "the right to be left alone." But that concept is inconsistent with the individual's responsibility to society.

Each man, woman, and child in our society reaps benefits from being a member of that society. Of course, these benefits vary from place to place and within the subgroups of our society. Yet the individual obviously derives benefits from dwelling among his fellow beings.

It is axiomatic that we never get anything for nothing. What, then, is the trade-off when it comes to the individual and society? The obligations of an individual living in a highly complex, densely-populated industrial civilization are greater than any in history. Sometimes the price the individual pays is in money---such as taxes; in other cases, it is time---such as jury duty, a jail sentence, or duty in the armed forces when required. Sometimes it is establishing qualifications to do certain things---such as driving a car, or practicing certain occupations.

If we grant that we all operate in the context of human society, and that we have a responsibility to that society, we can arrive at a definition of the right of privacy along these lines: it is the right of the individual, to the extent possible, to control what information about himself he releases, to whom he releases it, and under what conditions.

All of which is a roundabout way of saying there is a right of privacy, but it is a right which may be circumscribed to allow the expression of other freedoms. Obviously, any limitation of our right of privacy must be made with extreme caution and only after careful consideration of the consequences. That then is our understanding relative to the right of privacy. What then about confidentiality?

In his role as Chairman of the Subcommittee on Constitutional Rights, Senator Sam Ervin has said "Somewhere a balance must be struck between the individual's desire to keep silent and the government's need for information. If it is proved necessary to invade certain rights, clearly it is the constitutional duty of Congress to establish precisely how and under what circumstances this may be done."

Congress has been doing exactly that for almost a century. Since the act which provided for the 1880 census, laws protecting the confidentiality of information given in response to census questions have been progressively tightened.

Up until 1910, census law required the Director to furnish on demand to governors or heads of municipal governments certain parts of an individual's return---the name, age, sex, birth place, and race.

The act for the 1910 census changed that wording to read that the Director could---at his discretion---furnish information for genealogical and other proper purposes.

1910 also marked the start of another tradition ---the presidential proclamation. The one issued by President Taft told the American people their replies to census questions were to be used only to compile general statistical information, and that their answers were protected by law. In part it read: "The census has nothing to do with taxation, with Army or jury service...or with the enforcement of any National, State, or local law or ordinance, nor can any person be harmed in any way by furnishing the information required."

The current law under which the Census Bureau operates is Title 13 of the U.S. Code, most of which dates from 1929. This law is very specific when it comes to personal information. It requires that information obtained from an individual be used only for statistical purposes. It also requires that published data be in such a form that it is not possible to identify an individual or a single business establishment. The law stipulates that no one other than sworn officers and employees may have access to individual information, and each census employee has signed an affidavit of nondisclosure to uphold the law.

The current law still has wording much like that of 1910, which allows the Director at his discretion to provide copies of individual information for genealogical and other proper purposes. The key word here is "discretion." Over the years the application of this rule has become restrictive rather than permissive.

In current Bureau practice, the term "confidentiality" represents nothing less than a clear extension of an individual's right of privacy. I think the best way of showing this is to review the Bureau's track record regarding the confidentiality of individual data.

Most people who follow the Bureau's activities closely have assumed that data from individuals have been held in strict confidence at all times. I must report that this has not been always the case. But looking at the way we did things in the past, and comparing them with today's practices, makes me even more certain that our current position is a very strong one.

Between 1900 and the mid-1920's, there were authorized releases of individual data considered proper that today would cause a storm of protest in the press, in the courts, and in Congress. As far as we know this practice caused no such outcry then. I say as far as we know because complete records do not exist.

We do have some information on one case which demonstrates the type of situation in which it was considered proper in the past to release data about individuals. This occurred in 1918, during World War One. Congress had passed a War Powers Act, and presumably this was the basis for such an extreme use of census data. Information about individuals was given to the Department of Justice for use as evidence in prosecuting young men who claimed they were too young to register for the draft. While we do not know the exact circumstances surrounding the release, we do know that personal information for at least several hundred young men was released to courts, draft boards, and the Justice Department.

The Bureau stopped such releases during the 1920's, a position which was made official in 1930 by an opinion from the Attorney General. His opinion said that even the name and address of an individual is confidential.

Now we jump to 1941. It's hard to imagine now, but with World War Two underway, there was near hysteria about the Japanese-Americans living on the West Coast---emotion which led to one of the most embarrassing moments in U.S. history, the internment of large numbers of these loyal Americans. At the height of this feeling, the Secretary of War requested that the Census Bureau supply the names, addresses, and ages of all persons of Japanese extraction living on the West Coast. This time---in spite of the national emergency---the Bureau held to its position on confidentiality of individual records and refused. The Bureau did supply summary data at the tract level, which is now part of the regular publication program.

In 1947, during the rising concern about possible communist infiltration and sabotage, the attorney general requested information about certain individuals in census records on behalf of the FBI. Again, the request was denied.

A loophole in the law turned up in a case in the early 1960's when the courts ruled that file copies of census forms not kept by the Bureau could be subpoenaed. This resulted in Congress amending the law to extend confidentiality to include even copies of census questionnaires which are kept by businesses for their own files.

That briefly is a summary of how confidentiality grew to be an integral part of census taking. Keeping that information in mind, and my earlier remarks about generalized fear of the computer, let's look at how a modern census is processed.

After all the forms are collected, the data on them must be transferred to computers. It used to be that the data on each form were manually transferred to punchcards, and the punchcards fed to computer tape. Now, we bypass this laborious process. The forms are microfilmed on highspeed page-turning machines and returned to storage. This is the last time each original form is handled until it is destroyed.

The first page of the census form is not microfilmed. This page has the address of the household. So---the rolls of microfilm, which have names and personal information, contain only a geographic code relating that information to the block on which the household is located.

Another sophisticated piece of machinery reads the microfilm and transfers the dots that originated with the citizens' pencilled-in circles directly onto computer tape. This machine cannot read handwriting, so the personal information about individuals is separated from their names at this point for the rest of the tabulation process.

Even this is not enough to guarantee that a person could not be identified in the statistical summaries. Some areas have such a small population that it would be possible by deduction to know whose characteristics are in the tables. Our computer program is set up so that if this would be the case, that information is suppressed---both on computer tape and in the printed publications. Some analysts---probably some of you in this audience---have had problems with these suppressed figures, trying to add up to the tract level from block data, for instance.

We are examining other techniques for protecting confidentiality. These include rounding numbers to the nearest five, and a "random noise" system, in which values of one and negative one are scattered throughout the tabulations, balancing to zero at certain geographic levels. Such a system would have no substantial effect on statistical analysis.

When it comes to suppression of data from the economic censuses even the cutoff points are confidential---because that information by itself could be used for deduction, since the numbers involved are so much smaller that population figures.

I also should mention here the Public Use Sample, which we established in 1960. These are not summaries, but individual census records minus certain data to ensure that the individual cannot be identified. The smallest area description for which these records are available is 250,000, and even then certain data have been truncated to avoid identification. An example would be extremely high salary figures, for instance. Everything over $50,000 is simply marked "$50,000 and over."

These samples have proven to be of great value to the academic and business communities for research, and for determining if special tabulations would provide the summary data desired.

Another example of the Bureau's efforts to strike a balance between increased usefulness of our product and at the same time maintain confidentiality, is the GBF/DIME system. G-B-F stands for Geographic Base File, DIME for Dual Independent Map Encoding. This is an automated file containing address ranges along streets in metropolitan areas. Essentially it is an automated map. Used in conjunction with other computer programs, this file can blend locally-gathered data with census data and provide information for local problem solving and planning. Where do the disabled in a city live? Where is the incidence of a particular disease highest? Do most runaway children come from certain neighborhoods? The applications are endless.

The GBF/DIME system can do this and still maintain confidentiality. Even though the process takes the actual address and assigns it to area units such as blocks or tracts, it eliminates the address itself, and also any information that would identify the respondent. In other words, instead of being the culprit, the computer in this case literally supplies anonymity by converting personal information to area information.

When the tabulation of a census is finished, the original paper forms which have been stored in guarded buildings on a government facility are destroyed. They are shipped in sealed boxcars and recycled, with Bureau officials watching until they drop into the pulping vats.

That leaves the microfilm. Where does it go after we are finished processing the data? The rolls are sent to the Personal Census Service Branch in Pittsburg, Kansas, which we commonly refer to as the Age Search Service. This is a unique self-supporting operation which has helped millions of people. Every day the Bureau receives about 1,300 requests from people who need to verify some item of information about themselves. Most are for substitute birth certificates which either never existed, or have been lost or destroyed. People need them to qualify for retirement, for Social Security, for Medicare, to get a passport, and many other uses. For a very small fee, the Age Search Service will search old census records and issue a certificate which has legal standing.

This service is provided only at the request of the person himself. For example, a son cannot ask about his father unless he has a power of attorney or a death certificate. This operation is the only use made today of the Director's authority to release personal information at his discretion.

Finding this type of information for those who request it is not an easy job. It takes an expert to utilize the microfilm. Since the census is based on addresses, not names, there is no such thing as a master list of records arranged alphabetically by name. For the correct reel of film to be located, the person making the request must supply information about where he or she lived at the time the census was taken.

The very size of the U.S. population helps to guarantee confidentiality. It took some 5,000 miles of microfilm to process the 1970 census. For us to make this process of working backward any easier would be extremely costly, and would in theory, weaken the protection of confidentiality.

Now---where does the microfilm of past censuses exist? The records of the counts from 1790 through 1880 are accessible to the public in the National Archives. Data in these enumerations were not gathered under laws of confidentiality. The census of 1890 is almost non-existent, having been mostly destroyed by fire.

The Pittsburg, Kansas unit has microfilm for 1900 through 1960, and late this year will have the 1970 records set up to be able to answer the requests which are already coming in.

Copies of the 1900 through 1950 records are also held by the National Archives. These were sent to the Archives for storage. But in December of last year, the Archives opened the 1900 census to limited access by qualified researchers, a move opposed by the Census Bureau. While this access is under controlled circumstances, we feel at the very least it violates in principle the rights of the estimated seven million persons still alive who were counted in the 1900 census.

The law under which the Archives operates says government records may be made public after 50 years, unless an interagency agreement stipulates a longer period of time. In 1952, the Director of the Census Bureau and the Archivist agreed that census records should remain closed for 72 years---or the average lifetime. The Bureau's position is that the 1952 agreement was in excess of the Director's authority, and therefore is invalid.

The obvious question is---how long does the law's guarantee of confidentiality apply? A lifetime? One hundred years? Or forever? The Bureau hopes Congress will close this final loophole in the laws of confidentiality.

Congress at the moment has a lot to consider in the area of privacy and confidentiality. Some 60 bills are pending in the House and Senate. The basic question seems to be not whether something must be done to insure privacy and protect it, but what, and by whom. Four of those bills deal with census information. Eight of them would establish a Federal Privacy Board or some committee or commission as an overall authority. Many of the bills would allow the citizen the right to inspect his own records, correct them, and bring suit for damages resulting from incorrect or misused records.

In his first speech before Congress, President Ford made it clear that privacy is very much on his mind. He said: "There will be no illegal tapings, eavesdropping, buggings or break-ins by my Administration. There will be hot pursuit of tough laws to prevent illegal invasions of privacy in both government and private activities."

You might well ask why we at the Census Bureau are so concerned, if our record is good and our intentions are clear. Aside from the moral implications, I'll give you a very practical answer---one which this audience especially should appreciate.

A census or a survey is only as good as the contract of trust between the people about whom information is obtained...and those with the mandate to obtain it. If the public feels we are not keeping our word that their answers will be kept confidential---or that even the potential for such violation of their trust exists---their answers will not be as accurate, or given as willingly.

If this occurs on a large scale, the quality of the summary statistics will deteriorate. And if this occurs the Nation has lost its prime decision making tool, and society will be the loser.

This would be a tragedy. It would come just as more and more decision makers in the public and private sectors are becoming aware of how valuable census data is to them---and it would come as the Bureau's main thrust is to increase the utility of the data it gathers.

To the Census Bureau, a promise is a promise. The calendar may read 1984 in ten years, but I want to make it clear that as far as the Bureau is concerned, George Orwell's 1984 will never come.

In his remarks to the most recent meeting of the Advisory Committee on Privacy and Confidentiality, Commerce Secretary Frederick Dent said: "I think perhaps the strongest brand name in America might be that of the Census Bureau."

Ladies and gentlemen of the A.S.A., you can rest assured...at the Census Bureau...we plan to keep it that way.

THE UCR PROGRAM: DEVELOPMENT OF A STANDARDIZED AUDIT [1]

Donna F. Brown, International Association of Chiefs of Police

I. INTRODUCTION

The need for amassing a reliable fund of criminal statistics for use in law enforcement administration, operation and management has never been greater. In order to deploy manpower and allocate resources to combat crime effectively, the nature and scope of the nation's crime problem must be known. The national Uniform Crime Reporting (UCR) Program was established to achieve this goal.[2] The Program is currently used to determine:

- the extent, fluctuation, distribution and nature of serious crime in the United States through presentation of data on seven Crime Index offenses
- the total volume of serious crime known to police
- the activity and strength of law enforcement agencies through arrest counts and police employee data[3]

The idea for systematic crime reporting is not new. At the 1871 convention of police officials held in St. Louis, a resolution was adopted which declared that the newly-founded National Police Association "procure and digest statistics for the use of police departments."[4] Over half a century passed, however, before significant action was taken to achieve this goal.

In 1927-1928 Commissioner William P. Rutledge of Detroit chaired the International Association of Chiefs of Police Committee on Uniform Crime Records and, with financial assistance from the Rockefeller Foundation, appointed an Advisory Committee and Technical Staff to develop a system for the collection of crime statistics. Results of these efforts included basic definitions of the nature of the data base (a count of offenses known to police and of crimes cleared by arrest or by exception) and a definition of Part I offenses (those classes of serious crimes which are brought to the attention of the police as a matter of routine). In addition, methodology for collecting these data as well as proposals for the design of basic report forms were decided by the Technical Staff.

Based on the program and suggestions developed by IACP's Committee on Uniform Crime Records, Congress authorized the Federal Bureau of Investigation to collect and compile national crime statistics. In September of 1930, the national Uniform Crime Reporting Program was begun.

Selection of the FBI as the national clearinghouse for the monthly statistical submission was a natural one since this agency was already operating a nationwide identification service in which thousands of local law enforcement agencies were participating. In addition, the network of Special Agents already in close contact with local law enforcement agencies and staff would be of value to offer personal assistance to local officials in preparing reports and revising records systems.[5]

The FBI has consistently emphasized the need to enlarge the crime reporting area to include urban communities, sheriff's offices and state police organizations. In the early years, however, the program was hampered by a lack of widespread participation: in January 1930, for example, the monthly report included only reports of "offenses known" in 400 cities. By 1938, however, 4,283 agencies were contributing crime reports. In 1973, the program had expanded to include 11,000 reporting jurisdictions which accounted for 93% of the total national population.[6]

Insuring quality control over these statistical submissions has been a matter of concern to those who manage the program and to those who utilize or are affected by its reports. As early as 1930-1931 field work conducted by various organizations[7] revealed that the quality of the UCR submission was directly related to the quality of the reporting agency's records system and that in addition, the establishment of uniform reporting standards for scoring and classifying Part I data was of immediate importance. It was readily apparent that law enforcement agencies must first record and maintain certain data in order to generate reliable information for the UCR Program. Recognizing this fact, the Committee on Uniform Crime Records designed record forms based on the experience and need of law enforcement organizations, and also compiled detailed explanations as to the manner in which these records might be maintained to insure the compilation of accurate statistical reports.

High reporting standards have consistently been encouraged by the IACP Committee on Uniform Crime Records, which continues to serve in an advisory capacity to the FBI in the administration of the UCR Program. Likewise, the National Sheriff's Association in 1966, established a Committee on Uniform Crime Records to serve in an advisory role to its membership and to encourage full NSA participation in the UCR Program.

Realizing the importance of a good records system the FBI developed, in 1955, a Manual of Police Records designed to assist in the improvement of the law enforcement agency's basic records procedures. To insure adherence to national UCR definitions and guidelines, the FBI provides copies of its Uniform Crime Reporting Handbook which outlines procedures for scoring, classifying and clearing Part I offense data. In addition, the FBI offers, upon request, in-service training in national UCR procedures to agency personnel responsible for the compilation of UCR data.

Recently the FBI has actively assisted individual states in the development of statewide programs designed to collect UCR statistics that are compatible with the national system. As a result of this activity more law enforcement agencies are participating in the UCR Program and the completeness and quality of the information submitted have improved as well.

Recognizing that the quality of the monthly statistical submission depends on a wide range of variables including:

- accuracy of the agency's reporting procedures
- organization of the agency's incident reporting system
- internal and external pressures to reduce crime
- human and mechanical error in report compilation
- the level of experience and training of personnel responsible for the aggregation and submission of UCR data

the IACP, in cooperation with the FBI and through a grant funded by the Law Enforcement Assistance Administration (LEAA), is designing a system of audits for the purpose of confirming the validity of UCR reporting practices.

II. IDENTIFICATION OF AREAS OF UCR ERROR POTENTIAL

The first logical step in conducting an audit of the UCR statistics (which are simply a byproduct of the overall incident reporting system) is a thorough examination of the reporting practices and procedures which produce these data.

For purposes of analysis the law enforcement agency's incident reporting system may be divided into three distinct phases:

Data Capture -- a three-stage process in which crime event data become known to police through report or discovery; the police make some disposition of the information received; and the police record or fail to record this information or describe the action taken;

Data Review and Verification -- a verification process intended to insure full and accurate reporting and recording of the crime event data previously entered in the incident reporting system;

Data Aggregation -- a mechanical process which includes extraction of UCR data from police records, compilation of these data, preparation of the appropriate UCR forms and forwarding all data to the respective state collection agency or Federal Bureau of Investigation.

Within each phase several areas of error potential have been identified as being particularly vulnerable to the loss or distortion of UCR data.

A fundamental prerequisite for an efficient incident reporting system is the early capture (both by telephone tape recording and in written format) of all complaint data. The possibility exists that the most subtle and deliberate manipulation of crime event data takes place in the field during the on-view discovery of a crime and at the telephone input mode. Since the vast majority of all complaints becomes known to police through these two modes of entry and since events lost during this phase are least likely to be discovered through routine supervision or periodic inspection, proper control to insure data retention at this stage cannot be over-emphasized.

The accurate recording of all crime event data known to police agencies is the ultimate aim of the data capture phase. If an adequate determination is to be made regarding the accuracy and completeness of an agency's UCR statistics it is necessary to document the nature and disposition of all reports of alleged crimes and crime discoveries regardless of the nature of the police response to these events. Without an adequate record or source document, filed in some reasonable fashion to permit easy retrievability, the accuracy of the UCR submission cannot be verified.

Following the capture and documentation of offense and clearance data a process of review and/or verification is conducted in many agencies to insure the validity of the complaint information. This procedure is one of quality control and must be accomplished before the statistical data are extracted for submission to the UCR Program. If the raw data captured in Phase I of the incident reporting system are immediately translated into UCR statistics without benefit of operational controls, the quality of the agency's submission can only be suspect. It is also at this stage that errors in scoring and classifying UCR data occur. Additional attention must therefore be given to this aspect of the review process to insure that such procedural errors are minimal.

Once crime event data have been reviewed and verified as Part I offenses or clearances which should be included in the monthly UCR Return, it is the responsibility of the law enforcement agency to extract and compile these data for submission in the national program. Error potential in this phase stems not only from the manual and/or mechanical systems responsible for processing the data, but also from intentional omissions. This phase is particularly crucial to the integrity of the overall reporting system since, in contrast to the admittedly vulnerable data capture phase where one error probably loses only one data element, one error or miscalculation in the aggregation phase may result in the loss or distortion of large amounts of data, much of which is irretrievable.

Once the major areas of vulnerability which contribute to UCR data loss or distortion have been identified and corrected the agency will be in a position to implement (as part of an internal audit/inspection procedure) or request (as administered by an external agency or group) an audit to confirm the validity of its UCR reporting practices.

III. UCR AUDIT TECHNIQUES

In order for a successful UCR audit to take place there must exist a taped record of all incoming complaints or calls for police service and a source document (e.g., complaint card, dispatch ticket, incident/offense report, blotter entry, log entry, activity sheet, etc.) which contains basic information describing the alleged offense and the complainant. In addition, these items must be filed in such a way as to permit retrievability. Finally, the agency must maintain a tally or register of Part I offenses and clearances (which are recorded according to specific offense number) to permit verification of the totals which are submitted on the UCR Return A. Without these basic tools there exists no identifiable documentation of the complaint and therefore a comprehensive audit of the UCR data is not possible.

It is recognized, however, that the incident reporting systems of all police agencies cannot be audited utilizing the same techniques. It is therefore recommended that a preliminary audit be conducted in order to identify the mechanical process, system, and capability of the agency to effectively capture, record, process, store and retrieve UCR data. Besides resolving the question of degree of auditability the preliminary audit will also be useful in gathering information essential to the planning and execution of the subsequent UCR audit. These corollary data include:

- quantification of complaint activity for the time period to be audited
- assessment of anticipated assistance and cooperation likely to be provided by agency personnel
- estimation of time and manpower requirements for implementation of the UCR audit

Once the degree of auditability has been determined the agency may choose to implement one of three proposed audit procedures. These include:

A. Standardized UCR Audit -- this audit requires that the agency capture, record and file data in some reasonable fashion to facilitate rapid retrieval.

Based on a system of forward and reverse cross-checks (see Table 1 below) five system points or stages are reviewed to assess the integrity of the system and its UCR output. These system stages are:

Stage I -- Telephone Complaints (Tapes)

Stage II -- Complaint Cards

Stage III -- Offense/Incident Reports

Stage IIIA -- Arrest Reports

Stage IV -- Data Aggregation (UCR Return A)

At Stages I - IIIA the audit is conducted by

- determining the time period to be audited
- determining the group size of all source documents and/or tapes to be audited
- determining the sample size[8]
- randomly selecting the sample units
- segregating the randomly selected sample units from the group
- reviewing all sample units for substantive errors
- cross-checking both forward and reverse[9] in the system all erroneous or suspicious sample units
- calculating the percent of error for each stage being audited

An additional quality control procedure will also be performed at Stage III of the UCR audit. To verify the accuracy of apparently "valid" Part I offense reports, a 10% sub-sample[10] of the original sample of offense reports will be tracked in reverse to the victim/complainant. Interviews will be conducted to determine the accuracy of the initial investigation and documentation.

Stage IV consists of a simple recapitulation or recounting of the data reported on the UCR Return A for the time period being audited.

The Standardized UCR audit provides a progressive, indepth, multi-facted approach in which each stage is first examined individually and then in relation to the next forward and/or reverse stage to insure accuracy and reporting continuity. This audit is by far the most comprehensive of the three and is the only technique capable of generating statements regarding the reliability of the entire UCR reporting system.

B. Provisional Audits -- these audits may be utilized by agencies which do not possess the mechanical, technical or procedural capabilities necessary for implementation of the Standardized UCR Audit. These limited audits, while based on the same random sampling techniques as the Standardized UCR Audit, are designed to verify the accuracy of only a segment of the agency's incident reporting system (i.e., the telephone tape stage, complaint card stage, offense report stage, or arrest report stage). The preliminary audit will have determined the agency's degree of auditability. If the agency cannot immediately undertake a Standardized Audit, it may, while continuing to improve its system in an effort to meet the UCR audit requirements, implement an audit of one or more stages of its incident reporting system. Although these audits fall short of the comprehensive techniques provided by the Standardized Audit, they nevertheless provide an

Re-Investigation	Victim/Complaint Interview	Phone Tapes	Complaint Cards	Offense Reports	Arrest Reports	Register Return A
X ←	X ←	Stage I Audit	→ X	→ X	→ X	→ X
X ←	X ←	X ←	Stage II Audit	→ X	→ X	→ X
X ←	X ←	X ←	X ←	Stage III Audit	→ X	→ X
		X ←	X ←	X ←	Stage IIIA Audit	→ X
			X ←	X ←	X ←	Stage IV Audit

TABLE 1

STANDARDIZED UCR AUDIT
Forward and Reverse Tracking Potential at Each Audit Stage

opportunity to make some limited assessments regarding the accuracy and integrity of UCR data as well as promote improvement of the existing system.

C. Recapitulation Audit -- This audit procedure simply consists of validating the total number of Part I offenses and clearances which appear on the monthly Return A by recounting the source documents which produced these data and by comparing these findings with those claimed by the agency being audited. A recounting of this nature may be requested by those agencies which lack the technical or mechanical requirements of the Standardized UCR Audit, but which wish to validate their numerical submission. This procedure will identify arithmetic or transcribing errors, however, its narrow scope prevents any assessment of the integrity of Stages I-IIIA of the incident reporting system.

IV. AUDIT RESULTS

Once an agency has implemented the audit techniques described in the previous section, several judgments can be made concerning the adequacy of the incident reporting system and the reliability of its UCR data.

The Initial Audit May Only Serve to Discover Deficiencies Within the Reporting System

If the agency does not meet the four requirements necessary for the conduct of the Standardized UCR Audit a preliminary and/or provisional audit may constitute the only procedure which can be effectively implemented. After the agency has attained certain procedural and technical capabilities which will minimize data loss the Standardized UCR Audit can then be conducted.

The Standardized UCR Audit May Demonstrate that the Incident Reporting System Seems to be Producing Reliable UCR Data

If the audit reveals that there is minimal data loss or distortion at each of the four stages of the system, then it can be concluded that the UCR data generated by that system are reasonably reliable and accurate for the time period being audited.

The Standardized UCR Audit May Provide the Agency With a Variety of Quality Control and Internal Management Techniques

By identifying areas of error potential within the incident reporting system the agency will be aware of the specific areas of vulnerability and be better equipped to implement remedial measures. Specific quality control features include:

- more efficient deployment of personnel
- implementation of an on-going internal audit/inspection program
- indication of need for in-service training for UCR personnel
- indication of need for more effective review and verification procedures

- determination of field unit response time

- determination of integrity of complaint reception personnel

- improvement of the records system

Conversely, the results of the provisional and recapitulation audits are much less conclusive or meaningful as compared to the results of the Standardized UCR Audit. An audit of only certain stages of the system for example, will provide no assessment concerning the amount of data loss which may occur at other unexamined input modes and points. Similarly, a recapitulation of the data submitted on the agency's Return A for a specified period of time will provide little insight into the reliability or integrity of the incident reporting system which produces these data.

The provisional or recapitulation audit procedures should not be confused with the Standardized UCR Audit. The latter represents the optimum method of audit which is designed to measure the accuracy and integrity of the overall system along with the UCR output.

V. SUMMARY AND CONCLUSIONS

Over a century ago far-sighted police administrators recognized the need for establishing a reliable fund of crime statistics. Acceptance and implementation of the proposed UCR audit procedures described in the preceding section may serve to achieve this goal by

- affording a definitive means of measuring the validity and accuracy of the UCR reporting process; and

- improving the records systems and UCR reporting components of individual law enforcement agencies.

The UCR Program is largely an untapped source of valuable data and should not be discarded simply because it may, as a whole, represent an inaccurate measure of the volume of crime known to police. The proposed UCR audit is viewed as a positive contribution in helping to clarify many of the misconceptions and problems associated with the program.

LIST OF NOTES

1 This paper represents a synopsis of work to date on the "Development of a Standardized System for Uniform Crime Report Audits" project funded by the Law Enforcement Assistance Administration Grant No. 74-55-99-3303.

2 The UCR Program was never intended to be a measure of all crime committed in the United States, but only of those offenses known to police. The Crime Index offenses (which include the seven Part I offenses of criminal homicide, forcible rape, aggravated assault, robbery, burglary, larceny/theft, and motor vehicle theft) were selected because of their seriousness, frequency in occurrence and likelihood of being reported to police.

3 Federal Bureau of Investigation, Uniform Crime Report - 1973 (Washington, D.C.: U.S. Government Printing Office, 1973), p. 51.

4 Uniform Crime Reporting, A Complete Manual for Police (New York: J.J. Little, 1929), p. 1.

5 Ten Years of Uniform Crime Reporting 1930-1939 (Washington, D.C.: U.S. Department of Justice, 1939) p. 18.

6 Federal Bureau of Investigation, Uniform Crime Report - 1973, pp. 52-53.

7 For example, the Ohio Institute studied the records systems employed by local agencies in Ohio. Ten Years of Uniform Crime Reporting 1930-1939, p. 48.

8 Sample sizes were selected on the basis of the computation of the standard error for a number of possible sample sizes. Those selected were computed at the 95% confidence level.

9 Reverse cross-checks include re-investigation of the original complaint and conducting victim/complainant interviews. Forward checks are generally made with the documents found at the subsequent stage (e.g., complaint cards are compared forward to their companion incident/offense report).

10 This 10% sample will be drawn from the quantity of offense reports that remains after all erroneous or suspicious reports have been resolved through the system of forward and reverse cross-checks.

COMPARING MEASURES OF CRIME: POLICE STATISTICS AND SURVEY ESTIMATES OF CITIZEN VICTIMIZATION IN AMERICAN CITIES

Wesley G. Skogan, Northwestern University

Introduction

The growing use of sample surveys to measure the volume and distribution of crime in the United States will provide social scientists and public administrators with valuable new data with which to test their theories and plan crime-reduction programs. In particular, the National Crime panel and city-level samples currently being monitored by the Bureau of the Census should produce a rich body of information on aspects of criminal and victim behavior which previously have escaped systematic analysis. A host of research problems for which current statistics are unsuitable may be confronted with this emerging data base.

The most immediate use of survey estimates of crime rates, however, has been to compare them to official statistics. Reports released by the Law Enforcement Assistance Administration have stirred public interest by their contrast with police figures on crime of the type summarized in the F.B.I.'s yearly Uniform Crime Report.[1] Such comparisons inevitably reveal wide gaps between rates registered by the two sources. National or city-level survey-generated figures usually overshadow official police statistics by a substantial margin. This type of analysis has been encouraged by the government's decision to calculate U.C.R.-compatible figures from citizen surveys, although this is perhaps the least useful application of the data. The observation that there are varying discrepancies between official and survey crime estimates does not tell us where the error lies. Every statistic is shaped by the process which operationally defines it, the procedures which capture it, and the organization which processes and interprets it. Survey and police crime-measurement procedures produce different figures, but the reasons for this and its implication require analysis. A discussion of how survey and official crime statistics differ and why we obtain these discrepancies may clarify both their comparability and their individual interpretation, and it may speak to their improvement in the future.

Measurement Error and Official Crime Statistics

The presence of error terms of considerable magnitude is not unique to measures of crime, although a half-century of continuous criticism has focused more widespread attention upon the errorful nature of crime measures than enjoyed by most social statistics. Measurement is the process of mapping an empirical system into a numerical system. It involves the application of definitions to delineate aspects of the empirical system which are of interest, and a series of "If...Then..." rules matching selected attributes of those phenomena to numbers. The resulting figures always map the richness of the referent system simplistically and inexactly.

In measurement terms, all observed scores are composed of two elements: they are partially "true score" (reflecting what we wish to observe) and partially error. Even rapidly repeated, identical measurements of the same phenomenon will produce different numerical readings. The degree to which they are similar -- our ability to reproduce our findings -- is referred to as the "reliability" of a measurement process. Reliability tests, for example, would gauge the ability of various police patrol teams to classify the same set of events in the same manner. While the ability to examine events twice and find the same thing is the *sine qua non* of good measurement, even reliable measures may not be useful. A researcher's procedures may not be measuring the object of interest, or the resulting figures may be artifacts of the measurement process. This is a validity problem. Police districts with ambitious commanders may consistently produce low crime totals. In order to obtain valid, non-artifactual measures we must employ multiple and differing techniques, cross-checking our findings at every turn.[2]

Disciplines with well-developed measurement traditions have evolved routine procedures for coping with these problems. Economists have stressed reliability; they require measures which are stable and comparable across time.[3] Psychologists emphasize validity. The intangibility of the psychological domain heightens concern that its apparent orderliness may be an artifact of specific methods of investigation. Sophisticated psychological measurement combines the fruits of interviews, projective evaluations, and physical observations.[4]

The measurement of crime is a substantive and methodological problem of interest to researchers in a variety of disciplines. Perhaps as a result, most of the effort expended upon measurement problems has been conducted outside of any coherent measurement model. Scattered validation studies of official statistics have been reported. Price compared state-level property-crime totals with insurance rates and uncovered only moderate correlations.[5] But such criterion validation requires a dependent measure which is relatively error-free, and in this case crimes known to the police are probably a better indicator of the underlying true distribution of events than the independent validator. A better example of criterion validation is the California Criminal Statistics Bureau's comparison of police and American Bankers Association's figures on bank robbery. The latter measure appears to be clearly defined and exhaustively enumerated, and it proved to be reflected quite accurately in official statistics.[6]

Validity studies of official measures of more typical events, those which are less clear-cut and involve more discretion on the part of police officers and administrators, have been less hopeful. Comparisons between official records and self-reports of delinquency or informal police "contact" reports indicate that official figures greatly underestimate the volume of events which might be uncovered in other ways.[7]

The development of our current system of gathering and publishing official statistics on crime was a response to these problems. The invalidity of local department's efforts at data collection and the limited reliability of the reported figures led to the development of the Uniform Crime Reporting system in the late 1920's. This system improved reliability and sacrificed validity. Standardized definitions, data-collection forms, and data-gathering techniques produced city-level crime totals which were usually comparable from year to year, and inter-city comparisons undoubtedly are vastly improved by the U.C.R. system. But several important compromises were made in the formulation of this statistical system. The data are still gathered by local authorities, participation in the network is not mandatory, and the F.B.I.'s only option in the face of fraud is not to publish the reported figures.[8] As early as 1931 the Wickersham Commission called for the creation of a centralized data-collection service and rigorous data-quality control.[9] The misreporting and underreporting apparently endemic in current official statistics has led to their widespread devaluation.

Survey Measures of Crime

Continuing dissatisfaction with official measures led to the development of alternative techniques to gauge the scope and distribution of crime in American society. The most important of these is the population survey, a measuring device (with its own characteristic reliability and validity problems) which has yielded striking new pictures of the crime problem.

The use of the sample survey to study crime reflects dissatisfaction both with the apparent accuracy of official figures and the paucity of information they purport to reveal. The yearly Uniform Crime Report does not speak to questions about the characteristics of victims of crime. Offender data is available only on arrestees, although victim testimony might shed some light on the characteristics of successful criminals. Finally, little data is reported on the physical and social circumstances under which most crimes occur, even though this has tremendous implications for their solution and deterrence.

It was apparent to the President's Crime Commission that population surveys potentially could speak to all of these inadequacies, and in the mid-1960's the Commission funded several pilot projects and a national sample survey to test their utility.[10] Since then, the federal government has inaugurated a regular surveying program on a national scale and has funded several local and state-level investigations.[11]

It was inevitable that the victim-based data gathered by these large-scale surveys would be used to gauge police-reported crime statistics. Suspicion of official statistics has become widespread and appreciation of the errors in crime data particularly well-known, much more so than the enormous insecurity felt by researchers who regularly employ attitude measures and self-reports of behavior. The latter deal skeptically with data and demand elaborately scaled, multiple item indicators of concepts before they test theories with any confidence. The items in the Crime Panel surveys elicited a much larger volume of events than reported by police, so it is widely assumed that they are "more accurate" measures of the true volume of crime in society. But such gaps are inevitable. Despite the surface similarity of the resulting figures, the measurement operations and their errors differ greatly when we compare police and survey procedures for estimating crime rates. The social and organizational processes which stand between events occurring in the world and our survey or official maps of them produce quite different kinds of crime statistics.

Sources of Measurement Error

In the course of mapping crime events into a numerical system, both official and survey measurement procedures generate considerable error. If we think of error as the gap between a true score and an observed score for an event, Figure 1 may be a useful summary of what we know about its sources. On the survey side, measurement error has been investigated intensively in a

----
Figure 1 goes here
----

series of pilot studies which began in 1966. Our knowledge of error-generative processes on the police side is older, but has been considerably enhanced by studies of victim behavior and systematic observations of police work during the past decade.

Ironically, the first stages in the official measurement process lie largely in the hands of civilians: the victims of crime, their relatives, neighbors, and bystanders. The first public filter through which events must pass is perceptual: someone must know that a crime has taken place. This is in part an information problem. For example, a great deal of larceny from commercial establishments (shoplifting and employee theft) is discovered only in the form of inventory shrinkage.[12] In this case we know that crime is taking place, but criminal events remain unknown and uncountable. The general difficulty is that discreet events may escape detection, while continuous indicators of their occurrence --like dollar losses per quarter or shortages at audit--cannot be enumerated under our current system of social accounts. The problem is also conceptual: people must define an event as falling into the domain of events about which "something must be done." This appears to inhibit the reporting of much consumer fraud, and it is the difference between crime and "ripping-off." Attitudinal studies of the ligitimacy of theft or fraud upon large private and governmental bureaucracies indicate that there is far from universal agreement about the labeling of some behaviors in our society. The problem of who does the perceiving is also of interest. Pilot surveys in Dayton and San Jose revealed that a surprising 25 percent of all personal crime and 20 percent of all property crime is reported by someone other than the victim.[14] Sample survey, victim based studies of reporting and non-reporting are not designed to cope with this.

The decision to call the police has been the focus of considerable research, for it is probably the most important factor shaping official statistics on crime. In the Dayton-San Jose pilot surveys conducted in 1972 respondents recalled that about 60 percent of all robbery, 56 percent of all larceny, and 40 percent of all household burglary was not reported to the police. Their reasons for failing to do so were numerous: the largest categories chosen were "not serious enough" (25-30 percent), "nothing can be done" (25 percent) or that the harm or loss was slight (10 percent).[15] Other analyses of the reporting problem have focused upon race, class, or even personality characteristics of victims rather than their manifest responses, although the utility of this approach is not particularly clear. It appears that the characteristics of the event are controlling: who did it (relative or stranger), why it was done (economics or passion), what was the damage to person, property, or propriety, and what were the participants' estimates of the burdens and benefits of evoking the police. Only a portion of the latter calculation--that involving the victim's fear of the police--would appear to be a straightforward race-and-class problem. Despite much discussion of this factor, neither Ennis' national survey nor the Dayton and San Jose studies revealed more than 2 percent giving that response.[16]

Observational studies of police behavior indicate that even after the police are called the outcome of the crime-measurement process remains problematic. Crime recording becomes a social and organizational activity. Reiss and Black's descriptions of police-citizen encounters in Chicago, Boston, and Washington, D.C. indicate that extra-legal factors greatly influence the decision to write a formal report.[19] The police are loath to file a report when the relational distance between the participants in a dispute is small, in part because they know that it is very unlikely that the case will be pursued in the courts. They tend to defer to the dispositional preferences of the complainant, who often mobilizes the police only to warn or threaten another party. Both complainants who are deferential to the police and higher-status victims are more likely to be successful in persuading the police to file a report. The police also act upon their own assessment of the complainant's culpability. Often responsibility for personal crimes or their outcomes may be apportioned among the parties, and police respond to the division of blame. Finally, in cases where juveniles are parties to a dispute the police tend to defer to the dispositional preferences of adults at the scene.

These observations suggest another reason why official statistics on crime should be lower than survey estimates. Unlike survey enumerations where the victim's claim ultimately must be recorded on his terms, police "measurement" takes place within the context of the event. Complainants are often surrounded by witnesses and bystanders who contribute their interpretations of events, and--surprisingly often--suspects themselves are present to offer countercharges and alternative explanations. The decision to file a formal report is almost "judicial" in the sense that an officer weighs claims and counter-claims before making a disposition in a case. Patrol officers quickly learn to be suspicious of the motives of complainants, for their authority is often evoked for private purposes.[18] Claims of victimization are not taken on face value. As the Uniform Crime Report does not present pre-disposition case totals, but only "founded" complaints for each city, we have no idea of the dimensions of this process. Scattered reports of large departments on hand indicate that the effect of "unfounding" be considerable: approximately 25 percent of rapes, 13 percent of robberies, and 19 percent of gun assaults reported to the police were discounted in these cities. They probably would generate self-reports of victimization, but they did not enter our social accounts.

Technical considerations, including difficulties with the classification scheme employed in gathering official statistics, may introduce measurement errors on the police side as well. The uniform crime reporting system imposes a set of definitions which usually do not match the criminal-code pigeonholes into which the police must daily sort events. The definitions are also neither mutually exclusive nor exhaustive. The translation from local to national terminology appears to vary from jurisdiction to jurisdiction, enhanced by local differences in training and data quality control.[19] Errors of this sort will shift over time within cities as well. The tremendous variation and apparently random distribution of "manslaughter by negligence" totals reported in the Uniform Crime Report, for example, appears to be a function largely of variations in local practice.[20] Survey studies of crime, on the other hand, utilize measurement operations which may vary considerably across cities. Because these error terms differ, further "gaps" will appear between figures from the two sources.

The final source of error on the police side is organizational and political. The ability of official records systems to "retain" information once it has been entered is problematic. In 1966 a department audit of stationhouses in New York City revealed 20-90 percent underreporting of events in their files.[21] These and other discoveries suggest that crime is an organizational problem in police departments. Especially in cities with a strong "stationhouse culture" or where district commanders are evaluated on their ability to reduce crime, we should observe a consistent tendency toward underreporting by police departments. Events also disappear individually in response to political influence or bribes, but this is less likely to skew totals in common crime categories.

The dramatic impact of variations in police record keeping procedures upon crime statistics is illustrated by "before-and-after" studies of cities which have overhauled their systems. Many of these were noted by researchers for the Crime Commission in their discussion of crime statistics.[22] New York City's 1950 reorganization, for example, boosted that department's robbery totals by 400 percent, larceny 700 percent, and assault with a weapon 200 percent.[23] The Commission correctly perceived such overhauls as part of a more general phenomenon: the increasing professionalism of big-city police departments. A working hypothesis would be that as departments centralize their administration, automate their information systems, and encourage more legalistic behavior on the part of beat patrolmen, error in the official measurement of crime may be significantly reduced.

The sources of measurement error on the survey side have been investigated in a series of national and city-level pilot studies. In some, alternative techniques are employed in different random samples of a population and the results are compared. In others, police records are sampled to locate respondents who are known to have been victimized. They are interviewed and their recall patterns analyzed. Each method gives us a different check of the reliability and validity of survey measures of crime.

These investigations suggest that the first question we must ask is, "Will the victim be interviewed?" This raises both data collection and sampling problems. In early pilot studies a randomly selected adult was used an an informant for his entire household. Interviewers quizzed a single respondent about the victimization experiences of each family member. In the Dayton-San Jose surveys, a random half of the sample in each city was completely enumerated; interviewers questioned every household member over the age of thirteen to elicit self-reports of victimization. Apparently informant fatigue or lack of information about other household members is a substantial problem, for individual questioning elicited significantly more events. The differences were so marked that future government surveys will employ only complete household enumerations despite their increased cost.

Sampling deficiencies, on the other hand, have not been remedied. In the city-level studies conducted by the Bureau of the Census the sampling frame is bounded by the territorial limits of the central city. But an average of thirteen percent of the daytime population of the nation's core cities are commuters.[24] In Chicago, for example, over 400,000 workers leave the city at sundown. Tourists and other transients account for another fraction. Although they may be victimized and can report their experiences to the police, they are not eligible for interviewing under current procedures.

Even if they enter the sample, victims of crime may not successfully recall the event. As Albert Biderman has noted, one of the most striking findings of the victimization pretests was the relatively low salience of many crime events. "In practice, most respondents seemed to find it difficult to remember incidents of victimization other than recent cases.[25] The problem of memory fade has been investigated in two ways. First, known victims have been selected from police reports and interviewed. Their recall rates have climbed from 62 percent (Washington) to 74 percent (San Jose), reflecting successive improvements in the Census Bureau's questionnaire. Second, respondents have been required to recall known events within time frames ranging from three months to one year. These tests reveal a sharply decreasing recall rate for temporally distant events. The same phenomenon may be observed by plotting the date of occurrence of each event recalled by randomly selected respondents. Monthly crime rates estimated from survey responses drop sharply as an inverse function of time.[26] The effect is so striking that accurate survey measurements require brief recall periods. This means that very large samples are required to provide yearly crime estimates. The current compromise for the National Crime Panel is six months; respondents in the city studies are asked to recall events for an entire year. Police estimates, of course, are subject of few of these difficulties.

Reverse checks of police records also indicate that recall rates in an interview setting are sensitive to variations among the events themselves. They suggest that responses may not be forthcoming even if an event is recalled. Victims appear to be unwilling to report clashes with friends or relatives, for example. In San Jose, those who the police noted had been victimized by strangers recalled the event 75 percent of the time; only 22 percent of the cases where the police recorded that the offender was a relative were recalled, and 58 percent of those cases involving an acquaintance. In general, property crime was much more fully recalled than personal crime. Rapes were revealed only tentatively; in the San Jose pilot survey all recalled rapes were described as "attempted." It should be noted that these variations are similar to those which appear to affect the willingness of victims to relate their experiences to the police as well. Disputes within families and rapes are both highly underreported. And, as it was noted above, the police appear to be less willing to file formal reports when disputants are acquainted. In this case, survey and official measures both systematically undercount the same classes of events.

As noted in Figure 1, the final step in the survey measurement of crime involves the coding and classification of reported victimizations. It is difficult to judge how successfully this process reflects the event. In his report to the Crime Commission, Ennis related a modest test of the inter-coder reliability of his classification scheme. Teams of lawyers and detectives were successful in classifying citizen-reported victimizations in the same U.C.R. categories as his research staff about 65 percent of the time. In a validity test of the more advanced San Jose Survey instrument, Census personnel classified 259 of 292 recalled victimizations into the same categories as the local police who initially recorded them. Since we have no confidence that police and the interviewer were told exactly the same story, it is a remarkable correspondence. This, coupled with the face validity of the current survey instrument--the items are drawn to tap the dimensions which define Part I offenses in the Uniform Crime Report--suggests that the classification stage of the process is probably less troublesome than most.

A final and potentially important source of error in both survey and official measures is the intrusion of other events into the observed score for a city or household. On the police side, fraudulent claims may be registered. People may misuse the police in personal vendettas, they may invent crimes to disguise their own culpability, or they may attempt to register excessive insurance claims. In addition, actual events which lie outside the domain of interest may be misclassified as falling within it. The most serious problem on the survey side is telescoping. Method checks of all kinds indicate that the tendency of respondents to "telescope in" events which occurred outside of the reference period of the survey and to claim that they occurred within the specified interval is quite strong. For example, known victims telescope events which police files place firmly beyond the reference period. Experiments with the Census' Quarterly Household Survey panels indicate that bounded interviews may avoid distortions of this kind. Respondents who are asked to recall events which have occurred since an interviewer's last visit report as few as one-half the number of victimizations recalled by those who are quizzed about the same period but who have not been previously questioned.[27] Given the low salience of most crime events and their steep forgetting curve, victims require signposts to guide their recall.

Estimating Error Magnitude

Like any measure, estimates of crime rates contain error. Given the magnitude of the sources of error discussed here, it is remarkable that official and survey measures of crime covary as closely as they do. The existence of these multiple measures enables us to estimate in rough fashion the magnitude of the error in each, and to generate some simple correction factors which may make them more useful. Additional methods, tests and analyses of existing data may contribute further to our understanding of the dimensions of error.

Crosschecks of recall errors in the survey measurement process indicate that the rate at which interviews "recover" events is fairly high. In the San Jose pilot survey of 1971, of the 394 known victims who were located for questioning, 292 recalled the event in some form. Table 1-A presents the recall rate for various sub-categories of events. Note that rates for frequent

-Table 1 goes here----

crimes, larceny and burglary, were higher than those for less frequent events. Table 1-A also presents the total number of personal and household victimizations recalled by the residents of San Jose proper in the standard population survey phase of the pilot study. These are then projected into "corrected" totals which take into account patterns of non-recall. As the column totals indicate, the San Jose survey may have recovered approximately 75 percent of the five classes of events of interest. This is a very rough indicator of the recovery power of the victimization survey instrument, one that requires further refinement.

Table 1-B examines the respondent's contributions to errors in survey measures of crime. The forgetting curve plotted in Table 1-B indicates that recall periods exceeding three months may lead to the substantial undercounting of offenses in the population. A test of the ability of those recalling events to place them in the proper month --an essential check of the ability of surveys to provide time-series crime estimates of the type anticipated--indicates that recall accuracy degrades sharply after about three months as well.[28] These curves, which were computed from data in the report of the San Jose pilot study, suggest that the six-month recall period used in the National Crime Panel and the twelve-month period bounding the city-level samples may contribute significantly to the error components of those measures. Because these curves were calculated from the same data used to estimate survey recovery rates in Table 1-A, it is impossible to untangle here the distinct contributions to error of the salience of events and their temporal distribution, however. The estimates of the magnitude of forward telescoping error presented in Table 1-B are based upon the Washington, D.C. pilot survey. There, seventeen percent of the victimizations recalled by selected respondents occurred before the indicated cut-off point when a six-month limit was specified, and 21 percent telescoped in events which occurred before a twelve-month limit. As I noted before, telescoping effects--which lead to an overcounting of events--can be controlled by "bounding" the recall period with a salient event. The National Crime Panel utilizes the previous visit of an interviewer, while city-level interviews must rely upon verbal instruction. The latter measures are much more likely to overestimate crime rates due to telescoping errors.

Error introduced in stages preliminary to the interview are more difficult to estimate. Sampling errors for individual cities are introduced by the systematic elimination of commuters, conventioners, and tourists from the sampling frame. The effect of this loss upon one crime statistic, motor vehicle theft, is very roughly estimated in Table 1-C, where motor vehicle statistics for the city of Chicago are presented. The recent victimization survey of Chicago estimated that residents there suffered about 38,700 vehicle thefts in 1972, or a loss probability of .03 per motor vehicle. Projecting commuter vehicle losses at only one half of the rate for city residents it appears that excluding commuters from estimates of city crime probably undercounts victimization by about 8 percent for this offense. Other crimes dealt with in the victimization surveys cannot be so easily projected. Commuters are susceptible to personal larceny, robbery, assault, and rape, probably in that order; conventioners and tourists may be more at risk than commuters in all of these categories. This systematic elimination of potential victims of crime is redoubled when we consider the distribution of known victimizations. The victimization surveys indicate that many offenses, most notably assault and robbery, disproportionately victimize young males; they are also a demographic group which is most difficult to enumerate in a population survey. A summary estimate is that we undercount by 5 percent or more due to sampling limitations. Together, sources of error on the survey side of Figure 1 probably accumulate to undercount events by 30 percent.

Reversing the analysis enables us to probe the magnitude of error terms on the police side as well. The most satisfactory test would reverse the record-check procedures utilized in the pilot surveys: follow-up studies of offenses which were apparently reported to the police would be conducted to determine which could be found in police records.[29] None of the police departments which have extended their cooperation to researchers has granted access to their records on this scale. A simple analysis of the marginal frequencies of reported and officially recorded events in some of the sample cities suggests that the gap between the two sources would be considerable. Many events which occur and are said to have been turned over to the police do not appear to survive police processing.

Without further cooperation on the part of city police departments, errors in the official measurement of crime must be estimated from aggregate totals. Officials and survey estimates of city crime rates consistently differ. The ratio of robberies recalled in interviews to robberies known to the police in Portland--one of the smallest cities analyzed here--was greater than 3-to-1 in 1971-72; in New York City, more than 2½ robberies were recalled to interviewers for every event recorded by the police. Table 2 presents official and survey robbery estimates for five cities surveyed in 1972 as part of the Bureau of the Census' Large City study.

Table 2 goes here

The victimization surveys also asked each victim of a crime whether the event was reported to the police. This reporting rate can be used to "correct" survey estimates of the crime rate in each of the five cities for citizen-induced errors in police figures. While it is socially desirable to respond under questioning that one reported a crime, which will inflate this figure somewhat, it is clear from Table 2 that substantial distance remains between citizen "reported" and police recorded crime. The gap was smallest in Detroit, where police accounts of robbery added up to 73 percent of the "reported" (by recall) by city residents. Philadelphia's extremely low robbery count, which amounted to only 37 percent of what her residents claim to have reported to the police, may be related to numerous charges that police there cheat on their statistics.

The figures presented in Table 2 suggest that official "crimes known to the police" are probably not very accurate indicators of the true volume of crime in a community. In the case of robbery, official totals accounted for an average of only 38 percent of the victimizations recalled by citizens of these five communities. That figure varied considerably across cities: official robbery totals added up to 48 percent of the survey figure in Detroit, and only 22 percent in Philadelphia. If the interview recovery rate for individual and commercial robbery in these cities approximated the individual rate in San Jose (76 percent - see Table 1), official robbery totals might amount to an average of only 28 percent of the true total, and that figure might drop to only 17 percent in Philadelphia. The correction for error induced by citizen reporting practices improves this picture somewhat--official figures may reach 63 percent of "reported" robbery--but it appears that organizational processes contribute considerably to error in police-recorded crime statistics.

NOTES

1. The New York Times featured survey and U.C.R. incident totals in Census surveys.

2. See G.W. Bohrnstedt, "Reliability and Validity Assessment in Attitude Measurement," in G.F. Summers (ed.) Attitude Measurement. Chicago: Rand McNally, 1970, 80-99.

3. O. Morgenstern, On the Accuracy of Economic Observations. Princeton: Princeton Univ. Press, 1963 (2nd ed.)

4. S.W. Cook & C. Selltiz, "A Multiple-Indicator Approach to Attitude Measurement," Psychological Bulletin, 62(1964), 36-55; D.T. Campbell & D.W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, 56(1959), 81-105.

5. J.E. Price, "A Test of the Accuracy of Crime Statistics," Social Problems 14(Fall, 1966), 214-221.

6. Bank Robbery in California: A 35-Year Comparison of California with the Rest of the United States and an Intensive Study of 1965 Offenses. Sacramento: California Criminal Statistics Bureau, 1967.

7. See: W. Chambliss & R.H. Nagasawa, "On the Validity of Official Statistics," J. Research in Crime and Delinquency, 6 (1969), 71-77; R. Quinney The Social Reality of Crime. Boston: Little Brown, 1970.

8. D.J. Pittman & W.F. Handy, "Uniform Crime Reporting: Suggested Improvements," in A.W. Gouldner & S.M. Miller (eds), Applied Sociology. N.Y.: The Free Press, 1965, 180-188; NIMH. Criminal Statistics. Rockville, Md.: Center for Studies of Crime and Delinquency, 1972.

9. U.S. National Commission on Law Observance and Enforcement. Report on Criminal Statistics. Washington: U.S. Gov't. Printing Office, 1931, 75.

10. These included: Washington, D.C.; Chicago, Washington, & Boston; and a national sample. All published by U.S. Gov't Printing Office, 1967.

11. Pilot studies included: "Washington Pretest;" "Baltimore Pretest," "San Jose Methods Test." Full scale survey of two cities followed: "Dayton-San Jose Pilot Survey." Since then two reports of multi-city studies: Chicago, New York, Philadelphia, Los Angeles, & Detroit; and Atlanta, Baltimore, Cleveland, Dallas, Denver, Newark, Portland, St. Louis.

12. See: R.W. Dodge & A.G. Turner, "Methodological Foundations for Establishing a National Survey of Victimization," American Statistical Association, August, 1971.

13. E.O. Smigel & H.L. Ross, Crimes Against Bureaucracy. New York: Van Nostrand, 1970.

14. "Dayton-San Jose Pilot Survey," Table 23.

15. Ibid,, 24.

16. Ibid.; Ennis, op. cit., 44-45.

17. D.J. Black, "The Production of Crime Rates," American Sociological Review, 35 (August, 1970), 733-748; D.J. Black and A.J. Reiss, Jr., "Police Control of Juveniles," American Sociological Review, 35 (Feb. 1970), 63-77.

18. J. Rubenstein. City Police. N.Y.: Farrar, Strauss, Giroux, 1973, Chapter 5.

19. F.B.I. Uniform Crime Report, 1972, 56.

20. New York City, for example, has twice the population of Chicago and reports 3½ times as much crime, but Chicago in 1972 recorded 260 negligent manslaughter cases while New York recorded 66. In general, manslaughter rates do not correlate with anything.

21. M.E. Wolfgang, "Urban Crime," in J.Q. Wilson (ed.). The Metropolitan Enigma. Cambridge: Harvard Unviersity Press, 1968, 254.

22.President's Commission on Law Enforcement and Administration of Justice. Task Force Report: Crime and Its Impact--An Assessment. Washington: U.S. Gov't Printing Office, 1967, 22-24.

23. Wolfgang, op. cit.

24. J.D. Kasarda, "The Impact of Suburban Population Growth on Central City Service Functions," American Journal of Sociology, 77 (May, 1972), 1122.

25. Biderman, op. cit., 31.

26. Ennis, op. cit., 97.

27. A.G. Turner, "Methodological Issues in the Development of the National Crime Survey Panel: Partial Findings," National Criminal Justice Information and Statistics Service, Statistics Division, Law Enforcement Assistance Administration, December, 1972, 8-9..

28. Three months also appeared to be the optimal recall period in the Crime Commission's national survey. Ennis, op. cit., 95-98.

29. The magnitude of this task is discussed in "San Jose Methods Test...," op. cit., 10.

FIGURE 1

SOME SOURCES OF MEASUREMENT ERROR

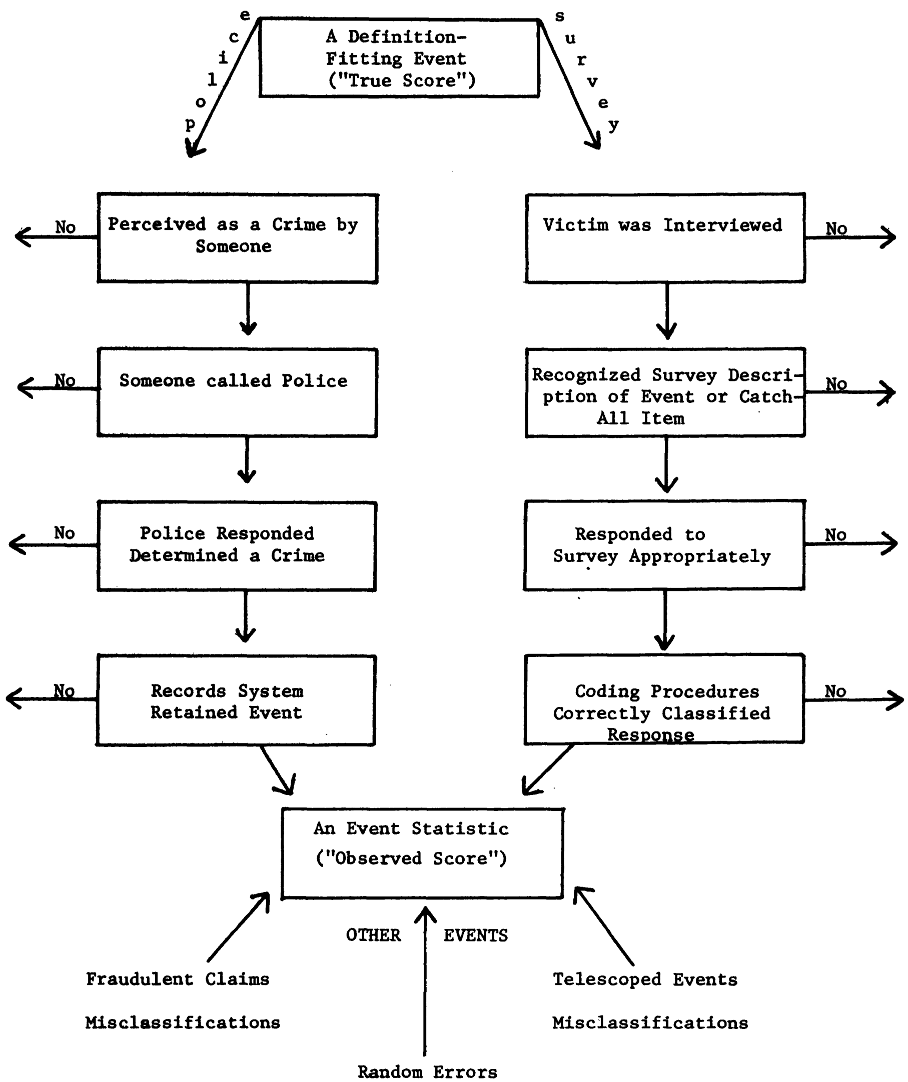


Table 1

Estimates of Survey Measurement Error

1-A: Interview Recall by Type of Household and Personal Victimizations: City of San Jose Only

Type of Crime	Total Recalled[1] Victimizations	Recall[2] Rate	Projected Victimizations
Rape	100	67	149
Robbery	2840	76	3737
Assault	8980	48	18708
Burglary	17610	90	19567
Larceny	45900	81	56667
Total	75430		98828

Event Recall = 75430/98828 = 76%

1-B: Recovery Rate and Accuracy of Interview Recalls by Time: San Jose Area and Washington, D. C.[3]

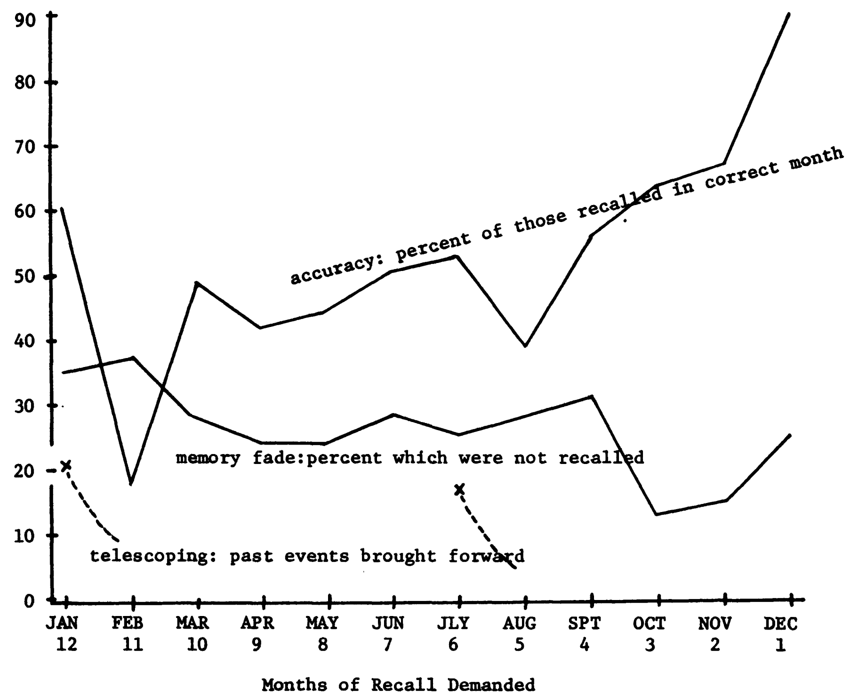


Table 2

Estimates of Official Measurement Error

City	Total[1] Official Robbery	Total[2] Survey Robbery	Survey[3] Measure of Reporting Rate	Estimated "Reported" Robbery	Official As Percentage of Survey "Reported"
Chicago	23531	64100	57.5	36881	64
Detroit	17170	36100	65.4	23638	73
Los Angeles	14241	36400	55.1	20064	71
New York City	78202	191400	59.5	113863	69
Philadelphia	9710	44000	58.9	25914	37
Average	28571	74400	59.3	44072	63

SOURCE: [1]Uniform Crime Report, 1972.

[2]"Crime in the Nation's Five Largest Cities," Washington: National Criminal Justice Information and Statistics Service, Law Enforcement Assistance Administration, April, 1974, Table 1. This includes both individual and commercial offenses.

[3]*Ibid*., recomputed from Table 8. This rate includes both individual and commercial reports.

Table 1-C Estimates of Survey Measurement Error (continued)

1-C: Sampling Frame Loss Estimate: Motor Vehicle Theft in Chicago, Illinois, 1972

Vehicle Registration	1,260,000[4]
Survey Theft Estimate	38,700[5]
loss probability	.03
Total Commuting Autos Daily Entering City	206,000[6]
Commuter Vehicle Loss at One-Half City Rate	3,090

SOURCE: [1]Tables 12 and 39, Crime and Victims: A Report on the Dayton-San Jose Pilot Survey of Victimization. Washington: Law Enforcement Assistance Administration, Department of Justice, June, 1974. The San Jose robbery total presented in Table 12 of that source is clearly incorrect--this is my estimate from the robbery sub-totals. Rape totals are not reported in raw form; 100 is estimated from percentages elsewhere in the report.

[2]"San Jose Methods Test of Known Crime Victims," Statistics Technical Report No. 1. Washington: National Institute of Law Enforcement and Criminal Justice Statistics Division, Law Enforcement Assistance Administration, June, 1972, Table C.

[3]Accuracy and memory fade figures were calculated from Table 4, "San Jose Methods Test...," *op. cit.*, 14; telescoping data were reported in: "Victim Recall, Pretest (Washington, D.C.)," Demographic Surveys Division, Bureau of the Census, June 10, 1970, Table G.

[4]Annual Report, Illinois Secretary of State, 1972.

[5]"Crime in the Nation's Five Largest Cities," Washington: National Criminal Justice Information and Statistics Service, Law Enforcement Assistance Administration, April, 1974, Table 1.

[6]U.S. Bureau of the Census. Journey to Work. Washington: U.S. Government Printing Office, June 1973, Table 2.

A GOALS ACCOUNTING SYSTEM*

Nestor E. Terleckyj, National Planning Association

I. INTRODUCTION

In this paper, I would like to give a brief summary of results of a study soon to be published by the National Planning Association under the title, Estimating the Possibilities for Improvements in the Quality of Life in the United States, 1973-1983, in which a goals accounting system is developed and a resulting set of estimates presented. Because the methodology of this work and its underlying rationale have already been described elsewhere[1] and are again discussed in detail in the forthcoming book, I shall give only a brief summary of the results and the current set of estimates. I will then discuss some of the conceptual issues which bear on the design of the goals accounting system and on the meaning of the estimates provided by it. These are: the content of economic analysis as applied to production of particular social changes, the basis for selecting the goals concerns and indicators, the treatment of time, and the applicability of results to policy-making.

The purpose of this work was to provide a method of estimating the production possibilities of particular social changes seen as goods.

The present results are highly experimental. While quantitative estimates are always made with incomplete information, the present estimates entail a substantial degree of uncertainty regarding the proper specification of the system as a whole and its variables and uncertainty at the estimation level regarding the statistics and other information used.

The system entails many abstractions and a set of very complex relationships. In order to obtain an even approximate understanding of the system as a whole and its relationships, a complete elaboration of the entire analysis is first necessary. Only then can judgments be made about desirable changes in the design and content of the analysis.

The goals accounting analysis rests on a series of specific assumptions. Of these assumptions, two should be mentioned here since they may help to distinguish the goals accounting analysis from other lines of endeavor. One assumption is that the social and economic trends are not completely determined; there exist possibilities for departures from these trends and in particular for achievement of levels of social output above those given by the trends. These possibilities are first broadly defined by technical opportunities, comprising the set of what may be possible within the scope of technology, within the limits of the environmental, social and demographic feasibility, and within the given limits of time. The feasible changes are further specified by a set of economic constraints, i.e., by the amount of resources which may be available to achieve these technically feasible additional changes which represent improvements on the ongoing trends.

The second assumption is that at any point in time in addition to fixed modes of operation there exists a discretionary margin in the use of resources and in the types of activities undertaken by the individual decision units in the private and in the public sectors. The relative size of this margin can be defined as a function of time such that the amount of resources which can be allocated to the discretionary activities while initially very small increases with time.

The present study has drawn on a number of related analytical developments from which many elements of the goals accounting system were derived. These related developments are mentioned briefly here and are discussed in greater detail later on in this paper.

Thus, the present work draws on the earlier goals research at the National Planning Association[2] which was aimed at relating a set of broadly defined national goals to the resources of the national economy. The critical element adopted from that line of analysis is the use of the resources base of the entire economy, and of activities of both public and private sectors. In all probability, at the level of output presently considered, a partial analysis confined to either the market economy or the households alone or only to public programs at the different levels of government, would be completely incapable of assessing the full set of production possibilities, because the inputs from the private sector and the public sector activities are highly complementary in production of improvements in the rates of social change.

The variables defining the goals outputs are in a large measure derived from the social indicator work, which in its modern embodiment was developed by social scientists in the late 1950s, and has continued since, finding expression in official governmental reports.[3]

Another component has been derived from new approaches to the analysis of public expenditure[4] and from the new theories and research in the field of consumer expenditures.[5] Both of these developments, in different ways, were concerned with the objectives for which the expenditures are made. These analyses helped in formulating the activity-output approach underlying the goals accounting system, and in selecting, from among the much wider range of social indicators, those that operationally could be considered to approximate most closely the objects of household and collective consumption. Current fiscal analysis also helped to define the concept of discretionary resources.

The goals accounting system is also related to work in economic projections and technological forecasting because it involves estimates of future resource supply and projections of the output coefficients, and to futurism in that it deals with contingent trends.

II. SUMMARY OF THE ESTIMATES FOR THE PERIOD 1973-83

1. Components of the Goals Accounting System. The analytical system for estimating the range of possibilities for producing discretionary social change consists of the following five elements:
(a) Selection of areas of social concern such as health and public safety and identification of quantitative indicators such as the average life expectancy and the rate of violent crime to

measure conditions that are the main objects of a given concern;
(b) Projection of ten-year trends in the indicators selected. These trends serve as the base from which discretionary changes may be produced in year ten;
(c) Identification of discretionary activities, their costs and their effects on the conditions measured by the indicators;
(d) Distinction between fixed and discretionary uses of economic resources on the part of individuals, private institutions and governments, and a ten-year projection of resources available for discretionary activities divided into two subperiods and into private and public sector components;
(e) Calculation of the maximum feasible output of combinations of discretionary activities that can be undertaken with the estimated resource supply.

The goals accounting system has been designed to be open-ended, so that it could readily accommodate changes in any of its components, such as addition or deletion of areas of concerns or of indicators, selection of activities and estimation of their possible effects and costs, and projections of resources. The estimates are tentative and preliminary.

As mentioned, the full rationale and discussion of this design and of the specific selection made is being provided in the forthcoming volume and cannot be described extensively here. Here I would like to summarize briefly the results obtained in the present round of estimates which were calculated for the period 1973-83 with the dollar magnitudes expressed in 1973 prices. These estimates supersede the quantitative results provided in the earlier articles.

Areas of Concern and Indicator Trends. In the present round of estimates, the set of social concerns selected for study and the indicators used to represent them were very similar to those used earlier. Table 1 lists the areas of concern and the indicators. It also shows the past, current and projected levels of the indicators used. (The present tabulation differs somewhat from those published earlier. The principal change from the past selection of goal categories is that the economic growth measured by GNP is no longer included as a goals category. This was done because economic growth requires different and a more complete analysis in its own right. But, the calculation of possible effects of the discretionary activities on economic growth is continued. An additional indicator is included for basic education representing dispersion of achievement of basic skills. Most indicators are now normalized for population size, and defined in terms of percentages or averages. In a few cases, however, it was too late to revise the entire set of estimates.)

The list of the areas of concern represents a set of domestic social concerns which require both individual and collective activity and resource expenditure for their support. The indicators are intended to represent the main dimensions of these concerns. The indicators represent a set of more basic and, therefore, probably also more stable variables than either quantities of goods and services purchased by the consumer units or the measures of volume of governmental programs and services, both of which serve the objectives represented by the indicator variables.

Table 1

A Summary List of Concerns and Corresponding Indicators, 1960, 1973 and 1983

Concerns	Principal Indicators	Indicator Levels 1960	1973 Estimate	1983 Projection
I. Health and Safety				
Health	Average life expectancy at birth years	69.7	71.3	72.7
	Percent of population with activity limiting disabilities	15.0	17.5	16.4
Public safety	Number of violent crimes per 100,000 persons per year	265	668	668
II. Education, Skills and Standard of Living				
Basic education	Index of performance in grade 12 based on standard tests--1973=100	n.a.	100	105
	Percent of students 3 or more years behind 1973 average	n.a.	24	19
Higher education	Number of persons completing college, thousands	392	957	1,342
Ability to earn	Number of workers not in the mainstream of labor force, millions	n.a.	11.1	8.8
General level of earnings	Median annual wage and salary earnings of individuals, thousands of 1973 dollars	4.8	5.9	7.8
III. Income				
Adequacy of income	Percent of population below the 1973 poverty standard	22.1	11.4	8.7
	Percent of population in near-poverty conditions based on 1973 definitions	8.1	4.8	3.5
Continuity of income	Percent of population with living standard loss of over 30%	n.a.	8.6	8.7
IV. Economic Equality				
General economic equality	Income ratio: 20th as percent of 90th percentile	20	25	28
Economic equality of races	Mean family income, Negroes as a percent of Whites	56	65	71
Economic equality of sexes	Hourly earnings of women as percent of earnings of men	n.a.	60	60
V. Human Habitat				
Housing	Percent of persons living in adequate houses	n.a.	88	92
Neighborhoods	Percent of persons living in satisfactory neighborhoods	n.a.	77	87
Pollution control	Percent of population exposed to bothersome pollution	n.a.	62	46
Outdoor recreation	Percent of persons 12 years and older taking part in outdoor recreation regularly	n.a.	21	54
Preservation	Index of preservation of life and natural forms	n.a.	100	110
VI. Art, Science, and Free Time				
Discretionary time	Discretionary time--hours per person per year	n.a.	2,111	2,199
Science	Number of scientists active in basic science, thousands	n.a.	81	139
The arts	Number of active artists, thousands	206	265	323

Discretionary Activities, their Potential Output and Cost. The distinction between discretionary and nondiscretionary activities and the resulting use of resources is reflected in the identification of discretionary activities which have a potential to produce indicator output and which could be undertaken over the period 1974-83, and in the projection of supply of discretionary resources over this period. The discretionary activities were derived from a survey of existing policy proposals and of analytical studies in the respective fields of social concern. Their list is meant to include all the major possible activity innovations or extensions of existing activities which are technically and culturally feasible and which could have some positive effects on any one or more of the indicators chosen. Activities with negative effects are not included. The particular effects of the discretionary activities and their ten-year costs, specified by subperiod and fiscal source, were estimated at their judged full capacity level beyond which they would not contribute productively. The estimates of activity effects and costs were arranged in two matrices which are shown as Tables 2 and 3, respectively.

Table 2 contains the activity-output matrix summarizing the effects of the 28 discretionary activities on the 22 indicators at full capacity. (GNP is an addendum item.) However, multiple activities sometimes may interfere with each other or one may be sufficient for achievement of a particular level of change. For these reasons the net effect of multiple activities on a given indicator may be less than the sum of the individual effects of the same activities. The maximum effect of all activities on each of the indicators is shown on the last line of Table 2.

The activity-cost matrix is shown in Table 3. The elements of this matrix are the private and public sector cost of the discretionary activities in each of the two subperiods. These distinctions are made in order to improve the realism of estimation because the resources clearly are not readily transferable between private and public sectors or between early and later periods. For many activities, substantial early resource uses are required in order to achieve the result at the end of the ten-year period. Two subperiods are defined to allow for these patterns.

The activity-output matrix and the activity cost matrix define the technical possibilities. The degree to which the technical possibilities are economically feasible is determined by the economic constraints which are estimated from a projection of discretionary resources.

The concept of cost of discretionary activities is based primarily on real resource costs. However, certain transfer amounts representing a 10 percent addition to the real GNP components are included because production of changes in certain goal indicators, such as reduction of the proportion of the population living in poverty conditions, may be accomplished by means of transfer of the economic output rather than its net production. The amount of these transfers included in the resource projection is constrained by the particular assumption made in the underlying economic projection.

Resource Projection. The first step in making the ten-year projection of resources which could be available for the discretionary activities was to identify within the national income accounts the portion of the economic totals which corresponds to expenditures in support of the goal categories included in the goals accounting system. The second step was to find the corresponding items within the current NPA economic projection. How this was done is described in full in the forthcoming volume. The third step consisted of identifying that portion of the total expenditures related to these goal categories which is necessary to support the projected base trends in the output indicators or which is in effect committed to ongoing activities and the residual which represents the discretionary margin. This residual is the estimated supply of resources available for performing the discretionary activities.

In the first step it was established that approximately 56 percent of the total economic resources (GNP plus the additional 10 percent of transfers) represented expenditure for social goals. By analyzing elements of the ten-year economic projections made by the National Planning Association, the past growth trend of the share of social expenditures relative to GNP and to the GNP plus transfer components was identified and projected. The discretionary proportion of the total expenditure for social concerns was estimated by an extension of the fiscal dividend analysis. A mathematical formula was developed which describes the momentum of the ongoing commitment and its weakening over time. The critical parameter of the formula which describes the rate at which the committed resources become released with extension of the time horizon, was estimated by reference to existing official and private[6] analyses of the federal government budget and private[7] analyses of state and local budgets. The public sector discretionary resources were estimated then from the formula. A further extrapolation was made for the private sector by reference to earlier and different studies of consumer expenditures for necessities and for service of fixed commitments.[8] The formula results are further constrained by a requirement that sufficient resources to support the base trends in the indicator levels remain nondiscretionary. The final estimates of discretionary resources and their composition by sector and subperiod are shown in Table 4.

In addition to this standard projection, alternative projections were made embodying various assumptions about the rate of economic growth and about shifts of resources between social and other objectives of spending. Neither seems to have any appreciable effect on the amount of discretionary resources available. However, changes in the rate which expresses the degree to which resource commitments are fixed could greatly affect the resource availability. The amount of discretionary resources is very small compared to total amount of resources available in the economy for the present group of objectives.

Estimating the Range of Feasible Changes. The final component of the analysis is calculation of the range of changes that are feasible both technically and economically. Technical feasibility is given by the coefficients in Table 2. The economic feasibility is limited by the resource availability which was estimated in Table 4.

In calculating the cost of simultaneous changes in diverse conditions represented by 21 different indicators, some rule is needed to select them systematically from the almost infinite variety of the possible combinations of changes. It is not possible given the present state of the art--and probably would not even be desirable--to collapse the distinct dimensions of the quality of life represented by the 22 indicators into a single index. The rule followed presently was to take uniform percentages of the technically feasible change, (shown in the last line of Table 2), and to calculate the least cost of that combination for different percentage levels of the technical potential first for all indicators and then for groups of indicators reflecting the more specialized concerns, such as with the environment or with education, skills and income.

The least cost was calculated for different levels of each combination until the cost of the

the combination approximated the amount of resources available.

The least-cost estimates were made by an adaptation of linear programming which was developed in the present study to allow for the non-additivity of effects of individual activities on an indicator.[9] The method consists of using an expanded version of the activity-output matrix (not shown here) with different columns applicable to different levels of indicator output.

Under the strict assumption that each subperiod and fiscal component represents a distinct resource constraint on how much change can be achieved, it was calculated that only 16 percent of the technical potential would be possible

Table 2

EFFECT OF ACTIVITIES ON GOAL OUTPUT INDICATORS, 1974-1983

GOAL OUTPUT INDICATORS

ACTIVITIES	Total Cost: 1974-1983 - 1973 dollars - billions	Health & Safety			Education, Skills, and Earnings					Adequacy and Continuity of Income			Economic Equality			Human Habitat					Arts, Science, and Free Time			GNP
		1. Average life expectancy at birth - years	2. Percent of population with major disabilities	3. Number of violent crimes per 100,000 persons per year	4. Index of mean performance in grade 12 based on standard tests	5. Percent of students 3 or more years behind 1973 average	6. Number of persons completing college - thousands	7. Number of persons not in the mainstream of labor force - millions	8. Median earnings of individuals, 1973 dollars - thousands	9. Percent of population below poverty standard	10. Percent of population in near-poverty conditions	11. Percent of population with living standard loss of over 30%	12. Family income ratio: 20th to 90th percentile	13. Mean family income, Negroes as a percent of Whites	14. Hourly earnings of women as percent of earnings of men	15. Percent of persons living in adequate housing	16. Percent of persons living in adequate neighborhoods	17. Percent of population affected by bothersome pollution	18. Percent of persons regularly taking part in outdoor recreation	19. Index of preservation of life and natural forms	20. Number of scientists active in basic science - thousands	21. Number of active artists - thousands	22. Discretionary time - hours per person per year	23. GNP - 1973 dollars - billions
Base 1973	--	71.3	17.5	668	100	24	957	11.1	5.9	11.4	4.8	8.6	25	65	60	88	77	62	21	100	81	265	2111	1275
Base 1983		72.7	16.8	668	105	19	1342	8.8	7.8	8.7	3.5	8.7	25	70	60	92	87	46	54	110	139	323	2199	2033
1. Change in health-related habits and patterns	$64	5.3	-3.3	-129								-.9							25				57	
2. Health services related to specific conditions	66	1.7	-3.1	-69								-.4											53	
3. Special health services for vulnerable population groups	91	2.5	-1.0																					
4. Improvement of law enforcement systems	26			-180																				
5. Employment and other opportunities for the young	51			-240	5	-4		-1.5		-.9	-.4			2										
6. Remedial and augmenting educational inputs	73				16	-11	50		.1				2	1										44
7. Improved educational technology and approaches	183				21	-14	50		.2				1											60
8. General day care for children	126							-1.8		-1.7	-1.3		2		3								117	43
9. Universal access to higher education	273						1050		.2		-.4		1								20	32		29
10. Structural improvements in higher education	70						350		.1												20	11		10
11. Maintenance, updating and improvement of job skills	342						650	-3.4	.2	-.9	-.9		2	4	6						10	16		34
12. Specialized training for those outside mainstream of labor force	94							-5.0		-1.3	-.4	-.4		3										21
13. Private savings, insurance, pension plans	200									-.9	-.6	-4.3	1	1										
14. Old age pensions at 40% of current median earnings	30									-1.7	-.6	-2.6	1	1										
15. Extended welfare program-tax and transfer to abolish poverty and near-poverty	76									-8.7	-3.5	-1.3		3										
16. Aid to depressed communities	171							-1.8	.2	-.9	-.9		1	2		4	5	-2	5					17
17. Construction and maintenance of houses	108									-.4	-.4			1		8	5	-2						
18. Design and testing of new neighborhood, city and regional environments	202															2	10	-9	3					
19. Innovations in cars, roads, and other transportation system components	155																	-4					71	
20. Pollution control	171																5	-29						
21. More basic environmental improvements	332																	-17	4					
22. Recreation facilities in neighborhoods	127	1.1	-.4														5		12					
23. Major parks and facilities	80																		5	10				
24. Preservation of wilderness and scenery	26																			50				
25. Pure science - institutions, education, communication	36																				51			
26. The arts - institutions, education, subsidies, new forms	28																					300		
27. Reduction in working time	107																						60	
28. Time-saving innovations	91								.4				1		2				4				319	50
Total listed (output not additive)	(3399)	(8.9)	(-6.8)	(-448)	(27)	(-16)	(1300)	(-7.2)	(1.4)	(-8.7)	(-3.5)	(-6.9)	(9)	(12)	(11)	(8)	(13)	(-37)	(28)	(60)	(81)	(315)	(677)	(160)

Table 3

ESTIMATED COST OF DISCRETIONARY ACTIVITIES: 1974-1983 TOTAL

Distribution by Subperiod and Private and Public Financing Requirements

ACTIVITIES	Total Cost: 1974-1983 -- billions of 1973 dollars	Components of Total: Public Outlays 1974-1977	Components of Total: Private Outlays 1974-1977	Components of Total: Public Outlays 1978-1983	Components of Total: Private Outlays 1978-1983	As Percent of Total Cost: All Outlays 1974-1977	As Percent of Total Cost: All Public Outlays
1. Change in health-related habits and patterns	64	17	16	15	16	52%	50%
2. Health services related to specific conditions	66	11	6	33	16	26	66
3. Special health services for vulnerable population groups	91	30	0	61	0	33	100
4. Improvement of law enforcement systems	26	10	1	15	0	41	96
5. Employment and other opportunities for the young	51	7	2	34	8	17	80
6. Remedial and augmenting educational inputs	73	11	1	55	6	17	90
7. Improved educational technology and approaches	183	7	0	158	18	4	90
8. General day care for children	126	17	4	84	21	17	80
9. Universal access to higher education	273	5	0	268	0	2	100
10. Structural improvements in higher education	70	11	2	53	4	18	91
11. Maintenance, updating and improvement of job skills	342	46	22	183	91	20	67
12. Specialized training for those outside mainstream of labor force	94	9	0	85	0	10	100
13. Private savings, insurance, pension plans	200	16	64	24	96	40	20
14. Old age pensions at 40% of current median earnings	30	0	0	30	0	0	100
15. Extended welfare program-tax and transfer to abolish poverty and near-poverty	76	0	0	76	0	0	100
16. Aid to depressed communities	171	18	3	136	14	12	90
17. Construction and maintenance of houses	108	14	17	36	41	29	46
18. Design and testing of new neighborhood, city and regional environments	202	16	16	85	85	16	50
19. Innovation in cars, roads, and other transportation system components	155	1	7	27	120	5	18
20. Pollution control	171	12	12	76	71	14	51
21. More basic environmental improvements	332	75	8	224	25	25	90
22. Recreation facilities in neighborhoods	127	15	10	61	41	20	60
23. Major parks and facilities	80	13	3	51	13	20	80
24. Preservation of wilderness and scenery	26	6	0	18	2	25	92
25. Pure science - institutions, education, communication	36	4	1	24	7	15	79
26. The arts - institutions, education, subsidies, new forms	28	3	1	16	8	15	69
27. Reduction in working time	107	0	0	21	86	0	20
28. Time-saving innovations	91	1	6	17	67	8	20
Total listed (output not additive)	3,399	375	202	1,966	856	17%	69%

across the board in all the indicators. The limiting resource for all indicators together, as well as for the specialized groups of indicators, is the amount of discretionary resources available in the public sector in the first subperiod (defined to consist of the first four years). In order to obtain an appraisal of the maximum amount of change that might be available,

Table 4

Resources Available for Discretionary Activities, 1974-1983, by Subperiod and Sector (in billions of 1973 dollars)

Period	Available in the Private Sector	Available in the Public Sector	Total
1974-1977	$82	$19	$101
1978-1983	1,208	399	1,607
Total for 1974-83	1,290	418	1,708

additional assumptions were made as to the extent to which the resources available in the public sector in the early subperiod could be augmented. After consideration of different possibilities, such as increases in taxation, borrowing or governmental productivity or development and substitution of private for public activity components, it was assumed that a total of $40 billion could be used as a limit on the achievement of outputs in 10 years. With the $40 billion in discretionary public sector resources as the effective constraint, 30 percent of the technically feasible improvement in each indicator was calculated as economically feasible. These 30 percent changes are shown in Table 5, together with the trend levels and the trend changes for the period. Because the activities have multiple output effects, more than 30 percent can be achieved for some indicators. But the table does not show such additional changes that would occur as a result of requiring achievement of a minimum 30 percent of the technical improvement potential in each of the indicators. The use of the matrix stage method to accommodate non-additivity of the effects in the least-cost calculation does not permit accurate calculation of these additional "slack variable" outputs.

But, even using only 30 percent, the amount of change that is both technically and economically feasible is quite substantial in relation to the trends. For 14 of the 22 indicators, it approximates or exceeds the amount of the autonomous trend improvements. When only some rather than all indicators are required to change, higher proportions of the potential are achievable. Therefore, the feasibility of achievement of subsets of goals such as those limited to health and safety or to the quality of the environment is at a higher percentage level. Those specialized estimations are not discussed here, but they are included in the forthcoming book.

However, it is very important to realize that, even if the numbers are assumed to be correct, the maximum feasible potential for change is not an estimate of the probable outcome. Indeed, in a non-market situation represented by large-scale activities which really are not within the scope of an optimizing decision mechanism, the possibilities for losses of efficiency are present at many stages of the implementation of activities as well as in choices of their components. The full range of possible outcomes then, with the base projection taken as given, is between an

Table 5

Changes in Levels of Goal Achievement: 1973 Base, 1983 Projection, and the Maximum 1983 Levels (Set at 30% of Technical Potential) Consistent with Judgmental Maximum Resource Supply of the Limiting Resources

Indicators	1973 Base Estimate	1974-83 Base Increment	1983 Trend Projection	Maximum 1983 Level Assuming Achievement of a Minimum of 30% of the Technical Potential: Increment over Trend Projection	Total
1. Average life expectancy at birth - years	71.3	+1.4	72.7	+2.7	75.4
2. Percent of persons with major disabilities	17.5	-1.1	16.4	-2.0	14.4
3. Number of violent crimes per 100,000 persons per year	668.0	0	668.0	-134.4	533.6
4. Index of mean performance in grade 12 based on standard tests	100.0	+5.0	105.0	+8.1	113.1
5. Percent of students 3 or more years behind 1973 average	24.0	-5.0	19.0	-4.8	14.2
6. Number of persons completing college - thousands	957.0	+385.0	1,342.0	+390.0	1,732.0
7. Number of persons not in main-stream of labor force-millions	11.1	-2.3	8.8	-2.2	6.6
8. Median earnings of individuals in 1973 dollars-thousands	5.9	+1.9	7.8	+0.4	8.2
9. Percent of population below poverty standard	11.4	-2.7	8.7	-2.6	6.1
10. Percent of population in near poverty conditions	4.8	-1.3	3.5	-1.0	2.5
11. Percent of population with living standard loss of over 30%	8.6	+0.1	8.7	-2.1	6.6
12. Family income ratio: 20th to 90th percentile	25.0	0	25.0	+2.7	27.7
13. Mean family income, Negroes as a percent of Whites	65.0	+6.0	71.0	+3.6	74.6
14. Hourly earnings of women as percent of earnings of men	60.0	0	60.0	+3.3	63.3
15. Percent of persons living in adequate housing	88.0	+4.0	92.0	+2.4	94.4
16. Percent of persons living in adequate neighborhoods	77.0	+10.0	87.0	+3.9	90.9
17. Percent of population affected by bothersome pollution	62.0	-16.0	46.0	-11.1	35.1
18. Percent of persons regularly taking part in recreation	21.0	+33.0	54.0	+8.4	62.4
19. Index of preservation of life and natural forms	100.0	+10.0	110.0	+18.0	128.0
20. Number of scientists active in basic science-thousands	81.0	+58.0	139.0	+24.3	163.3
21. Number of active artists-thousands	265.0	+58.0	323.0	+94.5	417.5
22. Discretionary time - hours per person per year	2,111.0	+88.0	2,199.0	+203.1	2,402.1

increment of zero and an increment of the 30 percent of the technically feasible change in all indicators, and of higher proportions for more limited groups of indicators.

This summary of the results contained in the forthcoming volume is quite brief. But rather than go into details which are or soon should become available, I would like to spend the remainder of time discussing some of the issues applicable to the system as a whole.

III. ECONOMIC ANALYSIS OF PRODUCTION APPLIED TO SOCIAL CHANGE

The goals accounting system is a result of application of economic analysis to the study of social change. Specifically, it is an application of economic theories of production and allocation to the questions of production of incremental social changes over a given time interval.

Application of economic analysis to a

noneconomic subject is not new. Formally, production of social change lends itself as much to economic analysis as any other kind of production. Indeed, present work draws on the considerable amount of economic analysis which has been applied already to many of the fields of social concern. The social concerns with income maintenance and with distribution and with aspects of human capital traditionally have been directly related to economic analysis and research. But, also the economics of some of the more specialized fields, such as health, has been reasonably well developed for some time.

The present study deals with the economics of production rather than of consumption and hence with the supply of rather than the demand for changes in the conditions represented by the 22 indicators. Consequently, the study is based on the economic theory of production. The production and the cost functions are defined by the two matrices which were shown in Tables 2 and 3. Technological conditions expressed in these functions are given exogenously by a survey of technological opportunities. This survey is described in detail in the forthcoming publication. The parameters of the production function include the activity-output and activity-cost coefficients, capacity limits of activities, and the interaction coefficients representing the nonlinearities of production by multiple activities which are included in the multi-stage matrices.

The present production function has a few specific characteristics. First, it is an incremental production function defined for a ten-year period for the discretionary increments of changes in the conditions measured by the indicators. These changes are in addition to changes given by nondiscretionary trends. These trends are treated as given in the use of resources, in the content of ongoing activities, and in the output of the production processes for which the resources are used.

The production function is an activity analysis function. For each activity a capacity level is defined. For activities taken singly, the output is proportionate to the level of activity from zero to the full capacity level. At levels beyond capacity the output remains the same as at the capacity level.

The production function deals with marginal products achievable over the time interval. With this focus of the question, there is no need to explain the base trends or the initial conditions represented in them.

Even though constant returns were assumed within each activity, diminishing returns prevail in production of change in every indicator between the activities even in the absence of interactions, because activities differ in the productivity (i.e., cost-output ratios), for each indicator output, and because more than one activity affects each output. Therefore, the least-cost expansion path for each indicator follows a diminishing returns path (though theoretically it could be constant for some) consisting of segments of progressively lower productivity. The diminishing returns to individual indicators are strengthened by the circumstance that for many indicator outputs there are negative interaction effects among activities.

Another feature of the production function is existence of pervasive joint products even at the present very high level of aggregation. The activity output matrix cannot be partioned at all into any submatrices.

The level of aggregation embodied in the production function is very high. Unlike in microeconomic analyses where the individual production processes are well within control of single firms or even departments of individual firms, there are no decision units corresponding to an activity in the present case. The activity is an analytical construct not a counterpart to an organization. In practice, many decision units would correspond to any one of the present activities and there would be much strategic behavior among them, as well as uncertainty, lack of information, time lags and other impediments to optimal equilibrium.

The budget constraint which defines the locus of feasible transformation possibilities is given by the projection of the discretionary resources.

Conceptually, present analysis addresses only the supply side of economic relationships, in that it deals with estimating the production possibilities frontier. It leaves out the demand side which could help identify optimal locus on the possibility frontier. The primary application of the goals accounting system, therefore, lies in efficiency analysis. The methods here developed (assuming that the data used is reliable or materially more reliable than alternative information) may help to establish whether or not a given point lies on the efficiency frontier or within it thus incorporating waste.

However, economic analysis can illuminate the demand side by contributing some criteria to definitions of categories of social concerns and output indicators or at least some criteria for distinguishing among alternative categorizations and measurements.

IV. IDENTIFICATION OF AREAS OF CONCERN AND OF THE INDICATORS TO MEASURE CONDITIONS IN THOSE AREAS

1. Level of Definition. The areas of concern and their indicators are defined at the level at which the objectives of collective and individual activities can be quantified, the resource uses identified, and the productive relationships estimated. Output defined at this level is presumably more fundamental and more closely related to utility of individuals than the output measured at the level of conventional goods and services exchanged in the market or at the level of workload items of public programs.

Most categories of social concern used in the goals accounting system are readily recognizable from other uses. They include, among others, health, basic education, safety, and recreation. Also, many of the indicators chosen as summary measures of the conditions in these categories are well known statistics, such as mean life expectancy at birth, average earnings of working individuals in constant dollars, or proportion of population below the poverty line. Others, such as the proportion of workers not in the mainstream of the labor force, or the adjusted rate of violent crime, are derived from more or less well-known public or private statistics. A few indicators were simply outlined or postulated, where no statistics exist. This group includes proportion of population affected by pollution, proportion of population living in satisfactory neighborhoods and the index of conservation.

At this level of definition the categories of concern are not oriented to the even more fundamental but intractable categories of values. The categories and the indicators do not address the fundamental, cultural or moral values or underlying principles of behavior. They are not on that plane. Rather, they represent definitions and measurements at a level intermediate between the conventional economic quantities (real dollar expenditures, man-hours worked, indexes of goods and services produced, etc.), on the one hand and either the fundamental principles of maximizing behavior of individuals (which in the economic literature has been called "utility") or basic and difficult intangibles (happiness, enlightenment, salvation) on the other.

The present system of categories and the indicators they represent attempts primarily to extend the application of economic analysis at least on the production side from the objects of market transactions and the amounts of taxes and public spending to a more generalized process of social production in which the objectives of individual and collective consumption which are measured by aggregate indicators are treated as goods and the market and public goods and services are treated as intermediate inputs.

The present analysis was developed at the time of and is based in part on analytical developments which were motivated by the dissatisfaction with the limitations of and desire to generalize such traditional formulations of the performance variables of the economy as the average family income or the economic growth rate. In going beyond real product and income, the indicator output variables of the goals accounting system attempt to represent the real quantities of some of the results of what the individuals and social institutions, including governments, presumably spend their income for.

Perhaps in the conditions of great economic scarcity, with consumption of necessities absorbing all or nearly all of income, measurement of income was, and is, a sufficient priority indicator for judgments about the level of economic and perhaps even social well-being. But, at higher levels of income, such as now attained in the United States, direct measurement of objectives of expenditure and a long-term view of these objectives and of income are appropriate. Many intellectual developments contribute to interest in direct measurement of the standard of living by means of the indicators of "quality of life."

There is now a much increased understanding of the difference between the expenditure of resources for a purpose and accomplishment of that purpose and of possible difficulties in translating expenditure into its purpose. This distinction is basic to two analytical developments which pursue it in private and in public sectors, respectively: the new economics of consumption and public program analysis. These developments had a large influence on the formulation of the activity-output and the activity-cost matrices included in the present system.

In the public sector, the initial concern was with attempts to estimate the economic value of benefits produced by public works in order to compare them with cost.[10] The analysis was extended to comparative costs and effects in defense expenditures and finally to the entire budget of the federal government with considerable adaptations in the state and local governments.[11] Underlying this movement was the growing recognition in and out of government that expenditure of money is not a sufficient definition of results of a public program but that these results should be measured in terms more directly relevant to the specific policy objective and presumably to the utility of individuals for whose benefit the program is conducted--in the case of public goods the entire population. While a satisfactory measurement of the public sector output may still be some time in the future, some beginnings have been made in developing useful real output indicators for some government programs.

New research in the consumer sector has been concerned with defining objectives of consumer expenditure. This analysis is much more empirically oriented and has a much firmer theoretical foundation than the analysis of output in the public sector.[12] The more traditional analysis of the consumption process stopped with the act of purchase by consumers of a given commodity. The new approach focuses on the consumption process which entails use of purchased goods along with work, time and other resources in activities on the part of members of households (which could consist of single individuals). The objective of this process is to transform the range of purchased commodities and other endowments and resources of the household into outputs or outcomes such as health or care of children. These outcomes or outputs are more directly related to the utility and hence the objectives of the households than the commodities and other inputs used in their production. This process has been called the household production process. At the same time, it was also realized that individual commodities possess multiple characteristics and that these characteristics when combined with other commodities embodying their own distinct sets of characteristics, can be used as inputs in producing the more desirable combinations of characteristics, which either directly or indirectly can serve in the household production activities.[13] The existence of multiple characteristics has also been used in the analysis of price differentials among the nonhomogeneous consumer and capital goods.[14] The present formulation of multiple activities with multiple outputs is analogous to this concept.

2. <u>Partial Coverage</u>. Even at the level of measurement which it addresses, the coverage of the goals accounting system is not complete. Less than the full range of national expenditures is represented.

As already indicated, the categories of concern and their output indicators were defined at one particular level. The present study has explored the extent to which the different types of social changes lend themselves to analysis within the framework of economic production. At the beginning of the present research, attempts were made to include categories of freedom, justice and harmony. Search was made for some at least partial indicators for these categories with which estimates of a set of productive relationships for conditions represented by these indicators could be developed. These attempts were not successful. The difficulty was encountered at the first stage of specifying an indicator which would satisfy the criteria of general or at least wide acceptability of its relationship to the object it was intended to measure. Therefore, the second level of difficulty which might arise in identifying productive relationships for the activity-output matrix and the activity-cost matrix was not encountered. It was possible, on the other hand, to specify indicators and, at least to some extent, the productive relationship for possible changes in the several dimensions of economic equality which were included as areas of social concern.

All output indicators are assumed to have a recognizable normative direction. As objects of household and collective production, they should have positive marginal utility. At this time, it was postulated that they do, but no validation test has been performed. The general, and perhaps even universal prevalence of positive marginal utilities can probably be argued, a priori, in case of objects of household production, such as health. Such hypotheses of positive marginal utility are testable by the now available empirical methods. But the goals accounting system also includes indicators representing the distributional changes as well as public goods (e.g., science) which are not in any major way objects of observable economic household behavior. Existence of a consensus for valuation of the direction of change is now argued ad hoc for each of these categories. It is not clear whether any empirical validation tests based on observable behavior could be designed at present for these categories.

Coverage of activities is also partial. The list of 28 activities included in the activity output matrix does not correspond to the full range of plausible futuristic scenarios. Not included among the present activities are changes with negative effects on any of the indicators or on basic social values. Also not included are any of the broad changes in the society and culture, such as, for example, changes in the general social climate regarding expectations for the future or in the extent of trust existing among the members of society. Level of trust present in the culture, it has been observed, contributed much to economic growth, for example, through the overwhelmingly voluntary compliance with contracts. Undoubtedly, changes in the level of trust would affect productivity in achievement of social outcomes here discussed. But such changes in culture are autonomous or at least not within the scope of economic production activities.

3. Selection of Categories of Concern. The specification and selection of categories is limited with regard to level and scope. The choice of level, intermediate between conventional goods and services and the basic values was already discussed.

In order to limit the amount of work, it was decided to focus on a group of interrelated concerns and to limit the selection of the areas of concern to domestic social fields, excluding international relations and national defense categories on the one hand and certain business investment and consumption items, mainly food and clothing, on the other. Also excluded were the concerns with the overall performance of the economy, i.e., employment, price stability, international balances, and economic growth for which well-developed analytical models already exist, which at some point could perhaps be linked with the goals accounting system. In any case the selection made was thought to have sufficient variety of concerns to permit study of achievement of multiple goals simultaneously.

The categories of concern were determined primarily by reference to earlier categorizations made by others on other occasions. These included categorizations used in the analysis of consumer expenditures and the functional categories of public budgets;[15] in earlier goals research at the National Planning Association which in turn followed categorizations in the report of the Commission on National Goals of 1960;[16] in the social indicator field,[17] and in discussions of public affairs.[18]

The choice of some of the categories, such as health, public safety, basic education, and higher education was almost obvious. The reason for including others was less clear-cut, especially for such newer fields of articulated public concern as environmental quality or supply of basic resources. Pollution control was included but the basic environmental improvements and continuity of supply of necessary resources emerged as articulated issues too late to be incorporated in the body of this work. However, as developed, the goals accounting system can readily accommodate future additions, reformulations or deletions of areas of concern.

Some fields which were included, such as conservation, the arts and science, while well-recognized as objects of collective concern and production, are at present not directly relatable to the framework of household production. One of the issues raised for the future work in goals accounting is how to treat public goods (including distributional) categories of concerns in a theoretical framework based essentially on the household production model aggregated to the national level. One possibility is to omit categories which are not directly objects of household production and perhaps add those, such as food, which were not presently included. The result would be a consumption model with goods redefined as objects of consumption as given by the household production processes permitting perhaps some new useful analyses, e.g., of living standards or cost-of-living defined in terms of the more basic variables than income and quantities of goods and services. Another possibility, the one followed presently would be to include collective or mostly collective categories alongside those which represent directly objects of household production and hence direct indicators of well-being of families and individuals. Still another possibility would consist of incorporating the collective concerns into a household utility model defined not as elements of governmental activities, such as biomedical research or public schools which contribute inputs to the already recognized household production processes but the public goods, and distributional concerns. Whether such an attempt would be successful is not clear. In any case, the task would be extremely difficult, but perhaps not hopeless. That useful analytic connections may be feasible is suggested by the circumstance that such a seemingly exogenous variable as population growth has been shown to be, at least in part, directly related to the desired family size and age structure which is an object of household production.[19] Such an approach would undoubtedly require a more complex analytical model than the one embodied in the present set of matrices and a more general concept of resources and cost than is presently used.

Within the field of household production, economic analysis can help in identifying categories of expenditure. The analytical tool of demand analysis may be of help in identifying groups of expenditures corresponding to groups commodities within which economic substitution and complementarity relationships are strong while they are weak or nonexistent between its members and commodities not in the group. The available methodology of econometric research permits in principle to establish which of the alternative categorizations reflect existing behavior better or worse.

As far as public expenditure is concerned, it is easy to identify the formal budgetary categories of governments. Some divergence exists between the detailed institutional decision-making structures, such as governmental agencies, appropriation accounts and congressional committees acting on them and the functional objects of governmental expenditure, but at the summary level the degree of correspondence is considerable. Those categories of public expenditure which already correspond to objects of household production probably should be examined together with private expenditures for those objectives. It is not clear, however, whether formulation of public goods categories should be derived from existing organizational and functional divisions in public budgeting, or whether the criteria for their definition should be sought directly in individual and household behavior.

4. Selection of Indicators of Social Change. The procedure followed in the choice of indicators for the fields of concern was to examine the existing general statistics and the social indicator literature for each field. The objective was to represent the principal or at least one or more of the important concerns of the field, so that the indicators selected would correspond in an important way to the objective of the field of concern. For some fields, such as health, or economic equality of races, it was comparatively easy to identify the statistical indicators which have had wide currency and which would correspond in a reasonably good way to the principal consideration, even if they did not completely cover all the important aspects of the concerns reflected in the field. In the field of health, two indicators rather than one were considered necessary because the state of health cannot be described adequately by survival alone. A person may be in importantly different conditions of health or disability, and the second indicator, prevalence of disabilities was therefore included. Another and sufficient reason for including the second indicator was that the correlation between causes of disability and causes of mortality is comparatively weak and potential reductions in disability are not highly correlated with potential increases in life expectancy. The degree to which the indicator selected was correlated with an alternative or additional indicator was another criterion for deciding on the number of indicators. If two are highly correlated, little insight is gained and much statistical noise is introduced into calculations by including both indicators.

For some of the other indicators, while some precedent existed, the statistics were very limited or not available at all. In the case of basic education, an index of the average scores of tests of verbal ability, writing and mathematics at the twelfth grade was postulated as the indicator along with the proportion of students more than three years behind the base year average. But, the actual trend data had to be pieced together from different sources, and comparability of different sets of tests had to be assumed. As an indicator for the concern with the public safety, it was decided to use the violent crime rate. But it was necessary to adjust the reported crime statistics for underreporting in order to obtain an estimate of crimes actually occurring. The adjustment was based on the data in the victimization survey conducted in the mid-1960s.

In still other fields, it was necessary to formulate conceptually the kind of indicator which would represent the basic concern and either attempt to develop the actual measurements, which was possible to do for recreation, or to derive a very rough estimate from rather tenuous information, as in the case of pollution control, or finally simply to postulate an index without even specifying its content and to assume rough orders of magnitudes for possible changes in this index as was done for conservation. This last procedure probably would not be very useful in the future except perhaps at an intermediate stage for facilitating inclusion of new fields for which information is typically very deficient, and rough judgments about the magnitudes may be preferable to omitting the field entirely.

Methods developed in economic research probably can be of some help in distinguishing among alternative indicators and in decisions on how many indicators may be appropriate for a field. In particular, the so-called "hedonic" analysis which attempts to explain differences in prices of complex commodities embodying many diverse characteristics, such as houses or automobiles, may be applicable. It may help in distinguishing, for example, between formulation in terms of one set of characteristics vs. another which would give rise to alternative indicators. For example, an analysis of automobile prices attempted to compare the explanatory power of physical specifications and of the performance variables as determined through consumer reports.[20] In terms of present categories, such an analysis may help, for example, in formulating an index of the quality of neighborhood reflecting judgments of the prospective buyers of houses and tenants derived from alternative formulations of definitions for neighborhood characteristics while holding all the other characteristics of housing constant and analyzing the actual behavior of a representative sample of renters and buyers.[21] Such an indicator would represent a great improvement over the presently assumed indicator. Methods of economic research together with the survey methods and other techniques of social research could within the scope of their applicability inform and test the judgments made by academic researchers and government statisticians in developing indicators and in choosing among them.

V. TREATMENT OF TIME

Time defines the scope of the goals accounting system. In the estimates presently developed, the time period is 10 years. The output is counted at only one point in time, at the end of the period. Outputs at times before and after the 10th year are not measured. Cost is measured by the flow of resources used over the entire period up to the 10th year, but not beyond that year, with further differentiation of this flow between costs incurred in the first four years and in the later six years of the ten-year period. This distinction corresponds to the distinction made in the resource constraints and in turn reflect judgment that the substitutability of resources between the time periods is limited.

1. Time Dimensions of Output. The output analysis is basically static, though it is incremental and historical in the sense that it is built into a time series analysis and utilizes existing time series statistics. Output is seen only at a point in time. This is an abstraction, perhaps useful, but it entails a limitation. The time of availability of output is clearly an important element in preferences and a truly dynamic analysis which would treat output as a future time path for each indicator, presumably would be more realistic. Such more generalized analysis, however, is totally impractical at the moment. Analytical complexities alone would preclude it. Problems of calculation and numerical analysis would be formidable and possibly insuperable. No basis exists for postulating any pattern of time preferences which might permit some simplification, such as use of discount rates.

It is possible, though, to establish the intertemporal tradeoffs within indicators and among indicators for two or more points in time. Of course, the amount of estimation and the empirical content of such analyses would represent a manyfold expansion of present scope of research.

For that reason, after the completion of the present estimates, a pilot analysis was begun of intertemporal tradeoffs within a single field of concern, in basic education. This work is still in progress and the results are not available, but the basic structure of that analysis can be sketched out. It consists of two output indicators, one the same as presently used for basic education, i.e., the mean level of achievement on the three tests, and the other, slightly different, representing the proportion of students one year or more below the base year mean, both defined at twelfth grade. It deals with the three points in time: 4 years, 10 years and 25 years from the base year which gives estimates for 1977, 1983 and 1998. Because the outputs at each of the three points in time are distinct there are in effect six different outputs in the model. Six activities are identified which can increase the level of one or more of these outputs above the trend level. Some of the activities yield output at one point in time while other activities yield output at other points. For example, activities consisting of increasing learning in early childhood yield output only in year 25. With the given economic constraint estimated from projections of school age population, economic growth, and revenues available for financing of education, the tradeoff rates in production (transformation rates) among the different pairs of outputs at different points in time can be calculated.

This approach is probably sufficient to form judgments about the structure of the intertemporal tradeoffs. Increasing the number of points in time, which would be a way to approximate a truly dynamic analysis in addition to being very laborious does not appear likely to add much to the insight that could be gained at this time, but this is only a surmise.

2. Time Treatment of Inputs. The inputs are aggregated over the entire period and are also specified as to the time of their use between two subperiods. Clearly, some activities require longer initiation than others or need preparatory phases in order to yield output in a given year. These early phases may include research or simply a build-up phase of a large activity such as removing the financial obstacles to access to higher education. The existing capacity of the higher education institutions would not permit instantaneous admission of all potential applicants, even if the funds were available.

In order to recognize the differences in the time shape of activities, two subperiods were defined of four and six years, respectively, and the expenditures were aggregated over those subperiods. The reasons for choosing only two subperiods and of these particular lengths are as follows: the scope of the present study could accommodate only a few subperiods. Little would be gained by attempting a year-by-year analysis; clearly the data does not permit that fine a resolution. In fact, the empirical data made it very difficult to distinguish among even three subperiods as was attempted initially; two of them invariably got blurred. And yet, one period would not be enough because the time shape of resources availability is highly unequal between the early and later years of the ten-year period. The disparity in the discretionary amounts estimated to be available, say, in the third year and in the eighth year is practically enormous. Because the activities do have different time-cost profiles, the time-shape of the availability of resources is an important determinant of the economic constraint. Using aggregates for two subperiods as distinct inputs, was virtually the only available choice consistent with making some allowance for the differences in the early resource requirements among the activities, on the one hand, and the tremendous differences over time in the supply of the discretionary resources on the other. The length of four and six years was chosen by first reviewing the activity time profiles. It was decided that the initiation phase should be shorter than the implementation phase but longer than two years, because the very short initial period would be very vulnerable to the uncertainties in the actual estimation of the early resource requirements. Also, the very small size of the discretionary resources in the public sector available over such periods could easily be lost within the margin of fiscal fluctuations. The same problem but in lesser degree applies to the three years. For that reason a four-year initial period was chosen.

The treatment of cost is thus not truly dynamic either but begins to approximate it. At this point, it is far too early to contemplate a dynamic analysis of both outputs and cost. The ongoing study of basic education to which I referred, will deal with cost over several time intervals and may throw some additional light on the implication of treating both outputs and inputs as time sequences.

VI. RELATION TO POLICY-MAKING

If the experience with the preliminary results is a good guide, the goals accounting estimates are likely to be used in a wide variety of situations as a reference source of information. The particular applications would vary quite a bit with the characteristics of particular users and their purpose, as well as with the type of organization involved. Perhaps the methodology and the data of the goals accounting system could also be helpful in construction of analytical policy models in government, business, and other organizations. But no experience is available with such an application. In any event, the goals accounting system itself is not a policy model. I dwell on this point because it has been often misunderstood.

In part, the goals accounting system consists of a formal optimization (cost minimization) model for social changes measured by indicators representing long term national averages and aggregates. Many formal decision or policy models have also been designed in this form but with different variables. This similarity of form may give a misleading impression that the goals accounting system could be viewed as a decision model. It cannot. It has no utility function and even its efficiency results are limited by the scope of the model. There are also three fundamental empirical differences between decision models applicable to real decision units as they exist and the present goals accounting system.

Differences in Output Variables. The output criteria or output variables in the goals accounting system are very different from the output variables of actual decision units, governmental or private. The output variables in the goals accounting system include such aggregate summary indicators as the national average life expectancy at birth or average earnings per worker while the outputs of actual decision units, such as governmental agencies or program divisions within them concern much more intermediate variables such as, at the national level, research budgets of the different National Institutes of Health or number of persons entering the nursing profession, or at the local level the number and staffing of the school health units.

Theoretically, one could express national life expectancy as a function of intermediate variables representing outputs of a decision unit and of variables exogenous to the decision system of this unit by means of some function such as:

$$L = L(X_i, E_k)$$

where L represents life expectancy, measured perhaps best as an increment in life expectancy that can occur over a specified time interval, the X_i's, $i = 1, \ldots n$; stand for the output variables of the programs of the organization for which the policy model is being developed and E_k's; $k = 1, \ldots m$, are all the other variables which could affect the change in life expectancy. Normally, the effects of the individual X_i's would depend on each other, and on the effects of the E_k's.

Analyses in which an aggregate social indicator is expressed as a function of policy variables would be useful in many cases, and the present estimates might be of some help in conducting them. But, the very few explicit decision models that actually exist do not make such connections between the output variables of the decision units and the macro indicators. There is a serious and real issue regarding who should be developing such decision models. The agency managers may be correct in their reluctance to go beyond their clear mandate, leaving the task of making these connections to interested outside evaluators. There exists, for example, no stated policy that the government of the United States is attempting to maximize life expectancy, which would permit, in principle, a government agency responsible for a health program to convert its direct output variables X_i into the effects of their output variables on life expectancy, $X_i \frac{\partial L}{\partial X_i}$. Aside from the formidable estimating problems resulting from the instability of the $\frac{\partial L}{\partial X_i}$'s because of all the other effects on L, there are still greater difficulties regarding the normative validity of this translation because of differences in the public goods content, equity effects, risk and uncertainty of the individual $\frac{\partial L}{\partial X_i}$'s. Also, it is in the practical interest of maintaining the consensus necessary for the program, to measure its effects by the direct output variables rather than in terms of their effects on social conditions which usually are multiple and often controversial,[22] as well as in the (bureaucratic, but very real) interest of income maximization of the producing unit, i.e., of the budget of the given governmental agency.

This discussion is not meant as an argument against analysis of existing connections between public programs and social conditions, but an attempt to point out that even in the case of large programs of the national government the differences between the objectives of decision units and the macro indicators are fundamental, and the connection between them may be complex, subtle and uncertain. Analogous problems arise from differences in the time horizons.

Differences in Time Horizon. The present analysis deals uniformly with a projection of possibilities of change over a ten-year time interval. The actual decision units have a different time phasing which, on the one hand, seems to involve a much shorter time horizon and, on the other, implies, at least informally and in some inarticulated way, a principle of dynamic optimization where the output is perceived as an entire path in time. Intent of dynamic optimization is often expressed in public policy-making at any level of government; in business, the short-term maximization of accounting profits has been shown to be a comparatively weak and inaccurate explanation of corporate behavior also suggesting dynamic optimization. Finally, studies of household behavior regarding important long-term decisions, for example, in building a family or in choosing jobs, offer evidence of dynamic optimization on part of the households.

Yet, regardless of any normative argument, one may maintain that as a matter of fact for many units in the public sector, the decision-making horizon is shorter than ten years. Then, the kind of activities which are undertaken represent extensions along short-term supply curves, which consist largely of changes in the size of the operating staff and budget. Over the longer period, such as 10 years, the activities would normally embody substantial capital investment and research and development components which are not available for short-term results. The activities in the present goals accounting system were formulated in the long-run period. Therefore, there is no direct correspondence between the time horizons of typical decision units and of the goals accounting system in this regard.

Differences in the Decision Content of Activities. The large-scale activities included in the goals accounting system have no coherent decision mechanism. For that reason, the present goals accounting system resembles more a set of economic projections than a structured decision model. The only way to determine the degree of possible correspondence between the activities as here defined and the outcome of behavior of actual decision units would be to develop a theory of multiple decision-making with different output variables, time horizons and organizational frameworks and behavior patterns within the categories of the present activities and to estimate the dependence of the aggregate system on these microrelationships. While this view may help in making useful conceptual distinctions and may even suggest a desirable direction for future research, such task cannot be accomplished empirically in the foreseeable future.

In defining an activity, attempts were made to picture a realistic scenario under which such an activity could actually occur. These scenarios are not policy scenarios; they are rather contingent trends. For example, the physical fitness component of activity one would occur if, for reasons of health and recreation, large numbers of persons decided to pursue physical fitness and the public and business sectors provided complementary facilities to support such a developing demand. It does not mean that provision of the facilities by the public sector would bring about the interest in physical fitness on the part of individuals. In general, activities rest on the assumption of existence of complementarities between components from different groups of decision units.

I do not believe it would be practical to develop a general purpose policy system model by expanding the present estimates to a greater level of detail. Even if the data could be obtained, its complexity would be overwhelming and its validity questionable. On the other hand, I consider estimation of selected points along the possibility frontiers to be useful, because it provides measurements and information which could be useful in a wide range of applications in research and policy analysis, without being specialized to any particular type of decision units.

REFERENCES

*The author wishes to acknowledge the helpful comments and suggestions from Everard Munsey and Eduardo Rhodes. The Goals Accounting Study, in the course of which this paper was prepared, is supported by the National Science Foundation Research Grant No. GS-29032.

[1]Nestor E. Terleckyj, "Estimating Possibilities for Improvement in the Quality of Life in the United States, 1972-81," Looking Ahead, Vol. 20, No. 10, January 1973; idem., "Measuring Progress towards Social Goals: Some Possibilities at National and Local Levels," Management Science, Vol. 16, No. 12, August 1970.

[2]Gerhard Colm, "National Goals and the American Economy," Financial Analysts Journal, November-December 1964; Leonard A. Lecht, Goals, Priorities, & Dollars (New York: The Free Press, 1966); idem., "Dollars for National Goals," Looking Ahead, Vol. 21, No. 8, January 1974.

[3]Raymond B. Bauer, ed., Social Indicators (Cambridge: Massachusetts Institute of Technology Press, 1966); Eleanor Sheldon and Wilbert Moore, Indicators of Social Change: Concepts and Measurements (New York: Russell Sage Foundation, 1968); U.S. Department of Health, Education, and Welfare, Toward a Social Report (Washington, D.C.: U.S. Government Printing Office, 1969).

[4]Otto Eckstein, The Benefit Cost Analysis of Public Projects (Cambridge: Harvard University Press, 1957); U.S. Congress, Joint Economic Committee, Subcommittee on Economy in Government, The Analysis and Evaluation of Public Expenditures: The PPB System, Vols. 1, 2, and 3 (A Compendium of Papers submitted to the Subcommittee on Economy in Government of the Joint Economic Committee, Congress of the United States, 91st Congress, 1st Session) (Washington, D.C.: U.S. Government Printing Office, 1969); Charles Schultze et al., Setting National Priorities: The 1974 Budget (Washington, D.C.: The Brookings Institution, 1974).

[5]Gary S. Becker, "A Theory of the Allocation of Time," The Economic Journal, Vol. 75, No. 299, September 1965; Kelvin J. Lancaster, "A New Approach to Consumer Theory," The Journal of Political Economy, Vol. 74, No. 2, April 1966; idem., Consumer Demand (New York: Columbia University Press, 1971); Robert T. Michael and Gary S. Becker, "On the New Theory of Consumer Behavior," The Swedish Journal of Economics, Vol. 75, 1973; Nestor E. Terleckyj, ed., Household Production and Consumption (Studies in Income and Wealth, Vol. 40) (New York: National Bureau of Economic Research, forthcoming).

[6]Executive Office of the President, Office of

Management and Budget, The Budget of the United States, FY 1974 (Washington, D.C.: U.S. Government Printing Office, 1973); Schultze et al., op. cit.

[7]Ibid.; National Planning Association, Economic Projections, Report No. 71-N-2, November 1971.

[8]William Franklin, "Discretionary Income," (Technical Paper) (New York: National Industrial Conference Board, 1957); Morton Ehrlich, "Discretionary Spending," (Technical Paper No. 17) (New York: The Conference Board, 1966).

[9]Neil J. McMullen, "A Piecewise Linear Adaptation of Linear Programming for the Estimation of the Costs of Achieving National Goals," (A Technical Staff Paper) (Washington, D.C.: National Planning Association, forthcoming).

[10]Eckstein, op. cit.; Roland McKean, Government Efficiency through Systems Analysis, 1960.

[11]Charles Hitch and Roland McKean, The Economics of Defense in the Nuclear Age (New York: Atheneum, 1965); U.S. Congress, op. cit.; Harry Hatry et al., Measuring the Effectiveness of Basic Municipal Services: Initial Report, The Urban Institute and the International City Management Association, 1974.

[12]Becker, op. cit.; Michael and Becker, op. cit.; Terleckyj, ed., op. cit.

[13]Lancaster, 1966, op. cit.; idem., 1971, op. cit.

[14]Zvi Griliches, ed., Price Indexes and Quality Change (Cambridge: Harvard University Press, 1971); Jack E. Triplett, "The Theory of Hedonic Quality Measurement and Its Use in Price Indexes," (Bureau of Labor Statistics Staff Paper No. 6) (Washington, D.C.: U.S. Government Printing Office, 1971).

[15]For current categorizations of consumer expenditure, see U.S. Department of Commerce, Survey of Current Business, July 1973, p. 29; the functional budget categories of the federal government may be found in The Budget of the United States, FY 1974, op. cit., p. 67.

[16]Colm, op. cit.; Lecht, 1974, op. cit.; idem., 1966, op. cit.; Goals for Americans: The Report of the President's Commission on National Goals (New York: Prentice-Hall, 1960).

[17]Sheldon and Moore, op. cit.; U.S. Department of Health, Education, and Welfare, op. cit.

[18]Kermit Gordon, ed., Agenda for the Nation (Washington, D.C.: The Brookings Institution, 1968).

[19]Robert T. Michael and Robert J. Willis, "Contraception and Fertility: Household Production under Uncertainty," and James J. Heckman and Robert J. Willis, "Estimation of a Stochastic Model of Reproduction: An Econometric Approach," in Terleckyj, ed., op. cit.

[20]See Makoto Ohta and Zvi Griliches, "Automobile Prices Revisited: Extensions of the Hedonic Hypothesis," in Terleckyj, ed., ibid.

[21]For an actual analysis along these lines, see A. Thomas King, "The Demand for Housing: A Lancastrian Approach," ibid.

[22]Charles Schultze, The Politics and Economics of Public Spending (Washington, D.C.: The Brookings Institution, 1968).

SOCIAL INDICATORS AND A FRAMEWORK FOR SOCIAL AND ECONOMIC ACCOUNTS

Richard Ruggles and Nancy D. Ruggles
Yale University and National Bureau of Economic Research

The need for social indicators as a basis for government policy

In recent years governments have become increasingly concerned with improving the economic and social performance of society. Beyond providing for law and order and national defense, governments are recognizing that they must bear responsibility for a wide variety of social and economic conditions. They must concern themselves with the health of the population, the equality of educational opportunity, the eradication of poverty, and income security for the aged. They have become increasingly aware of environmental problems such as the need to control waste disposal and reduce the amount of air pollution. In brief there is much greater concern with a variety of dimensions of the quality of life.

The social indicator movement has been a response to these needs for information. Specific social indicators have been developed to monitor various aspects of society and to give some guidance to policy makers. The publication of Social Indicators by the Office of Management and Budget represents a first effort in this direction by the Federal government. It follows closely the sort of information contained in Social Trends, published by the Central Statistical Office of the United Kingdom. Provided systematically over time, social indicators can not only monitor social conditions, but can also be used to evaluate the effectiveness of government policy in specific areas. This characteristic is very important since, unlike the enterprise sector of the economy, the government cannot rely on profitability to judge the efficiency of its performance.

As the name implies, social indicators are generally oriented to non-monetary aspects of society, including demographic and regional information. Discrimination, for example, needs to be considered in terms of age, sex, race, and ethnic origin. Air and water pollution by their very nature must have a regional dimension. Much of the interest in social indicators focuses on differences in the quality of life in different localities or among different social and demographic groups.

There is, nevertheless, considerable dissatisfaction with social indicators as they are presently conceived, deriving from the fact that they are a miscellaneous collection of tables on social statistics, with no clear guidelines as to what precisely should be included or excluded. Ideally, the social statistics should be so designed that they would fit into a common framework, and could be aggregated for the nation as a whole, and different types of social statistics could be related to one another.

The national economic accounts and the measurement of economic performance

In contrast with the apparent chaos of the existing social indicators, monetary transactions data have been systematically developed within a national economic accounting framework, where they are used to monitor the behavior of the economic system and to evaluate the performance of different sectors of the economy. The public eagerly awaits the release of the quarterly estimates of the gross national product and the measures of the amounts by which output and prices in the economy have changed. Business behavior, consumer buying intentions, government monetary and fiscal policy, and forecasts about the future of the economy are all highly sensitive to this information.

But the national economic accounts do more than provide current summary measures. The detailed disaggregations by sector and by industry show how change is taking place. Sales of new automobiles, farm output, the housing industry, and government spending on defense or social programs all feed into the aggregate figures, making it possible to assess their relative importance. The impact of changes also appears in the wages and salaries of individuals, the profits of corporations, and the revenues received by state and local governments. Together, the interrelated set of data facilitates an understanding of how the economy functions, and an assessment of the importance of different parts of the economic system in relation to the whole.

The national economic accounts have achieved this integrated, comprehensive framework by restricting their focus to the network of monetary transactions and excluding, for the most part, social and demographic content. Some limited information in terms of the distribution of income by size and for specific regions is provided, but it is recognized that inclusion of such information in any detail would cause the system to become unmanageable. There does not appear to be any simple way to link the social indicator statistics to the national economic accounts at the aggregative level.

Despite their omission of the kind of information needed for social indicators, the national economic accounts do have the considerable advantage that they provide a systematic and integrated reporting system for the economy as a whole. Because of the national accounts, the economist can speak about "the economy" and its behavior, and the definition and measurement of the economy and its performance are quite widely accepted. In contrast, the social scientist who wishes to talk in terms of "the society", has no corresponding system of social statistics to define or measure the concept which he has in mind.

The relation of national economic accounts to social indicators

The national economic accounts have been an evolutionary development. Adam Smith conceived of the wealth of nations as the income which those nations could generate. Pareto's preoccupation with the size distribution of income was based upon his feeling that the size distribution of income was the most important descriptor of the structure of a society. Pigou, examining the economics of welfare, came to the conclusion that the national dividend does in some utilitarian sense reflect the welfare of a nation. In the early work of the National Bureau of Economic Research on the national income of the United States, the major concern was to show how the national income changed over time, and how it was distributed among the people. In all of these endeavors, the focus of concern was on the measurement of how well off people were, in a nation as a whole and in relation to one another. The elaboration of the national economic accounts into a comprehensive economic accounting system developed much more recently, after the publication of Keynes' General Theory, and the wartime experience of the usefulness of a full-fledged economic accounting system for mobilization for war and for monetary and fiscal policy.

Although monetary transactions that include the receipt and spending of income by individuals are obviously highly related to welfare, focusing solely on monetary transactions does omit important aspects of welfare. It has long been recognized that the exclusion of non-market activities of households such as housewives' services, education, and leisure activity leads to an understatement of both total output and welfare. Similarly, failing to take into account the deterioration of the environment and other costs related to maintaining the existing way of life leads to overstatement of output and welfare. If we are to take the total seriously as a welfare measure, both additions to and subtractions from the existing constructs are needed. Such extensions are being worked on at the present time by economists both in and out of government.

This extension of coverage to include certain kinds of non-market activity will make the national economic accounts more valid for many purposes, but it is not as easy to handle at the aggregate level the kind of social and demographic information which can neither be assigned an imputed monetary value nor be aggregated for different social and demographic groups. As already suggested, the national economic accounting framework cannot easily accommodate highly disaggregated information relating to distributions by social group, by region, or by social and demographic characteristics. Nevertheless, it is extremely important to be able to relate social indicators and social statistics directly to the national economic accounts. If this is not done, there can be no correspondence between the social statistics and the economic transactions data, and the linkage between the economic and social information will be lost. It must be recognized that for public policy the heart of the problem is the interrelation between economic transactions and social measurements. If, for instance, a satisfactory health level is to be achieved for specific regions or social and demographic groups, there will be certain costs, and it is essential to relate the expenditures to the social conditions to which they are addressed. Improvement in the quality of education will also have costs, and models need to be developed to determine these costs. Poverty is by definition both a social and an economic reality. Thus, although economic transactions constitute merely a portion of the relevant information, they are a highly significant portion, and they are central to questions of policy relating to the quality of life.

The development of a framework for economic and social measurement

The current NSF/NBER pilot project on the measurement of economic and social performance and its proposed extension have been designed to develop a coherent, integrated conceptual and statistical framework for the measurement of social performance with both aggregative and distributional dimensions. It is based on the further development of the national economic accounting framework in the direction of including non-market activity, improving the measurement of intermediate goods and services, and evaluating the impact of environmental factors. Microdata sets for specific sectors--households, enterprises, and governments--will be developed to provide information on the social and demographic characteristics of the population, regional and locational information, and the interrelation among the revenues and expenditure programs of the different levels of government. Such microdata sets are designed to provide representative samples of individual observations containing both transactions and non-transactions information. The transactions information in the microdata sets would when aggregated yield the economic constructs of the national economic accounts. Thus for example the income of individual households in the household microdata set would add up to personal income, and the value added by individual enterprises in the enterprise microdata set would add up to gross national product. The revenue and expenditures of the various units of government in the government microdata set would add up to total government revenues and outlays. The national economic accounts thus would provide the integrating framework and the control totals. The individual observations would, however, carry with them such social, demographic, and locational information as was needed for the construction of social indicators and the measurement of social performance. Thus for example, in the household microdata set, information on work-leisure, education, health, age, sex, and race for each individual would be the building blocks from which social measurements could be made.

This strategy of utilizing the national economic accounts in an extended form as the framework for a system of economic and social data rests upon the belief that we should take advantage of the progress which has been made to date in national economic accounting, and that systematic extension of the accounts is fully consistent with the past evolution of national economic accounting. Since the days of Wesley Clair Mitchell, the National Bureau of Economic Research has made significant contributions to the development of national economic accounting, and it is logical that the NBER should continue its efforts in this field.

There is, however, a more compelling reason why the national economic accounts should be used in developing the framework for economic and social measurement. As already indicated, although the transaction flows are not satisfactory by themselves as a measure of social welfare, the social and economic dimensions of our lives are so completely intertwined that a common framework is needed to embrace them both. Since considerable progress has been made in developing an integrated economic framework, it seems logical to extend what already exists in the direction of a more complete system of economic and social data.

Enlargement of the national income accounting framework

Three of the principal investigators associated with the project (John Kendrick, Robert Eisner, and Henry Peskin) will concentrate their efforts on enlarging the national income accounting framework to make it more comprehensive. While it is not thought possible to construct a unitary welfare measure, various welfare-oriented adjustments and additions can provide much additional quantitative information that will make the economic accounts more relevant and useful for appraising changes and differences in material welfare.

John Kendrick will focus primarily on imputing values of non-market activities by sector. He will build on his recent work on the formation and stocks of total capital to provide information on the flows of services from durable consumer goods not now included in national income. Additional imputations will be made for unpaid household and volunteer services, the opportunity costs of schoolwork by persons of working age, and of the frictionally unemployed, and the personal consumption of employees and the public financed by business through charges to current expense. In addition, there are other categories of non-market activities, such as time spent in commuting and in personal business which should and can be added to the body of the estimates.

Robert Eisner will be working along these same lines, but will in addition concern himself with problems relating to the valuation and revaluation of assets which give rise to capital gains and losses. The role of capital gains and losses in the generation and distribution of income and as a determinant of behavior has been neglected in conventional national income accounting. Special emphasis will be placed on estimating both depreciation and revaluations. The work done to date on revaluations of existing stocks has relied heavily on price indices relating to the cost of a current supply of similar assets. A major part of the further work, however, will entail the development of systematic and internally consistent revaluation and depreciation accounts for all capital assets. Eisner will endeavor to apply this approach to all forms of capital, tangible and intangible, in governments and households as well as enterprises.

Henry Peskin is involved in integrating into the conventional national income accounting structure measures of the flow of services from non-marketed environmental assets such as air and water. The principal problem is that the sources of data on such items as the watershed service of wooded land or ecological service of estuaries are widely scattered. Furthermore, rarely are such data so comprehensive as to provide a readily available national estimate suitable for national income accounting use. An attempt will be made to provide time series of such data using extrapolation methods already adopted by EPA for their long term projections. Additional sector detail will be provided, especially in those cases where the present 2-digit classification hides significant intra-sector distinctions in the use of environmental assets. The stock of environmental capital will be measured for the base year of 1968, and possibly estimates can be made of the levels of these stocks for one or two future periods as well. Regional breakdowns of environmental factors will be made, utilizing microdata sets for households, enterprises, and government. Finally, an effort will be made to improve the existing estimates of the social damages associated with air and water pollution. Using microdata sets to study the pollution impacts on individual households can be expected to improve the estimates.

The development of microdata sets for various sectors

The remaining members of the project are working on the development and use of microdata sets for specific sectors of the economy. Each of these microdata sets will be aligned with the macroeconomic accounts for the sector, and each will contain the relevant social, demographic, and regional information needed for social measurement.

The development of a microdata set for the household sector is the primary responsibility of Richard and Nancy Ruggles, assisted by Edward Wolff. The function of a microdata set for households is to provide a set of observations which contain social, demographic, and locational information in sample form, and at the same time are fully integrated with the national economic accounts in that the income and outlay transactions for the individual households, when aggregated, will yield the totals for households in the personal income account of the national economic accounts. Such a microdata set will achieve the desired linkage between micro and macro data, making possible distributions of the aggregates, e.g., size distributions of income for different social and demographic groups, or for different regions.

In developing an appropriate microdata set for the household sector, two problems arise. First, appropriate economic, social, and demographic data must be obtained, and it is unfortunately true that no single data source contains all of the needed kinds of information. Second, the microdata set needs to be aligned so that it actually does generate the control totals in the comprehensive national economic accounts. The pilot project has devoted considerable attention to these problems. Techniques for combining microdata from a number of different sources into a synthetically matched microdata set have been developed. At the present time work is under way on linking the 1970 Public Use Sample, the 1960 Public Use Sample, the Longitudinal Employer-Employee Data File (LEED) of the Social Security Administration, and the IRS Tax Model files to form a single synthetically matched microdata set.

The alignment problem arises due to biases in reporting or sampling of the different microdata sets, and adjustment factors must be introduced to make the information in the microdata sets consistent with what is known about the totals for the economy as a whole. Thus for example, transfer payments received by individuals from the government are notoriously underreported in any survey relying on individual respondents. From government sources, however, the total number of people receiving transfer payments and the total amount paid out is known. Corrections can be introduced into a microdata file to compensate for some of this underreporting. To the extent that additional information is available from yet other sources on the characteristics of those receiving transfer payments, furthermore, such information can be utilized in the adjustment process.

Milton Moss is carrying out a closely related projection of the distribution of lifetime income for different groups of persons and households over the span of their adult lives, to show how these patterns have been changing for successive birth cohorts. This project involves putting together both macro and micro data to provide a better understanding of the changes taking place in the life patterns of the American population.

A closely related collaborative study is being carried out by Thomas Juster at the Institute of Social Research of the University of Michigan. This project is involved in collecting time use patterns of U.S. families for a set of analytically relevant functional categories. It is hoped that it will be possible to introduce such time use information into the microdata sets for households, to generate information on non-market activity required for the macro studies of Kendrick and Eisner, and to facilitate micro studies relating to the distribution of real income.

For the enterprise sector, Robert E. Lipsey and Michael Gort are developing micro data for individual corporations and establishments. In these data, individual plants and offices and their characteristics, such as location, size, industry, and type of employee, would be linked to the enterprises that control and finance them and often provide technology or markets for them. With such a dataset one could trace the distribution by locality, region, and type of employee of any event that has affected a corporation. One can study the aggregate impact on a locality of an economic change affecting companies by aggregating the effects on different establishments within the locality. For several studies already under way at the National Bureau, financial data have been assembled for approximately 2000 corporations, and for about 1000 of them data exist on the employment in each establishment classified by detailed industry. A large variety of microdata sets on firms and establishments are available from private sources, and exact matching of data by firm and establishment is possible for larger firms and establishments. For smaller firms, sample data based upon published statistics may have to be developed in order to complete the microdata set and align it with the macro control totals.

Finally, a microdata set is being planned for the government sector. In addition to the Federal government, there are, of course, 50 state governments and approximately 78,000 units of local government, including counties, municipalities, townships, school districts, and special districts. Much of the needed information for a microdata set of governments is already gathered and available at the micro level. For example, the census of governments collects data on government organizations, property values, effective property taxes, other sources of revenue, local finance, and public employment. In contrast to the problems of data collection in the household and enterprise sectors, a complete census exists and the information is in the public domain. Thus there are neither problems of sampling nor problems of the disclosure of privileged information. There are, however, serious problems of designing the structure of the data set so that it will be consistent for the analysis at the micro level (i.e., so that the interactions among firms, households, and the local public sector can be modeled), and so that the accounts can be aggregated to the national level.

Interrelation of the micro and macro information

In a fully articulated national economic accounting system, the flows among the different sectors are consistent. Thus for example, the sum of wages paid by enterprises and governments is equal to the wage income received by individuals. Similarly, taxes paid by individuals and enterprises precisely correspond to the tax revenue received by the government. It has already been noted that aggregation of the microdata set for each sector will yield the macroeconomic constructs for that sector. In the same way, the microdata sets will of necessity be linked with one another. At this more detailed level, alignment will have to be even more precise. Wages paid by a given industry in the enterprise microdata set will have to correspond to the wages received by individuals in the household sector employed in that industry. Similarly, in locational terms, wages paid in a given locality must correspond to the wages received in that locality. Although this task of aligning data and providing a consistent articulated set of information is quite difficult, the large variety of statistical sources available makes it feasible. In a great many instances, the problem resolves itself into one of developing appropriate computer algorithms for reconciling differences among different data sources.

Social indicators and the use of the integrated system of macroeconomic accounts and microdata sets

As was pointed out initially, the function of social indicators is to provide for the monitoring and evaluation of various dimensions of the quality of life and the functioning of society. By their very nature, social indicators embrace both economic and non-economic elements of individuals' wellbeing. It is apparent that the selection of social indicators will in major part depend on the particular observer's set of values. In absolute terms there is no "correct" set of indicators, and as society's view of its problems changes, the particular social indicators needed will also change.

In contrast, the integrated system of macroeconomic accounts and microdata sets does provide a fully integrated framework into which many different kinds of information can be fitted and interrelated, and it can be used as the basis for the construction of a wide variety of meaningful and useful social indicators. An example may clarify more precisely how an integrated system of macroeconomic accounts and microdata sets can be used for such purposes.

There is considerable interest at the present time in the topic of air pollution, and measurements of how much air pollution is generated by society are needed. On the one hand, the production of goods and services causes air pollution. Industrial plants, the production of electricity, the transportation of goods over highways are all sources of air pollution. On the other hand, households also generate air pollution through the use of their automobiles, the heating of their homes, and even through the smoking of cigarettes.

Mere observation of the air pollution at various places and points in time would be difficult both to obtain and to interpret in terms of the contributions and interactions of various pollution sources relative to population density in different localities. By using the framework of integrated macro and micro data, however, it is possible to put the known pieces of information together into a comprehensive picture of air pollution conditions in different regions or localities.

In a paper recently presented to the Conference on Income and Wealth by Henry Peskin, Leonard Gianessi, and Edward Wolff, a microdata set for households was used to estimate for the year 1970 how much pollution each household was responsible for generating in terms of (a) that household's involvement in productive activity, and (b) its consumption of use of automobiles, heating, and other polluting activities. The 1970 Census Public Use Sample, which provides for a representative sample of households information on employment, ownership of automobiles, type of heating fuel and expenditure on heating fuel, and other information on personal habits relevant to pollution, was used to estimate the total amount of pollution each household was likely to generate. Computer programs were written to carry out this estimation process for each household in a given area, and to this information were added other factors known to affect air pollution within that area. Data on forest fires, climatic conditions, dust particles, etc., were added in at the regional level. Because the Public Use Sample of households is comprehensive, its use yielded estimates of air pollution for each county group and for each Standard Metropolitan Area in the country. Once the area air pollution due to both natural and human causes had been estimated, it was then possible to examine who was affected by air pollution to determine how many people were subject to the various levels of air pollution. In this connection, actual observations of air pollution can play an important role in determining the validity of the basic information relating to sources of air pollution. If actual observations of air pollution and estimates based on known sources generate different results, this strongly suggests that we lack a full understanding of the causal factors at work. Although it was not possible to validate the current study fully against measurements of air pollution, there did appear to be general agreement between the estimates and generally known air pollution conditions in specific areas.

Given such basic estimates of air pollution by type of pollutant, it is possible to make alternative evaluations of air pollution damage using different criteria. These in turn can be contrasted with estimates of the cost of reducing air pollution. In this manner the body of information which has been generated not only provides social indicators relating to the seriousness of air pollution in various localities, but it can link these indicators with an understanding of how such air pollution is generated and how it can be alleviated.

Social indicators for specific regions or localities are not the only possible use for the system of macro and micro data. Attention can also be focused on specific social and demographic groups. Thus for example social indicators can be developed relating to the status of the aged population. Information on household composition, length of life, condition of health, housing, and level of economic wellbeing can all be handled within the household microdata set. This does not mean that a microdata set will automatically have all of the information required to construct any desired social indicator. Rather it provides a basis for systematically recording information in its full detail and interrelating different bodies of information.

In more formal terms, microdata sets can be utilized for more analytic evaluations. Thus for example, a sample of tax returns integrated with a microdata set on households is being used at the Brookings Institution to analyze and evaluate the incidence of various taxes and to determine how changes in the tax system would affect the total amount of revenue generated and the tax burden of specific groups. Similar techniques could be used to determine how much different groups benefit from different government expenditure or transfer programs. Such analyses go far beyond what is normally considered to be the function of social indicators, but they are extremely important if we are actually to develop policies and programs which do improve the quality of life for specific groups.

A further potential use of the system of integrated macro and micro data is in the development of microanalytic models which introduce behavioral relationships into a dynamic analysis of change over time. At the Urban Institute a microanalytic model is being developed to study demographic change over time, including questions of intergenerational transfers that affect the economic and social condition of later generations. One of the advantages of such a microanalytic model is its ability to test the sensitivity of the outcome to various assumptions regarding behavioral relations or policy actions. Of particular importance for this use is the ability of the integrated macro and micro data system to portray the interaction of economic, social, and demographic factors.

MODELS FOR SOCIAL AND URBAN INDICATORS: TOWARDS AN INTEGRATED THEORY OF POLICY ANALYSIS*

Harvey A. Garn, The Urban Institute
Michael Springer, The Urban Institute

I. INTRODUCTION

In this paper we will describe a set of analytic models developed to structure the applied research being conducted by the Social and Urban Indicator Program of The Urban Institute. The distinction we draw between applied and basic research is not the distinction sometimes made between "practice without theory or theory without practice." Indicators are meant to convey information, not simply to repeat numbers. Therefore, in both applied and basic work some inference structures are required in moving from data to information. The relative neglect of conceptual structure has been, in our view, one of the major defects in applied work rather than being a necessary part of it.

The distinction we wish to make in describing our work as applied is that our research takes its origin in the problems of those choosing, operating, and being affected by public policies rather than that research which derives from knowledge aspirations typically associated with individual academic disciplines. Those pursuing the latter type of research may or may not intend that their results will be useful to those making and being affected by public policy. In applied research, as we define it, possible utilization of the outputs of the research by these individuals and groups is a major criterion in choosing what to do. In this context, therefore, it is necessary to form some judgments about the public policy scene and the customary modes of policy analysis and evaluation. Prior to describing our approach, some general comments about these matters seem relevant.

There are many indications that the American people are passing from the relatively easy acceptance of the expansion of public programming and action which characterized much of the 1960's into a more skeptical mood. There are indications that major concerns with, for example, improved health, housing, education, public safety, and employment have not been resolved or perceptibly ameliorated by the programs and policies which were established. Responsibility for outcomes seems diffuse and accountability correspondingly difficult to establish. Indeed, mea culpa has become almost a password among those who have proposed, administered, and analyzed policies and programs, but there are few signs of positive motion to create new and better initiatives. Some analysts, in fact, take the view that nothing or very little can or should be done through public intervention. At the same time, those who have been purported beneficiaries of social programs have not been exceptionally eager to come forward in a concerted defense of the programs.

This situation presents a very complicated mixture of opportunities and difficulties for those who wish to develop improved ways to enhance public well-being. On the one hand, the increased public skepticism may make possible serious debates about the scope, structure, mix, and efficacy of public policies and programs. It may result also, of course, in a withdrawal from public purpose. The outcome of the current public mood depends, in considerable part, on whether or not information relevant to policy choices and concerns can be developed in a sufficiently structured way to make such debates not only serious (that is, motivated by real concern) but productive of useful initiatives.

In short, we have come to the point where we now have to wrestle seriously with the confusion, frustration, and failure that are part of the legacy of the policy approaches of the 1960's. This puts a heavy burden on those who provide the analytical basis for policy choices and evaluation to develop conceptual structures appropriate to the potential seriousness of the debate. Such conceptual structures must be responsive to at least some of the common elements discernible in the lessons of the recent past. Among the lessons we would note are:

- Citizens diverge significantly in what they mean by "safe streets" or "quality education" or "good health" and are raising important questions about what should be produced and how it should be distributed.
- Improvements to the quality of life can no longer be viewed as the sole responsibility of government agencies or private organizations but must be achieved in concert with primary social units--families, neighborhoods, and communities.
- The delivery of public goods and services not only involves the provision of outputs, but also includes opportunities and constraints for people to use these outputs to enhance their own welfare.
- To increase responsiveness, the incentive structures of organizations and institutions involved in the delivery of public goods and services must reflect the values of their customers/clients as well as those that flow from the internal problems of management.
- No delivery system for particular goods and services--the police, a federal agency, or city government--does or can control either the perception and use of these goods and services by consumers or the full range of other outputs which contribute to welfare outcomes.
- Policy is implemented through large agglomerations of public and private organizations, professional associations, political units, and social groups. The capacities of these implementation systems

*This analysis was supported by funds from The National Science Foundation's Special Projects Program under Grant No. GS-38613. Opinions expressed are those of the authors and do not necessarily represent the views of The Urban Institute. The analytic models reported on below were developed by our entire research group which also includes Michael J. Flax and Jeremy B. Taylor.

are affected by complex patterns of interactions in geographic space and through institutional structures.

Virtually no one individual, group, or institutional representative commands sufficient information or resources to pursue effectively optimization or maximization strategies, although they can behave more or less functionally given their objectives.

These "lessons" have important implications for the ideas, notions, and analytical constructs employed in policy analysis and evaluation. We require intellectual tools that are sensitive to the fact that the United States is a highly pluralistic society having a great diversity of social and political values. We require tools that, for the sake of manageable analysis, do not make facile or misleading distinctions between political and administrative issues or between public and private spheres of activities, and which recognize the differences between institutional imperatives and those interests that stem from clients or citizens to be served by these institutions. Finally, we require tools that are sensitive to the fact that any given set of decision makers can make only modest contributions to improving the quality of life.

It is important then to have a sense for whether or not existing policy analysis models have these characteristics. The great majority of current conceptual models commonly used in policy analysis come from some aspect of economics, operations research, political science, management science, and organization theory. Policy analysis based upon economics and operations research focuses on the relationships between the inputs and outcomes of policy, but tend to treat implementation and organization only sketchily.[1] Analyses based upon political science approaches focus attention on political conflict and the overall structure of governmental processes, but tend to ignore how these processes affect the kind of outcomes or relationships between inputs and outcomes that are the basic information used by economists and operations researchers. Management scientists and organizational sociologists when they address policy issues tend to focus attention on intraorganizational issues related to decision-making processes and bureaucratic structure, but have not established analytical links between these phenomena and outcomes from the client/consumers perspectives or the divergent values and conflict the political scientists seek to understand.[2]

It may appear that this is an argument for organizing more multidisciplinary research. In a sense this is true; but another lesson from the recent history of policy analysis and evaluation is that such research cannot solve the problem if the representative of each discipline pursues the application of his own paradigms supplemented by a concluding or introductory statement by a grand summarizer. Rather, the requirement is for the development of bridging paradigms which representatives of different disciplines can learn and use. What we have in mind is a theoretical framework that defines important relationships between the key conceptual and methodological concerns of several research traditions. Such sets of bridging paradigms would allow for the effective integration of discrete pieces of analysis conducted within the narrower bounds of several disciplines as well as for the conduct of joint research within a framework that combines the analytic power of several disciplines. Additionally, the set of bridging paradigms can provide a common language for a heterogeneous group of analysts.

Beyond the development of a set of bridging paradigms, it is our perception that conceptual understanding of public policy programs and their effects requires also a constant balancing act between those who feel that differences in values, roles, and activities of people are so great that there is no hope for systematic analysis and generalization and those who feel that the assumptions required to convert scattered observations into systematic models (particularly those built on assumptions of maximization behavior and equilibrium) are both "true" and "complete" descriptions of the phenomenon under study. In executing this balancing act, one tends to be confronted with two not entirely satisfactory alternatives at the extremes. One can draw upon a bounded and well-developed set of intellectual constructs from a particular body of academic knowledge such as organization theory or microeconomics. Such conceptual frameworks quite clearly allow for ordered and systematic formulation and analysis of issues or problems. The virtues of these constructs for academic research--a parsimonious set of assumptions and a focus of a limited range of variables--tend, however, to be a liability for much applied analysis. By encompassing a carefully delimited range of variables, consideration of crucial interactions is often beyond the scope of analysis. The very neatness and elegance of conceptual models often conveys the impression of far greater certainty than is warranted.

The case study approach is often preferred as a means of avoiding these errors. When well-executed, these richly detailed descriptions of processes or problems can convey a wide range of important interactions as well as the complexity and ambiguities of real-world situations. Such case study approaches, however, simply do not provide a framework for the ordered formulation of questions, or for systematic analysis. In effect, they provide no reliable means for the accumulation of policy-relevant knowledge.

Consequently, a more appropriate framework to address policy and program issues should embody several important characteristics. First, it should integrate those analytic perspectives that have demonstrated a particular capacity to address important aspects of policy considerations. In this regard, it should encompass some of the key paradigms and concepts of microeconomics, political science, and management science. Second, while it should be sufficiently detailed to capture the flavor of a complex reality, it should provide also a framework to structure systematic analysis. Finally, it should assist in the ordered consideration of policy issues by public officials and citizens and for the translation of their perceptions of these issues to researchers and analysts.

Much of the recent work of the Urban Institute's Social and Urban Indicators Program is an attempt to achieve such a conceptual structure.[3] We recognize, of course, that there are many others approaching similar issues from a variety of perspectives. We will claim, therefore, neither exclusive validity nor completeness of our approach. We have found, however, that it is suggestive of a wide range of potentially useful applications in a context where no approach to policy analysis and evaluation based solely on the restrictive and highly formalized models of a single discipline seems adequate.

In establishing the major features of such a framework, we began with an elaboration of a traditional microeconomic model of production processes. Our model traces chains of interactions from resource inputs to outcomes in terms of the level of human welfare through two structurally independent transformations--the transformation of resources into goods and services by production units, and the subsequent transformation of the characteristics of the goods and services by consumers into various welfare outcomes. This model allows us to explore questions pertaining to the role of consumers in the generation of their own welfare and to specify important differences between the production and consumption of goods and of services. Next, this model is integrated with schema that identify important bureaucratic and political roles and institutional patterns. These roles and institutional patterns are viewed as influencing the generation of welfare by structuring the uncertainties associated with the purposive choices that must be undertaken in the processes of both production and consumption. This allows us to systematically address the information requirements of different kinds of decision makers as well as deal with the impacts of both incentive systems and institutional structure on either social or institutional performance. Our third extension involved an adaptation of the theory of economic clubs in order to place considerations of institutionalized production, consumption, and welfare generation in a spatial context. In this model, the processes and problems of urban growth are treated through a simultaneous consideration of the spatial behavior of producer and consumer groups and the geographic and organizational structure of governmental activities.

Implicit in the development of all of our models is that choices are not and cannot be made in a manner that allows for the maximization or optimization of a set of values. Rather, we believe the decision makers are confronted by large areas of uncertainty based not only on complexity and limited information but also on the slack and disequilibrium which tend to characterize both large-scale systemic relationships and relationships within small groups and institutions.[4] Therefore, we agree with those organizational and political theorists who argue that rational decision making involves sequences of learning about problems and searching for solutions in order to reach an incremental decision that achieves results that are at least satisfactory in terms of some set of objectives.[5] Our models are intended to facilitate these searching and learning processes by providing a broad and detailed analytic framework that will assist in the conceptualization of problems, the specification of crucial relationships, and the identification of lucrative questions for further research.

II. THE GENERATION OF WELFARE - FROM INPUTS TO OUTCOMES

Our initial model grew out of dissatisfactions with conventional microeconomic models of production for tracing chains of interactions from resource inputs to final outcomes in terms of the quality of people's lives. Such schemas often present the policy maker and the researcher with huge analytic gulfs between the level and mix of resources entering into a production process (such as the number of students per teacher, or the number of hospital beds per capita) and welfare outcomes (such as the rate of learning or the level of morbidity). We observed, also, that problems of this character are far more acute in assessing the production of services than of goods. The outputs of goods production are tangible, are easily counted, and can be associated readily with a mix of resources having market prices even though the welfare outcomes of their use are far from clear. The outputs of service production are not only intangible and tend to be perceived differently by various actors, but also are partially determined by the client or consumer. Additionally, relationships between the characteristics of services and resultant outcomes are particularly tenuous.

After attempting to address these problems with the development of increasingly intricate elaborations of a standard economic model of production, we came to realize that our difficulties rested in large measure with the analytic assumptions of our model. Specifically, economic treatments of consumption have tended to ignore the contributions of consumption units--individuals, households, communities--to the generation of their own welfare.[6] We believe a more complete examination of the roles, values, and activities of consumers can help bridge the analytic gap between the outputs of production systems (the frequency of police patrols or the number of new housing units) and the welfare of citizens (the rates of victimization or the level of satisfactions with homes and neighborhoods).

Work being done on the economics of household production and related areas of consumer theory provided a useful starting point in our elaboration of the standard microeconomic model.[7] In this research, the household (or individual members of a household) is viewed as a production unit which combines market and nonmarket goods and services to produce various commodities. Researchers in this area of inquiry examine, for example, how households combine formal educational services with time spent with children to produce levels of learning. There are four key notions that are central to such analysis: (1) welfare outcomes of production can be different because of the activities of consumers, (2) these outcomes are different because consumers select different bundles of goods, (3) the efficiency of these selections can be constrained by the information about the price and availability of goods available to the consumer, and (4) goods themselves do not directly provide utility to the consumer. They possess characteristics that give rise to utility.

In extending these notions, we assume that the relevant characteristics of goods are more than technical in character (that is, determined solely by the good). The characteristics of goods and services are assumed to be a function of consumer as well as producer behavior. The consumer actively shapes the characteristics. Thus, treatment of consumption efficiency is not limited solely to an examination of the information available to the consumer. It entails the broader question of the character of consumer technologies in converting characteristics of goods and services into satisfactions, a process directly analogous to technologies of production in converting resources into goods and services. Further, certain commodities (primarily service outputs) must be viewed as the products of the joint activities of producers and consumers.

The above assumptions help provide us with a schema having three significant features: (1) consumption and production are viewed as analogous but structurally independent activities; (2) analysis of goods and of services requires different sorts of analytic treatment; and (3) consumer satisfaction can be both extrinsic and intrinsic to the processes of consumption.[8]

Central to our schema is the notion that human satisfactions--the level of welfare or illfare--are the result of two structurally independent classes of purposive activities; the highly institutionalized activities by producers to transform a mix of resources into goods and services, and the subsequent and less institutionalized efforts by consumers to transform these goods and services into human satisfactions. Thus the outputs of production systems are treated as inputs to consumer transformation processes, in which consumers combine the characteristics of these outputs and transform them into their satisfactions (utility). Production system outputs do not directly provide satisfaction to consumers. Given that consumers may identify different characteristics in goods and services and employ different technologies in their consumption transformation processes, it follows that one cannot assume that the equivalence of a bundle of goods (or income) necessarily leads to an equivalent level of satisfaction for two consumers--even if they have identical tastes.

The schema is constructed in such a way that issues relating to the use of normative resources, such as values and sentiments, as well as traditional economic resources can be treated. In addition, the treatment of opportunities for or constraints upon production and consumption activities and the delineation of the differences in the interests and roles of producers and consumers are designed to permit the inclusion of variables relating to patterns of social organization and the sources of social conflict.

We have noted that traditional economic models treat the production and consumption of goods more adequately than the production and consumption of services. A primary reason for this is that production and consumption of goods are generally separated from each other in space and time. Also outputs of the production process are tangible and they usually have market prices.

Services tend to have different characteristics. Frequently the service delivery occurs in a face-to-face situation and depends on the direct interaction of the ostensible deliverer of the services and the receiver of the services. Outputs are not tangible and may not have a market price.[9] Thus, our schema views the outputs of service activities to be the joint products of consumption and production transformation activities. In the case of goods, joint products (somewhat analogous to investment) when present result from the interaction of the characteristics of goods with the transformation activities of consumers rather than from interaction of the consumer with the producer of the good.

Our schema views consumer satisfaction in social-psychological terms as a direct measure of utility. Satisfactions stem from two sources. First, satisfactions can be intrinsic to the act of consumption, i.e., people's utility is directly affected by using the characteristics of a good or interacting with service deliverers. Second, satisfactions might be extrinsic in character, i.e., they might stem from commodities jointly produced by consumer and producer activities. This distinction suggests that consumers should not be expected to be indifferent about how goods and services are provided and by whom. Additionally, it opens up the question of time lags associated with different types of satisfactions and situations in which intrinsically dissatisfying consumption activities might result in outcomes that are positively valued, e.g., going to a dentist.

Some Implications of the Model

Our model of the generation of welfare can help structure consideration of difficult questions that have long troubled policy makers and analysts. It can help in clarifying the confusion generally associated with the specification and valuation of the outputs of public programs, assist in the identification of the generally neglected issues of consumer technologies, and provide a framework to formulate thorny questions of accountability.

The problems of specification and valuation of output plague consideration of countless policy issues. While citizens and public officials tend to agree on abstract policy goals such as "good housing" and "quality education," they are referring often to quite different mixes of outputs. For some, quality education might mean a highly structured academic program, while for others it might mean a loosely structured program that emphasizes the inculcation of personal and social norms. The salient characteristics of a good home, for some, might be the physical characteristics of housing units, while for others, salient characteristics might be associated with the social status of the residents of an area. Such differences in perceptions and preferences are reflected in the great difficulties encountered by policy makers and analysts in attempting to define program and policy outputs.

Our model suggests that improvements can be made by recognizing the consumer's role more explicitly. From this perspective, it would be useful for both policy makers and analysts to begin to structure questions pertaining to the different perceptions people have of schools, or homes, or health facilities, how they value the characteristics which they perceive in them, and how efficiently they can transform these characteristics into their own welfare.

Further pursuit of the same questions will assist in the identification of important issues pertaining to the technologies consumers employ in transforming the characteristics of production outputs into their own welfare. Constrained by the availability of resources, how do consumers vary in their ability to rent or purchase a home that satisfies their preferences or in their ability to select and participate in an educational program that would contribute to some desired level of learning? Such considerations are central to the current debate over the usefulness of income or direct service strategies in housing, health care, or educational policy. It might be fruitful to develop programs that directly enhance consumer technology. Because most programs have been production-side oriented, it remains an open question, for example, whether programs that attempt to improve housing construction technologies can lead to greater contributions to welfare than programs that attempt to improve the ability of citizens to assess, select, and finance a home.

Extending our model further, we have a means of asking new questions about accountability for outcomes--how much of which welfare outcomes can be attributed to the activities of producers and to the activities of consumers. Such issues are at the heart of the controversies over whether educational performance is primarily a function of the schools or the social backgrounds of students.[10] Similar considerations pertain to assessments of what kinds of crimes can and cannot be deterred by the police and what health problems are largely a function of behaviors of individuals and households. Our model not only suggests a way to structure questions such as these in terms of the joint effects of producer and consumer interactions, it also suggests consideration of a much finer grained set of variables than are generally examined in the analysis of these questions. Our current failure to sort out the issues of accountability hinders both the formulation of sensible program strategies and the conduct of policy-level evaluations.

III. INSTITUTIONAL ROLES IN THE GENERATION OF WELFARE

The next stage in our conceptual work was to integrate our economic model of the generation of welfare with a schema that identifies the important social, bureaucratic, and political roles that are involved in making choices affecting consumption and production processes. This allows us to identify in a systematic fashion the information requirements of different kinds of decision makers. It provides also a framework to formulate questions relating to performance and accountability from the perspectives of both production systems and various consumption units including individuals, households, and social groups.

This conceptual elaboration required synthesis of two important traditions of policy analysis--economic analysis that focuses on the problems of choice under the conditions of scarcity and the analysis of choice under conditions of uncertainty which is central to the study of organizations and political science. The theoretical problem was to develop a schema to describe how the structure of political and bureaucratic institutions shape the uncertainties associated with the production and consumption processes identified by our model of the generation of welfare.

Our approach to this problem rests upon a conception of a societal division of labor to cope with the uncertainty of making purposive choices. This division of labor can be described in terms of the patterns of authority and control formed by three broad classes of social institutions--the state, formal organizations, and the technostructure. The state is viewed as institutionalizing the right to cope with decisions affecting universalistic values (pertaining to the interests of all members of society). Included here are the legitimized political control centers of policy--executive offices, legislatures, and judicial bodies. Formal organizations institutionalize the right to cope with decisions affecting particularistic values. Included here are bureaucratic organizations such as industrial corporations, hospitals, trade unions, and government agencies. Finally, the technostructure institutionalizes the right to cope with those classes of decisions requiring the mastery of specialized knowledge, experience, or expertise. Included here are professional and guild-like associations.

These institutions provide the structural underpinnings for different kinds of decision-making arenas--the social setting of purposive choice. Formal organizations and the state shape hierarchies of decision-making arenas, with operational, managerial, and institutional arenas identified within formal organizations and an elaborate network of political arenas commanding the authority of the state. Operational arenas direct standardized and regularized organizational tasks. Managerial arenas control the interdependencies among these tasks and the relationships of an organization with its regular suppliers of inputs and recipients of output. Institutional arenas mediate between organization and its broader environment. Political arenas deal with decisions of a more generalized character, involving efforts to manage conflict among a range of particularistic interests and/or efforts to shape relationships between social inputs and social outcomes. These hierarchies of organizational and political arenas are crosscut by patterns of authority and control emanating from the technostructure. Actors in the technostructure control or influence decisions that draw upon specialized bodies of skills or knowledge. Our model can account for key elaborations of institutional structure including various patterns of bureaucratic form, alternative social bases of political support, and relationships between basic and applied science.

These institutionalized decision-making arenas can be characterized by the uncertainties they present to decision makers. These uncertainties have three major sources: (1) the internal structure of an arena, (2) their external environment, and (3) the level of culturally available technology. Further, we assume that it is functional (rational) for actors to draw upon information to reduce these uncertainties. Information consists of two elements--data items and inference structures. By "data items" we mean descriptors of events and activities. By "inference structure" we mean models of relationships between goals and means. These can range from formal causal models to procedures to achieve incremental conflict-settlement decisions.

It should be emphasized that this decision-hierarchy model identifies only those subclasses of indicators useful to decision makers in well-structured settings. This focus made our initial analytical tasks more manageable by excluding from direct consideration decision making by individuals, such as the consumption choices made by members of a family or investment decisions made by a self-employed entrepreneur. This restriction, however, may not be as limiting as it might appear. Decisions within institutionalized settings have large and increasingly significant impacts on many important aspects of human welfare. Additionally, specification of the characteristics of decision-making arenas in highly structured situations provides a framework to assess some of the processes of choice in informal settings, i.e., we can assume that a family must perform the same kinds of decision-making functions that are performed through the highly structured divisions of labor with a fully bureaucratized organization.

Some Implications of the Model

The application of this schema in determining the information requirements of different kinds of decision makers is a relatively straightforward exercise. By describing the structural underpinnings of various loci of authority and control within an institutional division of decision-making labor, we can isolate those sets of production or consumption activities for which different actors are responsible and the uncertainties associated with these responsibilities. This analysis provides criteria to identify both the data items required by different actors, and the kinds of inference structures that appropriately link these data items in a model of the relationships between goals and means. We have developed prototypical models of bureaucratic and political arrangements that can identify information requirements in general terms. More detailed identification of the information requirements of specific actors can be accomplished by an analysis of the particular sources of uncertainty that characterize a decision-making arena, i.e., the level of available technology, the internal structure of the arena and the characteristics of its external environment.

Implicit in such analysis is that information of a general character is of limited usefulness to any particular decision maker. Consequently, the problem of developing more useful social indicators rests not in the construction of some set of all-purpose measures of social conditions, but rather of identifying those sets of information that help to reduce the specific uncertainties associated with the performance of different sorts of social roles. The same analytic tools are applicable to address problems that have long troubled the developers of information systems for managers and policy makers, i.e., how to identify the demand for information in a systematic and structured fashion.

Beyond the identification of information requirements our models have another important class of applications in dealing with complex issues of performance measurement and the assessment of accountability. A complete treatment of performance issues requires an analysis of arrangements for consumption as well as analysis of the ways in which the pursuit of consumer values are affected by the structure and activities of a production system. This analysis contrasts sharply with conventional treatments of performance which tend to be bounded to the values and activities of producer units and respond to questions about how to achieve accountability within producing organizations. Our approach to performance assessment raises the broader question of how to hold institutions accountable for social outcomes in terms of the pursuit of various welfare values, i.e., in which ways do formal organizations or political units facilitate or hinder the pursuit of welfare by individuals, families, or communities. To distinguish these approaches, we refer to the former as institutional performance and the latter as social performance.

Analysis of social performance begins with an identification of the consumer-side analogs of the decision-making arenas of the production units. For example, the individualistic activities and particularistic values of a family unit or an individual might be the appropriate units to assess operational level performance while the collective activities and the more universalistic values of the local community might be the appropriate unit to assess the performance at the institutional level. This analysis requires not only an examination of consumer/producer interaction, but also interaction among the various levels of a hierarchical institution in examining social performance. At the political level, producer and consumer values can merge and assessments of political decisions must confront issues of social as well as institutional performance.

Determining accountability--the assignment of variability in levels of welfare to specific actors or units--requires analysis of the roles and interdependencies within a system of bureaucratic and political relationships as well as an analysis of the range and content of the transactions between production and consumption units. Such treatments of accountability bring together for simultaneous consideration analysis of policy making and of the implementation of policy. This capacity is particularly useful in a period when primary constraints on policy often appear to involve questions of implementation rather than of political resistance. In all but the most simple system of relationships, analysis of this sort will require sequences in questioning and analysis in order to achieve a more satisfactory understanding of situations, rather than the application of a formalized model of causal relationships which is to yield, somehow, optimal solutions.

IV. THE GENERATION OF WELFARE IN A SPATIAL CONTEXT

The third element in our conceptual work was an effort to place the major considerations of our other models in a spatial context in order to provide a framework for addressing the complex interrelationships among large numbers of production units, consumption units and governments.[11] Problems of sorting out these interrelationships in large measure are the source of the extraordinary degree of confusion that tends to characterize both academic and public discussion of urban growth issues. These discussions tend to focus exclusively on either the incentives for movement behaviors or how government can be restructured in order to respond to these behaviors. Our model addresses the problems and policies of urban growth in an analytic context allowing simultaneous treatment of the patterns of governmental organization and incentive systems that shape the structure of urban areas.

In the development of this model, we drew upon the growing literature on economic "clubs."[12] This literature examines the incentives for group formation for the purposes of production and consumption. Martin McGuire has shown that the incentives for the formation of groups (which are called "clubs") are essentially the same for the purposes of both production and consumption and in the private and public sectors. These incentives involve increasing benefits and reducing costs as well as a taste for association. These factors help identify key considerations which determine the size and scope of public and private production and consumption clubs. In addition to the fairly straightforward consumption and production clubs identified in this literature, we have chosen to treat political units as "clubs of clubs." Clubs of clubs pursue values that are universalistic in character, i.e., values that reflect the aggregation of individual interest or the interests of some community.

Drawing upon all our conceptual work, the model focuses on needed definitions of pertinent indicators of the interrelationships among the activities of producer groups, consumer groups, the state, and the characteristics of places. It should be noted at this point that we do not treat places in such conventional terms as a geographic area bounded by jurisdictional lines. Rather we treat places in terms of the potential values they present to various collectivities, with jurisdictional boundaries being viewed as one of many factors that affect these potential values.

In developing the model, we first explored the interrelationships between producer and consumer groups and their experiential environments. Here the characteristics of places were viewed solely in terms of the collective expression of particularistic interests or adaptations to the exercise of these interests. Next, the model was elaborated by showing how the authoritative expression of universalistic values by the state further shapes the characteristics of places and influences the decisions of both producer and consumer groups, focusing on purposive choices which are made by individuals and groups. These groups are more or less formally organized in an attempt to permit choices which improve the situation of the group and group members. Individuals and groups interact both within and across organizations and political arenas. The relationships among the various levels within a given producer or consumer organization, and between such organizations and political arenas, are neither status nor "equilibrium-seeking." This is also true of the relationships between producer and consumer groups.

Each group operates in an "experiential environment" which is more or less insulated from, expandable into, or permeable by the effects of actions of other groups. For example, entry into or exit from groups varies in its ease. Marketing of products can occur over a much more flexible geographic area than a single governmental jurisdiction. A particular tax affects some groups more than others. In considerable part, such dynamic interactions among individuals and groups give rise to boundary problems--including those of governmental jurisdictions and functions--which are at the heart of urban growth issues.

In this context, clubs of clubs attempt to deal with the consequences for universalistic values brought about by the mobility behavior of clubs. We refer to these efforts as urban growth policies. Clubs of clubs can attempt to channel spatial movements by altering the value of places for individuals and clubs. Such activities can be assessed in terms of the relationship between clubs and the potential values of places. They can develop, also, structural responses to changing characteristics of places by creating new jurisdictions, altering the boundaries of existing jurisdictions, or changing the flows of authority or resources through networks of intergovernmental relationships. These structural adaptations to growth processes not only alter the value of places for the pursuit of the universalistic values of clubs of clubs, but also for the particularistic interests of production and consumption clubs. Such adaptations can, in turn, initiate chains of second- and third-order spatial movements and alterations in the characteristics of places that might nullify the initial objectives of a particular adaptation.

Implicit in this model is that consequences of slack and forces leading to disequilibrium should be taken seriously and that major alterations in either the incentives for movement behavior or in the structure of government be assessed in a broad analytic context. For example, an effort to provide a more equal distribution of fiscal resources among the various jurisdictions of a metropolitan area by creating one large jurisdiction having a common tax base may have subsequent impacts on intergovernmental fiscal flows, plant location decisions, and household mobility that could nullify the original policy objective. Therefore, the mobility of individuals and groups in metropolitan settings necessitates a public urban growth policy which is viewed as an ongoing adaptive process requiring both structural (boundary and functional) changes in patterns of governmental organizations and specific efforts to channel the spatial movements of individuals and groups.

V. CONCLUSION

If we could claim at the conclusion of this paper that any of the policy issues confronting the nation could be solved once and for all by the application of these or any other set of models, we would appear to be much more in the mainstream of policy analysis as it has been practiced. We, of course, can make no such claim. Nor are we really convinced that anyone should imagine that such an outcome is possible. One of the key constraints which seems to have affected our ability to come to grips with the complexities of such issues as urban growth policy, institutional responsiveness, or accountability for social outcomes, is an initial presumption that there is no clear policy unless a set of agreed upon final outcomes is established. We have argued throughout this paper that the complexity of interrelationships which must be understood, the legitimate differences in perspectives and values of consumers, producers, and decision makers in different arenas, as well as the slack and disequilibrium which characterize most real-world systems make such an expectation for policy unrealistic.

Policy setting relates not only to outcomes; it also relates to the creation of processes in which issues can be raised and relationships developed. Policy analysis must reflect the complexities of needed interrelationships if it is to be useful. It will not do to blame the world for failing to perform according to a tidy analytical perception. At the same time, it will not do to argue that there are no systematic forces at work in determining variations in factors which influence the generation of welfare. We are left with the balancing act mentioned at the beginning of this paper--between the highly structured and formalized models of specific academic disciplines and rich descriptions of events or situations that provide no basis for the systematic examination of issues. Our hope is for the continuing development of coherent conceptual structures which assist in isolating key issues and relationships to facilitate (incremental) improvements in our ability to enhance welfare.

REFERENCES

1. Some economists are attempting to rectify this problem. Illustrative of such efforts is Buchanan and Robert D. Tollison, eds., Theory of Public Choice: Political Applications of Economics (Ann Arbor: University of Michigan Press, 1972).

2. In this regard we agree with the general thrust of Vincent Ostrom's argument that important bodies of political and administrative theory provide little capacity to formulate questions pertaining to the interactions between public institutions and the citizens served by these institutions. See his arguments in The Intellectual Crisis in American Public Administration (University, Alabama: The University of Alabama Press, 1974).

3. These efforts are reported in Harvey A. Garn, Michael J. Flax, Michael Springer, and Jeremy B. Taylor, Models for Indicator Development: Tools for Applied Social Research (Washington, D.C.: The Urban Institute, forthcoming) and Harvey A. Garn and Michael Springer, "Formulating Urban Growth Policies: Dynamic Interactions Among People, Places, and Clubs," Urban Institute Working Paper 1206-16, October 1973 (to be published in a special issue of Publius, 1974).

4. Our approach to these issues was instructed by Michel Crozier's analysis of bureaucracy as an institution characterized by dysfunctional behaviors, and Albert O. Hirschman's discussions of the endemic slack and inherent inefficiencies of firms, government agencies, and political units. See Crozier, The Bureaucratic Phenomena (Chicago: The University of Chicago Press, 1967) and Hirschman, Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States (Cambridge: Harvard University Press, 1970).

5. Our analysis of these processes draws heavily upon James D. Thompson, Organizations in Action: Social Science Base of Administrative Theory (New York: McGraw-Hill, 1967).

6. William Alonso has commented:

   "Economics has concentrated on the process of production and virtually ignored those of consumption. It has treated the act of consumption as magic: final goods and services disappear as they arrive to the consumer, without even a wave of the cape. To be sure, certain areas of spatial economics deal with the consumer, most notably rent and central place theories, some studies in transportation, and certain recent developments in welfare economics which take into account travel or queing time. But even here the consumer is treated in a most abstract fashion, disembodied into a demand curve or a utility function. Yet the consumption activity is the principal concern of the most widespread of economic units, the household. It is an activity that itself consumes resources, and it is internally structured in terms of a technology and the availability and relative costs of factors." in "The Economics of Consumption, Daily Life, and Urban Form," Department of City and Regional Planning, University of California, Berkeley, December 1970, Working Paper No. 139, p. 1.

7. We have combined two innovative approaches to the analysis of consumption, the work of Gary Becker on the household as a production unit and Kelvin Lancaster's recent contributions to the theory of consumption. See Becker, "A Theory of the Allocation of Time," Economic Journal, Vol. 75 (September 1965) and Lancaster, Consumer Demand: A New Approach (New York: Columbia University Press, 1971).

8. This work is reported in Models for Indicator Development, op. cit.

9. Although it is true, of course, that many services are exchanged on the market, what one usually buys is access to a service delivery process rather than an identifiable output.

10. Aspects of this controversy are illustrated by the debates surrounding the publication of the Coleman Report on the determinants of educational performance. See James S. Coleman, et al., Equality of Educational Opportunity, Office of Education, U.S. Department of Health, Education, and Welfare (Washington, D.C.: U.S. Government Printing Office, 1966), and Eric A. Hanushek and John F. Kain, "On the Value of Equality of Educational Opportunity as a Guide to Public Policy," Harvard University, Program on Regional and Urban Economics, No. 36, February 1969.

11. An initial formulation of this model is presented in "Formulating Urban Growth Policies," op. cit.

12. Our work was aided particularly by Martin C. McGuire, "Private Good Clubs and Public Good Clubs: Economic Models of Group Formation," Swedish Journal of Economics, Vol. 74, 1972, pp. 84-99. Other important contributions to the "club" literature include Mancur Olson, Jr., The Logic of Collective Action (New York: Schocken Books, 1969); Charles Tiebout, "A Pure Theory of Local Expenditures," Journal of Political Economy, 1956; and James Buchanan, "An Economic Theory of Clubs," Economica, 1965.

SOCIAL INDICATORS, 1973: STATISTICAL CONSIDERATIONS[1]

Stephen E. Fienberg, University of Minnesota
Leo A. Goodman[2] University of Chicago

INTRODUCTION

A reasonable scenario for the development of a system of social indicators might consist of several stages. First, we might start with the identification of broad areas of social concern (e.g. public safety might be one such area). As the second step in the scenario we can focus our attention on specific phenomena that pertain to each area of concern (e.g., the incidence of criminal behavior and the risk of victimization); these phenomena, if properly measured by ideal indicators (e.g., by crime and victimization rates), would reflect or elucidate the relevant aspects of the state of society and the changes (if any) that are taking place in it. (We might also consider other relevant phenomena.) Then as the third step, we can compare the ideal indicators with the ones available (e.g., the rate of offenses reported to the police as given in the FBI Uniform Crime Reports) with an eye toward accepting substitutes for some ideal measures, and possibly insisting on the development of new measures (e.g., victimization rates from national and local surveys). Finally, we must decide how often to make measurements and how best to present or report the measurements we make.

Although the statistician in this scenario might focus on methods of data collection, on assessing and maintaining the quality of measurements, and on the manner in which they are reported, we choose to view the statistician's role more broadly, especially since it is difficult for us to separate the quality of measurements from what it is we actually would like to measure. Thus our comments on Social Indicators, 1973 will be concerned not only with data collection and sampling techniques, and with accuracy and reporting, but also with broader issues related to the social indicators enterprise.

One final preliminary comment is in order. Some of the remarks that we shall make in this paper are critical in the sense that we suggest a variety of ways in which Social Indicators, 1973 can be improved upon statistically. It is our belief that many of our criticisms are equally applicable (if not more so) to social indicator reports produced by other countries, such as France, Germany, and the United Kingdom. Moreover, in no way should our remarks be interpreted as being critical of the many achievements of the federal statistical system which are reflected in Social Indicators, 1973.

SOME REPORTING PROBLEMS

When preparing a social report such as Social Indicators, 1973, one faces at least the following kinds of statistical problems:

(a) How to abstract a salient summary of major ongoing studies, especially those with complex statistical analyses?
(b) How to check whether the abstract distorts or disguises important features of a more complete presentation?

(c) How to check whether the abstract reproduces uncritically inadequacies in the original sources--inadequacies in the reporting or in the related original analyses?

There is no prescription we can offer that tells how to solve these problems, but the following example may serve to illustrate them. Charts 3/8 through 3/11 in the Education chapter [pp.82-85] purport to show "which groups in the population are behind or ahead and by how much," based on the results presented by the National Assessment of Educational Progress. The National Assessment data come from a national sample of 17 year-olds, selected under the direction of professional statisticians using intricate sampling techniques. The analyses of these data were preformed under the supervision of distinguished statisticians and the analyses and graphical presentations in the National Assessment reports issued to date reflect this professional strength. Yet the information from the National Assessment project that is abstracted in Social Indicators, 1973 can easily be misunderstood.

The baseline for Charts 3/8 through 3/11 in each case is the median performance of all 17 year-olds, and each bar-graph illustrates median differences between or among groups, over all questions in the achievement test. What is misleading about these graphs? First, the scales from graph to graph are somewhat different. Second, the apparent apportionment of differences to groups is a function mainly of group size. To see this, we focus on White-Black differences in Science Achievement (see Figure 1), and for simplicity we work with means rather than medians.[3] We let

$\bar{X}_W$ = average for Whites; N_W = number of Whites
$\bar{X}_B$ = average for Blacks; N_B = number of Blacks
$\bar{X}$ = average for total population; $N=N_W+N_B=$ number in total population.

Then $N\bar{\bar{X}} = N_B\bar{X}_B + N_W\bar{X}_W$;

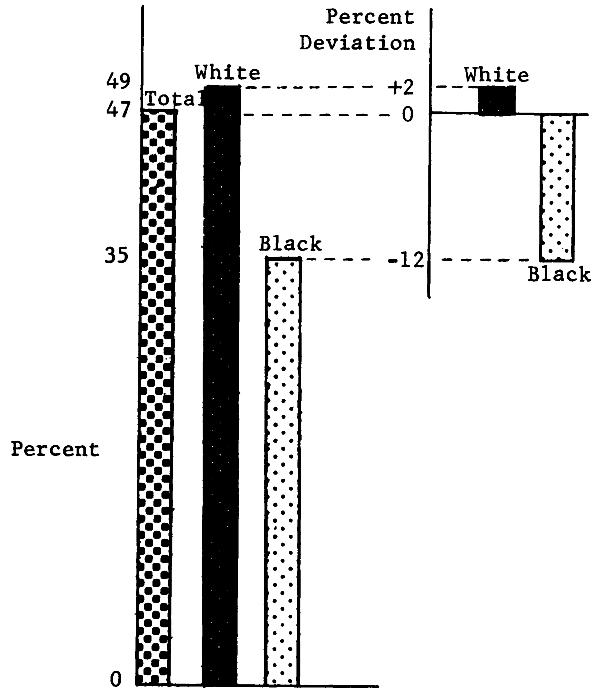
and we find that

$$\bar{X}_W-\bar{\bar{X}}=\frac{N_B}{N}(\bar{X}_W-\bar{X}_B),\quad \bar{X}_B-\bar{\bar{X}} = \frac{N_W}{N}(\bar{X}_B-\bar{X}_W).$$

Thus each of the bars in the two group comparisons simply represents the appropriate group difference (i.e., $\pm(\bar{X}_W - \bar{X}_B)$), magnified by the relative size of the other group. In this case, since the Blacks represent a smaller proportion of the population relative to the Whites, the graphs appear to show them as being more "disadvantaged." If Blacks had represented, say, 87 percent of the population and Whites 13 percent rather than the reverse, the overall average $\bar{X}$ would shift towards the Black average, and the picture for the deviations would look dramatically different (see Figure 2). In this case Blacks would appear only

Figure 1*

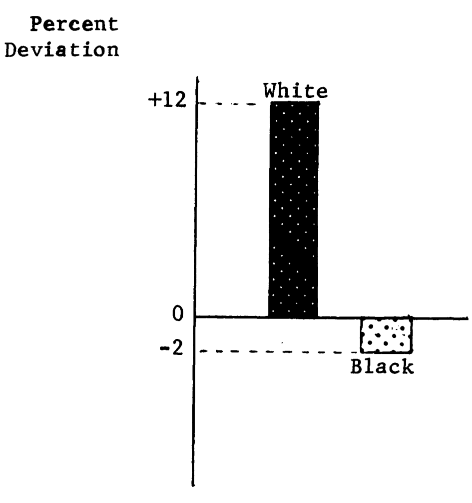
SCIENCE ACHIEVEMENT BY COLOR (1970)



*Actual Display is as in Chart 3/9 of Social Indicators, 1973 [p. 83]. Original Display is similar to that used in the unpublished report of the National Assessment of Educational Progress.

Figure 2*

SCIENCE ACHIEVEMENT BY COLOR



*Similar to Actual Display of Figure 1, but with population size of Black and White groups interchanged.

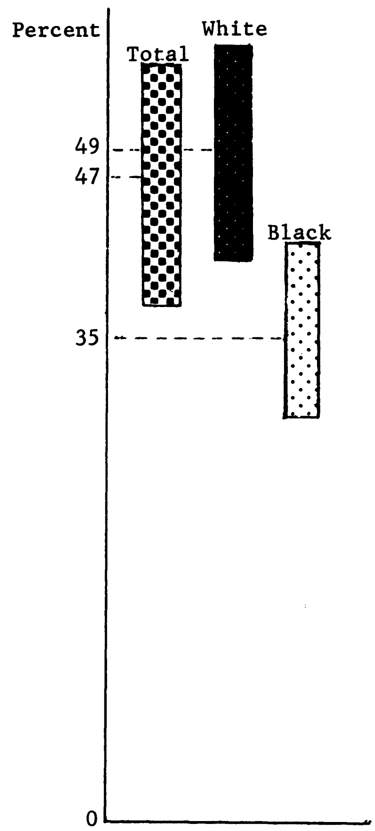
modestly disadvantaged while Whites would become more "advantaged." Thus the relative magnitude of each bar reflects only the relative size of the other group.

We note that the information in the National Assessment report correctly includes the equivalents of $\bar{X}$, $\bar{X}_W$, and $\bar{X}_B$, and not just the deviations, $\bar{X}_W - \bar{X}$ and $\bar{X}_B - \bar{X}$, as in Social Indicators, 1973. Of course, in some contexts the deviations themselves could be of substantive interest.

These graphical displays may easily be elaborated, to take advantage of the beautiful colors and graphics of Social Indicators, 1973. What we have here are displays of "typical" population values for Blacks and for Whites. But we can also present information bearing on the population distribution by using one or more bands to illustrate quartiles, and/or other suitable quantiles. In Figure 3 we give an example of such bands for Science Achievement, using hypothetical quartiles. These bands represent one point in time. How interesting they would be if used with the time series data that the National Assessment project will be collecting over the next few years. The purpose of plotting these bands is to introduce the variability associated with the population of achievement scores.

Figure 3*

SCIENCE ACHIEVEMENT BY COLOR



*The Display here is similar to Original Display of Figure 1, but with bars representing hypothetical interquartile ranges.

In addition to the variability considered above, there is another type of variability that we must also consider; viz., the variability or e error which results from estimating (from a sample) the typical values for a population. Later in this paper, we turn to this notion of variability and the ways we feel it should be considered.

Our earlier discussion of Charts 3/8 through 3/11 served to illustrate problems (a) and (b) as described at the beginning of this section. Our next example will serve as another illustration of problem (b). We consider now the display of grade enrollment by race, sex, and age in Chart 3/7 [p. 80]. The graph presents the percent of students below the modal grade. A more symmetric picture would have presented the percent in the modal grade and the percent above the modal grade as well as the percent below it. Although the percent below the modal grade is of special interest, information about the complete distribution may enable us to gain greater understanding of the meaning of the magnitudes displayed. The information presented in Chart 3/7 in its present form is potentially misleading.

While we are critical of some of the graphic presentations in Social Indicators, 1973, the graphics are among the best we have seen in such a report, not only because of the helpful use of color, but also because the authors have generally observed relatively high standards of presentation.

ACCURACY AND ERROR STRUCTURE

How accurate are the data and the series reported in Social Indicators, 1973? The Introduction does touch on accuracy as follows:

> "Most of the series included have been taken from Federal sources; their quality can be verified by those agencies. For data compiled by non-Federal sources, we have, wherever possible, relied on the judgment of those working directly with the data regarding their suitability for this publication" [p. xiv].

Yet in most cases Federal agencies have not prepared adequate studies of error for their series. (The Bureau of the Census is one of the important exceptions in this regard.) Thus agency verification can usually be only unsupported assertion. It is unfortunate that the Statistical Policy Division of the Office of Management and Budget has produced a major statistical publication without a serious discussion of error, especially in light of the recommendations and the general thrust of the President's Commission on Federal Statistics.

The absence of a discussion of data accuracy seems unfortunate as a matter of principle and statistical standards, and it may also lead to misunderstandings and mistakes. For example, the relatively innocent reader may note a difference between two tabulated values dominated by random variation and conclude that some real pattern exists when in fact this is not the case.

We are familiar with the notion of random sampling error; in order to keep such errors at low levels, it is necessary for us to have samples that are appropriately large and well-designed. While many of the Charts in Social Indicators, 1973 are based on sample data, few of the related Technical Notes give any details on sample design, and in only one case are we given an estimate of sampling error [(Note on Charts 2/20 and 2/21) p. 63]. In addition to the random component of sampling error, whenever nonprobability sampling techniques are used at some level of a survey, there enters the possibility of systematic sampling errors.

Whether or not sampling error is present in a study, there may be appreciable errors of measurement, such as nonresponse and various types of response errors (misunderstandings, failures of memory, deliberate falsehoods). These measurement errors may be random or systematic, and the taking of a census rather than a sample does not remove such systematic errors. For example, it is well-known and well-documented that the U.S. Census of Population has a systematic undercount (bias), the magnitude of which differs by race. For the 1970 census, it has been estimated that approximately 1.9 percent of the whites and 7.7 percent of the non-whites were not counted, and that the undercount of some non-white male five-year age groups was as high as 18.5 percent.[4] (These are estimated rates, however, and they too are subject to various sources of error.) This systematic measurement error is especially relevant for the Population chapter of Social Indicators, 1973, where unadjusted census figures are used for several charts, but it is also relevant for data presented in other chapters as well. For example, if adjustments to reported crime rates took the above mentioned undercount into account, the rates thus obtained might present a somewhat different picture from that associated with the unadjusted rates presented in Social Indicators, 1973.

Some examples of systematic measurement error in population data have been noted in Social Indicators, 1973. For example, in the Housing chapter, the Technical Notes [pp. 202-205] point out that the 1960 Census of Housing underestimated the number of sub-standard units in the U.S. by 536,000 or about 6 percent of all sub-standard occupied housing units. When these data are disaggregated by social and demographic classifications, however, no adjustments for this bias are made, even though the disaggregated data are susceptible to this bias. (An appropriate adjustment for each disaggregated class would, of course, be preferable, but if the information that would be needed to do this is not available, then a uniform adjustment for the disaggregated classes would be better than no adjustment.) Nevertheless, the discussion of the effects of measurement error with regard to this particular index in Social Indicators, 1973 is good, and we would like to see other thoughtful discussions of this sort.

The random component of measurement error often results from the imperfection of the measuring device, and the simplest way to get a handle on the magnitude of this measurement error is to repeat measurements independently and/or to compare measurements obtained using a "more accurate" measuring device. Neither of these approaches is necessarily easy to carry out, and nowhere in Social Indicators, 1973 is there a discussion of this random component of measurement error or of attempts by various agencies to measure it.

Many of the indices used in Social Indicators, 1973, both those used directly and those used indirectly in the construction of other indices reported in the volume, are subject to several sources and types of error. For example, the Consumer Price Index (CPI), which is used to adjust figures for across-time comparisons in the Income chapter, is based on a complex network of samples, not all of which are probabilistic. By instituting a replication design the Bureau of Labor Statistics has attempted to measure both types of random error (i.e., sampling and measurement error), but we have not been able to find any details on systematic errors, nor on how the errors may vary over time.[5] Furthermore, the Commissioner of the Bureau, Julius Shiskin, recently noted that:

> "The weighting of the CPI to take account of the proportion of disposable income spent in various items in the index, is based on a survey of consumer price patterns in 1960-61. A new survey is getting under way, but the results will not be available until 1977."[6]

Whether much or little is known about the error structure of a survey or a particular index, it is not sufficient to refer those wishing to examine in greater detail the material in Social Indicators, 1973 to the "quoted sources," many of which are unpublished reports or studies. Even the President's Commission on Federal Statistics was unable to obtain detailed information on error structures from a large number of federal agencies. The Commission's Report notes that

> "although there was considerable variation, both for different statistics in the same agency and across agencies, the responses to the [Commission's] survey showed disappointingly little knowledge of error structure. Sampling errors were estimated for most statistics based on probability samples, but there were, with only a few exceptions, very few analyses of response and other nonsampling errors, even in cases in which, because of long recall or the use of incomplete records, these were likely to be substantial."[7]

It is our hope that in conjunction with future editions of Social Indicators, the Statistical Policy Division of the Office of Management and Budget will compile in a form suitable for publication detailed information regarding what is known and what is unknown about the error structure for each of the series in the main report. This information should include descriptions (where relevant) of:

(a) sampling frame, sampling plan, and (effective) sample size,
(b) estimates of sampling error,
(c) any special or nonstandard aspects of questionnaire design or interview procedures,
(d) non-response rates, and treatment of missing observations (if the problem is substantial),
(e) degree of consistency and compatibility with related series of measures.

These descriptions need not be voluminous; they need only be summary in nature, with references to more detailed technical presentations.

This information on error structure, required by those who wish to draw inferences from the data in Social Indicators, 1973, should probably be published as a companion booklet, and only brief statements should be included in the Technical Notes in the main report. If it is unrealistic to expect such a companion booklet, then a brief checklist format could be included in the Technical Notes which would alert the reader to what is known or unknown about the error structure for each data set or chart. The Bureau of the Census, for example, provided the President's Commission on Federal Statistics with a detailed eight-page summary of information on the error structure of the Current Population Survey (CPS), and data from the CPS have been used in Social Indicators, 1973 in the chapters on Employment, Income, Housing, and Population. Something less detailed would suffice for the next edition of Social Indicators.

OBJECTIVE VS. SUBJECTIVE

Much discussion in the field of social indicators has focused on the use of objective versus subjective measures, and the Introduction of Social Indicators, 1973 touches on this point [p. xiii]. We believe that it is important to point out that there are two different senses in which a measure can be objective or subjective.

William Kruskal notes that phenomena may be subjective in the sense of being inside people's heads (attitudes, aspirations, happiness) or objective in the sense of being directly observable (dead-alive). Similarly, modes of measurement may be subjective (opinions about the magnitude of the crime problem in 1974) or objective (actual counts of reported crimes in various categories). Of course, there are philosophical difficulties as to what is objective and what is subjective, and there are intermediate, blurry cases, but roughly speaking, we can think in terms of the 2 x 2 cross-classification:

		Mode of Measurement: Subjective	Mode of Measurement: Objective
Phenomenon of Interest	Subjective	a	b
	Objective	c	d

In reading through Social Indicators, 1973, we have found examples of indicators that correspond to each of the four cells. Of course, the bulk of the measurements appear to be objective-objective, but a closer examination reveals that the mode of measurement in many of the so-called objective-objective cases might be classified as subjective. For example, are self-reports, such as those used in the compilation of disability data in the Health chapter objective or subjective? This depends in fact on instructions given to respondents, and we are not provided with these details in the Technical Notes. Indeed, the classification of various sets of data according to the 2 x 2 table described above is itself subjective in character.

Now let us turn to the other three cells in this 2 x 2 table (i.e., the a, b, and c cells). The indices of substandard housing and crowded conditions in the Housing chapter provide examples of subjective phenomena that are measured objectively [Charts 6/1 through 6/7, pp. 206-208]. The notion of "substandard" is a subjective one, but to measure this phenomenon, the Bureau of the Census has used criteria that are objective (more or less). Measures of this kind (i.e., where there is a subjective phenomenon and objective measurement) are often of interest to social scientists because they are related to social concerns and are available on a consistent basis over time. Indeed, in the present case, the index of substandard housing has been criticized, in part, as a consequence of its consistency over time. The definition has been consistent, but many complain that it is no longer meaningful.

Albert Biderman suggests another example of subjective phenomenon-objective measurement: the Uniform Crime Reports offense series, which he claims serve as important and quite accurate state-of-society indicators as they reflect people's perceptions of the magnitude of the crime problem. He would thus dispute the common view of these series as objective-objective.

Chart 1/26 [p. 21] provides us with an example of a subjective measure of a phenomenon that appears, at first glance, to be objective. This chart is based upon the respondent's assessment of the confidence he has with respect to his access to "good" medical care. If "good" had been defined for the respondents, then the phenomenon might be objective; if "good" were not defined then we would have the respondent's perception of what is "good," in which case the phenomenon would be subjective. The Technical Notes do not provide us with enough information to determine how "good" was actually defined in the study.

Far more prevalent than subjective-objective and objective-subjective indices are subjective-subjective indices in Social Indicators, 1973. For example, in the Employment chapter, there is a subjective evaluation of the highly subjective notion of "job satisfaction" [Charts 4/16 and 4/17, pp. 123-124]. This is a reasonable approach to one dimension of a quality of life index. Further examples of subjective-subjective measures can be found in the chapters on Public Safety [pp. 58-59] and Housing [pp. 200-201].

STATISTICAL ANALYSIS AND INTERPRETATION

One of the more striking features of Social Indicators, 1973 is the apparent lack of analysis and interpretation of the statistics included. Obviously, judgment has been exercised in the choice of which indicators to include, which to eliminate, what types of disaggregation to exhibit, when to show component elements of an index, and so on. Moreover, inferences are directed, at least implicitly, but the material selected and the manner of presentation. The reader is left to determine, without explicit guidance, what patterns are present in the data, what they mean, and what importance to attach to this meaning.

Information about the accuracy of the data is essential to proper statistical interpretation, and, as we noted earlier, the reader is not given this information in most instances. Although many of the indicators included are presented in the form of general purpose statistics, it is all the more difficult to discuss accuracy when we don't know what questions we wish our data to help us answer. It would have been helpful if Social Indicators, 1973 had indicated in more detail the purpose of including various charts and data sets.

Most of the measures reported in Social Indicators, 1973 have been drawn from existing statistical series produced by federal government agencies, and the bulk of these measures are reported in the same manner as in the official reports of the various agencies, or as in the Statistical Abstract. In addition to the measures reported in Social Indicators, 1973, the volume also includes impressive graphical displays. In some cases, such as in the Population chapter, the graphical displays include some statistical projections, which are based upon various assumptions regarding fertility, mortality, and migration. The graphical presentation of this material is highly informative.

We feel that more projections and more detailed statistical analyses are desirable in a social report. In addition, we believe that statistical analyses should be coupled with at least some statistical interpretation and comment.

An examination of the Public Safety chapter will allow us to focus on the type of analysis, interpretation, and comment that is both feasible and desirable. If you glance quickly at Charts 2/1 through 2/3, [pp. 44-45], and 2/15 through 2/17 [pp. 53-55], you cannot escape the (naive and perhaps erroneous) conclusion that the rate of criminal offenses has been increasing over time, although there is considerable variability in the rate of increase for different categories of crime. You may even notice the downturn in property crime for 1972, indicated in Charts 2/15 through 2/17. Moreover, Charts 2/13 and 2/14 [pp. 51-52] seem to indicate that the rate of commission of violent crimes (in urban areas) is highest for 15-24 year olds and for Negroes; and Charts 2/8 and 2/9 [p. 49] seem to indicate that the offenders (in a sample from 17 major cities) are mostly male and that they are to a large extent Negro rather than White.

Having drawn these apparent conclusions, we now ask whether they are warranted. If so, then we might ask what are the causes of the increases in crime and the means by which crime can best be prevented or controlled. To answer the first of these questions, we must know something about the error structure of the indices being used. Some of the limitations of the Uniform Crime Reports data are discussed [pp. 60-61], but little attention is given to accuracy in the statistical sense. There are many who argue that the statistical limitations of the data make it difficult to determine not only the magnitude but also the direction of changes in rates. Albert Biderman, for example, has stated:

> "I contend that most of the sources of error operate to inflate the newer figures relative to the older ones, resulting in a false picture of rapidly increasing lawlessness among the population. With respect to most of these sources of error, it is extremely difficult and sometimes impossible to give quantitative expression to the factor.
>
> Nevertheless, in examining several published criticisms of the index, and in subjecting it to my own critical examination, I believe that the following three conclusions emerge:
>
> 1. The errors and biasing factors affecting the Crime Index largely operate to show spurious increases, rather than decreases, in the rate.
>
> 2. The Crime Index does not provide a sound basis for determining whether criminal behavior is increasing, or decreasing, in the United States.
>
> 3. The Crime Index is highly sensitive to social developments that are almost universally regarded as improvements in the society. Thus, it is altogether possible that year-to-year increases in crime rates may be more indicative of social progress than of social decay."[8]

Are there other simple explanations for aspects of the apparent increases? Some explanation is suggested by a reworking of existing data relating to the size and age composition of the population. Since Chart 2/13 reveals that young people commit a disproportionate share of crime, even if the propensity to commit crimes remained constant over time for all age groups, an increasing proportion of individuals in the 15 to 24 year-old bracket could lead to an increasing crime rate.[9] Other demographic characteristics that obviously should be considered in a similar manner are race, sex, and geographical location. There is simple and straightforward statistical technique called "standardization," well known to demographers and epidemiologists, which adjusts rates or proportions for such factors. While the figures that we need in order to standardize the Uniform Crime Reports rates are not given in Social Indicators, 1973, some related direct standardization calculations have been carried out by the Commission on Population Growth and the American Future. They report that

> "About 28 percent of the reported increase between 1960 and 1970 in the number of arrests for serious crimes can be attributed to an increase in the percentage of the population under 25. Another 22 percent of the increase can be explained by the growing size of the population and other demographic factors. Thus, population change alone accounted for about half of the reported increase in the number of arrests for serious crime over the past decade."[10]

The inclusion of appropriate standardized rates in future editions of Social Indicators would be informative. If such analyses were applied to rates for various types of crimes, we might be able to decide if, for example, the 1972 dip in property crimes is spurious. More important for these future reports are new forms of data such as those now being produced via the National Crime Survey. These new data make possible more thorough statistical analyses. Hopefully, the Law Enforcement Assistance Administration will carry out the appropriate statistical analyses, and will provide the Office of Management and Budget with appropriate summaries.

It has been suggested that other factors that may possibly account for the increases in the crime rate over time are:

> "...(1) more widespread and intense identification with the norms of the national society, (2) greater integration and effectiveness of the economic and social systems, and (3) more effective operation of the formal agencies of control, such as police and courts."[11]

We have yet to come to grips with the possible causes of crime, and the policy implications for its control. What we need, besides more accurate reports on the incidence of crime, are real experiments. We also need measures of potentially related conditions such as the extent and dimensions of narcotics addiction, and also a variety of possible "leading" indicators (as the economists would say) for criminal activity such as, for example, school truancy rates. Exploring the interrelationships among such variables and their relationship with crime indicators would involve careful statistical analysis.

Other problems in the interpretation of crime statistics revolve around the issue of the incidence of offenses versus the prevalence of offenders, and the use of longitudinal versus cross-sectional data. The Uniform Crime Reports and the victimization data report cross-sectional incidence of crime, but they shed no light on the question of whether certain segments of society, after adjustment for relative size, are committing an increasing or a decreasing amount of crime. The development of an adequate longitudinal system of criminal statistics, as proposed by the President's Commission on Federal Statistics, might shed light on this problem. Again, these are statistical as well as substantive matters.

While we recognize that the exploration of causes and cures is not within the purview of

Social Indicators, 1973, we believe that such a report should present analytical measures that may aid the users of the publication who may have the obligation to interpret, or to take positions on, causes and cures.

For a final and somewhat more detailed example of statistical concern, we turn to the Technical Notes for the Public Safety chapter [p. 61], where we are told of the concern of crime analysts but not of statistical analysts:

> "... a serious problem with the NORC survey was the small sample. Of approximately 2,100 crime incidents identified from interviews carried out in 10,000 households, only 18 were forcible rapes. Crime analysts have questioned the validity of a national rape victimization rate on such a small number of incidents, particularly rates by race and age."

First, we feel that this statement conveys a misconception regarding the accuracy of estimated rates in sample surveys. The standard deviation of an estimated rate decreases as (a) the size of the sample increases, and (b) the size of the true rate decreases (for rates less than 0.5).[12] With respect to the NORC survey, the sample size (10,000 households) is relatively large in statistical terms, and the rape rate (estimated at approximately 0.002) is relatively small in statistical terms, so the standard deviation of the estimated rate will be relatively small. Although the 0.002 rate is relatively small in statistical terms, it is not at all small from some other points of view. Accuracy is a relative matter, and if we wished to compare the victimization rate from this survey with a rate produced from other sources, the accuracy of the estimated rate may not be high enough to detect small relative differences.

Second, the part of the statement dealing with the more detailed breakdowns of crime incidents by race and age seems to ignore the statistical activity of the past decade on the analysis of multiple cross-classifications. By using various statistical techniques, such as unsaturated log-linear models, we can smooth cross-classified data where many of the original cell counts are zero. The resulting smoothed data can then be used to get useful estimated rates.[13]

Such statistical methods could be applied to analyze more completely data reported in various chapters of Social Indicators, 1973, e.g., in the Public Safety and Education chapters, but the data in Social Indicators, 1973 itself were not sufficiently detailed for these analyses. In some attempts to go to the source, we found either that the data were from unpublished reports (unavailable to us), or that the cited sources still did not contain the relevant detailed information required. The problem here resides in the limitations of the statistical reports currently produced by some federal agencies and private research organizations which form the basis of Social Indicators, 1973.

SUMMARY

In this review of Social Indicators, 1973 we have focused on four major statistical considerations:

(a) the need for care in reporting data in graphical form,

(b) the need for detailed discussion of the error structure associated with each of the indicators reported,

(c) the need for careful statistical analyses of the data presented in the volume and of related data obtained from other sources,

(d) the need for statistical interpretation of the information reported.

We envision an increasingly important role for statistical analysis and related considerations in future editions of Social Indicators.

NOTES

1. This paper is a revised version of remarks prepared for the Review Symposium on Social Indicators, 1973, sponsored by the SSRC Center for Coordination of Research on Social Indicators, Washington, D.C., February 21-23, 1974. The paper appears in Social Indicators 1973: A Review Symposium edited by R.A.Van Dusen (1974), Social Science Research Council, New York, pp. 63-82, and is reproduced here by permission of the Social Science Research Council. It is based on coversations and written memoranda from several of our colleagues, as well as on our own conversations and observations. Specifically, we wish to thank Albert Biderman, O.Dudley Duncan, Morris Hansen, William Kruskal, Robert Parke, Joseph Waksberg, and Hans Zeisel for their help. In addition, we have drawn upon such sources as Federal Statistics: The Report of President's Commission, Volumes I and II (1971). U.S. Government Printing Office, Washington, and Mathematical Sciences and Social Sciences (prepared in connection with the survey on the Behavior and Social Sciences), edited by William Kruskal (1970), Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

2. This work was supported in part by Research Contract No. NSF 31967X from the Division of the Social Sciences of the National Science Foundation to the Department of Statistics, University of Chicago.

3. The comments we make are applicable in a slightly revised form to medians.

4. J.S. Siegel, (1974). "Estimates of coverage of the population by sex, race, and age in the 1970 Census." Demography 11, 1-23.

5. For various statistical discussions of the CPI see:
(a) W. H. Kruskal and L. G. Telser (1960). "Food Prices and the Bureau of Labor Statistics," J. Bus. Univ. Chicago 33, 258-285.

(b) BLS Handbook of Methods for Surveys and Studies (1971). Bulletin 1711. U. S. Government Printing Office, Washington, D. C.

(c) Zvi Griliches (ed.). (1971). Price Indexes and Quality Change. Harvard University Press. See especially pp. 185-197 and pp. 233-234.

6. Excerpts of a statement by J. Shiskin presented at the Washington Journalism Center, as reported in the Minneapolis Tribune, February 8, 1974. It is worth noting that, in his former capacity as Chief Statistician for the U. S. Office of Management and Budget, Mr. Shiskin was responsible for initiating the preparation of the document presently under review.

7. Federal Statistics: The Report of the President's Commission. Vol. II (1971), U. S. Government Printing Office, Washington, D.C. "How much do agencies know about error structure," by H. Grubert, pp. 297-334.

8. Albert Biderman, (1966) "Social Indicators and Goals," pp.68-153 in Social Indicators,,edited by R. A. Bauer, MIT Press, Cambridge, (see p. 115). The Crime Index referred to in the quotation is a composite index based on the Uniform Crime Reports series for criminal homicide, forcible rape, robbery, aggravated assault, burglary, larceny, and automobile theft.

9. Another factor having substantial impact on crimes committed by males in the 15-24 year-old age cohort is military service. This point deserves considerable attention, but we do not pursue it here.

10. Population and the American Future, (1972), U. S. Government Printing Office, Washington, D.C., p. 22. In Social Indicators, 1973 we are given rates per 100,000 population rather than the actual numbers of arrests referred to in this quotation. These reported rates already adjust for the growing size of the population and so standardization could not be expected to account for half of the increase as given in Social Indicators, 1973. Also, the standardized rates ignore interactions among the demographic factors.

11. Albert Biderman, op. cit., 115-116

12. If p is the true rape rate and n is the sample size, then the standard deviation of the observed rate or proportion is $\sqrt{p(1-p)/n}$. On the other hand, it is of course true that the coefficient of variation (i.e., the standard deviation divided by the mean), which is equal to $\sqrt{[p(1-p)/n]}/p = \sqrt{(p^{-1}-1)/n}$, will increase as p decreases.

13. For detailed discussion of these techniques see, for example:
(a) Y. M. Bishop, S. E. Fienberg, and P. W. Holland, Discrete Multivariate Analysis: Theory and Practice. Cambridge, Mass.: MIT Press, 1974.

(b) L. A. Goodman, "A general model for the analysis of surveys," Amer. J. Sociol: 77, (1972) 1035-1086.

SOCIAL INDICATORS AND POLICYMAKING
Some Comments on Social Indicators 1973

Richard C. Taeuber
Committee on National Statistics
National Academy of Sciences - National Research Council

Social Indicators 1973 is a first effort of limited purpose -- the first volume of its kind to be published by the Federal Government. Quite naturally, it is open to criticism -- but any praise, condemnation, evaluation, or criticism should remember this firstness, and thus be offered in a positive nature aimed at improving Social Indicators 1976. These remarks are intended to be taken in that light. To paraphrase Marc Antony, I come to praise indicators, not to bury them.

Chinitz has said that "the demand for instant wisdom where there is monumental ignorance is infinitely elastic".(1) In addition to seeming to respond to such a demand, Social Indicators 1973, especially in a policymaking sense, seems more of an overview of some prior social scene, rather than an indicator of actionable areas as one might expect from the title. Who was the expected audience? What was the expected purpose of the volume? Would OMB/SPD have been better advised to be less pretentious, or less ambitious, in putting out this volume which could be expected to be a definitive chronicling of the social status of the nation? Can one realistically produce a set of indicators or summary information measures, that are so universale that they can serve all audiences and all purposes simultaneously? The volume seems to have been shaped by data collectors and data presenters, rather than by researchers or policy analysts. This may be due partly to an intention to rely solely on available data. Since a social indicator can be constructed only if one knows how to construct it and if the data needed for its construction exist, many desirable indicators simply are not available.

Background

Social Indicators 1973 finally appeared in February 1974 but its history can be traced back at least to 1969 testimony by the then Bureau of the Budget that they were in the process of developing "a publication on social statistics" which would "contain regular series on the most significant" social topics. It would also serve "as a clearinghouse for the results of individual studies that should provide insights into social problems as well as some possible solutions. ". . . The first task is to organize what is available . . . over the years, as statistical gaps are identified, new statistical series will be developed and initiated to fill the gaps".(2) In a recent paper, Dan Tunstall stated that Social Indicators 1973 "is primarily a book of statistics. The statistics were chosen to be indicators of major national social concerns and to show change in social conditions over time. The focus of the publication is the American people. . . . Given a framework of national social concerns we went about the task of selecting one or more indicators for each concern. Social indicators are defined as statistical measures of the most important aspect of a concern."(3) Elsewhere in the paper, he went on to say that he interpreted "social reporting to mean something quite different. I defined social reporting as a process of informing the public about the conditions and trends of its society."

These two statements seem to fit the volume which has appeared in that the term "reporting," or the term "inform," seems to be much more applicable to this volume than does the term "indicators." Indicators, if they are to serve policymakers should, at a minimum, help delimit problems and problem areas and suggest action potential. If one raises the question: Is anything actionable from this volume or are the data of the volume only indicative of the social scene?, one has to state the latter to be the case. Its information areas are worthy of attention, but there is no indication derivable from the volume of any action or intervention strategies. Although this can be taken as a criticism, the fact that the volume is more information-oriented seems entirely consistent with the two quotations mentioned earlier.

In terms of its informational usage, approximately half of the first printing run of 10,000 were distributed to persons involved in policy formulation and execution such as members of Congress, senior committee staffs, agency heads, select assistant secretaries, Governors, state budget officers, mayors, etc. And, indeed, there have been stories of at least a few recipients of the volume having received some informational insights.

Definitions of Indicators

Although many persons feel that the use of the term indicators in the title of this volume is inappropriate, there is a major dilemma in substantiating that position. Both the term indicators and the term policy have a major commonality in that all of us know what each term means, but none of us can commit to writing a definition which will be universally acceptable.

The definitions of "indicator" put forth to date include that presented by Bauer in the 1966 volume he edited which suggested that social indicators are "statistics, statistical series, and all other forms of evidence that enable us to assess where we stand and are going with respect to our values and goals and to evaluate specific programs and determine their impact."(4) Biderman in an article in the same volume described his focus as being "quantitative data that serve as indexes to socially important conditions of society."(5) Another definition was put forth by Sheldon and Moore in 1968 -- "Such indicators would give a reading both on the current state of some segment of the social universe and on past and future trends, whether progressive or regressive, according to some normative criteria."(6) In 1969 Towards a Social Report included the statement that "a social indicator, as the term is used here, may be defined to be a statistic of direct normative interest which facilitates concise, comprehensive and balanced judgments about the condition of major aspects of society. It is in all cases a direct measure of welfare and is subject to the interpretation that, if it changes in the right direction, while other things remain equal, things have gotten better, or people are better off. Thus, statistics on the number of doctors or policemen cannot be social indicators, whereas figures on health or crime rates could be."(7) In 1972, Campbell and Converse saw two distinctive emphases associated with the definition of social indicators. "First, the term is intended to convey a stress on descriptive measurement which is much more dynamic than most social science research has been to date. . . . Second, and perhaps more noticeable, the call to arms represented by the social indicator movement lays a heavy stress on policy relevance."(8) In the volume under discussion, the authors say that their choice of indicators is based upon two criteria: "That the indicators measure individual and family (rather than institutional or governmental) well-being and that they measure end products of, rather than inputs into, social systems. In education, for example, the indicators were selected to measure individual achievement and attainment rather than inputs, such as school budgets, classroom construction, and the number of teachers."(9)

But every one of these definitions can be criticized in some manner or other. Use of the term normative, for instance, presents problems in that the normative of today may not be normative of tomorrow, or the normative of one policymaker or branch may not be the normative of others charged with action in the same problem area.

Whatever definition one chooses or might eventually evolve to lend more order to the indicator field, one has to agree that social indicators are constructs with a theory behind their choice or construction. The theory may be explicit and formally expounded as part of the discussion or presentation of the indicators, or it may be implicit in the minds of those who have selected the chosen indicators for whatever purposes they may be trying to advance. Indicators may be a simple presentation of data or they may be synthetic in that they are constructed from several series. They must have a sense of time and be based on observations, usually quantitative, whether they be objective in purporting to show what a position is or how it is changing or whether they be subjective in the sense of purporting to show how the objective is regarded by the community in toto or in constitutent groups. Social indicators have to relate to some area of social concern and have to be value- or action- or theory-oriented--at least in the eyes of those putting forth the indicator because even those espousing a purpose of satisfying curiousity or providing understanding undoubtedly have some end objective in mind, and thus some potential policy.

Role of Indicators

In commenting on policy implications or policy considerations of Social Indicators 1973, I think one has to face not only the question of what indicators are, but also the question of what the role of indicators should be. One view is that "The major purpose of social indicators is to affect the general image of society and the fund of knowledge about social changes intelligent people have. If the thousands of actors whose work on interactions aggregate to 'social policy' know what society is all about and if the publics to whom they relate share these understandings of what is important and why it is important, policy will improve."(10) This is somewhat different from the view that it is "meaningless to speak of a measure or an observation of condition, that is an indicator, which is value- or action- or theory-free."(11) That would seem to abide with the feeling that the role of indicators should be just that -- to indicate problems and interrelationships or even non-problems, assuming that there is potential to change the conditions which have produced that which is being indicated.

Springer specifies "two essential components of information: (1) specific items of data, and (2) inference structures that order these data items in some model of the relationships between goals and means" and further that "the characteristics of the demand for indicators -- the perspectives, the needs, and interests of our presumed clients -- should guide our work."(12) This seems to ties the two previous positions at lease closer together, for even those advocating indicators which serve solely a curiosity or understanding function require a theoretical or interpretational structure for the selection of a specific element or subset of the data series and constructs, actually or potentially available. And whatever the theory or action-motivation involved on the part of the producers or conveyers of summary information or indicators, they ignore at their peril the fact that a consumer of that selective bit of information will have his or her own attitudes, preconceptions, understandings, biases, etc. which have derived from personal knowledge and experience.

Economic indicators, which those in the other social sciences are attempting to imitate, have uses beyond the direct policy or direct action uses in that they are seen not just by policymakers or the Council of Economic Advisors but are released to the press and the public. Thus there are many persons outside of direct policy roles who can make their own attempts to suggest or impact on policy. Indeed "a social report or a set of social indica-
to tie the two previous positions at least closer
two ways. First, it gives social problems more visibility and thus makes possible more informed judgments about national priorities. Second, by providing insight into how different measures of national well-being are changing, it may ultimately make possible a better evaluation of what public programs are accomplishing."(13) This role would also serve well a dictum put forth by George Washington, applicable to policymakers even today, which says that the people must feel before they will see.

Social indicators thus seem to have roles in serving the needs of curiosity, of understanding, and of action or any combination of them. And, one can tie all three together by noting that curiosity and the wish for understanding frequently arise from a desire for action.

Research-Policy Interrelations

An aspect of the research-policy interaction which plagues the researcher is the problem of having findings considered. A classic case of positive and significant research findings (which could have been in the form of indicators) being ignored, is the inconsistency of interpretation of school-enrollment and teaching manpower data in the late 60's. All elementary school enrollees for the next five or six years were then alive, and projections showed an upcoming decline in enrollment; but despite the obvious implication of a decreasing need for teachers, policy encouraged the expansion of teacher training. By 1972 the supply of newly trained elementary and secondary school teachers was running twice what was needed -- up from about an even match in 1968. One aspect of the problem was that although this trend was very evident at the national level, most of the action potential is at the state and local level -- but the state population estimates and projections of the Census Bureau contained no age detail; there was no mid-decade census as a reference point; and the 1960 census data were too old to show the trend.

Another case in point is the Big Steel strike of the late 40's where a major issue was pensions. Big Steel rejected the idea out of hand rather than admit to an absence of information on the age structure of their work force, and thus the cost impact of yielding on the pension issue. Just over a year ago at a meeting of regional policymakers and planners a member of a state legislature mentioned the passage of a bill to provide tax breaks for the elderly residents of their state even though they were unsure of the financial impact -- although enthusiastic, they passed the legislation on faith, for they did not know how or where to ask about the number of elderly the state then had or were expected to have in future years.

There is more to having indicators, or information, or data then just creating them. They have to be communicated, as Social Indicators 1973 does so well. Statisticians need to be more alert to this need, but at least an equal portion of the problem lies with those who complain about the absence of data, or those who bemoan the supposedly excessive expenditures on data acquisition which then can not be used to answer any policy question. Both sides of a dialogue have to be aware of the interests, needs, and problems of the other if communication is to be effective.

Can we bridge the communication gap between the statistician or researcher and the policymaker? On one side of the gulf is the social scientist who is concerned with studying the existence of a problem and its causal and interactive aspects but who adheres to the traditional role of the academician by not including policy commentary, and not proposing remedies or evaluating the impact of alternate remedies. On the other side of the communications gap is the policymaker who is concerned more with action for a politically live problem area (possibly more so than with effective action actually to attack the real problem); or who may be concerned with producing change or improvement. (Or, as has been commented, policymakers are "concerned with changing society rather than understanding it."(14)) If social scientists or statisticians are to get involved with applied or policy research, there would appear to be a need to change some of the ground rules so that, at a minimum, questions can be properly posed and answers can have proper relevance (or the Type III error: the right answer to the wrong question can be avoided).

Policy research is different from discipline research in that:

o decision deadlines dictating that partial information at decision time is more useful than complete information two days later;

o elegance is less important than correctness of the predictions or results;

o redundancy is important; and

o the objectives are not a substantial contribution to existing knowledge but the modification of social policy.

As contrasted to disciplinary research, in policy research the problem should be formulated outside the discipline. As stated by Lord Rothschild, Head of the Central Policy Review Staff of the British Government:

> "The research worker should <u>not</u> formulate the objective, although he can and should help. The research worker should <u>not</u> decide that the objective requires research for its achievement. He should <u>not</u> decide that the research should be done, assuming it is necessary. He should <u>not</u> decide when to stop. Nor should he decide to change the objective in mid-stream, however desirable it may seem to him to do so."(15)

As true as this is or should be, there still remain some of the traditional roles of the researcher. Coleman defined the split between potential forensic activity and detached objectivity by stating, as a principle of policy research:

> "Those stages of policy research that lie in the world of action, formulation of the research problem, posing conditions for communication of the research results back into the world of action, and making policy recommendations based on the research results, should be governed by the investigator's personal values and appropriately include advocacy, though stages which lie within the disciplinary world, execution of the research and statement of the research results, should

be governed by disciplinary values and do not appropriately include advocacy."

I admit to being very mindful of the dictum that the role of statistics is to produce information, not to make decisions, and the concluding statement in Towards a Social Report says "that social reporting cannot make the hard choices the nation must make any easier, but ultimately it can help to insure that they are not made in ignorance of the nation's needs." (17) I am also aware of the recent statement of President Ford that he believes that truth is the glue that holds the government together, but it seems to me that policy-makers have the right and responsibility to argue or to ask whose truth; to ask who determines which is the real truth and how much reliance is to be placed on competing versions or values.

As to the role of the researcher, Nassau Williams Senior wrote in 1836 that "the business of a political economist is like a juryman, to give deliverance true according to the evidence and allow neither sympathy with indigence, nor disgust at profusion or at avarice, neither reverence for existing institutions nor detestation of existing abuses, neither love of popularity nor of paradox nor of system, to deter him of stating what he believes to be the facts or from drawing from those facts what appear to him to be legitimate conclusions. To decide in each case how far these conclusions are to be acted upon belongs to the art of government." (18)

Although the realities of internal and external politics may dictate that a report such as Social Indicators 1973 be as bland as it is, I would prefer to have seen alternative interpretations or evaluations included in the volume for the reader or policymaker to take under advisement. I happen to agree with those who are attempting to redefine the proper role of the researcher, especially the academic researcher to a forensic structure which permits or encourages advocacy. My beliefs here are best summarized in the words of Rupert Vance, who said, "thus, in spite of his modesty, the social scientist who uncovers and analyzes social facts will be asked, 'what do you recommend?' As an honest man who values his own integrity, as a citizen who admits of public duty, and as an expert in whose training society has made an investment, the social scientist, after admitting his reservations of ignorance and bias, must indicate his choices for policy, whatever they may be worth. Nor should he be overwhelmed by this assumption of high responsibility, for he may rest assured that even his facts will be discounted by practical men of affairs as impossible theory, while his cautious recommendations will be regarded as partisan statements by every faction whose interest they oppose. But if his facts are facts and still disregarded, he may take what consolation he can to himself in the knowledge that what they will also count in the long run to come." (19)

SOCIAL INDICATORS 1973

Turning more specifically to Social Indicators 1973, the extreme neutrality of the volume--except for the underlying assumptions implicit in the given data series and in their selection for inclusion in the volume--is bothersome, as is the absence of commentary or interpretation. People in general and policymakers in particular are expected to know what economic indicators mean, and eventually social indicators may have the same statement made of them. It is one thing to present trends which this volume does magnificently with charts and figures; and another to have them interpreted or perceived as having any relevance to problems facing decisionmakers. Failure in this sense may well derive naturally from the political realities with which the Statistical Policy Division staff had to live.

Additionally, interpretations of social data, trends, or indicators are always or continually befogged by ambiguity and the differences in the interpreter's background, capability, responsibility, or interests. To indicate possible methods of interpreting indicators does not require policy commentary, nor should it. But this should not be an excuse for avoiding commentary, for as Parke and Sheldon state: "if the careful documentation of methodology is the primary responsibility of the statistician, surely a close runner-up is the responsibility to utilize accepted analytical techniques and methods of data presentation which will enable the data to tell the story that will not be told in the absence of analysis." (20)

Use and Usefulness

Nathan Caplan, from a study of the use of scientific information by government executives (21), concludes that five out of ten Federal officials fall into a low-usage group and thus represent a sizable challenge to the producers of scientific information. Further, while such officials are eager for more information, they may not be able to assimilate it effectively. If one is to reach the low-usage half at all, and to help the other half of decisionmakers, there needs to be some discussion of either how to interpret the indicators or of what alternate interpretations might mean. To introduce interpretive analysis, possible or suggested methods of usage, and alternate implications for the future, one option might be a companion volume or quarterly journal--if policy implications are to be avoided in the basic volume. It could also report on actual uses of the information presented in volumes such as Social Indicators 1973.

Statistical Considerations

Another reason for advocating some commentary is that there should have been a fuller treatment of what is and what is not measured by any given indicator, and also the possible relevance to various sets of consumers of the information contained in the indicator. This volume must be criticized for the absence of measures of accuracy of the reported indicators. If policy makers are to use indicators effectively, the sources or structure of error in indicators must be discussed because the assessment of the accuracy of indicators has to be an important part of decision making.

Indicators, to be useful in a policy-making sense, have to address the future. That is, those who present indicators, for whatever their purpose, should also present the future or alternate futures under alternative policy assumptions. Policymaking is aimed at changing a trend and altering the future, and thus, to be really useful, indicators must tell not only where we are today as a result of policy but also where we might go as a result of policy changes. Both positions in time may involve projections and estimates, especially if the data base on which the social indicators are based is one or two or more years old. Policymaking by the Legislative Branch, with the larger time frame required to enact new laws, especially requires knowledge of the dimensions of the problem without action as well as the dimensions under possible aspects of the proposed legislations.

The question of currency of background information poses a special challenge for social indicators, especially uses for policymaking purposes where the user has to look to the future. This issue underlines the need for a quinquennial census--it is hard to imagine the usefulness of Social Indicators 1976 which will have to use the 1970 decennial census as its primary reference for major sections of the report, let alone Social Indicators 1980.

In summary I feel the presentation of interesting summary measures (which a good many of these are) without estimates of error, without the means or manners of alternative interpretations, without forecasts of alternative futures and the conditions and assumptions implicit in these forecasts, produces a volume of limited usefulness to policymakers.

Disaggregation

Towards a Social Report states "if the nation is to be able to do better social reporting in the future and do justice to all the problems that have not been treated here, it will need a wide variety of information that is not available now. It will need not only statistics on additional aspects of the condition of the nation as a whole, but also information on different groups of Americans. It will need more data on the aged, on youth, and on women, as well as on ethnic minorities. It will need information not only on objective conditions but also on how different groups of Americans perceive conditions in which they find themselves." (23) But those wanting indicators to be used for more than curiosity purposes need not only national data but also regional data and disaggregations beyond the age, race, and sex breakdowns used in Social Indicators 1973. If social indicators are to be used for social policy purposes, there must be actionable data for comparable areas. Any political jurisdiction contemplating action needs the ability to compare current conditions of life with standards and goals. Indicators need to indicate the status of population sub-groups in an ethnic or socio-economic sense and/or in a geographic sense. As has been demonstrated in the past, national rates for unemployment can be at acceptable levels, while local areas or specific sub-groups may have a serious problem--e.g., such as black teenagers or a particular occupation group. Finally, if the New Federalism concept continues or expands, it is important that there be comparable information at all the appropriate levels of policy action, especially for the states and cities.

In making this argument, one must concede, however, that this might not be the role of a national volume such as Social Indicators 1973. This does not lessen the need to insure that there is a potential for sub-groups of the population to have their status indicated in a similar manner. When information is available only at a national level and when the policy arena is at a lower political level, the measure could be a highly misleading "indicator", and, if not excluded, included to point out the need for additional detail. One way to produce the sub-national detail is exemplified by the national assessment of educational progress. While the program aims primarily at national assessment, some of the resources are used to provide technical assistance to states and counties to use the national model and technology to conduct their own assessments. Select states and other school jurisdictions have been able to assign their own priorities to assessment measures, and thus been permitted their own interpretation of the data.

Policy Areas

Social Indicators 1973, from its introduction, is as a book of statistics and describes eight major social areas, or "broad areas of interest or social concerns." The taxonomy used represents recognized areas of social concern but for policymaking there is no sense of perception of any problem in these areas; nor is there any parallel structure of policy mechanisms from which one can adduce concerns, an agenda, or the needs of policy and data.

In contrast to the discipline-oriented taxonomy used, one could have hoped for a problem-oriented taxonomy such as Kermit Gordon's listing of the problems facing this country in the next decade: inflation, performance of the public sector, distributive equality, and the inter-related questions of environment, energy, resource depletion, and economic growth. (23) Should not a volume of this ambition and and under OMB sponsorship have been more problem-oriented? For instance, the volume does not include any data dealing with the social or economic consequences of inflation. It could have explored, for instance, the consequences of inflation on diet patterns, consumption patterns, or on housing expenditures as a percentage of disposable income.

OMB's Role

The question must be raised as to the role of an indicator effort within the Office of Management and Budget. One can argue that OMB has an overall management responsibility in the federal government as well as a coordinative responsibility in multi-agency areas of interest, and thus is an appropriate home for an effort such as Social Indicators 1973 and the promised Social Indicators 1976. In looking to the future, one hopes that they can overcome the dual problems of (1) data which are, to be charitable, somewhat out of date at the time of presentation, and (2) the multi-year gaps between the production of this series of volumes. Many of the data, as has been mentioned earlier, were three to four years out of data by the time the volume appeared.

It would be interesting to know OMB's perspectives on why it was producing this volume--was it to produce assistive measures for the policy mechanisms of OMB, the Executive Office of the President, or the Executive Branch in general? Or was it just a volume of some interest to an ill-defined someone which was published with an attractive cover? Might there be a special analysis of the Federal Budget which ties social indicators to a budget cross-cut?

Whatever the motive, why was it not directly supported at an adequate level in the past? To produce Social Indicators 1973 most of the resources were borrowed from other agencies, rather than supported directly. Similarly, if Social Indicators 1976 is to appear, support for that volume will have to be obtained from outside of the Statistical Policy Division of OMB. Since a social audit crosses the interests of all Federal government departments, one has to ask what can be done to get direct support, sufficient in level to permit more timely and more useful volumes?

CONCLUSION

As John Dewey once pointed out, the ideal to be sought, is not a planned society but a continuously planning society. In such a society the various groups and individuals undergo a continual process of adjustment and readjustment, and without this continual adjusting and readjusting, the balance that produces integration and equilibrium will not be achieved. In a continually planning society there is a definite role for the production of summary information measures, or indicators, which either (a) represent qualitative social change, (b) measure public policies and programs for evaluative purposes, or (c) provide data for judging the effectiveness of the political process. Social Indicators 1973, unfortunately, attempts none of these. It thus contributes to a major problem facing the indicator movement (and indeed the whole of social sciences) in their attempts to impact on policymaking processes: the potential is well indicated, but delivery on the promise is lagging. The integration of indicators and policymaking must go forth, although neither can, nor should, dictate to the other its total content or rationale for existence. Policymaking is an art requiring "the compromise of conflicting claims or rival parties and groups in the interest of the total welfare" (24) and must not be totally dependant on any given set of quantitative information. On the other hand, indicators and their supporting data system have more uses than mere support of the efforts of a given set of policymakers at a given point in time.

In closing, I must agree with Dan Tunstall's comments in his June 1974 paper, that there is a need, not limited to OMB or the government, to inform Americans about the State of the Nation--not in terms of indicators selected to support administration policies, past, present, and future, but in terms of "social reporting that is as practical as the early censuses and as relevant to meeting national needs as economic reporting became after the depression." (25) The production of social information, its dissemination, and its use by individual citizens, by policymakers, and by the media which bridge the two groups should be an important part of any reshaping of the federal or any data system. If this volume can help trigger such considerations for expansion of the social information system and increased use and usefulness of its content, then Social Indicators 1973 will have more than justified its publication.

Footnotes

1. Chinitz, Benjamin, "The Management of Federal Expenditures for Research on Social Problems," in The Use and Abuse of Social Science, I.L. Horowitz (ed.), transactionbooks, New Brunswick, New Jersey, 1971, p. 80.

2. Full Opportunity Act, Hearings, July 7, 8, 10, 18; December 18, 1969; and March 13, 1970; p. 146.

3. Tunstall, Daniel B., "Social Indicators and Social Reporting," presented at joint American Marketing Association and British Marketing Resaerch Society Conference on Changing Values and Social Trends--How Do Organizations React, Oxford, England, June 6, 1974.

4. Bauer, Raymond A., "Detection and Anticipation of Impact: The Nature of the Task," Chapter 1 in R.A. Bauer (ed.), Social Indicators,The MIT Press, Cambridge, 1966, p. 1.

5. Biderman, Albert D., "Social Indicators and Goals," Chapter 2 in R.A. Bauer (ed.), Social Indicators, The MIT Press, Cambridge, 1966, p. 69.

6. Sheldon, Eleanor Bernert and Wilbert E. Moore (eds.), Indicators of Social Change: Concepts and Measurements. The Russell Sage Foundation, New York, 1968, p. 4.

7. U.S. Department of Health, Education, and Welfare, Towards A Social Report, Government Printing Office, Washington, D.C., 1969. p. 97.

8. Campbell, Angus and Philip E. Converse (eds.), The Human Meaning of Social Change, Russell Sage Foundation, New York, 1972, p. 2-3.

9. Executive Office of the President, Office of Management and Budget, Social Indicators 1973. Government Printing Office, Washington, D.C., 1974, p. xiii.

10. Attributed to Albert Biderman in Otis Dudley Duncan "Social Indicators 1973: Report on a Conference," in Roxann A. Van Dusen (ed.), Social Indicators 1973: A Review Symposium. SSRC Center for Coordination of Research on Social Indicators, Washington, D.C., 1974.

11. Ramsöy, Natalie Rogoff, "Social Indicators in the United States and Europe: Comments on Five Country Reports," in Roxann A. Van Dusen (ed.), Social Indicators 1973: A Review Symposium. SSRC Center for Coordination of Research on Social Indicators, Washington, D.C., 1974.

12. Springer, Michael, "Whose Indicators of What: Some Notes on Decision Making and Information." Working Paper 1206-12. The Urban Institute, Washington, D.C., June 1973.

13. Towards a Social Report, op cite, p. 4.

14. Gans, Herbert J., "Social Science for Public Policy," in I.L. Horowitz (ed.) transactionbooks, New Brunswick, New Jersey, 1971, p. 13.

15. Lord Rothschild, "A Report by Lord Rothschild." In The Framework for Government Research and Development: Memorandum by the Government. Command Paper 4814. Her Majesty's Stationery Office, 1971.

16. Coleman, James S., "Policy Research in the Social Sciences," General Learning Press, Morristown, New Jersey, 1972, p. 14.

17. Towards a Social Report, op. cite, p. 21.

18. Senior, Nassau Williams, An Outline of the Science of Political Economy, 1836.

19. Vance, Rupert B., All These People, University of North Carolina Press, Chapel Hill, 1945, p. 468-9.

20. Parke, Robert and Eleanor Bernert Sheldon, "Social Statistics for Public Policy," 1973 Social Statistics Section Proceedings, American Statistical Association, Washington, D.C., 1974.

21. From a summary of the study by Nathan Caplan, Director, Center for Research on the Utilization of Scientific Knowledge, University of Michigan, printed in the Institute for Social Research Newsletter, Spring 1974.

22. Op cite, p. 7.

23. Gordon, Kermit, "Some Conjectures on Policy Problems of the 1970's." American Economic Review, May 1974, pp. 125-128.

24. Vance, Rupert B., op cite, p. 467.

25. Tunstall, Daniel B., op. cite.

SOCIAL SURVEY 1974 - A NORWEGIAN COMPANION TO SOCIAL INDICATORS 1973

Petter Jakob Bjerve, Central Bureau of Statistics of Norway

Introduction

During the last few years a whole family of official and semi-official social statistical reports has been published. Although differing in detail and format these reports have important elements in common. They focus on groups of individuals or households; their main objective is to provide data on factors commonly assumed relevant to the evaluation of human welfare and they stress synthesis of scattered, previously published statistics. Special efforts have been made to make the reports intelligible to a large number of readers, both among policy-makers, contributors in mass communication and the politically and socially interested public. The communication of existing data rather than the development of new or improved social statistics has thus been the primary focus of these publications.

The veteran among these reports is, of course, the Social Trends prepared by the Central Statistical Office of the United Kingdom. The fifth issue of this publication will be published this year. Another prominent member of the family is the French Données Sociales - the first edition of which was published by the Institut National de la Statistique et des Études Économiques in 1973. A second edition is to be issued this year. Furthermore, the German Government has published Gesellschaftliche Daten 1973, and the Japanese Government a White Paper on National Life 1973. A Swedish and a Norwegian report will be published later this year by the Central Bureau of Statistics of Sweden and Norway, respectively.

This list of publications may not be exhaustive but indicates that the US contribution - Social Indicators 1973 - has many relatives and that the popularity of these reports is quite widespread. The purpose of my presentation is to inform you about the Norwegian report, whose English title will be Social Survey 1974, and to compare it with Social Indicators 1973 so as to provide a basis for evaluating the relative merits and limitations of the two publications. The comparison will be divided into three sections: Criteria for Selection of Statistics, Content, and Presentation. In conclusion, I shall make some observations on perspectives for future work in this field.

Criteria for selection of statistics

In Social Indicators 1973 eight major social areas are examined and within each of these several so-called social concerns have been identified. "The concerns have been defined and selected to reveal the general status of the entire population; to depict conditions that are, or are likely to be, dealt with by national policies; and to encompass many of the important issues facing the Nation. - The concerns thus embody widely held basic social objectives: .... For each of the identified social concerns, one or more indicators - statistical measures of important aspects of the concerns - have been identified. .... The choice of indicators is based upon two main criteria: That the indicators measure individual and family (rather than institutional or governmental) well-being and that they measure end products of, rather than inputs into, social systems." (Social Indicators 1973, p. XIII.) In short, this approach to the selection of statistics may be characterized as problem-oriented, at least as contrasted with the Norwegian approach.

In Norway there was also felt a need for principles to guide the selection of statistics for inclusion in our publication. We started with the same objective as the authors of the other reports listed, viz., to assemble in one publication the most important statistics available on individuals and households with the primary aim of illuminating conditions of living and the social structure. We tried to select data which large user groups would find convenient assembled in one volume, hoping thereby to achieve a better marketing of these statistics. However, how does one decide on the importance of individual statistics and guess the preferences of users? To do this we needed a conceptual framework. This framework can briefly be described as follows:

At the beginning of a time period an individual or a household is in some kind of an initial welfare situation, defined by personal characteristics such as age, health, educational level, housing situation, type of work, family situation, income and wealth. In addition, this initial situation of the individual or household is defined by certain institutional and environmental characteristics, such as the existing social security system, tax system, and political system. Determined partly by these characteristics and partly by the activities of governmental institutions, enterprises and other individuals, each individual engages in various activities during the period such as earning, learning, consumption, political work and perhaps even criminal activities. These activities yield certain results and create a new situation at the end of the period - likewise influenced by governmental institutions, enterprises and other individuals. The statistics included in a social survey should illuminate the various elements in this picture. Thus, the approach for our selection of statistics is rather comprehensive, and the amount of data from which the selection had to take place is correspondingly large. Accordingly, the selection of statistics for the Norwegian publication is not restricted to direct indicators of welfare, but aims also at providing data for the illumination of relationships and interrelationships, dependence and interdependence, of such social phenomena which strongly influence welfare.

Our main criterion for choice of data within this frame of thought was their relevance as indicators of human welfare in the wide sense. Of course, this criterion offers no clear guide. In many fields we do not even have a consensus on what is to be considered a desirable development in terms of human welfare. However, we tried to seek advice from representatives of a number of potential user groups, or to antecipate user preferences. We also sought guidance in corresponding foreign publications available at the time. Thus, within the general frame of reference just described, the method of choice was rather pragmatic and, of course, dependent upon the availability of data.

Similar to its American companion, our Social Survey 1974 emphasizes end products rather than inputs. However, within our conceptual framework the end product from one activity is often the input into another - both activities being relevant from a welfare point of view. Recognizing the sad fact that the end product of governmental activities is extremely hard and in some cases impossible to measure, we were not as reluctant as our American colleagues to use input data as substitutes for output data.

Similar to our American colleagues, we decided to restrict our choice of data almost entirely to statistics describing objective conditions - data on acts and facts. This is not only due to the fact that information on people's attitudes and opinions is not as abundant in Norway as in the US, but also because we feel that such data are extremely hard to interprete from a welfare point of view and that it in many cases would require a longer and more careful presentation than possible in a publication of this kind.

Finally, I like to mention that a major requirement for inclusion of a statistical table or diagram in Social Survey 1974 was that national averages were available and that the distribution of characteristics for the total population could be presented. In other words, data covering only certain groups were, as a rule, excluded. Also in this respect the criterion for choice seems to have been about the same for the two publications under consideration.

It is a matter of judgement which approach is "best" for selection of statistics to be included in a publication of this kind. We found that for the users of such a publication in a country like Norway, with a mixed economy and a relatively highly developed public welfare system, the comprehensive approach chosen would most likely be more appropriate than a problem approach. This does not imply, of course, that the Norwegian society is without problems, or that these problems are not reflected in the Social Survey 1974. Many of the problems describable by statistics are covered by our publication, but are presented as a part of a total which hopefully provides a broad picture of the Norwegian society - in welfare terms. The reactions of the public to our publication may give us some basis for evaluating the degree to which this hope has been realized.

Content

Our Social Survey 1974 contains about twice as many tables as Social Indicators 1973. This is partly due to the fact that the Norwegian publication includes a chapter on social services with as many as 40 tables, whereas its American companion lacks such a chapter. However, also other chapters of the Norwegian publication covering almost exactly the same social areas as Social Indicators 1973, contain considerably more information than those of the American publication. This is a consequence of the different approaches to the selection of data.

There are also differences in the kind of information included in corresponding chapters. Some topics are included in Social Survey 1974 but not in Social Indicators 1973, and vice versa. The topics included in the Norwegian publication but not in the American are to a large extent a consequence of a higher concern with the activities of government and institutions in general. I have already mentioned our chapter on social services containing data on those who receive benefits from the comprehensive social security schemes existing in Norway. Likewise, our inclusion of data on public consumption, membership in economic and civic organizations, and participation in civic activities, which are excluded in the American publication, reflects the different approaches in selecting statistics.

Some of the topics included in the American publication are not relevant for Norway. As examples I may mention the data on health insurance coverage, such insurance being compulsory for all Norwegian inhabitants, and the data on paid vacations which by law is fixed at four weeks a year for all Norwegian wage-earners. Some other data in the American publication are simply not available in Norway, e.g., the data on relationship between victim and offender, on achievement on education tests, and on daily use of time - which we got too late for inclusion. Finally, the different approaches in selecting statistics also appear to have entailed inclusion in the American publication of some data which are excluded in the Norwegian. As examples I can mention the statistics on job satisfaction and statistics on satisfaction with and assessment of neighbourhood.

In spite of these differences the main impression one receives from a comparison of the content of the two publications is that the topics included are basically the same. The differences are in detail and emphasis and in the total amount of information, rather than in the relative allocation of information on social areas.

Presentation

As statisticians we all recognize that statistics in its numerical reflection may make even the most fascinating subject boring to the uninitiated. We also know how easy it is to misuse or even lie with statistics, inadvertedly or by design. This poses a major communication problem to anyone preparing the kind of report which we are discussing - intended as it is for a wide distribution among non-experts. It seems that the authors of all the reports referred to have been aware of this problem and have tried to tackle it. The Norwegian method for solving the problem differs somewhat from the American and it may be worthwhile to discuss the differences in method of presentation.

The organization of the two publications has three major differences: the sequence of chapters; the number of sections within chapters; and the sequence of the verbal, graphical, and numerical presentation.

The Social Survey 1974 is introduced by a chapter on population which is designed to provide a background for the subsequent chapters, hopefully facilitating the study of relationships between the social areas represented by these chapters. Furthermore, the chapter on crimes and criminals is placed towards the end instead of in the front, next to the health chapter, as it is in Social Indicators 1973. The chapters on employment and income are placed next to each other in both publications, but in the opposite sequence. These variances seem to follow from the differences in approach. The same applies to the organization within chapters. All chapters are divided into more sections in the Norwegian publication than in its American companion, reflecting the fact that sections of the former do not refer to separate social concerns but rather to themes.

The authors of Social Indicators 1973 have chosen to present the charts and the tables in separate parts of each chapter, with "Technical Notes" in between. A few additional comments are made at the beginning of chapters and sections. This means that to find information presented on a particular phenomenon you may have to look at four different pages; e.g., on life expectancy at the age of 30 and 50 information is presented on pages 2, 3, 23, and 28. In the Norwegian publication, efforts have been made to locate both the verbal, graphical, and numerical presentation of a particular point as closely together as possible.

Readers interested in further details on organization of the Social Survey 1974, as compared with Social Indicators 1973, are referred to the Appendix on: Headings of Chapters and Sections in Social Survey 1974 and Social Indicators 1973.

The layout of Social Indicators 1973 is admirable. I am sad to say that we have no hope of reaching anywhere near to the high level of the American achievement in this respect. Lack of resources and staff have also prevented us from utilizing graphical presentation to the same extent as in Social Indicators 1973. Naturally, the authors of both publications have recognized the need for supplementing the tables and diagrams by verbal presentation. However, in this respect there is a considerable difference which may deserve a more detailed description.

In our attempts to promote the use of statistics in Norway, we have repeatedly found that most people have difficulties in extracting information from a statistical table. Graphical presentations may help, but as a rule words seem to be more easily understood. Surprisingly many of those who take interest in statistics, are unable to utilize the data without being led more or less by the hand through the table and explained in words what the statistics tell or do not tell. Therefore, aiming at making the Social Survey 1974 useful for a wide circle of readers, we deemed it necessary to supplement the tables with some text, restricting ourselves to pointing out some of the most interesting numerical information to be found in the tables and diagrams and issuing warnings against tempting and not quite obvious pitfalls and misinterpretations. This verbal presentation is made as close as possible to the relevant tables and diagrams. In addition, we have smuggled into this presentation some important methodological information and institutional information which in the American publication is presented in the "Technical Notes". Finally, we have attempted to make references and cross-references to other tables, inter alia, to point out relationships between data to be found in different tables and chapters. Nevertheless, there is no doubt that the Social Survey 1974 represents a statistical publication with supplementary comments rather than a verbal presentation with supplementary statistics.

The approach described above implies that there are no separate sections with "Technical Notes" in the Norwegian publication. This may be a disadvantage for advanced users. However, the Social Survey 1974 is not specially tailored for them, but rather for a much larger group of far less sophisticated users. Advanced users can utilize the many references that have been made to special publications which contain substantially more data as well as extensive and detailed technical notes.

To indicate the difference more concretely, I shall quote in English a few paragraphs intentionally selected from some chapters of the Norwegian publication.

The chapter on population is introduced by the following pedagogic remark: "Changes in the number of births - for example - determine variations in the demand for maternal care and the payment of family allowances. Changes in the age structure of the population have consequences for the growth of the economically active population, the demand for education, and the need for homes for the aged. This chapter thus serves as a background for the rest of the publication."

A combined numerical and methodological commentary is made in connection to a table and a disgram in the section on households and families: "The average number of children per married couple has declined for a long period of time. The married couples in 1920 had on the average 4 children, while those of 1960 had 2.3 children on the average. These figures are influenced by the duration of the marriages as the number of children tend to increase with the duration. However, from 1920 to 1950 the average number of children per married couple decreased for all classes of duration. From 1950 to 1960 the number declined for all marriages of more than 8-9 years' duration, but increased slightly on the average for marriages of a shorter duration."

In the section giving a general survey of the health situation, we combine information on health with a cross-reference to related data: "There is not much information available on the number of people who have permanent or prolonged illness, e.g., are blind, deaf or disabled for other reasons. According to the Health Survey 1968 39 per cent of those interviewed had a permanent illness or injury, but this includes even slight cases which caused no reduction in normal activity. In Chapter 5 on Social Services there is information on persons receiving disability pensions, which gives indication of the incidence of permanent cases of illness among persons below the age of 70."

Presenting numerical information on the level of education, both methodological and institutional information are added: "In the Population Censuses of 1950, 1960, and 1970, persons were asked to give their highest general education and all vocational educations. There was only small changes from 1950 to 1960 in the proportion of the population which had general education beyond primary school. This was partly due to the smallness of the relevant age groups. That there was no change from 1950 to 1960 in the proportion having vocational education, is mostly due to a change in definitions. In 1950 vocational education of 5 months or more was to be included, whereas in 1960 the lowest time limit was 10 months. - The returns from the 1970 Census show a substantial increase in the amount of general education beyond primary school (7 years) in the population during the period 1960-1970. This must be seen in conjunction with the large young age groups of this period, and the introduction of 9-years compulsory education in Norway."

The only data presented on attitudes are described in the following manner - in a section on housing environment: "About 70 per cent of all households (in the Housing Survey 1967) thought that the available playground was a safe one for their children. About the same proportion thought it was of good quality. The street or the sidewalk was thought to be unsafe by about 80 per cent of the households where the children only had such a playground. .... The inner courtyard as a playground was thought to be safe but of bad quality. Other types of playgrounds were mostly thought to be safe and of good quality."

While presenting numerical information on changes in the income distribution, we point out the potentially misleading influence of institutional factors: "Over the period 1950-1971 there has been a marked reduction in the relative differences between the average incomes in urban and rural municipalities. Some of this equalization may be due to the fact that some rural municipalities have become urban without being reclassified or that some rural municipalities have been merged with urban ones. This is clearly illustrated by the figures for the Hordaland and Bergen counties 1970 and 1971. Four municipalities which were formerly classified as part of Hordaland, were in 1971 classified as part of Bergen. This caused the average income of Hordaland relative to the rest of the country to decline substantially from 1970 to 1971, and that of Bergen to rise."

In the chapter on social services we give a survey on general public pension and assistance schemes. From this I quote the following: "Disability assistance was introduced by law January 1, 1961, and now forms part of the National Insurance System. There is a graduated disability pension scheme for persons with 50 per cent disability degree or more." From the numerical comments in the same chapter I quote: "By New Year 1967 4.1 per cent of the population in the age groups 18 years to 69 years received disability pension. The proportion was the same for women and for men. In 1972 this percentage had increased to 6.3."

These quotations may suffice to demonstrate how verbal presentation of numerical information has been combined with comments of a pedagogic, methodological or institutional nature. In a number of cases the verbal comments merely repeat some of the most interesting information provided by a table. User reactions will show whether this method of presentation has promoted careful interpretation and extensive utilization of the data presented.

Of course, we realized that commenting on data from a number of fields is a rather risky venture. In the first place, however, we abstained completely from explanation of the phenomena pointed out in the text. Secondly, we tried to take advantage of available expertice both inside and outside the Central Bureau of Statistics. The publication was prepared by a small project group within the Central Bureau of Statistics of Norway, but this group worked in close co-operation with the various subject matter divisions of the Bureau. A first draft of the manuscript was discussed informally with the specialized government agencies concerned and also with other experts outside the Bureau. Valuable information and comments were received and as far as possible were taken into consideration when preparing the final manuscript.

Permit me at this point to engage in some sales promotion by informing you that all tables and diagrams (including notes) as well as the lists of tables and diagrams are described in English as well as in Norwegian. The methodological and technical parts of the main text are summarized in English. Moreover, I may mention that the planning of a new edition has already started. It will probably be published late 1976 and will contain the results from a number of important statistical investigations which now are being carried out. We have some reason to hope that for anyone being interested in the Norwegian society whether as such, as part of Scandinavia, or Europe or the World, the Social Survey 1974 and subsequent editions of this publication will provide a treasury of information accessible for English speaking readers.

Perspectives for the future

As pointed out in the introduction, work on Social Survey 1974 was restricted to presentation of already available data. No attempts were made to develop new or improved social statistics. However, we hoped that experience from the work on this publication would provide feedback for developmental work. Having completed the work on the first edition, we see that this hope was not quite unrealistic.

Another hope was not fulfilled. At the outset we thought that one objective of this publication should be to describe relationships between social data in different areas beyond cross-references between chapters and tables and the obvious adjustments made for changes in size and age distribution of the population. However, this idea was soon abandoned as being too ambitious at this stage. There is no chapter aiming at combining data from all or several social areas on the basis of standard definitions, classifications, and definitional relationships similar to those underlying the national economic accounts. The explanation is, of course, obvious. As yet, we have not been able to develop a logical system providing a similar synthesis of social and demographic statistics as represented by the national economic accounts for economic statistics.

Most of you presumably know that quite extensive work is going on internationally under the guidance of the United Nations statistical agencies developing a system of integrated social and demographic statistics which, it is hoped, will provide a much better basis for analysis of relationships and interrelationships than the present system. The central statistical agencies of a number of countries, including Norway, intend to make considerable efforts to fit social and demographic data into this system as soon as it is sufficiently developed.

This developmental work will not be sufficiently advanced to apply fully already in our Social Survey 1976. Nevertheless, I hope that this publication can be introduced by a new chapter containing some tables covering all or most of the social areas for which Social Survey 1974 provides data. We have already experimented with some tables where a standard classification by age provides a link between the various statistical areas. In these tables, the Norwegian population is visualized being cross-classified by age and by characteristics such as health, educational attainment, housing standard, occupation, income, and perhaps even criminality. However, there are numerous methodological difficulties and limitations involved in the construction of such tables, particularly if one aims at making the classification by welfare characteristics comprehensive.

In my opinion the most urgent progress to be made in the next few editions of our social report is the presentation of an integrating introductory chapter. If we succeed, the Social Survey 1974 will represent a first step toward presentation of an integrated system of social and demographic statistics, in a similar manner as the data presented in Social Indicators 1973 are characterized as "a first step toward development of a more extensive social indicator system".

ACKNOWLEDGEMENT

I am indebted to Mr. Eivind Hoffmann, editor of the Social Survey 1974, for much help in the elaboration of this paper and to Mrs. Susan Lingsom for improvement of language.

HEADINGS OF CHAPTERS AND SECTIONS IN SOCIAL SURVEY 1974 AND SOCIAL INDICATORS 1973

(In parenthesis ( ) are added some words to further indicate the content of the sections.)

SOCIAL SURVEY 1974

1. POPULATION

Size and composition of population
Households and families
Vital statistics
Births
Deaths
Migrations
Marriages
Adoptions

2. HEALTH

Health (episodes of illness, reduction in activity, physical mobility, in-patients)
Expectation of life
Diseases
Injuries
Cause of death
Mortality by marital status and occupation
Maternal and infant mortality
Contact with physicians and health institutions
Health personnel, hospitals, pharmaceuticals
Public dental care
Abortions

3. EDUCATION

Level of education in the population
Persons educationally active
  Primary schools
  Special schools
  "Folk high schools"
  Secondary schools, upper stage
  Vocational schools and colleges
  Universities and colleges
  Adult education and popular education

4. HOUSING

Stock of dwellings and building of new dwellings
Size of dwellings
Equipment in the dwellings
Environment of dwellings
Tenure status to dwelling
Building costs, financing

5. EMPLOYMENT AND EMPLOYMENT CONDITIONS

Employment (labour force, participation)
Economically active foreigners
Unemployment
Working hours
Membership in Trade unions and Employers' associations
Labour conflicts, curtailment of operations, government measures to promote employment

SOCIAL INDICATORS 1973

8. POPULATION

Population Growth
Population Distribution

1. HEALTH

Long Life (life expectancy, death rates/causes, infant mortality)
Disability
  Long-Term Disability - Institutional (patients, admissions)
  Long-Term Disability - Non-Institutional (limited activity - chronic conditions)
  Short-Term Disability
Access to Medical Care (health, insurance, expenditures)

3. EDUCATION

Basic Skills - Attainment (level of education, enrollment, graduation)
Basic Skills - Achievement
Higher and Continuing Education (enrollment, degrees earned, adult education)

6. HOUSING

Housing Quality
  The Housing Unit (standard)
  Living Space (crowding)
  The Neighborhood (satisfaction, assessment)

4. EMPLOYMENT

Employment Opportunities (unemployment, labor force, participation)
Quality of Employment Life
  Job Satisfaction
  Working Conditions (earnings, hours worked, transportation to work, paid vacations, work injuries)

SOCIAL SURVEY 1974

6. INCOME AND CONSUMPTION

Wages
Income of persons (level and distribution)
Income of households (level and distribution)
Personal property
Private consumption
Prices
The Public Sector

7. SOCIAL SERVICES

Children and parents
Illness, rehabilitation and disability
Assistance to old people
Social support
Public expenditure on social security

8. CRIMES AND CRIMINALS

Crimes investigated by the police
Results of investigations
Victims of crimes
Self-reported crimes
Persons charged for crimes
Recidivism among charged persons
Sanctions
Recidivism among offenders
Misdemeanours of drunkenness, Traffic offences

9. PARTICIPATION IN POLITICAL, SOCIAL AND CULTURAL ACTIVITIES

Participation in elections
Correcting the list of candidates
Votes by party
Unpaid municipal offices
Membership in associations and organizations
Meetings and performances
Reading and musical activities. Use of radio and television
Amount and division of work in the household
Holidays

SOCIAL INDICATORS 1973

5. INCOME

Level of Income (families)
Distribution of Income (families)
Expenditure of Income (consumption, wealth)
The Low-Income Population

2. PUBLIC SAFETY

Safety of Life and Property from Crime
- Violent Crime (crimes, victims, offenders)
- Property Crime (crimes, victims)
- Freedom from Fear of Crime

7. LEISURE AND RECREATION

Leisure Time (daily use of time)
- Outdoor Recreation
- Television Viewing

USE OF LOSS FUNCTIONS TO DETERMINE SAMPLE SIZE IN THE SOCIAL SECURITY ADMINISTRATION

Thomas B. Jabine and Rudolph E. Schwartz, Social Security Administration

1. INTRODUCTION

In this paper, we illustrate the determination of optimum sample size by minimization of an appropriate loss function. The theory is straightforward and well known [3, 8]; however, opportunities to apply it do not appear frequently and examples in textbooks tend to be contrived. Surprisingly, the authors have recently found in the Social Security Administration (SSA) several applications of sampling which lend themselves to this technique. After a brief description of the basic idea and the general conditions required for its application, we describe in detail two applications in the newly enacted Supplemental Security Income (SSI) program. The paper concludes with a general discussion of some possible extensions of the technique, as well as some of the problems and limitations associated with it.

The traditional textbook approach to the determination of sample size starts with the specification of the desired variance of a sample estimate of some population parameter, expresses this variance as a function of the sample size and of other population parameters assumed to be known, i.e.

$$\sigma^2_{\hat{x}} = f(n, \theta_1, \ldots \theta_r) \quad (1)$$

and solves this equation to determine the required sample size.

Since there are normally a large number of alternate sampling and estimation procedures available, we may establish relationships like (1) for several possible sample designs, solve each for n, and choose the design which gives the smallest value of n. Often, the variable cost attached to each sample unit differs for different designs. When this is the case, we will choose the design with the lowest variable cost.

For some applications, the total budget is fixed. In that case, assuming that fixed costs and variable unit costs are known for each design, we will calculate the size of the sample that we can afford, calculate the variance from (1), and choose the design that minimizes the variance.

This procedure leaves unanswered, however, the fundamental question of how to establish the appropriate variance requirements or budget for a particular survey or other investigation carried out by sampling. Commonly, the sampling technician for a project proposes certain reliability requirements, based on his experience with similar applications, and if the corresponding costs look reasonable to the project manager, he accepts these specifications. Alternatively, the project manager may specify the budget and the sampling technician will try to maximize reliability within that cost.

People trained in decision theory, cost-benefit analysis and other tools of management find the whole process to be rather arbitrary; and, as a consequence, administrators of statistical programs often find themselves under pressure to develop more objective bases for allocating their resources among various data collection programs. Unfortunately for this end, no one has found, nor are they likely to find, any satisfactory way of quantifying the benefits of general-purpose statistical programs. We have no solutions to this dilemma. However, there are situations where a more objective approach can be used. These situations arise when

1. The purpose of the sample is to obtain estimates of one or more population parameters which will be used, according to some specified rules, to determine an amount of money to be disbursed or collected.

2. The cost of collecting and processing the necessary data for the sample units is known or can be estimated.

3. The loss resulting from estimates which differ from the population parameters being estimated can be defined in such a way that its expected value is not zero.

2. THE BASIC METHOD

Let $C = f(n)$ be the cost of collecting and processing the sample data. We will call this the cost function.

n = number of units in the sample.

Let $\hat{A} = g(\hat{Y}_1 \hat{Y}_2, \ldots, \hat{Y}_s)$ be the amount to be disbursed ($\hat{A} > 0$) or collected ($\hat{A} < 0$) by an entity, with $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_s$ representing sample estimates of population parameters which enter into the determination of this amount, and

Let $L = h(\hat{A}-A)$ be a function representing the loss (or gain) to the entity when $\hat{A} \neq A$, where $A = g(Y_1, Y_2, \ldots, Y_s)$. We will call this the payment error function.

Then our loss function is

$$\theta = E(L) + C$$

If there were no constraints on the sample size n, we could determine its optimum i.e., the value that minimizes the loss function, by differentiating the loss function with respect to n and solving for n in

$$\frac{\partial\theta}{\partial n} = 0$$

However, in this instance, we will introduce the restriction that n be an integer in the interval

$$1 \leq n \leq N$$

where N = number of units in the population from which the sample is to be selected

so that in some instances the value of n which minimizes the loss function will have to be determined by other methods.

One obvious form of the payment error function would be

$$L_1 = \hat{A} - A$$

However, $E(L_1) = 0$ if $\hat{A}$ is an unbiased estimate of A, and this would lead to n = 1, which is not very helpful.

There are at least two ways to resolve this difficulty. One would be to define a function L_2 which is always positive when $\hat{A} \neq A$, for example, a function involving $|A-A|$ or $(\hat{A}-A)^2$. This would reflect a philosophy which says there is always some economic loss when we err in estimating the amount to be disbursed or collected, regardless of the direction of the error. This seems like a reasonable position, but it is very difficult to quantify. For example, suppose the issue to be resolved is how much of the cost of a particular program is to be borne by the Federal and State governments, respectively. If the Federal government pays more than its share, and the State governments pay less, or vice versa, how do we evaluate the overall loss to the economy? This question is almost as difficult to answer as one requiring the quantification of the consequences of errors in a general purpose statistical survey.

A second approach is to look at the situation from the point of view of one of the two parties involved, and to take the position that the principal concern is with losses resulting from errors of estimation which have unfavorable consequences for that party. For example, using the illustration from the previous paragraph, suppose we represent the Federal government, and we want to minimize losses from estimates which result in overpayments to the States. We may then define a payment error function

$$L_3 = \hat{A}-A \text{ when } \hat{A} \geq A$$
$$= 0 \quad \text{when } \hat{A} < A$$

This function has a positive expected value, and in the important case where we can assume $\hat{A}$ to be normally distributed, we have

$$E(L_3) = \sqrt{\frac{1}{2\pi}}\ \sigma_{\hat{A}}$$

where $\sigma^2_{\hat{A}}$ = population variance of the estimate $\hat{A}$

This is the kind of payment error function we have used to determine optimum sample size in the illustrations which follow.

3. ILLUSTRATIONS

We now present descriptions of two applications of this technique in the Supplemental Security Income (SSI) program [2]. In the first one, which we call estimation of the adjusted payment level (APL), the cost per unit of obtaining and processing the data was relatively low, and the amounts of potential overpayments were substantial. Use of the loss function technique led to the conclusion that there should be no sampling, i.e., that the calculation of adjusted payment level should be based on all eligible cases.

In the second application, which we call estimation of Federal fiscal liability (FFL), the cost per unit of obtaining data was quite large. Here, application of the method led to recommendations for sample sizes considerably smaller than had been proposed on other grounds.

Application No.1 - Estimation of Adjusted Payment Level for the Supplemental Security Income Program

The Supplemental Security Income Program (SSI), in effect since January 1, 1974, provides assistance to people with low incomes who are aged (65 and over), disabled or blind. Eligible individuals and couples receive basic Federal payments, the amounts depending on their living arrangements and on how much income, if any, they receive from other sources. The current basic payment is $146 per month for an individual living alone with no income and $219 for a couple.

SSI also provides for supplementary payments by the States. In some cases, these payments are mandatory, in order to assure that persons who had been receiving benefits from the prior Federal-State assistance programs will continue to receive benefits at essentially the same level. In other cases, the supplementary payments by the States are optional. In either case, the State may elect to have its supplementary payments administered by the Federal government.

A provision of the SSI legislation known as "hold harmless" assures that no State electing Federal administration of its supplemental payments will have to spend more on this program than its share of the total expenditures for public assistance to recipients in these categories in calendar 1972. All costs in excess of this amount will be borne by the Federal government.

However, in order to limit Federal liability under the hold harmless provision, it was further specified that State supplementary payments would be protected only to the extent that these payments did not exceed, on the average, an amount called the adjusted payment level (APL). The APL is defined as the average money payment by the State, in January 1972, to individuals who had no income and were living alone.

In general, the States did not have available tabulations or tape files from which they could readily calculate the APL, so it was necessary to obtain the data for eligible individuals (those living alone, with no income, in January 1972) from case folders. For States with small numbers of recipients this was not difficult, but for States with large workloads it appeared that locating the case folders for cases qualifying for inclusion in the APL calculation, transcribing the data and making the calculation would be a substantial undertaking.

This situation led to consideration of the possible use of sampling. Initially several States asked for and received permission to estimate APL from a sample of cases, with the requirement that the estimate be made with a standard deviation of not more than \$2.50 or a coefficient of variation no greater than 1.5 percent, if the latter condition permitted a larger standard deviation.

Subsequently, however, concern developed about the possible effects of sampling error of the estimated APL on the size of the Federal liability under the hold harmless provision. It was at this point that the loss function approach was applied for the first time.

Without going into the details of the hold harmless calculation, we may assert that an appropriate payment error function of the type L_3 for this situation was given by:

$$L_3 = 12\, W\, (\hat{Y}-Y) \text{ for } \hat{Y} > Y$$
$$= 0 \text{ otherwise}$$

where Y = APL based on calculation including all eligible cases

$\hat{Y}$ = estimate of Y from a simple random sample without replacement

W = the program workload, i.e., the total number of persons currently receiving payments under the program.

The factor of 12 was used to convert the loss to an annual basis, since the APL is an average monthly payment. We arbitrarily chose one year to represent the loss, even though it can in theory go on for an indefinite period once the APL is established. As will be seen, extending this period would not have changed our conclusion.

Assuming the estimates $\hat{Y}$ from repeated samples to follow a normal distribution, we have

$$E(L_3) = 12\, W\sqrt{\frac{1}{2\pi}}\, \sigma_{\hat{Y}}$$
$$= 12\, W\sqrt{\frac{1}{2\pi}}\left(\frac{N-n}{Nn}\right)^{1/2} \sigma_Y$$

where N = number of persons eligible for inclusion in the APL calculation

n = number of eligible persons in the sample

σ_Y = population standard deviation of the January 1972 payment level for eligible persons.

In general, N<<W, since W represents the entire current program workload, whereas N is the number of individuals living alone, with no income, in January 1972.

Then we have for our loss function

$$\Theta = K\left(\frac{N-n}{Nn}\right)^{1/2} + cn$$

where c = unit cost per eligible case of locating the case folder and transcribing the data

and $K = 12\, W\sqrt{\frac{1}{2\pi}}\, \sigma_{\hat{Y}}$ is independent of n

Differentiating with respect to n, we have

$$\frac{\partial\Theta}{\partial n} = g(n) = -\frac{K}{2n^2}\left(\frac{N-n}{Nn}\right)^{-1/2} + c$$

Keeping in mind the restriction that n be an integer in the interval $1 \le n \le N$, analysis of g(n) shows

1. For $N < \frac{4}{3}\left(\frac{K}{c}\right)^{2/3}$

   g(n) is negative and Θ uniformly decreasing in the interval $1 \le n \le N$. Therefore, Θ is minimized by taking n = N, i.e., include all eligible cases in the sample.

2. For $N = \frac{4}{3}\left(\frac{K}{c}\right)^{2/3}$

   g(n) = 0 at $n = \frac{3}{4}N$, and is negative in the remainder of the interval, so $n = \frac{3}{4}N$ is an inflection point and the constrained minimum for Θ is at n = N.

3. For $N > \frac{4}{3}\left(\frac{K}{c}\right)^{2/3}$

   g(n) has its maximum at $n = \frac{3}{4}N$. The equation g(n) = 0 has 2 real roots in the interval $0 \le n \le N$, one to the left of $n = \frac{3}{4}N$ and one to the right. The root to the left is a local minimum and the one to the right is a local maximum. In order to determine the constrained value of n which minimizes Θ we must calculate Θ for each of the 2 integer values of n surrounding the local minimum and for n = N and select from this triad the one which minimizes Θ.

In practice the approximate optimum value of n was easily determined by computing Θ for all values of n spaced at some reasonable sized interval, say 500, between 0 and N.

Chart 1 illustrates the behavior of the loss function Θ for changing values of K. In this illustration, we have used fixed values

N = 36,000 c = \$10

and allowed K to vary in the interval \$10 million to \$80 million. For these values of N and c, we have

$$N = \frac{4}{3}\left(\frac{K}{c}\right)^{2/3}$$

at K ≐ \$44.4 million.

For each of 6 values of K, values of the loss function Θ were calculated for varying sample sizes, n, starting with n = 1500 and continuing at intervals of 750 to n = N = 36,000.

For the first 3 values of K, all in excess of \$44.4 million, we observe that the loss function decreases monotonically, and the constrained minimum occurs at n = N. For K = \$35 million, there is a local minimum in the neighborhood of n = 18,750, but the constrained overall minimum continues to be at n = N.

For K = \$20 million and \$10 million, however, the overall constrained minima are the local minima in the neighborhoods of n = 11,250 and n = 6,750, respectively.

It is interesting to note that at K $\doteq$ \$34.15 million there occurs a threshold value of K for which it is indifferent whether we choose n = N or n $\doteq$ 18,000. For all smaller values of K, n_{opt} will be the local minimum, and this will decrease continuously with decreasing K.

It is also of interest to observe how n_{opt} behaves if we keep K and c constant and vary N. Up to the threshold point, we have $n_{opt} = N$, so that n_{opt} increases with N. Beyond this point, however, n_{opt} will be the local minimum to the left of $n = \frac{3}{4}N$, and we find that this decreases as N increases, with

$$\lim_{N \to \infty} n_{opt} = \left(\frac{K}{2c}\right)^{2/3}$$

Coming back to the APL application, there were 5 of the larger States which had initially estimated APL from samples. For these States, we had reasonably good estimates of W, the program workload and, from the earlier sample calculations, N, the number of persons eligible for inclusion in the APL calculation and σ_y^2, the variance of their January 1972 payments. We had only a rough idea of c, the unit cost, so values of Θ, the loss function, were calculated for values of c in the range \$1 to \$10.

These calculations showed that $n_{opt} = N$ for all States, categories (APL was estimated separately for aged and for disabled, including blind) and unit costs in the range considered. In other words, sampling should not be used to determine the APL. As a result, the APL's for these States estimated from samples were accepted only on a provisional basis, and arrangements were made for new calculations based on all eligible cases.

Application No.2 - Determination of Federal Fiscal Liability to States for Errors in the Administration of Supplementary Payments

Under the SSI program, 33 States and the District of Columbia have selected Federal administration of supplementary benefit payments. The Federal Government (SSA) determines eligibility and benefit amounts for both basic and supplementary payments, makes the payments and is reimbursed by the States for the amount of the supplementary payments not covered by the "hold harmless" provision. All administrative costs are borne by SSA.

Federal fiscal liability (FFL) is the liability which the Federal government, represented by SSA, has agreed to assume for excessive errors in making supplementary payments on behalf of the States. These errors are of two kinds

- Eligibility errors, which occur when an individual is incorrectly certified as eligible for supplementary payments.

  Overpayment errors, which occur when an individual has been correctly certified as eligible, but the amount of the supplementary payment has been set too high.

Errors may occur either because incorrect information was obtained about factors used to determine eligibility and benefit amounts, or because correct information was processed incorrectly.

Errors affecting supplementary payments are to be identified as part of a broad quality assurance program, in which a sample of cases will be reviewed for both substantive and procedural errors relating to the basic and/or supplementary payments. The reviews are extensive, comprising an examination of materials in the SSI files, interviews with recipients, and contacts with collateral sources of information, such as banks and insurance companies.

The unit cost of these reviews is high and the use of sampling is clearly indicated. For the 31 States with Federal administration of supplementary payments, where a primary objective of the sample reviews will be to determine the FFL, the cost-benefit approach has provided useful guidance in setting reasonable sample sizes.

The liability determination period for FFL has been set at 6 months. Total liability to a State during a 6-month period is estimated, using the results of the sample reviews, by

$$A = \frac{Y}{\bar{y}} [\bar{y}_e (p_e - t_e) a_e + \bar{z}_o (p_o - t_o) a_o]$$

where the terms are defined as follows:

Y = total State supplementary payments

$\bar{y}$ = mean supplementary payment in the sample

$\bar{y}_e$= mean supplementary payment to ineligible cases in the sample

p_e= proportion of ineligible cases in the sample

t_e= tolerance limit for eligibility errors

a_e= 1 if $p_e > t_e$
 = 0 otherwise

$\bar{z}_o$= mean overpayment of the supplementary payment to cases with overpayments in the sample

p_o= proportion of cases with overpayments in the sample

t_o= tolerance limit for overpayments

a_o= 1 if $p_o > t_o$
 = 0 otherwise

In this case our payment error function was simply

$$L_3 = \hat{A} - A \text{ for } \hat{A} > A$$
$$= 0 \text{ otherwise}$$

where A is analogous to $\hat{A}$, with population values substituted for sample estimates.

To simplify matters, we restricted our analysis to situations where the eligibility and overpayment error rates were substantially in excess of their respective tolerances ($t_e = .03$, $t_o = .05$), so that we could let $a_e = a_o = 1$. This was a conservative approach, since the potential loss from overestimating A is clearly greatest in this range.

With this restriction, we could assume estimates of A from repeated samples to be normally distributed so that

$$E(L_3) = \sqrt{\frac{1}{2\pi}}\,\sigma_{\hat{A}}$$

By the method of propagation of variances [4, p.585], we obtained:

$$\sigma^2_{\hat{A}} = Y^2 \left[\sigma^2_{\bar{y}} \left(-\frac{b^2}{\bar{Y}}\right) + \sigma^2_{\bar{y}_e}\left(\frac{P_e - t_e}{\bar{Y}}\right)^2 + \sigma^2_{P_e}\left(\frac{\bar{Y}_e}{\bar{Y}}\right)^2 + \sigma^2_{\bar{z}}\left(\frac{P_o - t_o}{\bar{Y}}\right)^2 + \sigma^2_{P_o}\left(\frac{\bar{Z}_o}{\bar{Y}}\right)^2 + \text{covariance terms}\right]$$

where $b = \bar{Y}_e\,(P_e - t_e) + \bar{Z}_o\,(P_o - t_o)$

It seemed reasonable to neglect the covariance terms, since the variables involved are largely independent of each other. From our knowledge of the relevant population parameters, it appeared that $\sigma^2_{\hat{A}}$ would be dominated by the terms involving $\sigma^2_{P_e}$ and $\sigma^2_{P_o}$. Assuming the use of a simple random sample, with replacement, of n cases, and letting $\bar{Y}_e/\bar{Y} = .8$ and $\bar{Z}_o/\bar{Y} = .4$, we arrived at the approximation

$$\sigma_{\hat{A}} \doteq \frac{Y}{\sqrt{n}}\,[.64\ P_e(1-P_e) + .16\ P_o(1-P_o)]^{1/2}$$

The overall loss function was defined as

$$\Theta = E(L_3) + C$$
$$= Y\left[\frac{.64\ P_e[(1-P_e) + .16\ P_o(1-P_o)}{2\pi n}\right]^{1/2} + nc$$

where c = unit cost of a case review.

Differentiating Θ with respect to n, setting the result equal to 0 and solving for n, we found

$$n_{opt} = \left(\frac{Y}{2c}\right)^{2/3} \left[\frac{.64\ P_e(1-P_e) + .16 P_o(1-P_o)}{2\pi}\right]^{1/3}$$

Thus, at a given level of error, the optimum sample size is directly proportional to the total amount of supplementary payments, raised to the two-thirds power, and inversely proportional to the cost per sample case, raised to the two-thirds power.

Table 1 shows, for a unit cost c = $200, the optimum sample sizes for various combinations of Y, P_e and P_o. Since the average supplemental payments per recipient over a six-month period are on the order of $400, it can be seen that the optimum sample size is a small fraction, generally less than 1 percent, of the population of recipients. This is in strong contrast to the APL application where, with a much smaller unit cost, the indicated solution was to include all eligible cases in the calculation.

The above findings were applied directly to determine sample sizes for 18 States with mandatory supplementation only under Federal administration. A minimum sample size of 100 cases per State was established, because we felt it would be difficult to persuade all parties involved that it made sense to estimate FFL from a sample any smaller than that.

For the 13 States with mandatory and optional supplementation under Federal administration, somewhat different criteria were adopted, in order to be consistent with procedures currently in use to control the level of error by States in making payments, partly financed by Federal funds, under the program of aid to families with dependent children (AFDC). Nevertheless, the assigned sample sizes, with one or two exceptions, did not differ greatly from those indicated by use of the approach just described.

4. DISCUSSION

One can imagine many other applications and extensions of the basic method presented here. With respect to areas of application, any situation where data are needed to determine amounts of money to be transferred from one entity to another lends itself to this approach. In SSA, for example, when there are indications that a provider of services to Medicare enrollees may have overcharged for these services, there are provisions for reviewing a sample of cases to estimate the total extent of overcharging and hence the amount that the provider must return to SSA. Present procedures for determining sample sizes for these reviews are being reexamined using the technique described in this paper.

There would appear to be no special difficulty in using the method for sample designs other than simple random sampling. For example, in a situation calling for stratified sampling to estimate a single variable, we could proceed as follows:

(1) Determine optimum allocation of the sample for fixed n in the form

$$n_h = w_h\, n, \ \sum_h w_h = 1$$

where w_h is a function of various population parameters and costs for the h^{th} stratum and is independent of n[1/]

(2) Express the variance of $\hat{A}$ as a function of n and the stratum weights. Suppose, for example, that

$$\hat{A} = K\,\bar{x}$$

so that $\sigma_{\hat{A}} = K\,\sigma_{\bar{x}}$

We have, for stratified sampling without replacement

$$\sigma^2_{\bar{x}} = \sum_n \left(\frac{N_h}{N}\right)^2 \frac{N_h - n_h}{N_h} \frac{S^2_h}{n_h}$$

which becomes, when we substitute $w_h n$ for n_h

$$\sigma^2_{\bar{x}} = \frac{1}{n} \sum_h \left(\frac{N_h}{N}\right)^2 \frac{S^2_h}{w_h} - \sum_h \left(\frac{N_h}{N}\right)^2 \frac{S^2_h}{N_h} \quad .$$

hence $\sigma_{\hat{A}} = K \left(\frac{a}{n} - b\right)^{1/2}$

(3) Solve for the optimum n in the usual way. In the illustration, we would have

$$\Theta = K' \left(\frac{a}{n} - b\right)^{1/2} + n\, c$$

where $K' = K \left(\frac{1}{2\pi}\right)^{1/2}$

and $c = \sum_h w_h c_h$

and when we differentiate and set equal to 0, we find that n is the solution to a 4th degree equation.

(4) Allocate the sample of size n to the strata using the weights w_h. Since these weights were determined independently of n (within a specified range), we will have a solution which gives both the optimum sample size, n, for a stratified design and optimum allocation of the sample among the strata.

Similar solutions should be possible for designs involving cluster and multistage sampling.

Should we try to take into account the effects of nonsampling errors on the estimates of population parameters for use in payment or reimbursement formulas? This should offer no particular theoretical difficulty; however, in practical terms it would be much more difficult to develop a formal, predictable relationship between expected losses due to sampling and nonsampling error, and the resources expended to get the data. Nevertheless, for some applications nonsampling error may dominate total error, and every possible effort should be made to consider its effects.

What are the limitations of the cost-benefit technique for determining optimum sample size? Mostly they revolve around the difficulty of defining an explicit payment error function, i.e., the term that reflects the losses resulting from errors in estimating the relevant population parameters. The solution we have adopted is admittedly somewhat artificial, in that expected losses are based on errors in one direction only. It can be rationalized on the grounds that what we are trying to do is to protect the agency we represent against incurring losses which would expose it to the charge of failing to act in a prudent and responsible manner. Errors in other directions are considered as producing windfall "profits", and these should not enter into the determination of the optimum design. On the other hand, if we consider the interests of society as a whole, then clearly losses can result from errors in either direction. However, the losses to society will in most cases not be as large (assuming a value can be placed on them at all!) as the actual amounts of underpayment or overpayment.

It can be argued that the agency, even when given a fairly precise determination of the optimum sample size, may not wish to invest its resources in exactly that way. The administrator cannot consider this particular problem in isolation from all others confronting him. His problem is more likely to be one of considering relative opportunity costs of alternative uses of more or less fixed resources. Thus, it is possible that some of the resources that would be needed to take care of the optimum sample could be applied in another area where the payoff was greater. This kind of consideration might be built into the model explicitly by introducing utility functions, i.e.,

$$\Theta = u_1\,[E(L)] + u_2[C]$$

This would allow us to give different weights to real dollars which would have to be spent to collect and process the sample data, and expected dollars lost from errors in estimating the payment amount.

Another question to be considered is the extent to which the concepts involved in determining the optimum sample size by this method can be readily explained to those responsible for making decisions. In our experience, the idea of the expected loss, especially for a nonsymmetric function like L_3, has been somewhat difficult to explain. The use of tables showing values of the loss function and its components for various sample sizes and values of the relevant population parameters has been very helpful in providing a clearer view of the whole problem and the relationships of the variables. The general idea of applying a cost-benefit approach to determining sample size has been received enthusiastically, and we have been asked on various occasions why we can't apply the technique to general-purpose statistical surveys!

Despite the limitations we have described, we have found this method to be considerably more satisfactory than approaches used previously for similar types of problems. Even in cases where the solution is not precise because some of the parameters needed could not be estimated accurately, the results still provide useful guidelines to a general course of action.

FOOTNOTE

1/ If sampling without replacement, independence will hold only in the range of n for which $n_h < N_h$ for all strata.

REFERENCES

1. Blythe, R.H., "The Economics of Sample Size Applied to the Scaling of Sawlogs." Biometrics Bulletin, 1, (1945), 67-70.

2. Callison, James C., "Early Experience Under the Supplemental Security Income Program." Social Security Bulletin, 37 (June 1974), 3-11, 30.

3. Cochran, William G., Sampling Techniques. New York: John Wiley & Sons, Inc., 1953.

4. Kish, Leslie, Survey Sampling. New York: John Wiley & Sons, Inc., 1965.

5. Nordin, J.A., "Determining Sample Size." JASA, 39, (1944), 497-506

6. Sadowski, Wieslaw, The Theory of Decision Making, London: Pergamon Press, 1965.

7. Willis, Raymond E., "Confidence Procedures and the Cost of Sampling", The American Statistician, 27 (December 1973), 219-221.

8. Yates, Frank, Sampling Methods for Censuses and Surveys, London: Charles Griffin & Co., Ltd., 1949.

CHART 1 - Loss Function (Θ) for selected Values of K, with c = $10, N = 36,000

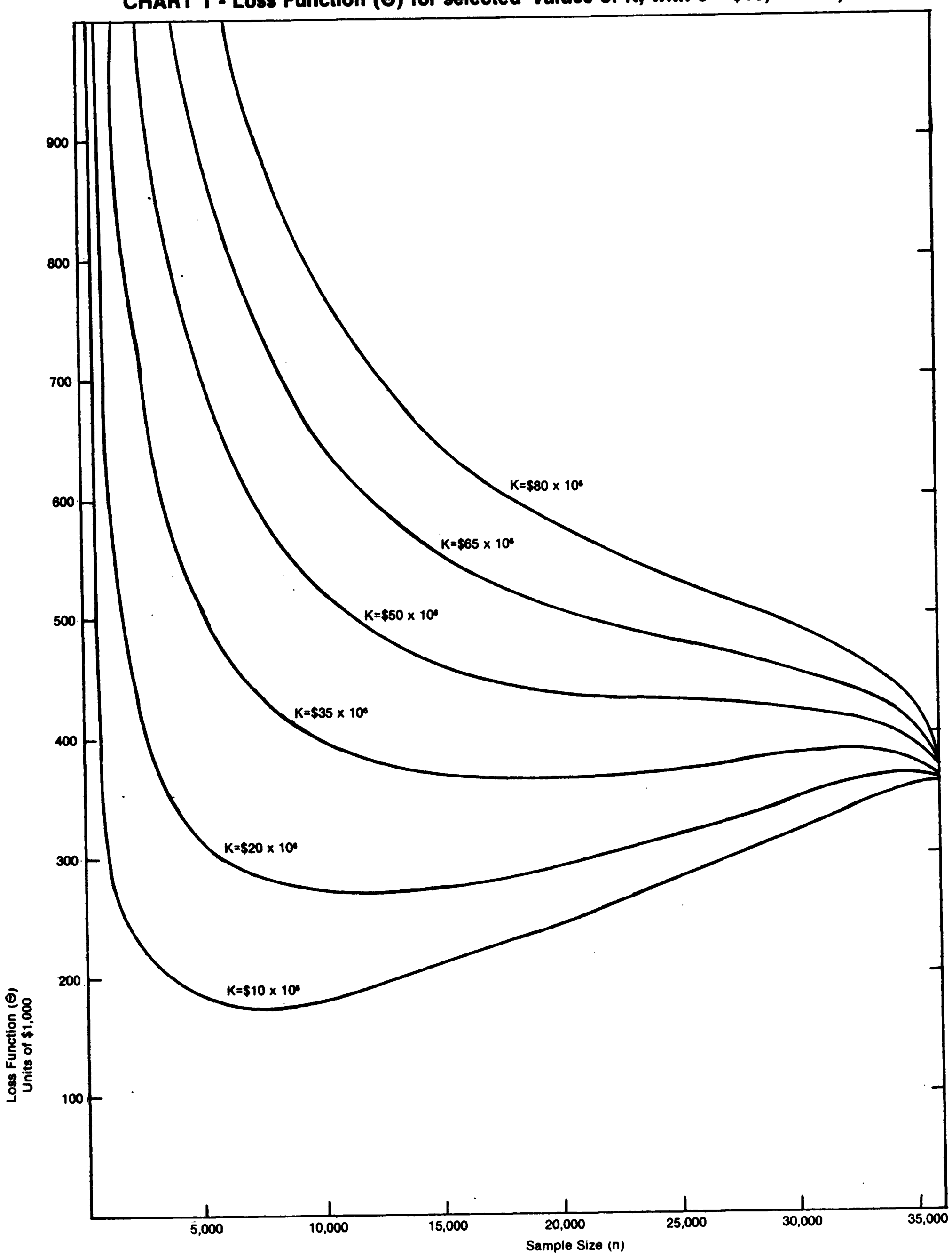


Table 1 - Optimum Sample Size for Determining Federal Fiscal Liability (FFL) for Errors in Determination of State-Funded Supplementary Payments

C = $200

(1) State payments subject to FFL	(2) Eligibility error rate	(3) Overpayment error rate	(4) Expected Federal Liability	(5) Optimum sample size	(6) Cost of selecting and reviewing sample	(7) Standard deviation of estimated liability	(8) The chances that the FFL will not be over-estimated by more than the amount shown are approximately	(9)
Y	P_e	P_o	A	n_{opt}	c n	$\sigma_{\hat{A}}$	95 percent	99 percent
$ 1,000,000	.05	.07	24,000	34	6,800	34,403	56,765	80,159
(N~2,500)	.10	.15	96,000	43	8,600	42,697	70,450	99,484
	.15	.21	160,000	48	9,600	47,613	78,561	110,938
$ 10,000,000	.05	.07	240,000	160	32,000	159,622	263,376	371,919
(N~25,000)	.10	.15	960,000	199	39,800	198,105	326,873	461,585
	.15	.21	1,600,000	222	44,400	220,913	364,506	514,727
$ 50,000,000	.05	.07	1,200,000	469	93,800	466,614	769,913	1,087,211
(N~125,000)	.10	.15	4,800,000	581	116,200	579,107	955,527	1,349,319
	.15	.21	8,000,000	648	129,600	645,781	1,065,539	1,504,669
$100,000,000	.05	.07	2,400,000	744	148,800	740,617	1,222,018	1,725,638
(N~250,000)	.10	.15	9,600,000	923	184,600	919,169	1,516,629	2,141,663
	.15	.21	16,000,000	1,029	205,800	1,024,995	1,691,241	2,388,238
$200,000,000	.05	.07	4,800,000	1,181	236,200	1,175,521	1,939,610	2,738,964
(N~500,000)	.10	.15	19,200,000	1,466	293,200	1,458,921	2,407,219	3,399,286
	.15	.21	32,000,000	1,634	326,800	1,626,890	2,684,369	3,790,654

OPTIMAL AND PROXIMAL MULTIPURPOSE ALLOCATION

Leslie Kish, The University of Michigan

1. BASIC DEFINITIONS AND AIMS

In the literature of survey sampling diverse problems of optimal allocation are treated separately. Yet they can usefully be viewed as distinct examples of the same simple expressions for the total variance and cost of the sample statistic $\bar{y}$:

$$\mathrm{Var}(\bar{y}) = V + V_o = \Sigma V_i^2/m_i + V_o \quad (1.1)$$

and

$$\mathrm{Cost}(\bar{y}) = C + C_o = \Sigma c_i m_i + C_o . \quad (1.2)$$

These linear forms occur in stratified, multistage and multiphase sampling, and other related techniques. Of several applications in Section 7, consider two specific examples. (a) For a stratified sample of elements the variance of the mean $\bar{y} = \Sigma W_i \bar{y}_i$ is:

$$\mathrm{Var}(\bar{y}) = \Sigma(W_i S_i)^2/m_i - \Sigma(W_i S_i)^2/M_i ,$$

where m_i, M_i, W_i and S_i^2 are respectively the sample and population sizes, weights and element variances of the ith stratum. The first term, V, depends on the allocation of the m_i; the second, V_o does not. (b) For two-stage random subselection of b from B elements from each of a random selections of A clusters, the variance of the mean is:

$$\mathrm{Var}(\bar{y}) = (1 - a/A)S_a^2/a + (1 - b/B)S_b^2/ab$$

$$= S_u^2/a + S_b^2/ab - S_a^2/A.$$

Here $S_u^2 = S_a^2 - S_b^2/B$; V comprises the first two terms, with $m_1 = a$ and $m_2 = ab$; the last term $-S_a^2/A = V_o$ does not depend on the m_i. The cost is $c_a a + c_b ab + C_o$.

Definitions and restrictions seem desirable here.

(1) The statistic $\bar{y}$ denotes an estimate of a mean or of an aggregate. Possible extensions to other estimates are not attempted here.

(2) The ith component of the variance, V_i^2/m_i, denotes a constant V_i^2 in the design, a unit variance, divided by the number m_i of sampling units for that component. We prefer V_i^2 to V_i to denote unit variances that are commonly defined with squared values.

(3) The ith component of cost, $c_i m_i$, denotes the unit cost c_i multiplied by the same number m_i of units as in (2).

(4) Components may refer to strata or stages or phases of sampling: generality is the essence of our approach. Components here represent additive sources of variation and cost.

(5) The constants V_i^2 and c_i are parameters for which values are assumed or guessed for numerical solutions of allocation problems. We take $V_i \geq 0$ and $\sqrt{c_i} \geq 0$ (hence V_i^2 and c_i) for allocating the m_i. For nontrivality two pairs at least of the V_i and c_i should be positive. Negative values of V_i^2 may be encountered, as with S_u^2 above; we then redefine the problem to facilitate a practical solution; for an example see Section 7C.

(6) The constants V_o and C_o do not affect optimal allocations of the m_i; their effects on losses in proximal allocation are shown in Section 3. C_o is nonnegative, but V_o is often negative, as above.

(7) For practical values of m_i we want positive integers. Also $0 < m_i \leq M_i$, where M_i denotes the number of units in the population for the component; and $m_i \geq 2$ for computing variance components. Frequently allocation formulas yield some optimal values of $m_i^* > M_i$; when these are reset to $m_i = M_i$ the other optimal values of $m_i^* < M_i$ can be recomputed with (5.4 - 5.5).

(8) It would seem more realistic to guess distributions for V_i^2 and c_i, rather than single values, and a Bayesian treatment of design will probably be worthwhile. But that is beyond our scope here, and I dread a complex procedure out of the reach of survey practitioners. Furthermore, its relative losses would probably not differ much from ours, because losses are insensitive to moderate departures from the guesses.

(9) In some applications, especially for some stratified samples, differences between the c_i are disregarded. Hence, the cost constraint becomes $C/c = m = \Sigma m_i$. Then the $\sqrt{c_i}$ should be omitted from the allocation formulas. Instead of C_o use C_o/c, where c is a common (average) unit cost.

(10) This last point calls attention to the dimensional (unit) homogeneity of all the formulas.

To find optimal values $m_i^* \propto V_i/\sqrt{c_i}$ for the m_i we minimize the product

$$VC = (\Sigma V_i^2/m_i)(\Sigma c_i m_i) , \quad (1.3)$$

when either V or C is fixed at V_f or C_f. This results in the same optimal values as

$$\mathrm{Var}(\bar{y}) \times \mathrm{Cost}(\bar{y}) = (V + V_o)(C + C_o),$$

* Support from the National Science Foundation grant 3191X and suggestions from several colleagues were gratefully received. Curtailed version of contribution to JRSS(A) 1975, which includes the references.

because in $(V + V_o)C_f$ or in $V_f(C + C_o)$ the second terms are unaffected by optimal allocation; their effects on proxima are more easily treated separately (Section 3). To use the product VC rather than some other function seems reasonable: An increase (or decrease) in cost by some factor should be equivalent to a decrease (or increase) in variance by the same factor. The product form leads directly to expressions for loss functions (1 + L) and relative losses (L) that are our goals here. For brevity I use "loss" for L that represents relative increase of variance or cost, without limits.

Our principal aim goes beyond optimization of linear forms, to a simple and coherent treatment of their proximization. We provide convenient forms, in terms of useful parameters, for relative losses incurred by proxima achieved with proximal allocations.

For statisticians "The perfect is the enemy of the good" (proximized from Voltaire). Conflict appears frequently; optimization for one convenient variable often usurps the place of proximization for multipurpose allocation. Proximal methods are seen to be particularly adaptable to multipurpose allocation in Sections 6 and 7F, and fulfill our second aim.

Third, we also present Section 5, a compact, simple and general formulation of optimal allocations for diverse sampling methods. Instead of solving each separately, we merely substitute appropriate symbols for the optimal values $m_i^* \propto V_i/\sqrt{c_i}$. This is obtained with the simple Cauchy inequality. This unified and simple treatment has heuristic and pedagogic merit. Applications in Section 7 cover the diversity of sampling methods. Sections 2 and 3 develop methods of proximization, and Section 4 contains convenient tables for relative losses L.

2. GENERAL FORMULATION

Our principal result (2.3) expresses the relative loss (L) in two parameters: U_i, the relative "sizes" of the components; and $k_i \propto m_i^*/m_i$, the relative departures of the sample sizes m_i from optimal allocations m_i^*. First (2.1), the product VC to be minimized is divided by $(\Sigma V_i\sqrt{c_i})^2$; this ratio will be shown to have minimal (optimal) value of 1. It expresses the relative loss L for any allocation of the m_i (>0), by compensating for the units of measurement of the V_i^2 and c_i. Next (2.2) the m_i are stated in terms of relative departures $k_i \propto m_i^*/m_i$ from their optimal values m_i^*; these will be shown to be $m_i^* \propto V_i/\sqrt{c_i}$. Hence we substitute $m_i \propto V_i/\sqrt{c_i}k_i$ to obtain (2.2); the factors of proportionality cancel. Finally (2.3) for generality and brevity we substitute the relative "sizes" $U_i = V_i\sqrt{c_i}/\Sigma V_i\sqrt{c_i}$.

$$1 + L = VC/(\Sigma V_i\sqrt{c_i})^2 = (\Sigma V_i^2/m_i)\ (\Sigma c_i m_i)/(\Sigma V_i\sqrt{c_i})^2 \tag{2.1}$$

$$= (\Sigma V_i\sqrt{c_i}k_i)\ (\Sigma V_i\sqrt{c_i}/k_i)/(\Sigma V_i\sqrt{c_i})^2 \tag{2.2}$$

$$= (\Sigma U_i k_i)\ (\Sigma U_i/k_i), \tag{2.3}$$

where $U_i = V_i\sqrt{c_i}/\Sigma V_i\sqrt{c_i}$ and $k_i \propto V_i/\sqrt{c_i}m_i$. We take the k_i and U_i to be positive and finite. We have $\Sigma U_i = 1$; and we may also use any convenient $U_i' \propto U_i$, if we divide by $\Sigma U_i'$. Note that we need only the relative values of k_i, and we can use Ak_i, with A any positive and finite constant. Furthermore, the form of (2.3) shows that the k_i may be replaced by their reciprocals; they may refer to ratios of oversampling, as well as undersampling. With this flexibility we can use $\min(k_i) = 1$, as we do in Table 4.1 for convenience.

The minimal value of 1 for (2.3) is obtained with all $K_i = k^*$ equal. This may appear obvious or seen with the Lagrange Identity in (5.1).

Examples may be useful here.

(a) Consider the variance of the mean $\Sigma W_i\bar{y}_i$ for two strata where $W_1 = 0.2$, $W_2 = 0.8$, $S_1^2 = S_2^2 = S^2$, and $c_1 = c_2 = c$. Then $U_i = W_i \propto V_i$, and $U_1:U_2 = 1:4$.

This implies (5.3) that optimal allocation of sample sizes should be in the ratio of stratum sizes W_i, hence $m_2 = 4m_1$. If samples of equal sizes, $m_1 = m_2$, are taken, this implies a departure factor of 4; we can use simply $k_1 = 1$ and $k_2 = 4$. The consequent relative loss L would be given by (2.3) as $1 + L = (0.2\cdot1 + 0.8\cdot1/4)(0.2\cdot1 + 0.8\cdot4) = 1.360$.

(b) To illustrate the effect of the U_i on the loss L: suppose now $S_1^2 = 4S_2^2$, and $c_1 = 4c_2$. Since $S_1^2/c_1 = S_2^2/c_2$, optimal allocation is still 1:4. But now $U_i = W_iS_i\sqrt{c_i}/\Sigma W_iS_i\sqrt{C_i}$, hence $U_1 = U_2 = 0.5$. Therefore the relative loss L from equal sample sizes now would be given by $1 + L =$

$$(0.5\cdot1 + 0.5\cdot1/4)(0.5\cdot1 + 0.5\cdot4) = 1.5625.$$

(c) To illustrate a conflict in allocation: suppose that as in (a) $S_1^2 = S_2^2$ and $c_1 = c_2$, but that now we want to minimize the variance of the difference of means $(\bar{y}_1 - \bar{y}_2)$. Now $U_1 = U_2 = 0.5$. Optimal allocation is at $m_1 = m_2$. Departure from this in the ratio 1:4 to satisfy (a) would result in $1 + L =$

$$(0.5\cdot1 + 0.5\cdot4)(0.5\cdot1 + 0.5\cdot1/4) = 1.5625.$$

Note that these answers can also be found in Table 4.1 in column K = 4 for relative departures. The size U of one component is 0.2 for (a), and 0.5 for both (b) and (c), in the top two rows. Results for (a) and (c) illustrate common conflicts between totals and domains, treated in Section 7F, and in Table 7.2A.

The weights U_i are convenient for design; based on population parameters, we may call them population weights. However, when dealing with sample results it may be more convenient to use sample weights, based on sample sizes: $u_i = U_i/k_i$. Then (2.3) may be written as:

$$1 + L = (\Sigma u_i k_i^2)(\Sigma u_i) \quad (2.4)$$

$$= 1 + (\Sigma u_i k_i^2/\Sigma u_i - \bar{k}^2)/\bar{k}^2 = 1 + C_k^2 \quad (2.5)$$

$$= 1 + \Sigma u_i (k_i/\bar{k} - 1)^2/\Sigma u_i \quad (2.6)$$

$$= 1 + \Sigma (k_i/\bar{k} - 1)^2 \, m_i c_i/\Sigma m_i c_i. \quad (2.7)$$

C_k^2 is the relative variance of the k_i with sample weights u_i around their mean $\bar{k} = \Sigma u_i k_i/\Sigma u_i = 1/\Sigma u_i$. Here larger $k_i > 1$ represent larger weights to compensate for undersampling proportionately to their reciprocals. The $u_i = U_i/k_i$ are proportional to $c_i m_i$ because the $U_i = V_i\sqrt{c_i}/\Sigma V_i\sqrt{c_i} \propto c_i m_i^* \propto c_i m_i k_i$.

3. ON PROXIMAL ALLOCATION

Extreme departures from optimal values of m_i^* can result in large relative losses measured in either cost or variance. However, small or even moderate departures from the optimal m_i^* lead only to negligible or small relative losses. These vague precepts of practicing statisticians are given formal and practical expressions (2.4 - 2.7) in terms of the relative loss (L) compared to optimal allocation.

When the frequencies for the k_i are given or estimated in sample proportions u_i, then (2.5 - 2.7) yield readily the loss L in terms of the relvariance C_k^2 of the k_i values. One of these formulas may be most convenient for judging the losses from actual sample results.

However, for comparing designs of planned samples the frequencies may be more conveniently stated in terms of the population weights U_i. Formula (2.3) can be readily computed for moderate numbers of components. Furthermore, the simple models of Table 4.1 can often give instant answers for approximate distributions. I have often found these answers close and adequate for planning designs.

Computations of the relative loss L in our formulas and tables take account of the factors V and C in the minimized function VC, but they neglect the constant V_o or C_o in the total variance and cost (1.1 and 1.2). However this neglect may be corrected with translations of L into L' that does take into account the constant factors V_o and C_o. If V_{min} is the optimal V for fixed C_f^o then the ratio of the attained proximal variance to the optimal variance is

$$\frac{(1+L)V_{min}+V_o}{V_{min}+V_o} = 1 + L/(1+V_o/V_{min}) = 1 + L'. \quad (3.1)$$

Thus the adjusted actual relative loss L' differs from that indicated by L; since V_o is often negative, L' can be somewhat greater than L. For a C_{min} found for a fixed V_f, the adjusted relative loss L' may be somewhat less than L, due to a positive C_o in

$$L' = L/(1 + C_o/C_{min}) \quad (3.2)$$

4. TABLES OF LOSSES FOR MODEL DISTRIBUTIONS

For a variety of simple models we can give instant answers about expected losses. Actual population distributions can usually be matched against one of these models so as to provide useful approximations of the expected losses.

The losses are given in terms of departures k_i from optimal allocations for the relative weights U_i in the models, and the k_i range from $\min(k_i) = 1$ to $\max(k_i) = K$. The simplest model consists of two components U and (1-U), where the relative departures from optimal sample sizes are in the ratio $k_1:k_2 = 1:K$. The loss for two components may be expressed (7.4) as

$$L = U(1-U)(K-1)^2/K. \quad (4.1)$$

The dichotomous models represent maximal losses for ranges of departure fixed at 1 to K. Thus losses for large values of K are much greater in the top three rows of Table 4.1 than further down where five other models are shown.

The five models represent diverse frequency distributions for the population weights U_i; and for each model both discrete and continuous versions are shown. In the discrete versions the relative departures k_i take K integral values from 1 to K, and the relative weights U_i are concentrated at those values. In continuous versions the departures k_i and relative weights U_i vary continuously from 1 to K. Frequencies are divided by their sums to produce relative frequencies U_i.

Note that the loss L is both very small and uniform for all models for small K; for K = 1.3 L_d = .017 and L_c = .006, and for K = 1.5 L_d = .04 and L_c = .014. (Note that for L_d the k_i take only two values 1 and K = 1.3 or 1.5). From K = 2 to about K = 5 the losses are moderate and fairly similar for the five models. The L_c are lower than the L_d, though in an irregular ratio. Below K = 10 we can make fairly good guesses about L just from the range 1 to K, without knowing much about the U_i -- if this is not dichotomous or U-shaped.

However beyond K = 10 the losses L increase and diverge. Three of the models show rather similar losses, but for the model $U_i \propto 1/k_i$ the losses are much larger. And this model may often resemble actual frequencies. The fifth has much lower losses, but it is not realistic, I think.

Table 4.1 Relative Losses (L) for 6 Models of Population Weights (U_i); for Discrete (L_d) and Continuous (L_c) Weights: for Relative Departures (k_i) in the Range from 1 to K.

Models	K	1.3	1.5	2	3	4	5	10	20	50	100	500	1000
Dichotomous U(1-U)													
(0.5)(0.5)		.017	.042	.125	.333	.562	.800	2.025	4.512	12.005	24.50	124.5	249.5
(0.2)(0.8)		.011	.027	.080	.213	.360	.512	1.296	2.888	7.683	15.68	79.7	159.7
(0.1)(0.9)		.006	.015	.045	.120	.202	.288	.729	1.624	4.322	8.82	44.8	89.8
Rectangular	L_d	.017*	.042*	.125*	.222	.302	.370	.611	.889	1.295	1.620	2.403	2.746
$U_i \propto 1/K$	L_c	.006	.014	.040	.099	.155	.207	.407	.656	1.036	1.349	2.120	2.461
Linear Decrease	L_d	.017*	.040*	.111*	.203	.283	.353	.616	.940	1.437	1.917	2.879	3.333
$U_i \propto K+1-k_i$	L_c	.006	.014	.040	.097	.153	.205	.409	.680	1.127	1.514	2.507	2.956
Hyperbolic Decr.	L_d	.017*	.040*	.111*	.215	.312	.404	.807	1.466	3.014	5.076	16.802	28.342
$U_i \propto 1/k_i$	L_c	.006	.014	.041	.103	.171	.235	.528	1.011	2.138	3.621	11.998	19.915
Quadratic Decr.	L_d	.016*	.036*	.080*	.150	.211	.264	.460	.696	1.048	1.333	2.026	2.331
$U_i \propto 1/k_i^2$	L_c	.006	.014	.040	.099	.155	.207	.407	.656	1.036	1.349	2.120	2.461
Linear Increase	L_d	.017*	.040*	.111*	.167	.200	.222	.273	.302	.320	.327	0.330	0.333
$U_i \propto k_i$	L_c	.006	.013	.037	.083	.120	.148	.223	.273	.308	.320	0.331	0.332

Dichotomous $1 + L = 1 + U(1-U)(K-1)^2/K$

Discrete $1 + L_d = (\Sigma U_i k_i)(\Sigma U_i/k_i)$, with $k_i = i = 1,2,3,\ldots K$

Continuous $1 + L_c = \int Uk.dk.\int(U/k)dk$, with $1 \leq k \leq K$.

Only 2 Values, 1 and K, were used for L_d for K = 1.3, 1.5 and 2

From the models one can also make conjectures about actual distributions that differ somewhat from them. For example, a rectangular distribution for integral values of k_i from 1 to 5 has $L_d = 0.370$; more than 5 values evenly spaced within the same range to 1 to 5 would have a loss between that value and the continuous loss $L_c = .207$. On the other hand for only three values of k_i from 1 to 5 and $U_i = 1/3$, the loss (actually 0.533) is above 0.370, but below the dichotomous value of 0.800 in Table 4.1.

When sample weights $u_i = U_i/k_i$ seem more convenient, the relative loss L may be estimated by the relvariance C_k^2 of the k_i with weights u_i (2.7). Formulas and tables can be constructed for such relvariances, if we begin with the means M and variances σ^2 of convenient distributions from 0 to 1 [Kish, 1965, p.262]. To obtain the relvariances C_k^2, those variances are multiplied by the $(\text{range})^2 = (K-1)^2$ and divided by the new $(\text{mean})^2 = [M(K-1)+1]^2$; thus $C_k^2 = \sigma^2(K-1)^2/[M(K-1)+1]^2$.

5. ON OPTIMAL ALLOCATION

The Lagrange identity is a basic tool of great utility, and it may be stated here simply. Assume x_i and y_i (i = 1,2,...n) finite and real; but here we need only nonnegative values. Then

$$(\Sigma x_i^2)(\Sigma y_i^2) = \Sigma x_i^2 y_i^2 + \sum_{i \neq j} x_i^2 y_j^2$$

$$= \Sigma x_i^2 y_i^2 + \sum_{i \neq j} x_i y_i x_j y_j + \sum_{i \neq j} x_i^2 y_j^2 - \sum_{i \neq j} x_i y_i x_j y_j$$

$$= (\Sigma x_i y_i)^2 + \sum_{i<j} (x_i y_j - x_j y_i)^2. \qquad (5.1)$$

The second term has a minimum of 0, when $y_i = Fx_i$, F constant. The first term alone is the lower bound of the Cauchy - Schwartz inequality. If we take in (5.1) $x_i = \sqrt{(U_i/k_i}$ and $y_i = \sqrt{(U_i k_i)} = k_i x_i$ we now rewrite (2.3) as

$$1 + L = (\Sigma U_i k_i)(\Sigma U_i k_i) = \Sigma y_i^2 \Sigma x_i^2$$

$$= 1 + \sum_{i<j} \frac{U_i U_i}{k_i k_j}(k_i - k_j)^2, \qquad (5.1')$$

with $(\Sigma x_i y_i)^2 = (\Sigma U_i)^2 = 1$. The minimal value is 1, when the second term is 0, because all k_i are equal.

Now let $x_i = \sqrt{(V_i^2/m_i)}$ and $y_i = \sqrt{(c_i m_i)}$, with V_i^2 and c_i as assumed parameters and m_i as variables (all ≥ 0). The minimal value of

$$VC = (\Sigma V_i^2/m_i)(\Sigma c_i m_i) \geq (\Sigma V_i \sqrt{c_i})^2, \qquad (5.2)$$

a Cauchy - Schwartz inequality is obtained when $\sqrt{(c_i m_i^*)} = F\sqrt{(V_i^2/m_i^*)}$.

Then

$$m_i^* = FV_i/\sqrt{c_i} \qquad (5.3)$$

are the optimal values of the m_i^* that obtain the

$$\text{minimal VC} = (\Sigma V_i \sqrt{c_i})^2. \qquad (5.3')$$

The constant F can be determined from either C_f or V_f fixed. With $C_f = \Sigma c_i m_i^* = F\Sigma V_i\sqrt{c_i}$ one uses $F = C_f/\Sigma V_i\sqrt{c_i}$. For $V_f = \Sigma V_i^2/m_i^*$ note that $V_i\sqrt{c_i} = FV_i^2/m_i^*$ and $\Sigma V_i\sqrt{c_i} = FV_f$; hence $F = (\Sigma V_i\sqrt{c_i})/V_f$.

6. MULTIPURPOSE ALLOCATION

Sample surveys are typically multipurpose in nature, and it seems imperative to extend the methods of allocation to multipurpose designs. For lack of these methods univariate allocation dominates our literature and theory of sampling; practical work is also affected, but less often. The methods for optimization and proximization developed here seem particularly adaptable to multipurpose design. The general form $\Sigma V_i^2/m_i$ for variances can serve well the many purposes of a sample survey; for the gth purpose the variance will be denoted by $\sum_i V_{gi}^2/m_i$.

The many purposes of a single survey may have several sources. (1) A single variable may result in several statistics; e.g. the mean and median of incomes can benefit from different allocations [Kish, 1961]. (2) Most surveys obtain results for several variables on a single subject. (3) Furthermore, some surveys are multisubject in character; e.g. with economic, demographic, social variables. (4) Results for subclasses and for their comparisons may be as important as results based on the entire sample. Designs for subclasses often point to different designs and allocations than those for the entire sample. (5) The common but neglected conflict between designs for comparisons between domain means and for the combined mean for the entire sample is developed in Section 7F.

Suppose a sample is allocated optimally for variate $\bar{y}'$ with m_i' proportional to $V_i'/\sqrt{c_i'}$, but optimal allocation for another variate $\bar{y}$ would be $m_i \propto V_i/\sqrt{c_i}$. The loss incurred for $\bar{y}$ can be measured with the departures $k_i = m_i/m_i' = (V_i/V_i')(\sqrt{c_i'}/\sqrt{c_i})$ and with weights $U_i = V_i\sqrt{c_i}/\Sigma V_i\sqrt{c_i}$ in formula (2.3). We are mostly concerned with allocation of the m_i within one survey sample, so that $\sqrt{c_i'}/\sqrt{c_i} = 1$. Then the loss function for $\bar{y}$ due to optimization for $\bar{y}'$ may be represented by

$$1 + L(m_i') = (\Sigma V_i^2\sqrt{c_i}/V_i')(\Sigma V_i'\sqrt{c_i})/(\Sigma V_i\sqrt{c_i})^2 \quad (6.1)$$

$$= \sum\left(\frac{V_i\sqrt{c_i}}{\Sigma V_i\sqrt{c_i}}\right)^2 \Bigg/ \left(\frac{V_i'\sqrt{c_i}}{\Sigma V_i'\sqrt{c_i}}\right).$$

This may be regarded as the relvariance of $k_i = V_i/V_i'$ with weights $u_i = V_i'\sqrt{c_i}$ (2.7). Often the cost factors are constant or disregarded, and (6.1) has a particularly simple form

$$1 + L(m_i') = \Sigma(V_i/\Sigma V_i)^2/(V_i'/\Sigma V_i'). \quad (6.1')$$

If the m_i allocated for one survey with c_i are used for another with $c_i \neq c_i'$, then we rewrite (6.1), with $V_i'\sqrt{c_i}/\sqrt{c_i'}$ in place of V_i', as

$$1 + L(m_i') = (\Sigma V_i^2\sqrt{c_i'}/V_i')(\Sigma V_i'\ c_i/\sqrt{c_i'})/(\Sigma V_i\sqrt{c_i})^2. \quad (6.2)$$

Now consider a loss function for several variates indexed with g(= 1, 2, 3...). The loss function, for a fixed cost $C_f = \Sigma c_i m_i$, may be expressed for each as

$$1 + L_g = (\Sigma V_{gi}^2/m_i)/V_{gmin},$$

where the denominator denotes the minimal variance attainable and computed for the gth variate. Assign the weights I_g $(\Sigma I_g = 1)$ to denote the relative importance of the lost precision on the gth variate. Then consider the total expected loss as a linear function of the quadratic loss functions (for a fixed set of m_i) of the variances

$$1 + L(m_i) = \sum_g I_g(1 + L_g) = 1 + \sum_g I_g L_g(m_i)$$

$$= \sum_g I_g \frac{\sum_i V_{gi}^2/m_i}{V_{gmin}}$$

$$= \sum_i \frac{1}{m_i} \sum_g \frac{I_g V_{gi}^2}{V_{gmin}} = \sum_i \frac{Z_i^2}{m_i}, \quad (6.3)$$

where $Z_i^2 = \sum_g I_g V_{gi}^2/V_{gmin}$. Changing the order of summation permits defining this ith component that can be computed. For the multipurpose joint allocation we may compute (5.3) the

$$\text{optimal } m_i^{**} = \frac{Z_i}{\sqrt{c_i}} \cdot \frac{C_f}{\Sigma Z_i\sqrt{c_i}} \quad \text{and} \quad (6.4)$$

$$1 + L(m_i^{**}) = V_{min} = (\Sigma Z_i\sqrt{c_i})^2/C_f. \quad (6.5)$$

From the multipurpose optimal allocations m_i^{**} we may compute the loss function $1+L_g(m_i^{**})$ for the gth variate considered separately. For each of these we can use (6.1) with $V_i = V_{gi}$, $V_i' = Z_i$, $k_{gi} = V_{gi}/Z_i$ and $U_{gi} = V_{gi}\sqrt{c_i}/\Sigma V_{gi}\sqrt{c_i}$. These may be averaged with the weights I_g to obtain the joint loss function (6.3) of $1 + L(m_i^{**})$ with the multipurpose optimal allocations m_i^{**}.

This, however, may be obtained more directly from (6.4) or (6.5). Thus

$$1 + L(m_i^{**}) = \sum_i \frac{Z_i^2}{m_i^{**}} = (\sum_i Z_i\sqrt{c_i})^2/C_f$$

$$= \frac{1}{C_f}\left[\sum_i \sqrt{\left(\sum_g \frac{I_g V_{gi}^2 c_i}{V_{gmin}}\right)}\right]^2. \quad (6.6)$$

When we accept (from 5.2) $V_{gmin} = (\Sigma V_{gi}\sqrt{c_i})^2/C_f$, we obtain a simpler form,

because $V_{gi}^2 c_i / V_{gmin} = (V_{gi}\sqrt{c_i}/\Sigma V_{gi}\sqrt{c_i})^2$.
Thus the jointly determined minimal loss function becomes

$$1 + L(m_i^{**}) = \{\Sigma_i \sqrt{[\Sigma_g I_g (V_{gi}\sqrt{c_i}/\Sigma_i V_{gi}\sqrt{c_i})^2]}\}^2.$$

$$= \{\Sigma_i \sqrt{(\Sigma_g I_g U_{gi}^2)}\}, \qquad (6.7)$$

The minimal and optimal values may be unobtainable, due chiefly to the constraints $m_i^* \leq M_i$ (Section 5). In that case the above loss function overestimates the losses incurred over obtainable values of V_{gmin}. Note also that using these leads to $Z_i\sqrt{c_i} = \sqrt{[\Sigma_g I_g U_{gi}^2]}\sqrt{C_f}$, hence to

$$\text{optimal } (m_i^{**}) = \frac{\sqrt{[\Sigma_g I_g U_{gi}^2]}}{\Sigma_i[\Sigma_g I_g U_{gi}^2]} \cdot \frac{C_f}{c_i}. \qquad (6.8)$$

These can be seen applied in 7F to the important and frequent conflict between allocations for weighted totals and comparisons of domains. Two examples are shown in Table 7.2. Note in the last column of 7.2B how encouragingly insensitive are the values of (6.8) for moderate differences in the assignments of I_g.

The weighted mean of relative quadratic losses (6.3) is a modified version of a function proposed by Dalenius (1957, Ch.9). Another version (Yates, 1960, Cochran, 1963) uses $\Sigma_g I_g' \Sigma_{g_i} V_{gi}^2/m_i$, the weighted average of variances. Our (6.3) can be easily adapted by using $T_i^2 = \Sigma_g I_g' V_{gi}^2$ instead of Z_i^2; in this formulation the weights $I_g' = I_g/V_{gmin}$ include the minimal variances. This may appear simpler, but it is less explicit.

The optimal allocation of $m_i^* \propto Z_i/\sqrt{c_i}$ can also be obtained with Lagrange Multipliers applied to the function

$$F(m_i) = \Sigma_g I_{g_i} \Sigma V_{gi}^2/m_i V_{gmin} + \lambda \Sigma c_i m_i. \qquad (6.9)$$

With Lagrange Multipliers we also investigated two other loss functions: the product, $\Pi(1 + L_g)$, and the sum of the relative precisions, $[\Sigma(1 + L_g)^{-1}]^{-1}$. But the results seem less crucial than good choices for the weights I_g of relative importance.

Our methods here aim to minimize the first term of $V + V_o$ for fixed C_f. In situations where V_o is considerable, the actual loss should be modified to $L' = L/(1 + V_o/V_{min})$, as noted in Section 3. Furthermore, I consider fixing C_f more practical than trying to fix values for a set of V_g and then to minimize C_f. This problem seems to have been solved with "convex programming" on several separate occasions, [Srikantan, 196?, and Huddleston, 1970, were not the first]; but I do not find this approach useful.

7. $\text{Var}(\bar{y}) = \Sigma V_i^2/m_i + V_o$ IN SEVERAL APPLICATIONS

A) Stratified Sampling: $V_i = W_i S_i$

$$\Sigma\frac{(W_i S_i)^2}{m_i} - \Sigma\frac{W_i^2 S_i^2}{M_i} \qquad \text{opt } m_i^* \propto \frac{W_i S_i}{\sqrt{c_i}}$$

B) Multistage Random Selection of Equal Clusters:

2 stages $$\frac{S_a^2 - S_b^2/B}{a} + \frac{S_b^2}{ab} - \frac{S_a^2}{A}$$

$$\text{opt } b^* = \sqrt{\left(\frac{c_a}{c}\frac{S_b^2}{S_a^2 - S_b^2/B}\right)}$$

$m_1 = a$, $m_2 = ab$, $m_3 = abc$

3 stages $$\frac{S_a^2 - S_b^2/B}{a} + \frac{S_b^2 - S_c^2}{ab} + \frac{S_c^2}{abc} - \frac{S_a^2}{A}$$

C) Two-Phase Sampling: Cost $= \Sigma c_i m_i + c_L n_L + C_o$

for Stratification:

$$\Sigma\frac{(W_i S_i)^2}{m_i} + \frac{\Sigma W_i(\bar{Y}_i - \bar{Y})^2}{n_L} - \frac{\Sigma(W_i S_i)^2}{M_i} + \frac{\Sigma W_h(\bar{Y}_i - \bar{Y})^2}{\Sigma M_i}$$

$$= (\Sigma W_i S_i)^2/\Sigma m_i + \Sigma W_i(\bar{Y}_i - \bar{Y})^2/n_L \text{ when } m_i \propto W_i S_i$$

for Regression: $$\frac{S^2(1 - R^2)}{\Sigma m_i} + \frac{R^2 S^2}{n_L}$$

D) Subsampling (1-P)m/k of Nonresponses:

$$\text{Cost} = (c_o/P + c_p)Pm + c_q(1-P)m/k$$

$$\frac{P^2 S_p^2}{Pm} + \frac{(1 - P)^2 S_q^2}{(1 - P)m/k}$$

$$\text{opt } k^* = \frac{S_p}{S_q}\left[\frac{c_q}{c_o/P + c_p}\right]^{1/2}$$

E) Weights in Estimation: $\Sigma W_i = \Sigma W_i^* = 1$

$$V^2 = \Sigma W_i \sigma_i^2 \qquad \text{opt } W_i^* \propto 1/\sigma_i^2 \qquad V_{min} = 1/\Sigma 1/\sigma_i^2$$

$$1 + L = V^2/V_{min}^2 = \Sigma W_i^2/W_i^*$$

$$L = \Sigma W_i^*\left(\frac{W_i}{W_i^*} - 1\right)^2$$

7F. Allocation Conflict Between Totals and Independent Domains

Serious conflict often exists between reducing the variance for the combined mean $\Sigma W_i \bar{y}_i$, and equal precision desired for the means $\bar{y}_i$ of H independent domains that differ greatly in relative sizes $W_i (\Sigma W_i = 1)$. The domains may be the regions or provinces of a country, etc. This common example of multipurpose allocation deserves special attention.

The combined mean variance $V_c = \Sigma W_i^2 S_i^2 / m_i$ is minimal when the optimal $m_{ci}^* \propto W_i S_i / \sqrt{c_i}$. However $m_{di}^{**} \propto S_i / \sqrt{c_i}$ are optimal for obtaining equal precision for each of the H domain means; also to obtain equal precision for the H(H-1)/2 possible comparisons of domain means. Thus we can denote an <u>average domain variance</u>

$V_d = (\Sigma S_i^2 / m_i)/H^2$ for the variance of $\Sigma \bar{y}_i / H$. The conflict between the purposes is represented in the above two optimal values for m_i^* by the presence of the weights W_i for the combined mean, their absence for the domain means. Thus the loss function (2.3) for the combined mean, due to allocation $m_i \propto S_i / \sqrt{c_i}$, has the departures $k_{ci} = m_i^* / m_i = W_i$, and the weights $U_{ci} \propto W_i S_i \sqrt{c_i}$. The loss function for the average domain means, due to allocations $m_i \propto W_i S_i / \sqrt{c_i}$, has the departures $k_{di} = 1/W_i$ and the weights $U_{di} \propto S_i \sqrt{c_i}$.

To see clearly the effects of variation in the domain sizes W_i, we make some simplifying assumptions that are often approximated in practical situations. Assume that the S_i^2 incorporate the effects of complex designs, and that they are constant across domains, as are the c_i. Further, suppose that $m_i^* \le M_i$ in all domains. We shall also neglect effects of the constants V_o and C_o on the loss functions.

Under these conditions we may omit, for brevity, the constants S^2 and c from the formulas, and we allocate the total sample size $m = \Sigma m_i$ among the domains. For $\Sigma W_i \bar{y}_i$ the optimal $m_i^* = mW_i$, with departures $k_{ci} = m_i^*/m_i = mW_i/m_i$ and weights $U_{ci} = W_i$, the loss function $1 + L_c = m\Sigma W_i^2/m_i$ is minimal at $mV_{cmin} = 1$. For $\Sigma \bar{y}_i/H$ the optimal $m_{di}^* = m/H$, weights $U_{di} = 1/H$, with departures $k_{di} = m/Hm_i$ and the loss function $1 + L_d = mH^{-2}\Sigma 1/m_i$ the minimal at $mV_{dmin} = 1$ also.

In Table 7.1 for loss functions $(1 + L_c)$ the minimal value 1 appears with $m_i \propto W_i$ for $\Sigma W_i \bar{y}_i$, and with $m_i \propto 1/H$ for $\Sigma \bar{y}_i / H$. The other allocations produce relative losses (L>0) that increase with diversity among the relative sizes W_i; and C_w^2 denotes their relative variance, $H^2 Var(W_i)$.

Jointly for the two purposes, we can find optimal allocation and the loss function with (6.3). For any allocation m_i, the joint loss function is

$$
\begin{aligned}
1 + L_j(m_i) &= I_c m\Sigma W_i^2/m_i + I_d mH^{-2}\Sigma 1/m_i \\
&= m\Sigma \left[I_c W_i^2 + I_d H^{-2} \right] / m_i \\
&= mH^{-2} \sum_i [I_c D_i^2 + I_d]/m_i \\
&= mN^{-2} \sum_i [I_c N_i^2 + I_d \bar{N}^2]/m_i \\
&= m \sum_i t_i^2/m_i . \qquad (7.23)
\end{aligned}
$$

Here $0 < I_c < 1$ is the relative importance for the combined mean variance and $I_d = 1 - I_c$ for the mean domain variance. We may find it convenient to use $D_i = HW_i$ with mean $\bar{D} = 1$, or $N_i = NW_i$ when these denote domain sizes.

We find the joint optimal allocations $m_i^{**} = mt_i / \Sigma t_i$ where the

$$t_i = \sqrt{(I_c W_i^2 + I_d H^{-2})} = \sqrt{(I_c D_i^2 + I_d)}/H = \sqrt{(I_c N_i^2 + I_d \bar{N}^2)}/N.$$

The m_i^{**} may be found simply with (5.3) but also as an illustration of (6.8).

The multipurpose allocation m_i^{**} can also be shown (5.2) to produce the multipurpose minimal variance

$$V_{min} = (\Sigma t_i)^2 / m. \qquad (7.24)$$

When we use the multipurpose optimal $m_i^{**} \propto t_i$ we can determine the loss functions $(1 + L)$ incurred for the variances of $\Sigma W_i \bar{y}_i$ and $\Sigma \bar{y}_h / H$; we use (6.1) or (6.2) with

$$k_{ci} \propto V_{ci}/V_i' \propto W_i/t_i$$

and $\quad k_{di} \propto V_{di}/V_i' \propto 1/Ht_i \quad$ respectively.

These (1 + L) are shown on the bottom row of Table 7.1. The last column shows the effects of the three different allocations on the joint multipurpose loss function $1 + L_j(m_i)$.

Two numerical problems illustrate the method in Table 7.2. In A, for two domains having sizes $W_1/W_2 = 4{:}1$ are shown the loss functions for three purposes --- total, domain and joint --- under diverse allocations. In B the method is applied to the 133 countries of the world, omitting the four largest, over 200 millions, and a few smallest, under 0.2 millions. Including them would be more dramatic but less realistic.

Conflict of Combined Mean ($\Sigma W_i \bar{y}_i$) and Average Domain Mean ($\Sigma \bar{y}_i/H$)

(S_i^2 and c_i are assumed constant and omitted.)

Table 7.1 Loss Function (1 + L) for the Combined Mean, for the Average Domain Mean, and for a Weighted Joint Function.

Note $t_i = \sqrt{(I_c W_i^2 + I_d H^{-2})} = \sqrt{(I_c D_i^2 + I_d)}/H$

	Loss Functions (1 + L) for		
	$\Sigma W_i \bar{y}_i$	$\Sigma \bar{y}_i/H$	$I_c \Sigma W_i \bar{y}_i + I_d \Sigma \bar{y}_i/H$
$(1 + L) = mV^2$	$m\Sigma W_i^2/m_i$	$mH^{-2}\Sigma 1/m_i$	$m\Sigma t_i^2/m_i$
Allocation of m_i			
mW_i	1	$H^{-2}\Sigma 1/W_i$	$I_c + I_d H^{-2}\Sigma 1/W_i$
m/H	$H\Sigma W_i^2 = 1 + C_w^2$	1	$I_c H\Sigma W_i^2 + I_d H^{-1}$
$mt_i/\Sigma t_i$	$(\Sigma W_i^2/t_i)(\Sigma t_i)$	$H^{-2}(\Sigma 1/t_i)(\Sigma t_i)$	$(\Sigma t_i)^2$

Table 7.2 Loss Functions (1 + L) for Two Populations

Allocations m_i	(A) (1 + L) for $W_1/W_2 = 4$: $\Sigma W_i \bar{y}_i$	$\Sigma \bar{y}_i/2$	Joint	(B) (1 + L) for 133 countries: 0.2 to 100 mm: $\Sigma W_i \bar{y}_i$	$\Sigma \bar{y}_i/133$	Joint with weights 1:1	I_c/I_d:1
mW_i	1	1.56	1.28	1	6.86	3.93	
m/H	1.36	1	1.18	3.34	1	2.17	
$\propto\sqrt{W_i}$	1.08	1.125	1.102	1.35	1.54	1.44	
$\propto\sqrt{(W_i^2 + H^{-2})}$	1.116	1.080	1.098	1.31	1.28	1.295	
$\propto\sqrt{(0.5W_i^2 + H^{-2})}$				1.47	1.17	(1.32)	1.27
$\propto\sqrt{(2W_i^2 + H^{-2})}$				1.20	1.44	(1.32)	1.28
$\propto\sqrt{(4W_i^2 + H^{-2})}$				1.12	1.66	(1.39)	1.23

In (A) there are two strata and domains ($W_1 = 0.8$ and $W_2 = 0.2$); note that the allocation $m_i^* \propto \sqrt{W_i}$ does almost as well for the joint loss as the optimal.

In (B) we have the populations of 133 countries, ranging in size from 0.2 to over 100 millions, a range of 500 in relative sizes. From this problem of allocation (for the World Fertility Survey) we omitted, for practical reasons, the four largest countries and a few under 0.2 millions. Their inclusion would raise the variance of relative sizes, W_i, from 2.5 to 12, and would make the results more dramatic. Note that the $\sqrt{W_i}$ allocation reduces losses quite well. Some compromise is better than none. But the optimal allocation, $\sqrt{(W_i^2 + H^{-2})}$, is considerably better. Different values of I_c/I_d (= 1/2, 2/1 and 4/1) increase slightly the variance of the joint loss function with (1:1) weights; but they remain steady for joint loss functions with their own weights I_c/I_d:1.

THE COUNTING RULE STRATEGY IN SAMPLE SURVEYS

Monroe G. Sirken, National Center for Health Statistics

1. INTRODUCTION

Frequently, the enumeration units in sample surveys are not identical to the population elements whose parameters are being estimated. For example, Table 1 on the following page lists four illustrative surveys in which the enumeration units are different from the population elements. Clearly, in surveys of this type, it is necessary to specify rules for linking elements to the enumeration units where the elements are eligible and required to be counted in the survey. The counting rule adopted in the survey should be viewed as a design factor for improving the efficiency of the survey. We have investigated [4], [5], [6] some of the statistical properties of counting rules. In this expository report, we summarize some of those findings, and we describe the strategy for selecting counting rules in sample surveys.

2. COUNTING RULES

Counting rules specify the conditions for linking enumeration units to population elements. It is understood that elements are eligible and required to be reported by the enumeration units to which they are linked by the rule and that the elements are ineligible to be reported by any other units. Conventional counting rules have the property of uniquely linking every population element to one and only one enumeration unit. For instance, the de jure residence rule is a conventional counting rule in surveys in which households are the enumeration units and persons are the population elements. The rule links every person to one and only one household, namely his de jure residence.

Conventional rules are appealing because they assure that every element is eligible to be counted once and only once. However, as we shall see, survey estimates based on conventional rules are sometimes subject to intolerably large sampling and measurement errors. Consequently, we have been investigating the statistical properties of estimators based on multiplicity rules that would make them subject to smaller errors than conventional rules. Multiplicity rules specify conditions that permit more than one enumeration unit to be linked to the same elements.

Of particular interest is a class of multiplicity rules which have the property of supplementing the condition of a conventional rule with other conditions for linking elements to enumeration units. Multiplicity rules of this type have two desirable properties. First, they assure that every element is linked to at least one enumeration unit. Second, they permit the survey to produce several sets of estimates--one set based on the conventional rule and each of the other sets based on a multiplicity rule that incorporates the condition of the conventional rule. In this manner, the estimate based on the conventional rule is preserved. For instance, the rule that links persons to their own residence as well as the residences of their siblings and children would produce four sets of estimates based on the following rules:

(1) persons are linked to their own residence;

(2) persons are linked to their own residence and to the residences of their siblings;

(3) persons are linked to their own residence and to the residences of their children;

(4) persons are linked to their own residence and to the residences of their siblings and children

The first is a conventional rule. The others are multiplicity rules.

Table 2 presents conventional and multiplicity rules which have been reported in the literature [7], [2], [3], [9] for each of the illustrative surveys listed in Table 1. It will be noted that in each example, the multiplicity rule supplements the condition specified by the conventional rule. Table 2 also compares the population elements that are eligible to be counted at an enumeration unit in compliance with the conventional and the multiplicity rule.

3. COUNTING RULE ESTIMATORS

In surveys based on conventional rules, all elements have the same probability of being selected in the sample but in surveys based on multiplicity rules they do not. Unbiased estimators for multiplicity rules adjust for the different probabilities by appropriately weighting the elements. Several kinds of unbiased estimators have been proposed [1], [8], each estimator being based on a different system of weights. Some estimators require matching the elements enumerated in the survey to eliminate duplicate reports. One unbiased estimator [11] that does not eliminate duplicate reports, we shall refer below to as the multiplicity estimator.

The multiplicity estimator assigns a weight to every element everytime it is enumerated. It assigns a weight of unity to elements that are linked to a single enumeration unit, and to elements linked to multiple units, it assigns a weight, greater than zero and less than unity, everytime they are enumerated. The multiplicity estimator is unbiased if the sum of the weights assigned to the multiple linked elements is equal to unity. The multiple linked element may be assigned the same or a different weight each time it is enumerated. If it is assigned the same weight each time, its weight is the inverse of the number of units it is linked to.

Table 1. Illustrative Surveys in Which Enumeration Units are not the Same as Population Elements

Survey	Enumeration Unit	Population Element	Variate
1	Housing unit	Decedent	Hospital utilization in last year of life
2	Areal segment	Farm	Number of farms, and acres of farmland
3	Medical source	Patient with a disease	Medical costs of treating disease
4	Line of text of a statistical report	Statistical statement	Statements failing statistical standards

Table 2. Conventional and Multiplicity Rules in Illustrative Surveys

Survey	Rule		Elements Counted at an Enumeration Unit	
	Conventional	Multiplicity	Conventional Rule	Multiplicity Rule
1	Decedents are linked to their former *de jure* residence.	Decedents are linked to their former *de jure* residence and to the *de jure* residences of surviving relatives.	Deaths of persons who resided in the housing unit	Deaths of persons who resided in the housing unit or whose relatives reside in the unit
2	Farms are linked to the segment in which the farm headquarters are located.	Farms are linked to all segments overlapped by the farms.	Farms whose headquarters are located within the segment's boundaries	Farms entirely or partially overlapping the segment's boundaries
3	Patients are linked to the medical source with primary responsibility for treating their disease.	Patients are linked to all medical sources who treated them for the disease.	Patients for whom the medical source had primary responsibility for treatment	Patients treated by the medical source for the disease
4	Statements are linked to the line of text which contains the first word of the statement.	Statements are linked to all lines of text overlapped by the statement.	Statements starting on the line of text	Statements overlapping the line of text

In a formal sense, the estimator based on the conventional rule is a special case of the multiplicity estimator in which each of the elements is assigned a weight of unity. However, there is an important difference between the conventional and multiplicity estimators. The conventional estimator presumes that the weight assigned each element is equal to unity because by definition every element is uniquely linked by the conventional rule to one enumeration unit. On the other hand, the weights assigned elements by the multiplicity estimator are rarely known beforehand and usually require collecting ancillary information in the survey.

The procedure for determining the weights that are assigned to elements and the source of the ancillary information needed to calculate the weights varies for surveys based on multiplicity rules. For instance, Table 3 below presents this information for the surveys that are based on the multiplicity rules described in Table 2.

4. COUNTING RULE STRATEGY

The objective of the counting rule strategy is to select the optimum rule for linking elements to enumeration units and to assign the optimum sets of weights to enumeration units that are linked to the same elements. The rule and weights are considered to be optimum in the sense that they minimize the combined effect of sampling and measurement errors on the survey estimates subject to the cost constraints of conducting the survey. Although we have focused on surveys in which the population elements are different from the enumeration units, it is noteworthy that the counting rule strategy is also applicable to surveys in which the population elements are the enumeration units.

In this section, we describe some conditions which may make multiplicity rules preferable to conventional rules. We consider in turn, the effect of counting rules on sampling errors, on measurement errors, and on the costs of conducting the survey.

Sampling Errors

Since the counting rule specifies the conditions for linking population elements to enumeration units, the distribution of the population elements among the enumeration units is a function of the rule. Changing the counting rule modifies the population distribution and thereby alters the sampling distribution and the sampling variance of the survey estimate. It would be a mistake to conclude, however, that the estimators based on multiplicity rules are necessarily subject to smaller sampling variance than estimators based on conventional rules. They are not, except under special conditions such as when the multiplicity rule links at most one element to an enumeration unit. Nevertheless, when the population distribution based on the conventional rule is highly skewed, multiplicity rules offer a strategy for redistributing the elements among the enumeration units and thereby decreasing the variance of the population distribution. Two conditions contributing to the skewness of the population distribution based on the conventional rule are: (1) few enumeration units are linked to population elements and (2) the distribution of the variate among the population elements is highly skewed. When either or both of these conditions exist, multiplicity rules may also exist which would have the effect of redistributing the weighted elements among the enumeration units so that the resulting distribution would be less skewed.

Table 3. Weights Used in Illustrative Surveys Based on Multiplicity Rules

Survey	Weight	Source of Information about Weight
1	Inverse of the number of housing units linked to the death.	Unit where the death is enumerated reports the number of different housing units which are either residences of surviving relatives or the former residence of the decedent.
2	Fraction of the farm area overlapped by the boundaries of the segment	Areal maps
3	Inverse of the number of medical sources treating the patient	Source where patient is enumerated identifies the referral and referring sources which are followed up in the survey
4	Inverse of the number of lines overlapped by the statistical statement	Clerk counts the number of lines of text that are overlapped by the statement

A multiplicity rule that links every enumeration unit to every population element would not be subject to any variance. Unfortunately, this rule would be subject to unacceptably large measurement errors in virtually all surveys. Although the rule is impractical, it serves to illustrate the point that consideration must be given to the joint effect of sampling and measurement errors in selecting a counting rule.

Two kinds of measurement errors associated with counting rules are coverage errors and response errors.

Coverage Errors

Coverage errors occur when enumeration units either fail to report population elements to which they are linked or erroneously report elements to which they are not linked by the counting rule. In most surveys underreporting appears to be a far more serious problem than overreporting.

Multiplicity rules are often apropos when it is difficult to implement conventional rules.

If some characteristic of the population leads to multiple links of the population elements with enumeration units, it is natural to think in terms of multiplicity rules. Thus, if patients have encounters with several physicians, it may be easier to implement a multiplicity rule that links the patients to all their medical sources than to implement a conventional rule that defines a condition that would uniquely link every patient to a single medical source.

If the attribute being measured in a household survey defines a population of persons that does not reside in the area covered by the survey, it is extremely difficult to implement conventional residence rules. Thus, it would be virtually impossible to collect emigration statistics in cross sectional surveys using de jure residence rules. Multiplicity rules based on consanguine relations are often more apropos than conventional rules in these surveys. For instance, the conventional rule which links the deceased person to his former residence is subject to gross underreporting of deaths in single retrospective surveys of population change because the death of an individual is often followed by the dissolution or emigration of the household. It appears [10] that a multiplicity rule that links decedents to the residences of surviving relatives misses fewer deaths than a conventional rule that links decedents to their former residence.

Multiplicity rules may sometimes be subject to smaller coverage errors than conventional rules in population surveys which collect sensitive information. For instance, a person with a stigmatized attribute, such as an alcoholic or a drug addict might be less likely to report himself than he would be to report his close friends or relatives living elsewhere who were drug addicts or alcoholics since he could report them anonymously without identifying them either by name or by their place of residence.

Response Errors

Response errors represent the invalid or incomplete reports by the enumeration units of the characteristics of elements to which they are linked by the counting rule. Sometimes one of the enumeration units is "best" in the sense of having more complete and accurate information about the variate being measured than any other unit that is linked to the same element. The "best" unit may or may not be the unit linked to the element by the conventional rule. The surviving spouse, for example, might be a best informant for the decedent, whether or not she lived with the decedent when he died. If every unit linked to an element by the multiplicity rule knows the identity and location of the "best" unit, it might be cost effective to introduce a supplementary strategy to determine the value of the variate for the elements which are enumerated at units which are not the "best" units. The strategy would involve reenumerating these elements at their "best" enumeration units.

Survey Costs

A multiplicity rule is not necessarily more efficient than the conventional rule it incorporates even though for a fixed size sample of enumeration units, it is subject to less combined sampling error and measurement error. This is so because the survey based on a multiplicity rule is more costly. There are two reasons why the survey costs are greater for a multiplicity than for a conventional rule: First, the enumeration unit cost is greater because the average number of elements linked to an enumeration unit is greater for the multiplicity rule than for the conventional rule. Second, the estimator based on the multiplicity rule, and its variance require ancillary information to determine the weights assigned to elements which is not needed by the conventional estimator. Since this ancillary information is frequently collected in the survey, it adds to the survey cost.

5. SUMMARY

The counting rule strategy treats the counting rule as a design factor for improving the efficiency of the survey. The strategy selects the optimum counting rule for linking population elements to enumeration units and assigns the optimum sets of weights to the elements that are linked to more than one enumeration unit. We have indicated the types of survey conditions in which the sampling errors and measurement errors of the estimate based on a conventional rule may be reduced by selecting a multiplicity rule. However, the cost of conducting a survey for a fixed size sample of enumeration units is invariably larger for a multiplicity rule than for a conventional rule. To apply the counting rule strategy, therefore, requires information that compares the cost coefficients as well as the error components associated with different counting rules.

References

[1] Birnbaum, Z. W. and Sirken, M. G., "Design of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates," Vital and Health Statistics, P.H.S. Publication No. 1000-Series 2-No. 11, Public Health Service, Washington, D.C.: U.S. Government Printing Office, 1965, pp. 1-8.

[2] Hendricks, W. A., Searls, D. T., and Horvitz, D. G., "A Comparison of Three Rules for Associating Farms and Farmland with Sample Area Segments in Agricultural Surveys," Estimation of Areas in Agricultural Statistics, Food and Agriculture Organization of the United Nations, Rome, 1965, pp. 191-198.

[3] Kramm, E. R., Crane, M. M., Sirken, M. G. and Brown, M. L., "A Cystic Fibrosis Pilot Survey in Three New England States," American Journal of Public Health, December 1962, pp. 2041-2057.

[4] Sirken, M. G., "Household Surveys with Multiplicity," Journal of the American Statistical Association, 65 (March 1970), pp. 257-266.

[5] ___________, "Stratified Sample Surveys with Multiplicity," Journal of the American Statistical Association, 67 (March 1972), pp. 224-227.

[6] ___________, "Variance Components of Multiplicity Estimators," Biometrics, 28 (September 1972), pp. 869-873.

[7] ___________, "Design of Household Sample Surveys to Test Death Registration Completeness," Demography, 10 (August 1973), pp. 469-478.

[8] ___________, "Effect of Counting Rules on Sampling Errors and Survey Costs," unpublished manuscript.

[9] ___________, and Levy, P. S., "Multiplicity Estimation of Proportions Based on Ratios of Random Variables," Journal of the American Statistical Association, 69 (March 1974), pp. 68-73.

[10] ___________, and Royston, P. N., "Underreporting of Births and Deaths in Household Surveys of Population Change," Proceedings of the Social Statistics Section, American Statistical Association, (1973), pp. 412-415.

DISCUSSION
Joseph Waksberg, Westat, Inc.

I. Use of Loss Functions to Determine Sample Size in the Social Security Administration

This paper correctly points out that sampling textbooks and sampling theory usually start out in the middle of a problem, that is they assume there is advance knowledge of the desired variances for the sample survey under consideration. This is, of course, one of the early issues that arise in developing plans for surveys, and in my experience it is one of the most difficult ones to resolve in a satisfactory manner; mostly because the consequences of errors generally have not been quantified. Similarly, in teaching courses in sampling theory and methods, I have always found it difficult to give students a "feel" for the size of sampling errors that might be desirable under different circumstances.

I am therefore delighted to see examples in which the problem is approached rationally, and in a way that permits the power of mathematics to be used to derive the variances that appear appropriate for specific surveys. I think it would be useful to establish a file of similar cases, both to have as examples in teaching and to help a consultant direct the thinking of survey sponsors, who are generally perplexed about how to come to grips with the problem of the precision needed for their studies.

It would probably be naive to expect that the planning of all (or perhaps even most) surveys can be approached by establishing a loss function and comparing the loss with the cost of conducting surveys of various sizes. Typically, surveys are conducted to add to general knowledge, or to permit an administrator to make better decisions by use of the results, with or without other, external information. It's unlikely that in those circumstances anyone can even guess at the monetary value of the improvement in decisions, or for that matter even whether there are any improvements as a result of the surveys. However, I believe examples like the ones in the paper can help direct thinking in even these cases.

I suspect that considerable more work needs to be done on the construction of appropriate loss functions. Two aspects of the function used bothered me. First, expressing the loss function as $\Theta = E(L) + C$, implies that the cost of the survey is of the same level of importance as the same amount of money distributed in error. This implies an indifference to how an equivalent amount of money is spent. If I were administering a program and was told I could spend an additional $1,000,000 on a survey to get zero error, or risk an error of $1,000,000, I would prefer to skip the survey and just add the $1,000,000 to the money allocated.It would be interesting to give a higher weight to the cost of the survey, and see the effect on the results.

The other aspect is the one discussed in the paper, whether the loss should ignore the direction of error or just consider the situation from one of the parties concerned, and minimize errors for that party.This principle appears troublesome. If carried to a logical conclusion, it says that in the program the Federal Government should strive to make all its errors in one direction -- at the expense of the States. Even at the risk of introducing more arbitrariness in the construction of the loss function, I wonder if including the effect of errors in both directions would not be more realistic. The loss functions are bound to have a certain element of subjectivity but hopefully they can be expressed so that most analysts agree that they are reasonable.

I have one final comment about other areas in Government programs in which it would be useful to apply similar types of analyses. In the past few years a number of very large revenue sharing programs have been developed, both general and special-purpose. They involve the transfer of billions of dollars a year from the Federal Government to States and localities. They have certain characteristics in common; they rely on allocation formulas which are based on statistical data, and there has been a general reluctance to spend much money to produce current, reliable estimates of the parameters required for the formulas. The staffs of the agencies involved would perform an important public service if they established reasonable loss functions and examined the implications for survey operations, in a way similar to that done by Social Security Administration.

II. Effect of Counting Rules on Sampling Errors and Costs

Dr. Sirken's paper on counting rules and the papers published earlier on the same subject have, I think, served a useful purpose in indicating a direction in which greater flexibility can be exercised in survey planning, with a potential improvement in efficiency. I was glad to see an example which provided guidance on the conditions for which multiplicity rules would be expected to be more efficient than conventional enumeration rules. The published papers had left me with a feeling that it would be extremely difficult to carry out the theory in practice,due to inability to estimate the necessary parameters. The example in this paper helped in this,although I still am not sure how I would tackle specific problems. Some more general guidelines on this would be helpful.

In the absence of such guidelines, or of a general body of knowledge that might become available as more surveys use multiplicity rules, I can see two important areas that lend themselves to the use of multiplicity. The first is in the estimation of a rare item in the population. Some of the earlier work by Sirken and others, using this technique, involved estimating the prevalence of certain rare health conditions. Similarly, one might think of studies aimed at estimating the characteristics of specific ethnic groups, persons with specialized educational background or training, or other low-frequency elements of the population for which establishment of a separate sampling frame is not possible. Such surveys are generally considered as comprising a screening operation to identify the subset of a sample that belongs to the group being studied, and an enumeration phase. Multiplicity rules should have a useful effect on the screening operation, certainly in cases where the enumeration can be performed at the initial sample unit, but even where additional travel is necessary for interviewing purposes. Obviously, this can only be done where there is a reasonable assurance that response errors would be kept under control.

The second area is where multiplicity rules serve to reduce response errors. Dr. Sirken mentioned one example in which this happens, estimating births and deaths. Another example involving work done at the Census Bureau, may be instructive. This was developed with a simpler concept in mind, but it can be viewed as an example of a counting rule application.

The example parallels the problem of measuring births and deaths, but with application to housing. As part of the past few Censuses of Housing, there have been sample surveys designed to measure, among other things, the change in housing arising from mergers,in which two or more apartments in a building are combined to form a single unit, or from growth when one housing unit is subdivided into several units. Associating such changes with unique housing units, although theoretically possible, is difficult for interviewers to perform properly. It is much more accurate to view the building as a whole and compare the number of units now present, with the number at an earlier point in time. The Census practice has thus been to select a sample of housing units,and essentially to perform a complete enumeration on all buildings in which the sample units are located. In estimating universe totals, the probabilities of selection are, of course, taken into account.

This can be considered an example of a multiplicity rule. It is a rather simple type since each enumeration unit is linked to only one network. At the Census Bureau, this procedure has been viewed as an example of PPS selection rather than as a case of multiplicity, but in this simple case, the two merge.

III. Optima and Proxima in Linear Sample Designs

There isn't very much that I can add to Dr. Kish's paper. He has shown, in a rather elegant manner, that a common approach is possible for analysis of what I had always treated as somewhat different problems.

Based on my own experiences, I suspect that the paper will be of greater interest to teachers of courses in sampling methods than to practicing statisticians. Working with problems of sample design, it did not take very long for me to realize that it is rare to get good estimates of either unit costs or variance, and possibly even more important that the models used for cost functions are crude approximations to reality. I imagine that most statisticians quickly become accustomed to dealing with approximations and don't worry too much about it.

I've found it more difficult to get a "feel" for this across to students. It's obviously important for them to learn about optimization and the applicable formulae. However, they also need to find out that in the real world, crude guesses about the parameters are frequently necessary. I have never been satisfied with the vague words I have had to use in indicating that considerable variation from the optimum allocation is usually possible without great increases in variances. There are few actual examples in the literature. The Hansen-Hurwitz-Madow textbook does have several tables for specific designs indicating the range of variation in parameters that can exist without important additions to variance. I can think of very few other examples.

A comprehensive discussion of this subject is thus very welcome.

THE ROLE OF THE SOCIAL SECURITY NUMBER IN MATCHING ADMINISTRATIVE AND SURVEY RECORDS

JOSEPH STEINBERG
Chairman
Bureau of Labor Statistics

GENERAL INTRODUCTION TO THE SESSION ON THE ROLE OF THE SOCIAL SECURITY NUMBER IN MATCHING ADMINISTRATIVE AND SURVEY RECORDS

Because of the interrelatedness of the papers given at this session, the authors felt that readers of the proceedings would find it easier to follow the presentations if an introduction were provided first. To this end, a number of the remarks made by the individual speakers (including remarks made by the session chairman, Joseph Steinberg) have been brought together here.

GENERAL BACKGROUND

The Social Security Administration (SSA) and the Census Bureau have, for quite some time, engaged in joint efforts to improve the quality of statistical output in the areas of income distribution and redistribution. One of these studies, which is currently underway, involves matching information on the March 1973 Current Population Survey (CPS) with earnings and benefit data from Social Security records. The work being reported on at the session was done in connection with this "1973 Match Project."

1973 Match Project.--The 1973 Match Project differs in several respects from earlier linkages between Census Bureau surveys and Social Security administrative information. Three of the major differences are:

1. The sample involved, consisting, as it does, of over 100,000 individuals 14 years or older, is many times larger than that used in any previous joint project. (Matching studies made to evaluate decennial census data [e.g., 60,104,139] have been on the order of one-fifth as large or less. 1/Previous linkages completed between CPS and SSA [e.g., 102] have been based on samples only about one-sixth as large.)

2. The process used to bring together the data from the various sources is more automated than formerly. This was one reason a larger sample of cases could be matched. Also, the fact that the major components to be linked are machine-readable promises to make it possible to publish at least some of the principal findings from the project in 1975. (However, as the second paper at the session makes clear, there are still certain manual-clerical steps which are essential.)

3. Prior joint Census Bureau-SSA "exact" match studies have focused principally on the analysis of response, nonresponse, and coverage errors. In the 1973 work, considerable emphasis is also being placed on obtaining a microdata file in which CPS reporting has been "corrected" or, more properly, calibrated. What is planned are not only modifications to the survey income amounts made possible by the presence of a comparable administrative figure obtained by means of an exact match (calibration), but also adjustments will be introduced by using synthetic or "statistically" matched information. (In the past, the statistical matching work done with the CPS has essentially been conducted independently of exact matching efforts. 2/)

The subject matter content of the 1973 Match Project is quite similar to that in earlier CPS-SSA linkages. The items being extracted from SSA's benefit and earnings files, for example, are about the same as in the efforts directed by Joseph Steinberg [e.g., 118,121] when he was at Social Security. Also, just as in some of the previous studies, information from income tax returns will eventually be included on the files, making a three-way linkage. The Internal Revenue Service (IRS) data that will be available, 3/ however, is far more limited than in the past (so limited, in fact, that it will be necessary to supplement it with IRS data introduced by using statistical matching).

Confidentiality Arrangements.--One of the things that the 1973 Match Project has in common with earlier linkage efforts is the great care that is being taken to insure the confidentiality of the shared information. The laws and regulations under which the three agencies operate impose very definite restrictions on such exchanges, and special procedures have been followed throughout so as to adhere to these provisions.

Information on the confidentiality requirements which governed prior linkage projects involving the Current Population Survey can be found in [102] and [119], which were available as handouts at the session. 4/ The 1973 work has operated under procedures which are at least as stringent as those imposed in the past. This is particularly the case for the IRS data. The details of the 1973 arrangements were also available as a session handout and are incorporated in Roger Herriot's discussion comments which appear at the end of the proceedings for this session.

SESSION FOCUS AND ORGANIZATION

The Social Statistics Section has had over 20 invited and contributed papers5/ the last decade or so devoted, either in whole or in part, to matching and data linkage. This, however, is the first time an entire session has been given to a matching study in which the primary piece of identifying information was the social security number (SSN).

Matching with the SSN.--The problems which arise when using the SSN to link Current Population Survey interview schedules to Social Security records differ in degree, but not in kind, from the problems other "matchmakers" have had. The three major factors to consider still are: (1) reporting differences in the identifying information being used to bring the files together, (2) omissions or incompleteness in the identifiers, and, finally, (3) nonuniqueness of identifiers.

In the 1973 study, as in prior CPS-SSA linkages, the chief difficulty that had to be faced was incompleteness in the identifying information. The first two papers at this session describe the situation that existed in this regard and what has been done about it so far. Next in importance were reporting errors in the social security number or in the other identifiers (name and date of birth, etc.). These are the subject of the third paper at the session.

The problem of nonuniqueness also exists with the social security number. It is estimated that more than six million people have two or more SSN's. In well over half of these cases, SSA has cross-referenced the numbers so the multiple reports for an individual can be brought together rather routinely. For most of the remaining persons, the numbers were issued in the early days of the program and probably are no longer active.6/ There are also some instances in which more than one person uses the same SSN. Fortunately, however, this situation is quite rare. 7/ Thus, compared to reporting errors and omissions in the identifying information, the nonuniqueness of the SSN's poses a relatively minor problem for the 1973 Match Project. In any event, the papers given at the session do not deal directly with the procedures that will be followed to mitigate its effects.

Nature of Papers.--The papers are reports on work in progress. For the most part, they are descriptive and nontheoretical. No attempt has been made in the presentations to set forth in a systematic way all of the procedures that have been followed in the 1973 Match Project. Just some of the important highlights which were felt to be of general interest have been given.

At the session itself, extensive tabular material supporting the results in the papers was provided as a handout. For reasons of space, these tables cannot be included here; however, they are available on request. 8/ The papers, as shown in these proceedings, follow the remarks of the participants quite closely, except for comments which appear in footnotes. The footnotes have been used to introduce parenthetical information which, in many instances, was not part of the actual presentations, to clarify points about which questions were raised during the general discussion which followed the talks, and to cite the relevant literature when this could not be conveniently done in any other way.

ACKNOWLEDGEMENTS

The authors would like to conclude this introduction by acknowledging the extensive assistance given them by Fritz Scheuren, who organized the session, and H. Lock Oh, who prepared the tabular material. Denton Vaughan oversaw the production of the copy for these proceedings on the IBM/360 Administrative Terminal System (ATS). Lois Gale, Alda Seubert, Catherine Murphy, and Tillie Mazor adapted their typing skills to specialized ATS procedures necessary for putting the material in galley form. Shirley Carter provided valuable technical assistance in preparation of the charts.

The authors would also like to thank Joseph Steinberg for his thoughtful participation as chairman and Roger Herriot, who is in charge of the Census side of the 1973 Match Project, for his role as discussant.

FOOTNOTES

1/ The citations given in square brackets here and elsewhere refer to references listed in a bibliography on matching studies which was handed out at the session and which is included in these Proceedings.

2/ For a brief historical sketch of the statistical (and exact) matches which have been done with the CPS , see Benjamin Okner, "Data matching and merging: an overview," in the Annals of Economic and Social Measurement, vol. 3, 1974, pp. 347-352.

3/ For this project, IRS made available to the Census Bureau magnetic tape abstracts of limited income information from tax returns, subject to the confidentiality arrangements discussed in this introduction and in remarks of the discussant. The dollar items abstracted consisted of total income, salaries and wages, dividends, and interest. Codes were also included to indicate the type of return filed (e.g., joint, surviving spouse, etc.), the types of schedules used (e.g., Schedules C, D, F, etc.), and the number of exemptions claimed.

4/ Also handed out at the session were two other reports which deal more generally with various aspects of Social Security's statistical research work. These were a paper by Joseph Steinberg and Heyman Cooper, "Social Security statistical data, social science research, and confidentiality," which appeared in the 1967 Social Security Bulletin (pp. 2-14) and a 1973 SSA publication entitled Some Statistical Research Resources Available at the Social Security Administration.

For an historical summary of the Census's general provisions with regard to confidentiality, see the address given by the Bureau's current director, Vincent Barabba, which appears elsewhere in these Proceedings. See also, Robert Davis, "Confidentiality and the Census, 1790-1929," in Records, Computers and the Rights of Citizens, Report of the Secretary's Advisory Committee on Automated Personal Data Systems, U.S. Department of Health, Education and Welfare, 1973.

5/ Of particular note is the 1963 session, "Matching of medical, social and economic records for research purposes," [39,95,107]; and, also, the 1965 session entitled, "Matching of Census and vital records in social and health research: problems and results," [58,91,94,98, 112,113].

6/ It is possible, by request, and for good cause, to have more than one number; however, individuals may forget that they have a number or forget what it is, and apply again. When such inadvertent multiple issuances occur, routine administrative processes usually detect them eventually.

7/ The most important cases where more than one person is using the same SSN arise because social security numbers have been employed on occasion by advertisers in promotional schemes. Perhaps the best known such instance is the number 078-05-1120. It first appeared on a sample social security number card contained in wallets sold nationwide in 1938. Many people who purchased the wallets assumed the number to be their own. It was subsequently reported thousands of times on employers' quarterly reports; 1943 was the high year, with almost 6,000 wage earners listed as owning the number. Even today the number is still being reported at least 10 times a quarter.

8/ To obtain the tabulations, write to Dr. Benjamin Bridges, Chief, Long Range Research Branch Division of Economic and Long-Range Studies, Office of Research and Statistics, Social Security Administration, 1875 Connecticut Avenue, Washington, D.C. 20009.

CURRENT POPULATION SURVEY REPORTING OF SOCIAL SECURITY NUMBERS

Linda Vogel and Terry Coble
Social Security Administration

The Social Security Administration and the Census Bureau are engaged in a project in which SSA earnings and benefit records are being matched to Current Population Survey data. (See the introductory comments to this session.) The goal of this effort is an improved data base for use in studies on income distribution and redistribution. In order to locate SSA earnings and benefit data for an individual, one must have, or be able to determine, his social security number (SSN). Of primary importance in the 1973 Match Project, therefore, was the extent of missing SSN's in the Current Population Survey.

This paper examines the reporting and non-reporting of SSN's from various vantage points. In section 1, a description is given of how the SSN is collected in the CPS and the possible sources of error that can be introduced. Section 2 discusses the overall response rate to the SSN question and compares our results to previous studies. The remaining sections analyze some of the characteristics of the individuals in the survey which may influence the likelihood of an SSN being reported (section 3); examine the characteristics of the household respondents which affect whether a number is reported (section 4); and present some other aspects of SSN reporting, such as differences by data collection centers and by rotation groups (section 5).

1. COLLECTION OF SOCIAL SECURITY NUMBERS IN THE CURRENT POPULATION SURVEY

Background.--The structure of the CPS since the early 1950's has been such that each sample household is eligible for interview on eight separate occasions: four consecutive months one year, followed by the same four calendar months a year later, with an eight month "rest" in between. These interviews are staggered so that for any given month's survey--March 1973, for instance--one-eighth of the cases are being enumerated for the first time, one-eighth for the second, and so on. Thus, each month's sample actually consists of eight subsamples, or "rotation groups," which can be distinguished by the number of times the addresses they contain have been eligible for interview. 1/

A "control card" is prepared for a household when it first enters the survey. The control card is a basic repository of data and lists for each household member such items as name, date of birth, age, race, sex, etc. In 1963, in connection with the first SSA-CPS linkage project [102, 118, 119, 121, 147, 148], the control card was revised to obtain social security numbers for persons 14 years or older. Since then, SSN information has been routinely requested just of rotation groups which enter the sample during the months of December, January, February, or March. The reason for this is that interest in linking SSA and CPS data, as mentioned in the overall introduction to this session, has centered on comparing income responses in the CPS with the corresponding SSA administrative figures. 2/ Rotation groups entering in these four months are the only ones in the survey in March when the supplementary income questions are asked.

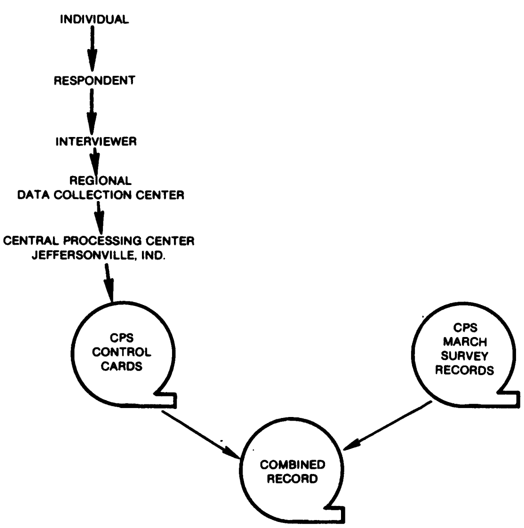
The exact phrasing of the SSN question on the control card is, "What is . . . .'s Social Security (Railroad Retirement) number?" Interviewers are instructed to tell respondents, if asked, that the SSN's are necessary to help evaluate the accuracy, consistency, and comparability of some of the statistics produced from the CPS; that this evaluation is conducted using various techniques, one of which is through the use of governmental administrative records; and that these evaluation programs are carried out by sworn employees of the Census Bureau under the strictest controls, to protect the confidentiality of the information secured. As an example, respondents are told about the possibility of matching the survey data to the records of the Social Security Administration.

Usually, in the CPS, one member of a household responds for the entire household. If the household respondent knows, or thinks he knows, the SSN's for the other members, these are recorded by the interviewer on the initial visit. If the respondent reports that a person has no number, "none" is recorded. The respondent may, at this time, of course, refuse to provide numbers. 3/ When the respondent is unaware of the SSN's of other household members, a form is left to be filled in and mailed to the census offices. If any numbers are still missing on the second or later months of interview, they are again requested.

SSN Collection Process.--The major steps in the collection of social security numbers are set forth in figure 1. At each of these steps, errors of response or nonresponse can be introduced which would affect the quality and completeness of the SSN reporting. The individual, himself, is obviously a possible source of error. He may not know or may misremember his SSN. Another source of error can arise if the individual does not respond for himself, but the information is obtained by proxy. This was the case for approximately 57 percent of the individuals in the survey who did not answer the SSN question directly--it was answered for them by someone else in the household. Despite careful training and control, the CPS interviewers may also have introduced some errors, either by failing to ask the question at all or by transcription mistakes and the like. Differences among the Census Bureau's regional data collection centers in the emphasis placed on the SSN question are, also, a possible cause for concern.

Up to this point, the steps, as shown in figure 1, to obtain SSN data in the CPS do

Figure 1.—PROCESSING FLOW OF SSN'S



not differ very much from the steps that would be taken for any other item included in the survey. Unlike much of the other information on the control cards, however, SSN's are <u>not</u> routinely transcribed onto the monthly survey questionnaires. For this project, therefore, the control cards had to be specially encoded into machine-readable form (at the Bureau's Central Processing Center in Jeffersonville, Indiana). They were subsequently matched to the tape files of the March 1973 interview schedules in order to have both the CPS income data and the SSN information available for an individual on one combined record. These latter steps are, of course, also subject to clerical and other errors. For example, lost control cards would result in missing SSN's.

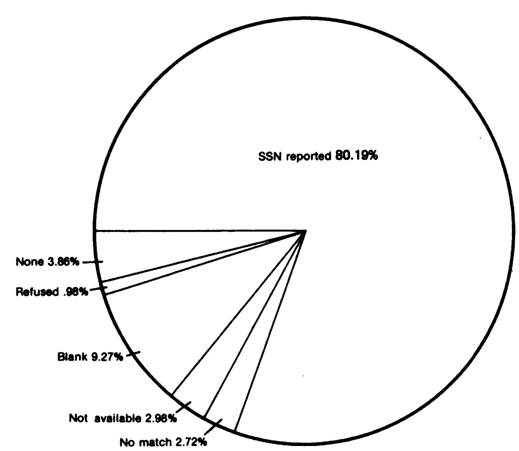
In this paper, as was previously stated, we will be confining our attention to comparisons between persons who have had an SSN recorded for them in the CPS and those who have not. The question of the correctness of the SSN responses given is not taken up here. (The steps taken to validate numbers and to search for missing and misreported SSN's are covered in the subsequent papers at this session.)

For our purposes, all numerical responses to the SSN question, whether valid or invalid, have been treated as equivalent. Since essentially the entire U. S. population 17 years or older has an SSN, 4/ instances in which no SSN was reported in the CPS for such persons are, for the most part, cases in which an error of some sort was made. While we cannot always relate our hypotheses regarding the completeness of the SSN responses to specific steps listed in figure 1, nonetheless, the reader may find it useful to keep the various steps in mind during the discussion which follows.

2. OVERALL RESULTS AND COMPARISONS TO OTHER STUDIES

Four out of five of the over 100,000 CPS persons 14 years or older in households interviewed in March 1973 reported a social security number (see figure 2). Of

Figure 2.—OVERALL RESPONSE TO SSN QUESTION



the remaining 20 percent who were not shown to have reported SSN's, the response "none" was recorded for 3.9 percent, while only one percent were listed as refusals. Blank responses, possibly due to the interviewer not emphasizing or asking the SSN question at all, totaled approximately 9.3 percent. Answers of "not available" are recorded for almost three percent of the cases. This situation might arise if the respondent failed to mail in the form left by the interviewer for numbers not obtained at the time of interview or if SSN's on the returned forms were never incorporated on the control card. Finally, missing numbers resulted for about 2.7

percent of the March supplement records because no match could be made to the source of the SSN, the control card.

The percentage of social security numbers reported in the CPS, about 80.2 percent, represents a successful effort on the part of the Census Bureau in collecting the numbers and is an improvement over previous comparable projects. In the 1963 Pilot Link Study, 72 percent of the 15,533 persons 14 years or older initially responded with an SSN [147]. A similar 1966-67 Census Bureau study by Ono, Patterson, and Weitzman [93] found that approximately 76 percent of the 4,500 persons 14 years or older comprising the sample reported an SSN.

While it is enlightening to compare the positive response rate of the Match project with previous studies, it is necessary to remember that the present study should yield an increase in the percentage of SSN's reported for two reasons. First, the SSN is more widely used now than ever before, for such diverse purposes as automobile driver's licenses and student identification, etc. One would expect, therefore, that more people are able to report their social security number than in the past. Secondly, a larger proportion of the U. S. population now has SSN's. This increase is especially noticeable among women, as this segment of the population has joined the labor force in greater numbers. Although for the above-mentioned reasons we would expect a higher SSN reporting rate in the current study, there is, on the other hand, another factor involved that should be taken into account; namely, the growing tendency of individuals to be reluctant to divulge what they consider to be confidential information and to regard such questions as an invasion of privacy. Fortunately, judging from the low refusal rate, this still does not seem to be a significant factor in Census Bureau conducted surveys involving the social security number.

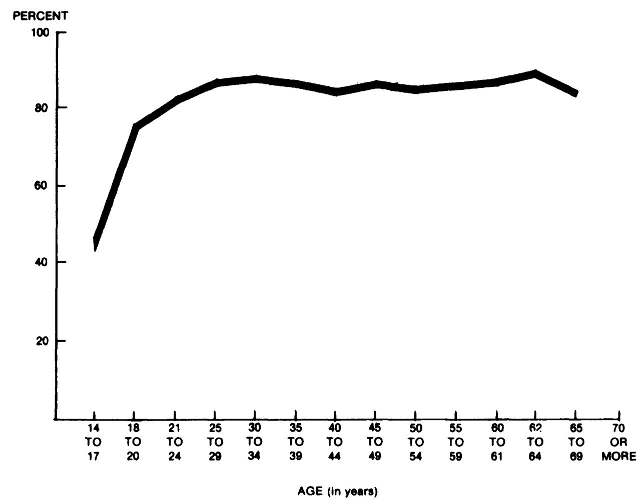
3. SSN REPORTING BY AGE, RACE, SEX, AND RESPONDENT STATUS

Individual's Age.--Age is found to be, by far, the most significant characteristic in determining whether an SSN is reported for an individual. 5/ Persons 14 to 20 years old appear to be much less likely to show a positive SSN response; this was especially true of those in the 14 to 17 year-old category, where only 46.0 percent had an SSN reported (see figure 3).

Age also has a strong influence on whether or not an individual shows up in the "none" or "blank" groups, but it seems to have little impact on the proportion of "no matches" or answers of "not available." Approximately 23.0 percent of all 14 to 17 year-olds were shown to have an answer of "none" for the SSN question, compared to 1.5 percent for the other age groups. The age effect on the number of blank entries is also striking; for persons younger than 21 years of age, almost 20 percent were found to have blanks, while the percentage drops to seven percent for persons 21 or older.

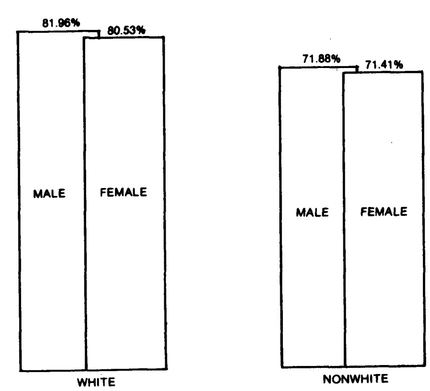
Refusals, though a small proportion of the total sample, behave in an interesting way. Those over the age of 35 are more likely to be listed as refusals (1.3 percent) than are persons in the younger age groups (0.7 percent).

Figure 3.—SSN REPORTING BY AGE



Individual's Race and Sex.--Proportionately more whites than nonwhites have an SSN reported (81.2 percent for whites versus 71.6 percent for nonwhites). Differences by sex are not nearly as large, however. Males are only slightly more apt to report an SSN (81.0 percent for males as opposed to 79.5 percent for females). Figure 4 clearly illustrates this marked disparity in reporting rates between the races and the only marginal difference between the sexes.

Figure 4.—SSN REPORTING BY RACE AND SEX



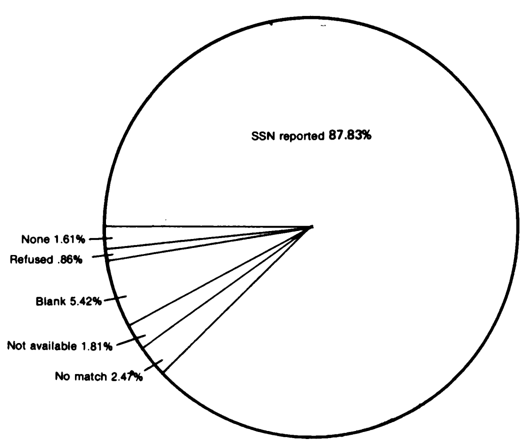
Both race and sex exert some influence on whether persons respond with the answer "none." As expected, females were slightly more likely than males to answer "none" (4.9 percent as opposed to 2.8 percent) and nonwhites were more likely than whites (5.9 percent compared to 2.6 percent). Race, but not sex, seems to determine, to some extent, whether the SSN question is left blank or filled in with "not available." Blank entries were approximately 13.0 percent for nonwhites and, for whites, only nine percent. The percentage of "not available" answers for nonwhites (5.5 percent) is about twice the number for whites (2.7 percent).

Differences in the age distributions of whites and nonwhites account for some of the reporting patterns that have been remarked on. It is fair to say, though, that, for much of the disparity, we must search for other causes.

Individual's Respondent Status.--Because one potential source of distortion in the reporting of SSN's does not exist for those individuals who responded directly to the interviewer's question, it is of interest to compare the response rates for all persons with those for the household respondents alone. 6/ We would, of course, expect respondents to be more complete and more accurate about themselves than about others, and, in fact, this seems to be the case.

The SSN reporting rate among respondents for each age-race-sex category is much higher than the reporting rate for all persons in that same category. For all age-race-sex groups combined, about 87.8 percent of the household respondents reported an SSN (see figure 5), as opposed to 80.2 percent for all persons. 7/

Figure 5.—RESPONSE TO SSN QUESTION FOR HOUSEHOLD RESPONDENTS ONLY



When comparing the response rates between the set of all persons and the subset of household respondents, it should be noted that the age distribution between the two groups differs radically. This is because of a reluctance by Census interviewers to accept younger members of households as household respondents. Slightly more than 18 percent of all persons 14 years or more in the CPS are 14 to 20 years old; however, less than four percent of the household respondents were from this age group. Even more extreme are the figures for persons 14 to 17 years old. They represent 11.1 percent of all the adults in the survey but are respondents only 0.8 percent of the time. Given this disparity in the age distribution, a sensible approach to comparing respondents and nonrespondents is to look at them by age, as in the following table:

Table 1.--Proportion Reporting An SSN

	Respondent	Nonrespondent
Total	87.8%	74.4%
14 to 17 year-olds	61.0%	45.5%
18 years or older	88.0%	81.1%

The improvement in reporting is pronounced for the 14 to 17 year-olds. One of the possible reasons for this may be that the older members of the household are not reliable sources of the SSN for the younger family members.

Replies of "none" occur about 1.6 percent of the time for respondents; the proportion for nonrespondents climbs to 5.6 percent. Virtually all of this difference is due, however, to differences between the age structure of respondents and nonrespondents, as can be seen from the following breakdown:

Table 2.--Proportion Reporting "NONE"

	Respondent	Nonrespondent
Total	1.6%	5.6%
14 to 17 year-olds	23.0%	23.2%
18 years or older	1.5%	1.5%

As we expected, the number of blank entries was lower among respondents than among nonrespondents (5.4 percent versus 12.2 percent). The younger groups had a large influence on this; the incidence of blanks among all persons 14 to 20 years of age is 19.8 percent, whereas the incidence of blanks among respondents of the same age is 9.1 percent.

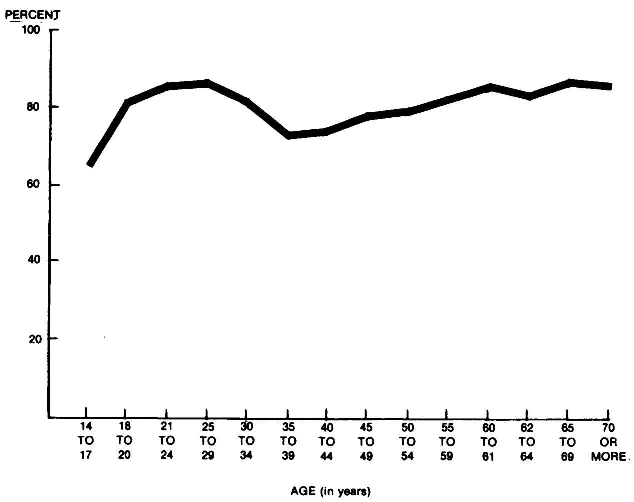
The percentage of answers of "not available" was also lower for respondents than for nonrespondents (1.8 percent as opposed to 3.9 percent). However, the incidence of refusals (0.9 percent for respondents compared to 1.1 percent for nonrespondents), and of nonmatches (2.5 percent versus 2.9 percent), do not represent significant differences.

4. HOUSEHOLD REPORTING BY THE CHARACTERISTICS OF THE HOUSEHOLD RESPONDENT

We have seen in section 3 that the respondent status was a strong determinant affecting the SSN reporting for individuals. In this section, we attempt to analyze how the SSN reporting of the entire household varies according to selected attributes of the respondent.

Household Reporting of SSN's by Respondent's Age.--Figure 6 shows a graph of SSN reporting for all persons based on the age of the respondent. As can be seen, the proportion of SSN's obtained for

Figure 6.—SSN REPORTING FOR ALL PERSONS BY RESPONDENT'S AGE



the household is noticeably smaller when the respondent is either 14 to 17 or 35 to 44 years old. The percentages of SSN's obtained are 65.4 percent and 73.8 percent when the respondent is 14 to 17 years old, and 35 to 44 years old, respectively, versus 82.5 percent for the other age groups. This variation is caused in part by the nature of the households for which these particular age classes were respondents; i.e., these households contained a higher proportion of individuals between the ages of 14 and 17, when they are less likely to have SSN's. In fact, when these age groups (14 to 17 and 35 to 44) are household respondents, the whole household shows a much higher incidence of answers of "none" (8.1 percent and 7.0 percent compared to 3.0 percent for the other age groups).

As mentioned before, the CPS interviewer is instructed to interview the household head or the spouse of the household head, if possible. Therefore, when a younger member of the household is chosen as a respondent, it is generally because older members are not at home at the time of the interview. However, additional household members may be present when the head of household is the respondent. This situation may account for the high incidence of "not available" responses for households with 14 to 17 year-old respondents (6.2 percent versus 2.9 percent for the other ages). Blank entries are also more prevalent among households having a respondent who is 14 to 17 years of age (15.0 percent as opposed to 9.2 percent overall).

Just as an individual over the age of 35 was found to be slightly more likely to refuse to report his number, respondents over the age of 35 are also slightly more likely to refuse to provide numbers for other household members (1.1 percent versus 0.6 percent).

Household Respondent's Race and Sex.--As was the case when SSN reporting was examined by the characteristics of the individual, the race of the respondent had a greater influence than did sex in determining the proportion of SSN's reported for the household. About 81.4 percent of the persons 14 years or older in households of white respondents gave an SSN versus 71.9 percent of the households of nonwhite respondents. Whether the respondent was male or female made a difference of only 1.4 percent (reporting rates by sex of the respondent are 81.4 percent for males and 80.0 percent for females).

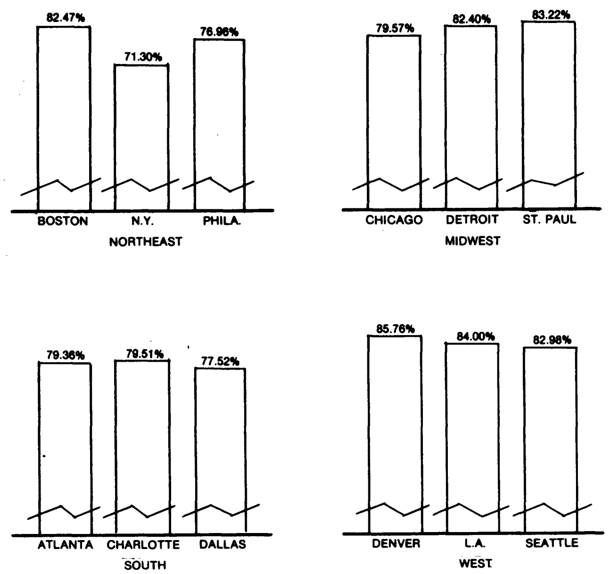
5. OTHER ASPECTS OF SSN REPORTING

Data Collection Centers.--The twelve regional offices of the Census Bureau both train and supervise the CPS interviewers and, thus, exercise some influence on the quality of the data collected. Also, the final manual processing of the control card--that of its being keyed onto tape for computer use--was divided into work units within data collection centers.

The SSN reporting rate does vary between the centers; New York's rate of 71.3 percent is significantly lower than the others, while Denver's performance of 85.8 percent is over five percent better than the overall rate of 80.2 percent for all the centers (see figure 7).

The low reporting rate of New York is due, in part, to a much higher proportion of blanks (13.7 percent versus 8.9 percent when New York is excluded) and almost twice as many "not available" responses as the other centers (5.5 percent as opposed

Figure 7.—SSN REPORTING BY REGIONAL DATA COLLECTION CENTER



to 2.8 percent). The somewhat lower reporting rates of Philadelphia and Dallas, at 77.0 and 77.5 percents, respectively, also reflect slightly higher blank and "not available" entries. In addition, all three centers have higher proportions of answers of "none" to the SSN question than do the other centers.

Rotation groups.--Examining the data by rotation groups, we find there is a direct relationship between the incidence of SSN reporting and the number of months that the rotation group was in the survey during the time frame that the SSN question was asked. (This was from December 1972 to March 1973.) The table below illustrates this association.

Table 3.--SSN Reporting Rate by Number of Months SSN Question Asked

Rotation Groups	Months Question Asked	SSN Reporting Rate
1 and 5	December-March	82.8
2 and 6	January-March	81.8
3 and 7	February-March	78.7
4 and 8	March only	77.5

Rotation groups 1 and 5 were eligible for interview during all four months and show the highest response rates. The two rotation groups (4 and 8) which were in the survey for only one of the four months have the lowest SSN reporting rates.

Conclusion.--The steps taken by the Census Bureau in conducting the survey are seen to be carefully controlled and administered; nevertheless, chances for human and machine error exist. Although not an exhaustive study of all the possibilities, we have in this paper identified some variables affecting the likelihood of SSN's being reported in the CPS. Age is a very strong factor, and race, more so than sex, plays an important role in determining whether a number is provided. If an individual was the household respondent, chances are improved that his number will be reported. Also, the characteristics of the respondent affect to some extent the reporting rate of his entire household.

FOOTNOTES

1/ Detailed information on the Current Population Survey is available in many Census Bureau and Bureau of Labor Statistics publications. A particularly good source on the CPS as redesigned after the 1970 Census is a paper in the Annals of Economic and Social Measurement, "The Current Population Survey: an overview" by Marvin Thompson and Gary Shapiro (vol. 2, 1973, pp. 105-119).

2/ Linkages between the survey and administrative records have been attempted during most of the years since 1963. However, unlike the 1973 Match Project, these earlier linkage attempts were made only with a subsample of the cases in the survey. In some instances, like the Pilot Linkage work referenced in the text, the subsample was representative of the total civilian noninstitutional population. In other instances, only certain subgroups were represented, such as persons 60 years or older.

3/ It was possible, as is described in the second paper at the session, to find about half of the SSN's which were not reported initially. SSN's found for persons who refused, however, could not be returned to the Census Bureau, since to do so would violate the respondent's expressed wish in this regard. Instead, at Social Security, the characteristics of the refusals are being analyzed as a group and adjustments to the overall match results will be made accordingly.

4/ Our best estimate is that about 98 percent, or perhaps a little more, of the U. S. noninstitutional population 17 years or older have been issued SSN's. For persons 14 to 17 years old, between 80 and 90 percent have been issued numbers.

5/ Age is determined as of March 17, 1973. This differs from the age variable used in the other two papers, which is in terms of age attained in 1972.

6/ The March household respondent is being identified here. Usually, but not always, this was the same individual who responded to the SSN question. The SSN information may have been obtained at any time from December 1972 through March 1973, depending on the rotation group in which the household was a part.

7/ Recall that, in the discussion of previous studies, it appeared that our SSN reporting rate improved to about 80 percent over the 1966-67 Census study rate of 76 percent. It would be more valid to contrast our 87 percent SSN reporting rate for respondents, since each person in the 1966-67 study was to respond for himself.

SEARCHING FOR MISSING SOCIAL SECURITY NUMBERS

Beth Kilss and Barbara Tyler
Social Security Administration

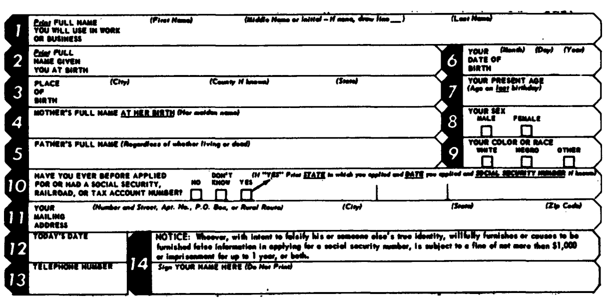
As we just learned from the previous paper, social security numbers were obtained for about 80 percent of the CPS sample. This paper focusses attention on what we did about the other 20 percent; that is, those with missing social security numbers. 1/

In cases where there was no response to the SSN question, a search of social security files was made by name and date of birth. The search procedures employed are described in section 1 of this paper. Results of that search are analyzed and evaluated in sections 2 and 3. "Find" rates are presented by age, race, sex, and commonness of last name. An examination is also made of the impact on the find rates of errors and omissions in the variables used in searching.

1. SSN SEARCHING PROCEDURES

Information Needed for Searching.--There are a number of manual-clerical search procedures routinely employed at Social Security to locate the SSN's of individuals who fail to report them. 2/To make optimum use of these procedures, the following items from the application for a social security number, Form SS-5, are considered desirable (see figure 1).

Figure 1.—APPLICATION FOR A SOCIAL SECURITY NUMBER, FORM SS-5


1 Print FULL NAME YOU WILL USE IN WORK OR BUSINESS (First Name) (Middle Name or Initial – If none, draw line ___) (Last Name)
2 Print FULL NAME GIVEN YOU AT BIRTH
6 YOUR DATE OF BIRTH (Month) (Day) (Year)
3 PLACE OF BIRTH (City) (County if known) (State)
7 YOUR PRESENT AGE (Age on last birthday)
4 MOTHER'S FULL NAME AT HER BIRTH (Her maiden name)
8 YOUR SEX MALE FEMALE
5 FATHER'S FULL NAME (Regardless of whether living or dead)
9 YOUR COLOR OR RACE WHITE NEGRO OTHER
10 HAVE YOU EVER BEFORE APPLIED FOR OR HAD A SOCIAL SECURITY, RAILROAD, OR TAX ACCOUNT NUMBER? NO DON'T KNOW YES (If "YES" Print STATE in which you applied and DATE you applied and SOCIAL SECURITY NUMBER if known)
11 YOUR MAILING ADDRESS (Number and Street, Apt. No., P.O. Box, or Rural Route) (City) (State) (Zip Code)
12 TODAY'S DATE
14 NOTICE: Whoever, with intent to falsify his or someone else's true identity, willfully furnishes or causes to be furnished false information in applying for a social security number, is subject to a fine of not more than $1,000 or imprisonment for up to 1 year, or both.
13 TELEPHONE NUMBER
Sign YOUR NAME HERE (Do Not Print)

Listed in their general order of importance, these are:

1. Full name
2. Date of birth
3. Mother's maiden name
4. Father's name
5. Place of birth
6. Sex
7. Race

If all this information is provided, the correct account number can be located for most individuals. With just the name and date of birth, SSN's can still be determined in the majority of the cases. However, for persons with fairly common last names, it is often necessary to have at least the mother's maiden name or the father's name or the place of birth or, in some instances, all of these, in order to find the correct SSN.

Of the items mentioned above, only name, date of birth, race, and sex were available from the CPS control cards. This lack of information is believed to have been detrimental to the find rates, as will be seen later in this paper.

Files Employed in Searching.--The National Employee Index microfilm file is the basic source used in general-purpose searches for account numbers at Social Security. It contains information from the SS-5, or updates of the SS-5, for all persons who were ever issued an SSN. This was the file most heavily relied on in the Match Project. All CPS persons 14 years or older with a missing SSN were searched for in this file. 3/ In addition, because of the importance of obtaining complete data on benefits, persons 62 years or older were also sought in another file as well, one restricted just to beneficiaries.

Both of the microfilm sources employed are in sort by last name, first name, middle initial, year and month of birth. For the Beneficiary file used, name is in strict alphabetical order. In the National Employee Index file, only the first and middle names are alphabetized; the surnames are grouped by the Russell Soundex Code, which brings together names which "sound" alike. 4/

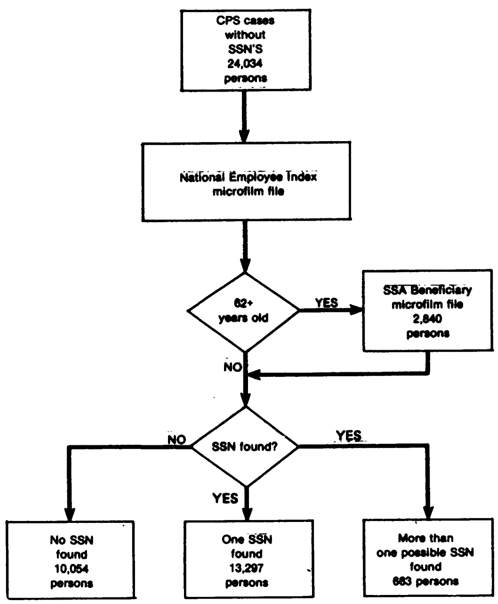
Search Procedure.--The first step in the search procedure for individuals with missing SSN's was to check clerically against the microfilm, looking for exact agreements on name and date of birth. If an account number could not be located in this way, reasonable variations in the information were allowed for--such as common variations of first names, dropping middle initials shown, substituting different middle initials, transposing digits in the date of birth, etc.

Standard Social Security administrative practice calls for "re-searching" cases where the first attempt to locate an account number fails. However, this was not done in the Match Project because of cost and timing limitations; only a single search of the microfilm files could be conducted.

2. BASIC SEARCH RESULTS: GROSS FINDS

Altogether, there were 24,034 persons with missing SSN's for which a name search was conducted. 5/ Of these, 2,840 individuals (or about 12 percent) were searched for in both the National Employee Index file and the Beneficiary microfilm file (see figure 2).

Figure 2.—OVERALL SEARCH RESULTS



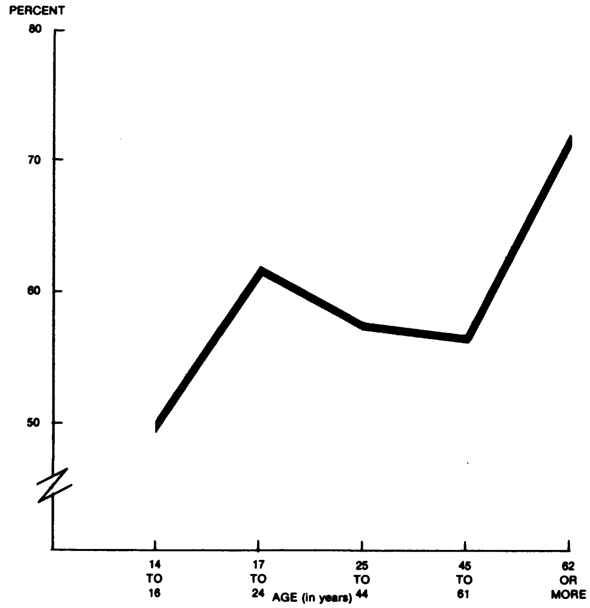
The overall results of these searches are as follows: for 10,054 persons (or 42 percent), no SSN was located; for the remaining 13,980 cases, at least one possible SSN was found. For 13,297 persons, or 55 percent of the total sent, exactly one SSN was found; there were also 683 persons (or three percent of those sent) where more than one possible SSN was returned. 6/ Thus, the overall gross find rate for the persons with missing account numbers was 58 percent.

Of course, finding a "possible" SSN for an individual is not the same as finding the "correct" SSN for that person. In this section, however, we will examine the gross find rates; that is, the proportion of the cases for which a possible SSN was found. The last section of the paper will look at the extent to which a possible SSN turned out to be the correct one.

Age.--Persons 14 to 16 years old 7/ had the lowest gross find rates of the five age categories examined (see figure 3). These persons were found just slightly less than 50 percent of the time. Many in this age group, of course, have not yet filed for an account number and, therefore, are not in our files.

For persons 17 to 24, there was a 12 percent increase in the gross find rates over the 14 to 16 year age group (to 61.8 percent). The rates decline somewhat for the next two age groups (to 57.4 percent for the 25 to 44 year age group and 56.5 percent for the 45 to 61 year age group). The downward trend could be attributable

Figure 3.—GROSS FIND RATES FOR ALL PERSONS WITHOUT SSN'S BY AGE.



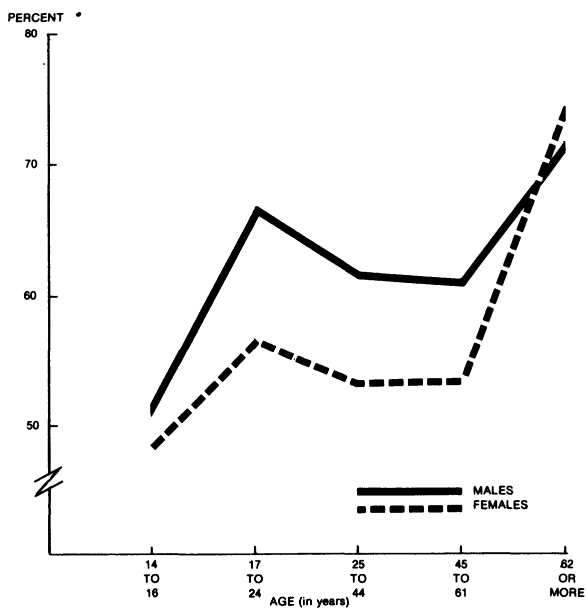
to the length of time since the original information was given on the Form SS-5. It would seem that for younger persons the original information given to obtain the SSN is more likely to be in agreement with that given at the time of the CPS survey.

Persons 62 years or older had a gross find rate of 71.6 percent, an increase of 15 percent over the previous age bracket. This high find rate was due to the fact that people in this age group were subject to two searches. For the basic search, the gross find rate was 51.4 percent; the additional Beneficiary file search yielded 20.2 percent more finds.

The fact that the find rate for the people 62 years or older is the highest of all the age groups and that they were searched for twice has implications for the find rates for all other ages. It is virtually certain that the other four age groups would have had higher find rates if there had been time to conduct additional searches for them. The two searches conducted for persons 62 years or older were done concurrently, since they involved separate microfilm files. Time constraints in the project did not allow for additional searches for persons under 62 years old, as they would have had to be carried out successively.

Race and Sex.--The difference between gross find rates for whites and nonwhites was very small on an overall basis. Race does not seem to affect whether or not a searcher would be successful in finding a person without an SSN. The gross find rates for whites were only about 0.6 percent higher than for nonwhites. These search results are in contrast to the large differences in gross find rates between males and females. Being male improved the chances that a person would be found by about 5.4 percent overall. There are particularly large differences in the gross find rates for males and females between the ages of 17 and 61 years (see figure 4). The gross find rates

Figure 4.—GROSS FIND RATES FOR MALES AND FEMALES BY AGE



for males in this age group are about 8 to 10 percent higher than for females of the same ages. Most women 17 to 61 years of age are married or have been married. One possible explanation for the higher gross find rates of males versus females is that often a female reports her married name in the CPS survey but has failed to report this change of name to Social Security. Therefore, she appears on the National Employee Index file under her maiden name and cannot be found.

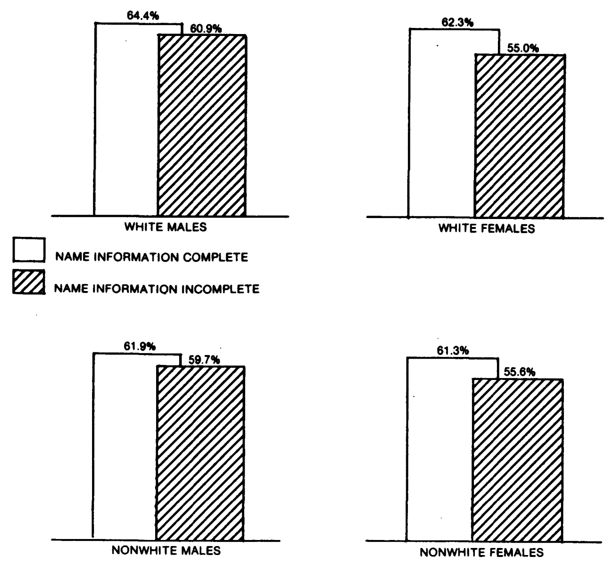
Males did not have higher gross find rates in every age group. At age 62 years or older, the gross find rates for women are about 2.3 percent higher. A possible explanation for this could be that there are more women who are beneficiaries at an earlier age than men; therefore, we were more successful in locating these persons in the Beneficiary file.

The overall results of the name searching seem to indicate that gross find rates were affected more by sex than by race and that we were more successful in locating males without SSN's than females.

Completeness of Name.--As we stated earlier, the CPS control card provides only some of the information necessary to successfully search for a person. A further difficulty arises in that the information that was available was not always complete. Of the total number of persons for whom a search was made, full name information was available for only 14.4 percent of them. For most of the remaining persons the first name was present, but middle names or initials were missing.

The results of the name searching indicate that improved find rates are associated with the presence of a middle name or initial. For all persons without an SSN for whom there was a full name, gross find rates were about 5.4 percent higher than for those persons with an incomplete name. Among the four race and sex groups, gross find rates were always higher when the full name was provided (see figure 5). The

Figure 5.—GROSS FIND RATES BY COMPLETENESS OF NAME INFORMATION, RACE, AND SEX

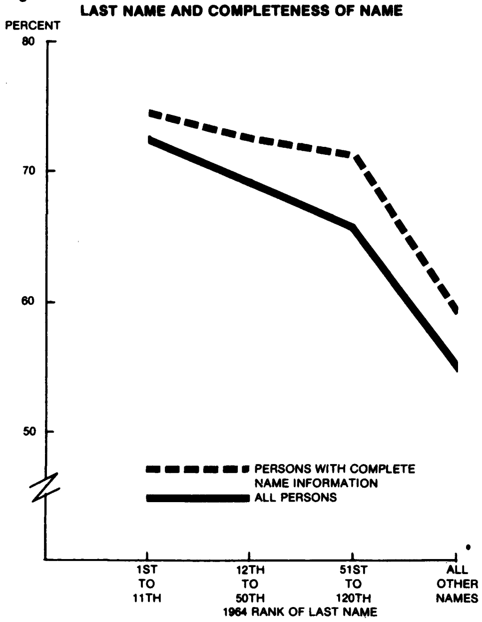


find rates improved more for females than for males (from 55.0 percent to 62.3 percent for white females and from 55.6 percent to 61.3 percent for nonwhite females; from 60.9 percent to 64.4 percent for white males and from 59.7 percent to 61.9 percent for nonwhite males). Thus the net increase was 7.0 percent for females as compared to 3.4 percent for males. By race there was also some difference in the amount of the improvement, but this was not as great as by sex--5.9 percent for whites versus 4.1

percent for nonwhites. It is not obvious why improved find rates for complete name information are associated with females in particular. However, it is clear that gross find rates would have improved substantially had full name information been available for all persons.

Commonness of Last Name.--In addition to completeness of name information, commonness of surname is an important variable in searching for social security numbers. In fact, it may be the most important of all the variables we are considering in this paper. Gross find rates for persons with very common names are higher than for persons with less common ones, as can be seen in figure 6 (see the appendix for discussion and table of common names). It turns out to be much

Figure 6.—GROSS FIND RATES FOR ALL PERSONS BY COMMONNESS OF LAST NAME AND COMPLETENESS OF NAME



more likely that a possible SSN or SSN's would be found for a Mr. Smith or Ms. Johnson than for someone with a less common name.

The find rates show a strong downward trend with decreasing commonness of surname. The gross find rate for persons having the eleven most common names is 72.7 percent. For persons having surnames which ranked 12th to 50th, the find rate drops to 69.1 percent and, then, to 65.8 percent for surnames which ranked 51st to 120th. For persons with all other names, the find rate is the lowest (55.4 percent).

Figure 6 also shows the improvement in the gross find rates when the full name was available. The completeness of name improved the gross find rates by two percent for persons with the 11 most common names. For the other groups, the improvement ranged from 3.6 percent to 5.4 percent, with the maximum increase occurring for persons whose surnames ranked 51st to 120th.

3. RINGER SAMPLE RESULTS: NET VERSUS GROSS FIND RATES

Ringer Sample.--A stratified sample of 1,930 persons reporting valid SSN's was selected and searched for in essentially the same way and at the same time that the searching was done for persons without an SSN. We refer to these persons as the ringer sample. Like the persons without SSN's, these "ringers" were also 14 years or older. The purpose of processing this sample was to provide an internal measure of the quality of the SSN searching. The sample was drawn proportionately (by age, race, and sex) with some oversampling among persons having one of the 120 surnames most common in 1964.

The overall gross find rate of 71 percent for the ringer sample is higher than the 58 percent gross find rate for persons who did not have an SSN originally. The difference between these search results arises for the following reasons:

1. A higher gross find rate was expected for the ringer sample, as it included only persons who actually had been issued an SSN.

2. Date of birth, race, and sex information were complete for the ringer sample but were not necessarily complete for persons without an SSN.

3. In some respects, the ringer sample is not representative of the basic cases. For example, persons age 14 to 16 make up 3.8 percent of the ringer sample, while 23.4 percent of the basic cases are age 14 to 16.

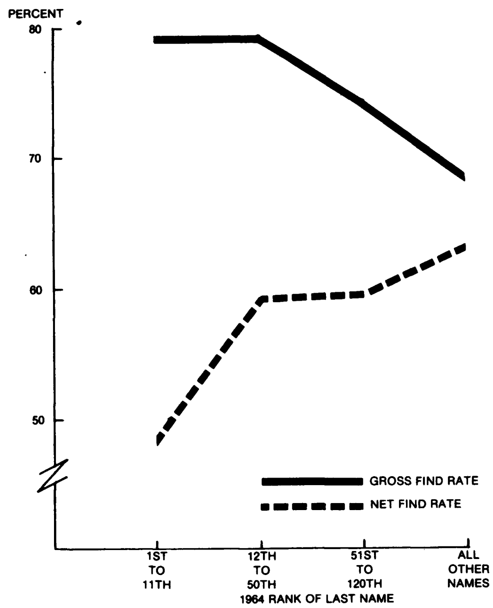
Net Find Rates Overall.--For about 62 percent of all the ringers, the "correct" SSN was found. 8/ This net find rate is a measure of the number of persons for whom the social security number found was the same as the number originally reported in the CPS.

The net find rate is nearly nine percent lower than the overall gross find rate for the ringers. One of the reasons searchers did not always locate the correct SSN becomes apparent when examining the gross and net find rates for the ringer sample by commonness of last name (see figure 7).

Net Find Rates by Commonness of Last Name.-- Higher gross find rates for persons with one of the more common last names continued to be true for the ringers, just as it was for the overall search. In contrast to the gross find rates, the net find rates decrease with increasing commonness of last name. Although it is easier to find at least one possible SSN for people with common names, there appears to be considerable difficulty in selecting all the possible SSN's for such people. Thus, finding the correct SSN occurs less often in these cases.

For persons with names among the 11 most common, the net find rate is about 30 per-

Figure 7.—FIND RATES FOR RINGER SAMPLE BY COMMONNESS OF LAST NAME



cent lower than the gross find rate. However, there was only about a 5.5 percent difference between the gross and net find rates for those persons with names other than the 120 most common.

Net Find Rates by Age, Race, Sex and Completeness of Name.--While differences between the pattern of gross and net find rates are at their greatest by commonness of last name, there are also some differences by age, race, sex, and completeness of last name:

1. Age.--Gross find rates exceed the net find rates by at least 5.4 percent for all age categories, as can be seen below:

Age (in years)	Ringer Find Rates: Gross Finds	Net Finds
14 to 16	68.9%	63.5%
17 to 24	66.6%	58.4%
25 to 44	67.6%	58.8%
45 to 61	67.5%	58.4%
62 or older	87.1%	75.8%

The smallest difference occurs for the 14 to 16 year olds and increases as a function of increasing age to a maximum of 11.3 percent for persons age 62 or older. 9/

2. Race and Sex.--Net find rates were always at least 8.5 percent lower than gross find rates for each race and sex group (see figure 8). For nonwhites, the difference between the rates was the largest (13.1 percent). Males had the highest net find rate (64.7 percent), while nonwhites had the lowest (52.7 percent).

   There appears to be a race effect for ringer gross finds which does not exist in the basic sample; however, it is hard to judge how important this is because of the smallness of the ringer sample for nonwhites (184 cases).

3. Completeness of Name Information.--The gross and net find rates for ringer persons with complete name information are 72.9 percent and 65.7 percent, respectively. The corresponding rates for ringer persons with incomplete name information are 69.9 percent and 60.4 percent. In other words, when complete name information was available, there was a 2.9 percent increase in the gross find rate and a 5.3 percent increase in the net find rate.

Conclusion.--Social security numbers were originally reported for about 80 percent of the CPS sample. We attempted to improve this reporting rate by modifying Social Security's searching procedures in order to obtain additional SSN's. Possible SSN's were located for about 58 percent of all persons without an SSN.10/ Based on the results of the ringer sample, correct SSN's may have actually been found for about 50 percent of all the persons for whom a search was made. This effort has, in effect, cut in half the number of SSN's originally not obtained from the survey data.

APPENDIX ON COMMONNESS OF LAST NAME

According to a 1964 study done by the Social Security Administration,11/ the eleven most common surnames at that time each occurred over 500,000 times. In fact, there were nearly 1.7 million account numbers for the name Smith alone.

Table 1.-- 1964 Surname Summary

Rank of Surname	Minimum Number of Records per Name	Percent of File (percent)
1 to 11	500,000	5.89
12 to 50	209,000	7.28
51 to 120	117,000	6.39
121 to 2183	10,000	33.15
All other names	1	47.29

As can be seen from table 1, although the file contained over a million different surnames, the top 120 names accounted for 20 percent of all the records on the file, and the top 2,183 names accounted for about 53 percent of the file.

Since 1964, approximately nine million new records have been added each year, including new account numbers and changes of name. By the time of the 1973 CPS, there were nearly 210 million account numbers. Comparison of the 120 most frequent surnames from the 1973 CPS with the results of the 1964 study shows that most of the names did not change (see table 2). In particular, the eleven most common names stayed the same, although the ordering within the group changed somewhat. For the names ranking 12th to 50th, there is considerable reordering within these 39 names and six name replacements. Among the 51st to 120th, there are 10 additional name replacements.

In other words, six names cross from the 12th to 50th group into the 51st to 120th, and 10 names move from the 120 most common to a rank below 120. Despite these changes, it appears that the 1973 findings are substantially in agreement with the 1964 results.

FOOTNOTES

1/ The searching procedures described here were only possible for persons for whom a last name and a year of birth were available. Therefore, for March CPS cases which were not matched to a control card, no search could be conducted. Furthermore, it should be pointed out that the results described in this paper are not just for persons in households interviewed in March, as was true for the previous paper. In situations where an interview had been secured for an eligible household in a month other than March 1973, a search was also made for the SSN's of persons 14 years or older whenever these were missing. This additional step was taken in order to allow some analysis of the CPS noninterviews to be conducted.

2/ Electronic searching procedures are being developed at SSA to replace the existing clerical processing. Fully automated searches for account numbers will be installed over the next few years.

3/ No search was made for CPS persons under 14 years of age, because, for these individuals, no income information was available in the survey.

4/ The first character of the Soundex code is the first letter of the surname. The remaining three digits are obtained by assigning a numerical value to letters which produce equivalent sounds. See [95] or [147] for details.

5/ Name searches were also conducted for 3,941 cases where the original SSN reported in the CPS was suspected of being incorrect. Also searched for were an additional 1,930 cases where the SSN was judged to be "correct." This step is described in section 3.

6/ In most instances, when more than one SSN was returned from the search, the reason was that the searcher could not determine which, if any, of the possible SSN's was the correct one. In some cases, however, the multiple possible SSN's may all have actually belonged to the individual in question. (See the general session introduction for more discussion of the problem of nonuniqueness in the SSN.)

7/ Age, as measured here and elsewhere in this paper, is that attained in 1972.

8/ For the basic sample, the gross find rate was 58 percent. Taking into account the differences between this sample and the ringer sample, the net find rate for the basic sample is estimated to be approximately 50 percent.

9/ The net find rate for beneficiaries 62 years or older may be somewhat understated as, sometimes, the SSN found in the Beneficiary Microfilm file is not that of the person for whom a number is being searched, but of a spouse, for example. This is all that is needed, however, for, even though we did not find the individual's own number, SSA keeps a cross-reference system which links together persons who receive benefits

Table 2.--COMPARISON OF RANKINGS OF COMMON SURNAMES: 1964 and 1973

Ranked According to Findings of 1964 Study						Ranked According to CPS 1973 Result					
Name	**1964 Rank**	**1973 Rank**	**Name**	**1964 Rank**	**1973 Rank**	**Name**	**1964 Rank**	**1973 Rank**	**Name**	**1964 Rank**	**1973 Rank**
Smith	1	1	Bailey	61	53	Smith	1	1	Murphy	47	61
Johnso(n)	2	2	Garcia	62	64	Johnso(n)	2	2	Long	77	62
Willia(ms)	3	3	Ward	63	87	Willia(ms)	3	3	Patter(son)	91	63
Brown	4	6	Cox	64	47	Jones	5	4	Garcia	62	64
Jones	5	4	Steven(s)	65	48	Miller	6	5	Evans	42	65
Miller	6	5	Howard	66	70	Brown	4	6	Gonzal(es)	55	66
Davis	7	8	Bennet(t)	67	96	Anders(on)	9	7	Rogers	51	67
Martin(ez)	8	9	Sander(s)	68	97	Davis	7	8	Nichol(s)	72	68
Anders(on)	9	7	Brooks	69	98	Martin(ez)	8	9	Ellis	107	69
Wilson	10	11	Watson	70	50	Harris(on)	11	10	Howard	66	70
Harris(on)	11	10	Gray	71	93	Wilson	10	11	Cooper	56	71
Taylor	12	17	Nichol(s)	72	68	Thomas	14	12	Bell	58	72
Moore	13	16	Sulliv(an)	73	73	White	16	13	Sulliv(an)	73	73
Thomas	14	12	Hughes	74	82	Thomps(on)	15	14	Reynol(ds)	90	74
Thomps(on)	15	14	Myers	75	55	Jackso(n)	17	15	Olson	111	75
White	16	13	Ross	76	78	Moore	13	16	Foster	81	76
Jackso(n)	17	15	Long	77	62	Taylor	12	17	Powell	84	77
Clark	18	18	Price	78	79	Clark	18	18	Ross	76	78
Robert(s)	19	20	Russel(l)	79	102	Allen	26	19	Price	78	79
Peters(on)	20	41	Fisher	80	60	Robert(s)	19	20	Kenned(y)	109	80
Lewis	21	26	Foster	81	76	Morris(on)	25	21	Jacobs(on)	94	81
Walker	22	22	Hender(son)	82	95	Walker	22	22	Hughes	74	82
Robins(on)	23	33	Daniel(s)	83	89	Baker	31	23	Marsha(ll)	108	83
Hall	24	27	Powell	84	77	King	28	24	Wallac(e)	92	84
Morris(on)	25	21	Perry	85	58	Young	27	25	Snyder	113	85
Allen	26	19	Butler	86	115	Lewis	21	26	Graham	93	86
Young	27	25	James	87	112	Hall	24	27	Ward	63	87
King	28	24	Jenkin	88	109	Scott	33	28	Kelly	59	88
Wright(son)	29	35	Barnes	89	101	Adams	34	29	Daniel(s)	83	89
Nelson	30	37	Reynol(ds)	90	74	Cook	50	30	Schmid(t)	112	90
Baker	31	23	Patter(son)	91	63	Green	36	31	Gibson	115	91
Hill	32	38	Wallac(e)	92	84	Edward	49	32	Hayes	106	92
Scott	33	28	Graham	93	86	Robins(on)	23	33	Gray	71	93
Adams	34	29	Jacobs(on)	94	81	Philli(ps)	40	34	Ford	114	94
Richar(ds)	35	36	Simmon(s)	95	100	Wright(son)	29	35	Hender(son)	82	95
Green	36	31	Colema(n)	96	108	Richar(ds)	35	36	Bennet(t)	67	96
Lee	37	42	Hamilt(on)	97	59	Nelson	30	37	Sander(s)	68	97
Mitche(ll)	38	45	Lopez	98	134	Hill	32	38	Brooks	69	98
Campbe(ll)	39	43	Stephe(ns)	99	129	Collin(s)	45	39	Cole	102	99
Philli(ps)	40	34	Rivera	100	623	Carter	41	40	Simmon(s)	95	100
Carter	41	40	Murray	101	105	Peters(on)	20	41	Barnes	89	101
Evans	42	65	Cole	102	99	Lee	37	42	Russel(l)	79	102
Stewar(t)	43	56	McDona(ld)	103	146	Campbe(ll)	39	43	Gordon	124	103
Turner	44	54	Alexan(der)	104	106	Wood	60	44	Rodrig(uez)	48	104
Collin(s)	45	39	West	105	141	Mitche(ll)	38	45	Murray	101	105
Parker	46	51	Hayes	106	92	Reed	57	46	Alexan(der)	104	106
Murphy	47	61	Ellis	107	69	Cox	64	47	Stone	136	107
Rodrig(uez)	48	104	Marsha(ll)	108	83	Steven(s)	65	48	Colema(n)	96	108
Edward	49	32	Kenned(y)	109	80	Griffi(n)	52	49	Jenkin	88	109
Cook	50	30	Jordan	110	188	Watson	70	50	Matthe(ws)	184	110
Rogers	51	67	Olson	111	75	Parker	46	51	Walter(s)	139	111
Griffi(n)	52	49	Schmid(t)	112	90	Christ(ian)	53	52	James	87	112
Christ(ian)	53	52	Snyder	113	85	Bailey	61	53	Meyer	129	113
Morgan	54	57	Ford	114	94	Turner	44	54	Berry	176	114
Gonzal(es)	55	66	Gibson	115	91	Myers	75	55	Butler	86	115
Cooper	56	71	Burns	116	161	Stewar(t)	43	56	Webb	128	116
Reed	57	46	Bryant	117	118	Morgan	54	57	Palmer	138	117
Bell	58	72	Hernan(dez)	118	144	Perry	85	58	Bryant	117	118
Kelly	59	88	Wells	119	193	Hamilt(on)	97	59	Lucas	258	119
Wood	60	44	Hoffma(n)	120	147	Fisher	80	60	Ryan	140	120

The letters in parenthesis following certain names are intended to show some of the more common surnames included. The 1964 results are based on a census of the account number files of SSA. The 1973 results are based on all sample persons in the March 1973 CPS who were 17 years of age or older. Persons living in Puerto Rico are excluded from the CPS but not from SSA's files; therefore some of the common Spanish names in the 1964 study tended to be less important in 1973.

on the same number, and, thus, it will eventually be possible to locate the required SSN.

10/ In searches for missing SSN's of special population groups (for example, the aged [69]), the gross find rates attained in past studies have tended to be better than those for the 1973 Match Project. However, search results for samples from the general population (e.g.,[147]) are comparable to those shown here.

11/ U.S. Social Security Administration. Report of Distribution of Surnames in the Social Security Account Number File, 1964 (processed).

VALIDATING REPORTED SOCIAL SECURITY NUMBERS

Cynthia Cobleigh and Wendy Alvey
Social Security Administration

In the first two papers at this session, the discussion has focussed on the problem of incomplete reporting of social security numbers in the March 1973 Current Population Survey and what we have done about it so far. This paper will examine the degree to which the SSN's reported in the survey are "correct." Comparisons are made between name, race, sex, and date of birth, as shown on the CPS, and the SSA values for these items. In some cases, it is fairly obvious that the account number reported in the survey was incorrect. However, the confirmatory information is, itself, subject to response error, so some tolerance must be allowed. This paper describes some of the considerations that need to be taken into account in deciding what tolerances will be necessary. 1/

Section 1 provides an overview of the validation process and a brief examination of certain obvious errors that occurred. After eliminating from further analyses definable cases containing such errors, there remain persons for whom it is reasonable to construct CPS-SSA comparisons of demograhpic characteristics. For these individuals, sex and race differences are examined in section 2. In section 3, surname comparisons are made. Sections 4 and 5 deal with date of birth and age, respectively.

1. AN OVERVIEW

Every available SSN "belonging" to a survey person had to be checked. 2/ The purpose of this procedure was to determine whether the same number on the SSA administrative record and the CPS record was referring to the same individual. The search for a matching SSN was conducted on the Social Security Administration's Summary Earnings (Magnetic Tape) File, which contains a record for each of the 210 million account numbers ever issued. The SSN's sent through this search came from various sources.

When the survey data was placed on tape in Jeffersonville, it was keyed twice. The first time a person's record was constructed, it went onto what is termed the "A" file. In the second round, done by more experienced operators, these "A" records were verified, rekeyed when differences arose, and put on the "B" file. We were not necessarily convinced that the "B" file contained the more accurate data. There are subjective elements involved in the process, such as the ability to read the handwritten survey forms, which made this judgment inadvisable without further investigation. So, initially, we processed both. After validation procedures were completed, we determined that the "B" file actually did contain much more accurate data; therefore, in all following discussion, "reported SSN's" will refer to the "B" file only.

For these reported SSN's, some preliminary comparisons on sex, race, name, and date of birth were examined. Based on gross dissimiliarities of names and dates of birth, a subset of these SSN's were processed through the manual search. In addition, for persons without an SSN in the survey, the manual search described in the preceding paper was done.

Figure 1.—ACCOUNT NUMBERS BY SOURCE

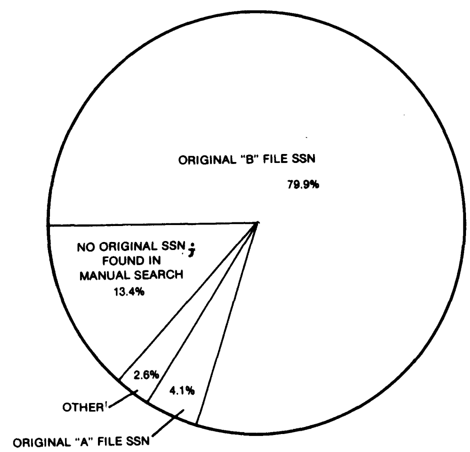


Figure 1 illustrates the various sources of all the account numbers that we now have for the 1973 Match Project. As can be seen, about four percent came from the "A" file; 80 percent came from identical "A" and "B" records or from a "B" record only. Another 13 percent of the total SSN's were obtained for persons who were missing a number in the survey. The additional manual search, carried out for SSN's judged to be incorrect, yielded 2.6 percent more numbers.

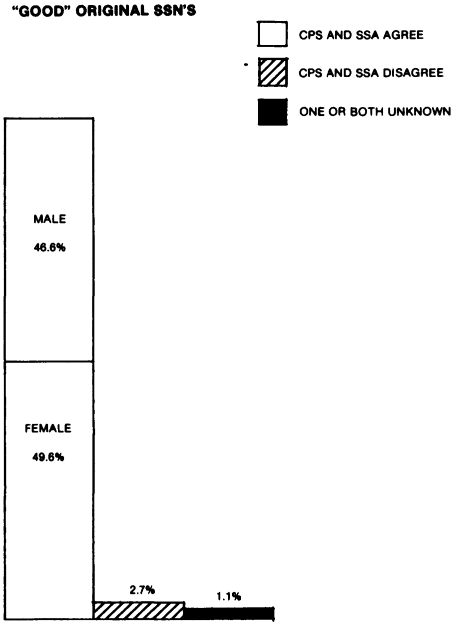
Originally Reported SSN's.--This paper will mainly concern itself with the subset of reported SSN's (i.e., "B" file cases) from which obvious errors have been excluded. Before limiting ourselves to this universe, though, some brief remarks about the types of errors encountered seem appropriate. About three percent of the SSN's are not now considered viable possibilities. Over half of them (2.06 percent) were found to be transposition or single digit errors. Furthermore, some persons reported an SSN belonging to another household member (.53 percent). This could occur if, for example, a woman reported the number under which she collected benefits, rather than her own. The remaining SSN's (.45 percent) could not be located in the administrative files.

The causes of these errors are varied. Keypunching and copying mistakes are only two of the several possibilities. In any event, it is not meaningful to conduct a CPS-SSA comparison of demographic characteristics in such cases. Eliminating these errors from further discussion, we will now focus our attention on what can be considered the potentially "good" segment of reported SSN's. The first two characteristics to be considered are sex and race.

2. SEX AND RACE AGREEMENT

Agreement on Sex.--When considering sex alone, figure 2 illustrates that for 96

Figure 2.—AGREEMENT ON SEX FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S

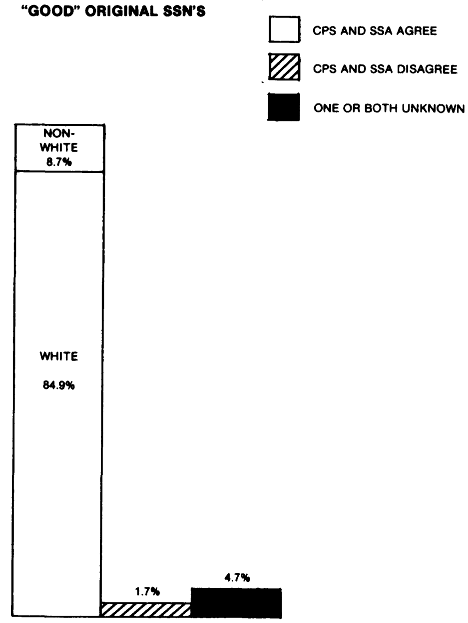


percent of the cases with potentially "good" original SSN's, CPS and SSA sex agree. In this group, females and males are represented in about the same proportions as in the population. Of the remaining persons, 1.1 percent reported sex as unknown on either one or both files, while 2.7 percent disagreed on sex.

An interesting pattern emerges among the cases whose CPS-SSA sex is in disagreement; almost two-thirds are SSA male/CPS female combinations. It can be ventured that a majority of these account numbers were reported by widows who gave their husband's SSN, under which they are receiving benefits. Since these women have their own number, the situation will be remedied when the benefit portion of the project is matched in. This step will substitute the correct SSN for the present one. 3/

Agreement on Race.--Considering race alone, figure 3 illustrates that, about 94

Figure 3.—AGREEMENT ON RACE FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S



percent of the time, the Census and SSA records agree on race, with whites and nonwhites being represented in the same ratio as in the population. The proportions of persons in the unknown and disagreement categories have reversed from what was noted for sex. Unknown race was reported on one or both files 4.7 percent of the time; only 1.7 percent had a CPS-SSA race disagreement.

Before examining agreement on sex and race combined, let us suggest some possible explanations for the different patterns between race and sex when they are unknown or disagree. As just mentioned, there is a much higher incidence of unknown race as opposed to unknown sex. The proportion of unknowns in the CPS and SSA files also differ, as can be seen below:

Table 1.-- Percentage of cases with unknown race or sex by source

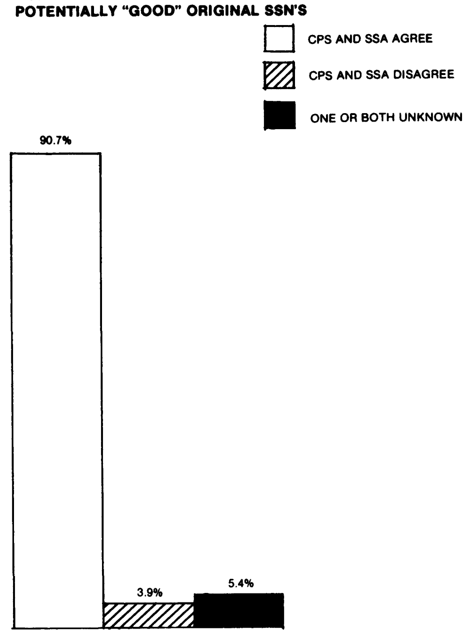
	Sex	Race
SSA unknown	0.14	2.90
CPS unknown	0.92	1.78
CPS and SSA unknown	0.00	0.04

The manner in which the SSA data was collected may be the source of some of the problem. Social Security information on race and sex is generally taken from the original application for a social security number, Form SS5. For a short period of time, IRS Form 3227 was used in cooperation with the Internal Revenue Service to provide SSN's for taxpayers who needed such identification. This form did not request race information. Although it was later changed to include this additional data, no effort was made to obtain a race for those who received account numbers in the interim; their race remained unknown.

It is likely that a substantial portion of the disagreement on race arises from the way it is collected in the CPS. Census interviewers do not ask the race question; the control card is filled in based on observation. Unless all household members are present to make another conclusion possible, the interviewer generally assumes that all related individuals are the same race. The respondent's race may be difficult to determine or other household members may be of a different race, and, as a result, race disagreements may occur.

Agreement on Sex and Race Combined.--Combining sex and race produces the results displayed in figure 4. Here, 91

Figure 4.—AGREEMENT ON SEX AND RACE COMBINED FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S

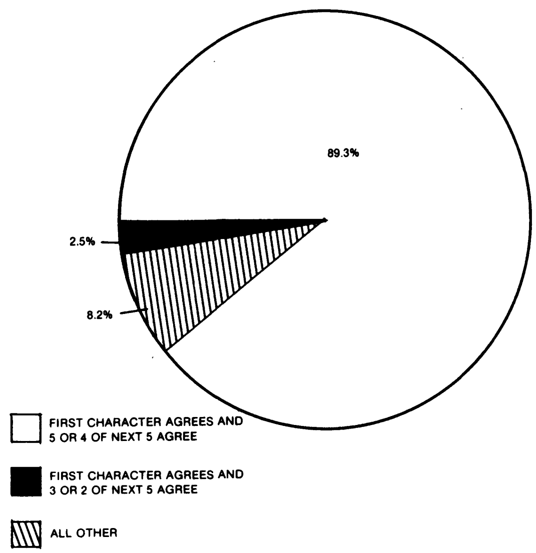


percent of the persons with potentially "good" original SSN's reported the same sex and race to both Census and Social Security. Unlike the distributions when sex and race are considered separately, the proportion of unknowns and disagreements are relatively equal. Sex-race combinations that disagree comprise 3.9 percent of the cases, while unknowns constitute the remaining 5.4 percent.

3. SURNAME AGREEMENT

In this discussion, the complete surname is not being compared; all references to surname pertain to no more than the first six characters. 4/ Figure 5 illustrates the results when names are examined character-by-character. Regardless of the other characteristics, we can say with relative certainty that the 89 percent who have surname agreement on the first character and four or five of the remaining characters are the same individual on both files.

Figure 5.—NAME AGREEMENT FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S



At first glance, it would appear as though all persons in the striped area (eight percent) should be discarded since the character-by-character surname disagreement is substantial. However, this particular group points out the problems that arise when only a single variable is used to evaluate a match. CPS women comprise 69 percent of this pie section; of these females, 59 percent are individuals whose other characteristics (sex, race, and date of birth) agree exactly. What has occurred for these women is that the name in SSA's administrative records is most probably a former one, and the SSN does refer to the same person on both files.

It is much more difficult to say anything definite about the 2.5 percent who lie between the two extremes. It will probably prove true that no single decision will resolve all the cases in this fuzzy in-between group. Those whose "demographics" strongly agree are likely to be victims of a spelling or keypunching error. However, obviously, as more disagreement on various characteristics appears, we become more inclined to conclude that the CPS and SSA persons are not the same.

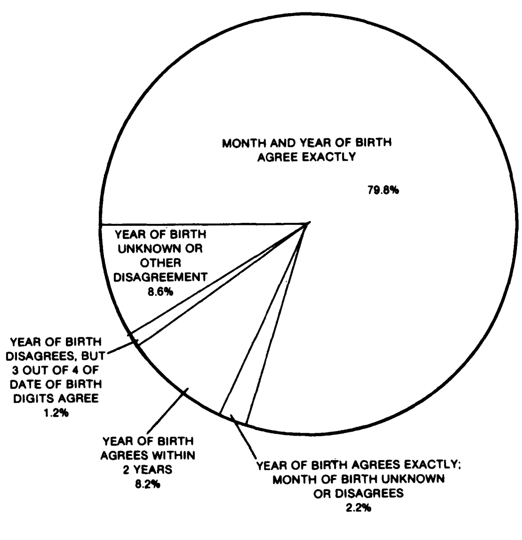
Instead of taking into consideration six characters of the surname, we could apply a number of other rules, one of which is to confine the evaluation to only the first four characters. Such a standard has been used by the Internal Revenue Service. 5/ According to the IRS rule, two surnames are judged the same if the first character agrees and, of the next three, there is no more than one difference or one transposition. Referring back to figure 5, if the first character agrees and, at most, one of the remaining five disagrees, this standard would be satisfied. Most of the cases in the remaining two pie sections would be rejected, however. The main exceptions are the previously-mentioned women whose former surname appears on the administrative record. IRS validates only the husband's SSN on a joint return, and, in actuality, these women would not even be examined by them.

4. DATE OF BIRTH AGREEMENT

Date of birth has been of use in evaluating surname. It merits some remarks as a variable in itself.

Figure 6 provides an illustration of the distribution of date of birth agreement. In 80 percent of the cases, persons reported exactly the same month and year of birth to both Census and Social Security. Even though this exact agreement is somewhat less than that cited earlier for sex and race, it still indicates a very respectable rate. In about two percent more of the cases, year of birth agrees exactly, but month of birth either disagrees or is unknown. The rest of the records contain disagreeing years of birth. However, at least half of them are similar enough to be acceptable for matching if other characteristics agree, such as sex, race, and surname (i. e., the eight percent which agree within two years and the one percent for which three out of four of the date of birth digits agree). The remaining nine percent of these cases either have an unknown year of birth or some other disagreement on year of birth.

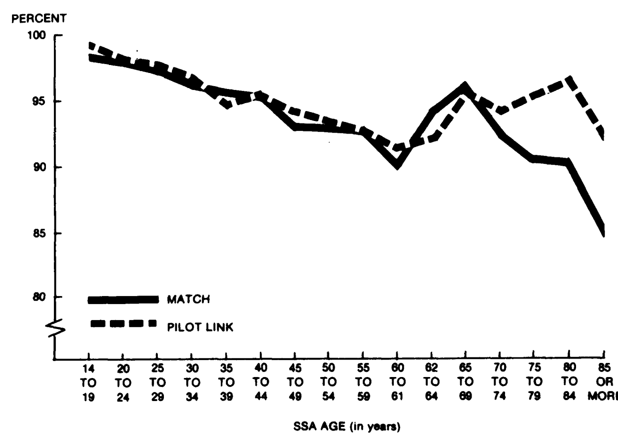
Figure 6.—BIRTH AGREEMENT FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S



5. AGE AGREEMENT

<u>Overall Agreement by Age Class</u>.--In this section, we will confine our examination of age agreement to those persons who reported the same race and sex to both Census and Social Security. Figure 7 presents the total distribution of such cases by age class agreement. 6/ Here the 1973 Match information (the solid line) is compared to an earlier Pilot Linkage Project (the broken line) which was also conducted jointly by SSA and the Census Bureau. SSA age group is plotted on the horizontal axis; the vertical axis shows the percentage of individuals in each SSA age group who were in the same age class in the CPS.

Figure 7.—AGE CLASS AGREEMENT IN MATCH AND PILOT LINK STUDIES FOR PERSONS WITH SSA AGE 14 OR OLDER HAVING POTENTIALLY "GOOD" ORIGINAL SSN'S AND AGREEMENT ON RACE AND SEX



It is quite evident from looking at the Match data that there is a marked downward trend in age class agreement as age increases. One possible explanation for this is that, as time elapses since the date of birth was originally reported to Social Security, awareness of SSA age and, hence, consistency in CPS-SSA age reporting decline. Three points, however, seem to vary from the trend. At ages 60 to 61, the amount of age class agreement drops below the expected level (to 90 percent). This may be partially due to the fact that this age class is smaller than the rest (and, hence, specifications for falling in the cell are more stringent). The next two age groups show agreement which is decidedly higher than the trend. This occurs at the time in these individuals' lives when they are applying for or, at least, inquiring about Social Security benefits. 7/ It is, therefore, reasonable to suggest that not only they, but also members of their households, would be more aware of their SSA age. Consequently, CPS-SSA age reporting would tend to be more consistent.

Up to the 65 to 69 year age group, the trends we have commented on in the Match data also occur for the Pilot Link Project. The two studies do not, however, behave in the same way for older persons: the agreement rates for persons 70 years or older in the Link study tend to be a good bit better than those in the Match. This may be due to the fact that SSA benefit data were used to improve the Pilot Link age information; the 1973 Match Project has not yet made use of this additional data.

Age Agreement by Race and Sex.--Figure 8 looks at those who agree on age class by race and sex. Here again, the Match and Pilot Link studies are compared. It is immediately evident that the age class agreement distributions by race and sex are practically identical for the two samples. In both studies, about 96 percent of the whites, both female and male, reported ages to SSA and CPS which fall in the same age group; about 90 percent of the nonwhites, again regardless of sex, agree on age.

Since there is so little difference in age agreement between the sexes, it seemed unnecessary to look at males and females separately by age group. We do, however, consider it worthwhile to look at age class agreement by race (see figure 9). As can be seen, the curve for whites (the solid line) is quite flat, with only a slight downward trend. The dip at ages 60 to 61 and the subsequent rise in agreement rates for 62 to 64 and 65 to 69 year-olds are still apparent, but they have become noticeably dampened as compared to what they were in figure 7. Nonwhites (the broken line), on the other hand, show a steeper decline in age class agreement as age increases. Furthermore, the three departures from the trend which occur in the 60 to 69 year-old age period are much sharper. This obvious difference in age reporting by race is, no doubt, due to multiple causes, and no one good explanation can be ventured at this time. However, it should be mentioned that the patterns shown here are similar to those found in other studies which have looked at differences in age reporting by race [e.g., 36, 101, 139, 147].

Figure 9.—AGE CLASS AGREEMENT FOR PERSONS IN THE MATCH STUDY WITH SSA AGE 14 OR OLDER HAVING POTENTIALLY "GOOD" ORIGINAL SSN'S AND AGREEMENT ON RACE AND SEX

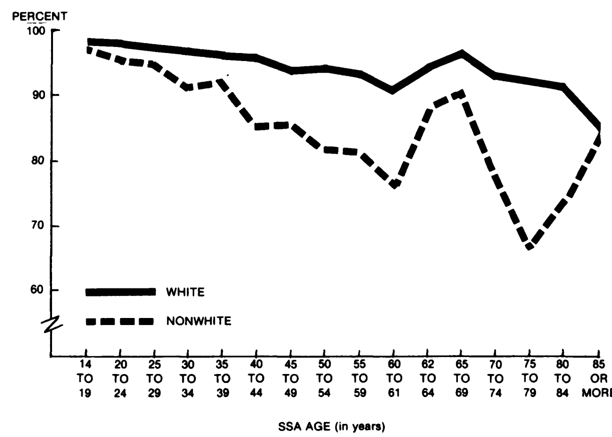
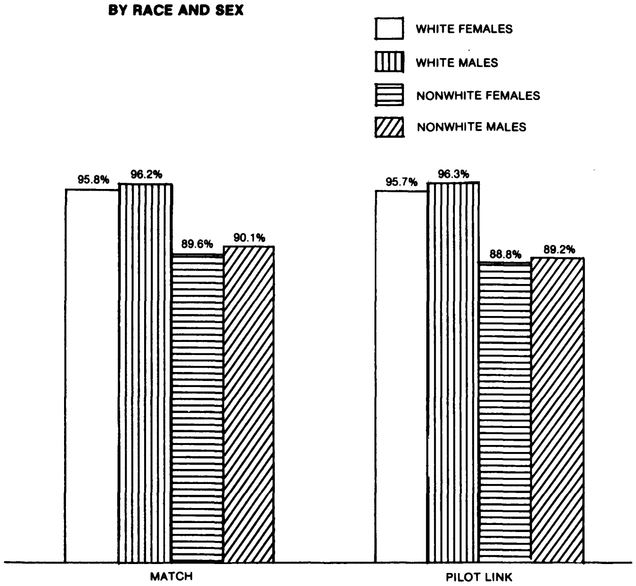


Figure 8.—AGE CLASS AGREEMENT IN THE MATCH AND PILOT LINK STUDIES FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S BY RACE AND SEX



Conclusion.--This paper's analysis of the "correctness" of originally reported SSN's is a preliminary one. Even so, several interesting patterns have emerged. Certain common demographic characteristics have been examined, including sex, race, surname, date of birth, and age. Overall, once obvious errors are eliminated from consideration, the extent of agreement between the CPS sample information and SSA's records is quite high:

1. SSA-CPS sex and race are the same 96 percent and 94 percent of the time, respectively.

2. For surname, about 89 percent of the persons whose SSN's were matched agreed closely enough to be considered the same individual. Because of the use of a former name, the amount of agreement for males and females differed significantly (92 percent versus 87 percent).

3. Exact agreement on month and year of birth was about ten percent less than that for name. When disagreements on date of birth were examined, it became evident that as age increases, agreement decreases. One major exception was observed: agreement improves substantially at the age of retirement. Furthermore, age agreement for nonwhites was considerably less than for whites, a pattern which exists in other studies.

The next step in the Project will be to look at all the common CPS-SSA characteristics simultaneously and decide exactly what tolerances should be established in the final matching.

FOOTNOTES

1/ Striking the balance between nonmatches and mismatches not only involves employing some of the theory on "optimum" matching rules [23, 27, 70, 125, 151], but also requires decisions as to what adjustment it will be possible to make for nonmatches and mismatches in the subsequent analysis. For dealing with nonmatches, one of the techniques we expect to employ is a "raking" procedure, presented in a 1974 SSA paper entitled "The Rake's Progress." For dealing with mismatches (which cannot be entirely eliminated no matter what matching rule is chosen), we hope to use adaptations of standard methods that are robust against mismatches [73]. One such adaptation that may have some merit is contained in a series of 1974 SSA papers entitled "Fitting Square Tables with Nonsquare Procedures." All of these papers are available upon request. (See footnote 8 in the session introduction for the address.)

2/ The validation included a few cases of persons less than 14 years of age with a reported account number.

3/ As these women generally seem to be using their husband's SSN rather than their own, it is likely that this is the number they reported to IRS. Thus, in order to obtain tax information for these individuals, it will probably be necessary to use their reported number rather than their correct SSN for matching to IRS' files.

4/ While the complete last name is available from the CPS, only the first six characters of the surname were available on the Summary Earnings File. When benefit data is matched in, we will be able to make a more extensive comparison for persons who are beneficiaries.

5/ As the Internal Revenue Service is the third source of data to be utilized in the Match project, it seemed beneficial to mention what effect its guidelines would have had. The purpose of the IRS standard for matching surnames is that the posting of social security numbers to the tax files can be checked during data processing of the returns. Surnames have also been compared elsewhere [e.g., 91, 95] by using procedures other than character-by-character checking.

6/ Both the CPS and SSA ages were calculated as of the end of 1972, using year of birth. (For the Pilot Link, age is as of the end of 1963.) Unfortunately, only two digits of the year of birth are available on the Summary Earnings File. Because of this, it was not always possible to distinguish between persons under 14 and those 100 years of age or older. Thus, we cannot be sure that the 85+ age class reflects the behavior of all sample persons with an SSA age over 84 years.

7/ Eligibility for retirement benefits begins at age 62, although most people do not retire until they reach the age of 65.

SELECTED BIBLIOGRAPHY ON THE MATCHING OF PERSON RECORDS FROM DIFFERENT SOURCES

This bibliography was selected principally from the following three sources:

> U. S. Bureau of the Census. Indexes to survey methodology literature, Technical Paper No. 34, 1974.
>
> Forsythe, J. List of References on Results and Methodology of Matching Studies, 1966 (Unpublished Census Bureau Memorandum).
>
> U. S. Public Health Service. Use of vital and health records in epidemiologic research, Vital and Health Statistics, series 4, no.7, 1968.

Scope

The references listed here are restricted essentially to published information on the results and methodology of matching person records. No material relating to establishment matching is included. Several compilations of articles exist on the confidentiality issues raised by record linkages. As a rule, therefore, we have excluded these citations from the bibliography. Two recent such compilations are:

> Report of the President's Commission on Federal Statistics, vol. 1, 1971, pp. 246-254.
>
> U. S. Department of Health, Education, and Welfare. Records, Computers, and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems, 1973, pp. 298-330.

Some other limitations should also be mentioned:

1. The listing does not contain citations to studies which began with an administrative record and then drew a sample of cases to be interviewed. (Excluded from the bibliography, therefore, are basically all reverse record check studies of financial characteristics, as well as prospective epidemiological studies.)

2. Only references to "exact" matching are included. Synthetic or "statistical" matching studies are not shown. (For citations to the literature on synthetic matches, see Radner, D. B. The Statistical Matching of Microdata Sets: the Bureau of Economic Analysis 1964 Current Population Survey--Tax Model Match, Yale Univ., New Haven, 1974. See also the Annals of Economic and Social Measurement, vol. 3, 1974, where several articles on synthetic matching are presented.)

Completeness of coverage

The bibliography's coverage of major U. S. studies involving linkages between survey (or census) schedules and administrative records is believed to be reasonably complete for the period 1950-1973. However, only a few references are given to work done outside the United States and to research engaged in before 1950.

FOOTNOTES

1/ This bibliography was compiled by Fritz Scheuren and Wendy Alvey with the assistance of Irene Jacobs, Office of Research and Statistics Library, Social Security Administration, and Patricia Fuellhart and her staff, Statistical Research Division, Bureau of the Census.

[1] Acheson, E. D. The Oxford Record Linkage Study, Report on the Second Year's Operations, Oxford Regional Hospital Board, Oxford, 1963.

[2] Acheson, E. D. Oxford Record Linkage Study, Brit. Jour. Prev. and Soc. Med., vol. 18, 1964, pp. 8-13.

[3] Acheson, E. D. The Oxford Record Linkage Study, a review of the method with some preliminary results, Proc. Roy. Soc. Med., vol. 57, 1964, pp. 11 ff.

[4] Acheson, E. D., and Evans, J. G. The Oxford Record Linkage Study, Biometrics, vol. 2, 1963, pp. 367 ff.

[5] Anderson, O. W., and Feldman, J. J. Family Medical Costs and Voluntary Health Insurance: a Nationwide Survey, McGraw-Hill, New York, 1956.

[6] Anderson, O. W., and Sheatsley, P. B. Comprehensive Medical Insurance, Health Information Foundation Research Series, No. 9.

[7] Anderson, R., and Anderson, O. W. A Decade of Health Services, Univ. of Chicago, Chicago, 1967.

[8] Bachi, R., Baron, R., and Nathan, G. Methods of record-linkage and applications in Israel, Bulletin of the International Statistical Institute, vol. 42, Proc. of 36th Session, Sydney, 1967, pp. 751-765.

[9] Bahn, A. K. Methodological study of population of out-patient psychiatric clinics, Maryland, 1958-59, Public Health Monograph No. 65., PHS Pub. no. 821, 1961.

[10] Balamuth, E. Health interview responses compared with medical records, U. S. Public Health Service, Vital and Health Statistics, series 2, no. 7, 1965.

[11] Bancroft, G. The American Labor Force: Its Growth and Changing Composition, Wiley, New York, 1958, pp. 151-175.

[12] Belloc, N. B. Validation of morbidity survey data by comparison with hospital records, J. Amer. Stat. Assn., vol. 49, 1954, pp. 832-846.

[13] Binder, S. Present possibilities and future potentialities, The Use of Vital and Health Records for Genetic and Radiation Studies, United Nations, New Ysrk, 1962, pp. 161-170.

[14] Bixby, L. E. Income of people aged 65 or older: overview from the 1968 survey of the aged, Social Security Bulletin, report no. 1, 1970, pp. 28-34.

[15] Brounstein, S. H. Data Record Linkage Under Conditions of Uncertainty, 7th Annual Conference of the Urban and Regional Information Systems Association, Los Angeles, California, 1969.

[16] Chase, H. C. A study of infant mortality from linked records: method of study and registration aspects, U. S. Public Health Service, Vital and Health Statistics, series 20, no. 7, 1970.

[17] Christenson, H. T., Cultural relativism and premarital sex norms, Amer. Soc. Rev., vol. 15, 1960.

[18] David, M., Gates, W., and Miller, R. Linkage and retrieval of microeconomic data. Heath, Lexington, Mass., 1974.

[19] Davidson, L. Retrieval of misspelled names in an airline passenger records system, Communications of the ACM, vol. 5, 1962, pp. 169-171.

[20] Densen, P. M., and Shapiro, S. Research needs for record matching, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1963, pp. 20-24.

[21] DuBois, Jr., N. S. D. On the problem of matching documents with missing and inaccurately recorded items (preliminary report), Annals of Mathematical Statistics, vol. 35, 1964, p. 1404.

[22] DuBois, Jr., N. S. D. A document linkage program for digital computers, Behavioral Science, vol. 10, 1965, pp. 312-319.

[23] DuBois, Jr., N. S. D. A solution to the problem of linking multivariate documents, Jour. Amer. Stat. Assn., vol. 64, 1969, pp. 163-174.

[24] Dunn, H. L. Record linkage, American Journal of Public Health, vol. 36, 1946.

[25] Dunn, H. L., and Grove, R. D. Completeness of birth registration in the United States, Estadistica, vol. 1, 1943, pp. 3-17.

[26] Elinson, J., and Trussell, R. E. Some factors relating to degree of correspondence of diagnostic information obtained by household interviews and clinical examinations, American Journal of Public Health, vol. 47, 1957, pp. 311-321.

[27] Fellegi, I. P., and Sunter, A. B. A theory for record linkage, Jour. Amer. Stat. Assn., vol. 64, 1969, pp. 1183-1210.

[28] Ferber, R. The Reliability of Consumer Reports of Financial Assets and Debts, University of Illinois, Urbana, 1966.

[29] General Register Office. 1951 Census of England and Wales, General Report, London, 1958, pp. 41-45, 5056.

[30] Goldberg, I. D., Goldstein, H., Quade, D., and Rogot, E. The use of vital records for blindness research, Amer. Jour. Pub. Health, vol. 54, 1964, pp. 278-285.

[31] Goldberg, S. Discussion of data storage and linkage, Bulletin of the International Statistical Institute, Proc. of 36th Session, Sydney, vol. 42, 1967, pp. 806-808.

[32] Grove, R. D. Studies in the completeness of birth registration, U. S. Bureau of the Census, Vital Statistics -- Special Reports, vol. 27, no. 18, 1943.

[33] Guralnick, L., and Nam, C. B. Census-NOVS study of death certificates matched to census records, The Milbank Memorial Fund Quarterly, vol. 37, 1959, pp. 144-153.

[34] Haber, L. D. Evaluating response error in the reporting of the income of the aged: benefit income, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1966.

[35] Haenszel, W., Loveland, D. B., and Sirken, M. G. Lung-cancer mortality as related to residence and smoking histories. I. white males, Journal of the National Cancer Institute, vol. 28, no. 4, 1962, pp. 947-1001.

[36] Hambright, T. Z. Comparability of age on the death certificate and matching census record, U. S., May-Aug. 1960, U. S. Public Health Service, Vital and Health Statistics, series 2, no. 29, 1968.

[37] Hansen, M. H. The role and feasibility of a national data bank, based on matched records and interviews, Report of the President's Commission on Federal Statistics, vol. 2, pp. 1-63.

[38] Hauser, P. M., and Kitagawa, E. M. Sscial and economic mortality differentials in the United States, 1960: outline of a research project. Proc. Amer. Assn. Soc. Stat. Sec., 1960, pp. 116-120.

[39] Hauser, P. M., and Lauriat, P. Record matching--theory and practice (abstract), Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1963, p. 25.

[40] Hedrich, A. W., Collison, J., and Rhoads, F. D. Comparison of birth tests by several methods in Georgia and Maryland, U. S. Bureau of the Census, Vital Statistics -- Special Reports, vol. 7, no. 60, 1939, pp. 681-695.

[41] Hoel, P. G., and Peterson, R. P. A solution to the problem of optimum classification, Annals of Mathematical Statistics, vol. 20, 1949, pp. 433-438.

[42] Hogben, L., Johnstone, M. M., and Cross, K. L. Identification of medical documents, British Med. J., 1948, pp. 632-635.

[43] Horn, W. Reliability Survey: a survey on the reliability of responses to an interview survey, Het PTT-bedrijf, vol. 10, 1960, pp. 105-156.

[44] Horn, W. Nonresponse in an interview survey: some specific phenomena (a case-study), Het PTT-bedrijf, vol. 12, 1963, pp. 11-19.

[45] Jaro, M. Unimatch--a computer system for generalized record linkage under conditions of uncertainty, Spring Joint Computer Conference, 1972, AFIPS--Conference Proceedings, vol. 40, 1972, pp. 523-530.

[46] Johnson, Jr., C. E. Consistency of reporting of ethnic origin in the Current Population Survey, Technical Paper No. 31, 1974, pp. 23-27.

[47] Kaplan, D. L., Parkhurst, E., and Whelpton, P. K. The comparability of reports on occupation from vital records and the 1950 census, Vital Statistics -- Special Reports, vol. 53, no. 1, 1961, pp. 1-27.

[48] Kelly, T. F. Factors affecting poverty: a gross flow analysis, The President's Commission on Income Maintenance Programs: Technical Studies, 1970, pp. 1-82.

[49] Kelly, T. F. The creation of longitudinal data from cross-section surveys: an illustration from the Current Population Survey, Annals of Economic and Social Measurement, vol. 2, 1973, pp. 209-214.

[50] Kelly, W. H. Methods and Resources for the Construction and Maintenance of a Navajo Population Register, a report prepared for the National Cancer Institute by the Bureau of Ethnic Research, Department of Anthropology, University of Arizona, Tucson, 1964.

[51] Kennedy, J. M. Linkage of Birth and Marriage Record Using a Digital Computer, Atomic Energy of Canada, Ltd., Chalk River, Ontario, 1961.

[52] Kennedy, J. M. The use of a digital computer for record linkages, The Use of Vital and Health Statistics for Genetic and Radiation Studies, United Nations, New York, 1962, pp. 155-160.

[53] Kitagawa, E. M., and Hauser, P. M. Methods used in a current study of social and economic differences in mortality, Emerging Techniques in Population Research, Milbank Memorial Fund, New York, 1963, pp. 250-266.

[54] Kitagawa, E. M., and Hauser, P. M. Differential Mortality in the United States: a Study in Socioeconomic Epidemiology, Harvard Univ., Cambridge, 1973, pp. 183-227.

[55] Klebba, A. J., Maurer, J. D., and Glass, E. J. Mortality trends: age, color, and sex, U. S. 1950-69, U. S. Public Health Service, Vital and Health Statistics, series 20, no. 15, 1973.

[56] Livingston, R. Evaluation of the reporting of public assistance income in the special census of Dane County, Wisconsin (May 15, 1968), Proc. Ninth Workshop on Public Welfare Research and Statistics, New Orleans, 1969.

[57] Livingston, R. Evaluating the reporting of public assistance income in the 1966 Survey of Economic Opportunity, Proc. Tenth Workshop on Public Welfare Research and Statistics, 1970.

[58] Lloyd, J. W. Discussion of matching of census and vital records in social and health research: problems and results, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, p. 141.

[59] Loeb, J. Weight at birth and survival of the newborn by age of mother and total-birth order, U. S., early 1950, U. S. Public Health Service, Vital and Health Statistics, series 21, no. 5, 1965.

[60] Mandel, B. J., Wolkstein, I., and Delaney, M. M. Coordination of old-age and survivors insurance wage records and the post -enumeration survey, Studies in Income and Wealth: an Appraisal of the 1950 Census Income Data, Princeton, vol. 23, 1958, pp. 169-178.

[61] Marks, E. S., Mauldin, W. P., and Nisselson, H. The post-enumeration survey of the 1950 censuses: a case history in survey design, Jour. Amer. Stat. Assn., 1953, pp. 220-243.

[62] Marks, E. S., and Waksberg, J. Evaluation of coverage in the 1960 Census of Population through case-by-case checking, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1966, pp. 62-70.

[63] Masi, A. T., Sartwell, P. E., and Shulman, L. E. The use of record linkage to determine familial occurrence of disease from hospital records (Hashimoto's disease), Am. J. Pub. Health, vol. 54, 1964, pp. 1887-1894.

[64] McCarthy, M. A. Comparison of the classification of place of residence on death certificates and matching census records, U. S. Public Health Service, Vital and Health Statistics, series 2, no. 30, 1969.

[65] Meltzer, J. W., and Hochstim, J. R. Reliability and validity of survey data on physical health, Public Health Reports, vol. 85, 1970, pp. 1075-1086.

[66] Miller, H. P., and Paley, L. R. Income reported in the 1950 census and on income tax returns, Studies in Income and Wealth: an Appraisal of the 1950 Census Income Data, Princeton, vol. 23, 1958, pp. 179-201.

[67] Moriyama, I. M. Uses of vital recsrds for epidemiological research, J. Chron. Dis., vol. 17, 1964, pp. 889-897.

[68] Morrison, F. S., and Blen, A. L. Method of testing in the 1950 census the completeness of birth registration, Estadistica, vol. 7, 1949, pp. 185-193.

[69] Murray, J. H., and Haber, L. D. Methodology and validation, Appendix A of the Aged Population of the United States: the 1963 Social Security Survey of the Aged, by L. A. Epstein and J. H. Murray. SSA/ORS research report no. 19, 1967, pp. 193-219.

[70] Nathan, G. On Optimal Matching Processes, Case Institute of Technology, Cleveland, 1964.

[71] Nathan, G. Outcome probabilities for a record matching process with complete invariant information, J. Amer. Stat. Assn., vol. 62, 1967, pp. 454-469.

[72] Neter, J., Maynes, E. S., and Ramanathan, R. The effect by mismatching on the measurement of response errors, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1964, pp. 2-8.

[73] Neter, J., Maynes, E. S., and Ramanathan, R. The effect of mismatching on the measurement of response errors, J. Amer. Stat. Assn., vol. 60, 1965, pp. 1005-1027.

[74] Newcombe, H. B. Detection of genetic trends in public health, Effect of Radiation on Human Heredity, World Health Organization, Geneva, 1957, pp. 157-168.

[75] Newcombe, H. B. Environmental versus genetic interpretations of birth-order effects, Eugenics Quarterly, vol. 11, 1964, p. 36ff.

[76] Newcombe, H. B. Panel discussion, session on epidemiological studies, Second International Conference on Congenital Malformations, The International Medical Congress, Ltd., New York, 1964, pp. 345-349.

[77] Newcombe, H. B. Pedigrees for population studies, a progress report, Cold Spring Harbor Symposia on Quantitative Biology, vol. 29, 1964, p. 21ff.

[78] Newcombe, H. B. Population genetics, population records, Methodology in Human Genetics, Holden-Day, New York, 1962, pp. 92-113.

[79] Newcombe, H. B. Risk of fetal death to mothers of different ABO and RH blood type, Am. J. Human Genet., vol. 15, 1963, pp. 449-464.

[80] Newcombe, H. B. Screening for effects of maternal age and birth order in a register of handicapped children, Ann. Human Genet., vol. 27, 1964, pp. 367-382.

[81] Newcombe, H. B. Untapped knowledge of human populations, Transaction of the Royal Society of Canada, vol. 56, series 3, section 3, 1962, pp. 173-180.

[82] Newcombe, H. B., Axford, S. J., and James, A. P. A plan for the study of fertility of relatives of children suffering from hereditary and other defects, Atomic Energy of Canada, Ltd., report no. 511, Chalk River, Ontario, 1957, p. 50.

[83] Newcombe, H. B., James, A. P., and Axford, S. J. Family linkage of vital and health records, Atomic Energy of Canada, Ltd., report no. 470, Chalk River, Ontario, 1957.

[84] Newcombe, H. B., and Kennedy, J. M. Record linkage making maximum use of the discrimination power of identifying information, Communications of the ACM, vol. 5, 1962, pp. 563-566.

[85] Newcombe, H. B., Kennedy, J. M., Axford, S. J., and James, A. P. Automatic linkage of vital records, Science, vol. 130, 1959, pp. 954-959.

[86] Newcombe, H. B., and Rhynas, P. O. Child spacing following stillbirth and infant death, Eugenics Quarterly, vol. 9, 1962, pp. 25-35.

[87] Newcombe, H. B., and Rhynas, P. O. Family linkage of population records, The Use of Vital and Health Statistics for Genetic and Radiation Studies, United Nations, New York, 1962, pp. 135-154.

[88] Newcombe, H. B., and Tavendale, O. G. Effects of father's age on the risk of child handicap or death, Am. J. Human Genet., vol. 17, 1965, pp. 163-178.

[89] Newman, S. M. Problem in mechanizing the search in examining patent applications, Patent Office Research and Development Report No. 3, 1956.

[90] Nicholson, W. The income data series in the graduated work incentive experiment: an analysis of their differences, Chapter 6, Part C, The New Jersey Graduated Work Incentive Experiment, 1974.

[91] Nitzberg, D. M., and Sardy, H. The methodology of computer linkage of health and vital records, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, pp. 100-106.

[92] Ohlsson, I. Merging of data for statistical use, Bulletin of the International Statistical Institute, vol. 42, Proc. of 36th Session, Sydney, 1967, pp. 751-765.

[93] Ono, M., Patterson, G. F., and Weitzman, M. S. The quality of reporting social security numbers in two surveys, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1968, pp. 197-205.

[94] Perkins, W. M., and Jones, O. D. Matching for census coverage checks, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, pp. 122-139.

[95] Phillips, Jr., W., and Bahn, A. K. Experience with computer matching of names, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1963, pp. 26-29.

[96] Phillips, Jr., W., Bahn, A. K., and Miyasaki, M. Person-matching by electronic methods, Communications of the ACM, vol. 5, 1962, pp. 404-407.

[97] Phillips, Jr., W., Gorwitz, K., and Bahn, A. K. Electronic maintenance of case registers, Public Health Reports, vol. 77, 1962, pp. 503-510.

[98] Pollack, E. S. Use of census matching for study of psychiatric admission rates, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, pp. 107-115.

[99] President's Committee to Appraise Employment and Unemployment Statistics. Measuring Employment and Unemployment, 1962, pp. 386-394.

[100] Rogers, P. B., Council, C. R., and Abernathy, J. R. Testing death registration completeness in a group of premature infants, Public Health Reports, vol. 76, no. 8, 1961, pp. 717-724.

[101] Schachter, J. Matched record comparison of birth certificate and census information: United States, 1950, U. S. Public Health Service, Vital Statistics -- Special Reports, vol. 47, no. 12, 1962, pp. 365-399.

[102] Scheuren, F. J., Bridges, B., and Kilss, B. Report no. 1: subsampling the Current Population Survey: 1963 Pilot Link Study, Studies from Interagency Data Linkages, Social Security Administration, 1973.

[103] Scheuren, F. J., and West, G. 1966 and 1967 Survey of Economic Opportunity: Computer Consistency Checks, Office of Economic Opportunity, 1971, pp. 113-124 (Processed).

[104] Schneider, P., and Knott, J. Accuracy of census data as measured by the 1970 CPS-Census-IRS Matching Study, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1973, pp. 152-159.

[105] Shapiro, S. Estimating birth registration completeness, J. Amer. Stat. Assn., vol. 45, 1950, pp. 261-264.

[106] Shapiro, S. Recent testing of birth registration completeness in the United States, Population Studies, vol. 8, 1954, pp. 3-21.

[107] Shapiro, S., and Densen, P. Research needs for record matching, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1963, pp. 20-24.

[108] Shapiro, S., and Schachter, J. Birth registration completeness, 1950, Public Health Reports, vol. 67, 1952, pp. 513-524.

[109] Shapiro, S., and Schachter, J. Methodology and summary results of the 1950 birth registration test in the United States, Estadistica, vol. 10, no. 37, 1952, pp. 688-699.

[110] Siegel, J. S. 1970 Census of Population and Housing Evaluation and Research Program, Estimates of Coverage of Population by Sex, Race and Age: Demographic Analysis, U. S. Bureau of the Census, PHC(E)-4, 1974.

[111] Siegel, J. S., and Zelnik, M. An evaluation of coverage in the 1960 census of population by techniques of demographic analysis and by composite methods, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1966, pp. 71-85.

[112] Silver, J. Discussion of matching of census and vital records in social and health research: problems and results, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, p. 140.

[113] Simpson, J. E., and Von Arsdol, Jr., M.D. The matching of census and probation department record systems, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, pp. 116-121.

[114] Sirken, M. G. Hospital utilization in the last year of life, U. S. Public Health Service, Vital and Health Statistics, series 2, no. 10, 1965.

[115] Sirken, M. G. Research uses of vital records in vital statistics surveys, Research Methods in Health Care, ed. by J. B. McKinlay, Prodist, New York, 1973, pp. 39-46.

[116] Sirken, M. G., Maynes, E. S., and Frechtling, J. A. The survey of consumer finances and the census quality check, Studies in Income and Wealth: an Appraisal of the 1950 Census Income Data, vol. 23, Princeton, 1958, pp. 127-168.

[117] Sirken, M. G., Pifer, J. W., and Brown, M. L. Design of Surveys Linked to Death Records, 1962.

[118] Steinberg, J. Some aspects of statistical data linkage for individuals, Data-Bases, Computers, and the Social Sciences, Wiley, New York, 1970, pp. 238-251.

[119] Steinberg, J. Some observations on linkage of survey and administrative record data, Studies from Interagency Data Linkages, Social Security Administration, 1973, pp. 1-14.

[120] Steinberg, J., Hearn, S., and Deutch, J. Social Security Administration's evaluation and measurement system, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1966, pp. 262-268.

[121] Steinberg, J., and Pritzker, L. Some experiences with and reflections on data linkage in the United States, Bulletin of the International Statistical Institute, vol. 42, 1967, pp. 786-805.

[122] Sunter, A. B. A statistical approach to record linkage, record linkage in medicine, Proc. of the International Symposium in Oxford, E. & S. Livingstone, Ltd., London, 1968.

[123] Tepping, B. J. Study of Matching Techniques for Subscriptions Fulfillment, Philadelphia, 1955.

[124] Tepping, B. J. Progress Report on the 1959 Matching Study, National Analysts Inc., Philadelphia, 1960.

[125] Tepping, B. J. A model for optimum linkage of record, J. Amer. Stat. Assn., vol. 63, 1968, pp. 1321-1332.

[126] Tepping, B. J. The application of a linkage model to the Chandrasekar-Deming technique for estimating vital events, Technical Notes 4, U. S. Bureau of the Census, Washington, D. C., 1971.

127] Tepping, B. J., and Bailar, B. A. Enumerator variance in the 1970 census, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1973, pp. 160-169.

[128] Tepping, B. J., and Chu, J. T. A Report on Matching Rules Applied to Reader's Digest Data, National Analysts, Inc., Philadelphia, 1958.

[129] Turner, Jr., M. L. A new technique measuring household change, Demography, vol. 4, 1967, pp. 341-351.

[130] U. S. Bureau of the Census. Infant enumeration study: 1950 completeness of enumeration of infants related to: residence, race, birth month, age and education of mother, occupation of father, Procedural studies of the 1950 Census, no. 1, 1953.

[131] U. S. Bureau of the Census The post-enumeration survey: 1950, Technical Paper, No. 4, 1960.

[132] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: Record Check Studies of Population Coverage, series ER60, no. 2, 1964.

[133] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by CPS-Census Match, series ER60, no. 5, 1965.

[134] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: the Employer Record Check, series ER60, no. 6, 1965.

[135] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses on Population and Housing, 1960: Effects of Interviewers and Crew Leaders, series ER60, no. 7, 1968.

[136] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: Record Check of Accuracy of Income Reporting, series ER60, no. 8, 1970.

[137] U. S. Bureau of the Census. 1970 Census of Population and Housing Evaluation and Research Program: Test of Birth Registration Completeness 1964 to 1968, PHC(E)-2, 1973.

[138] U. S. Bureau of the Census. 1970 Census of Population and Housing Evaluation and Research Program: the Coverage of Housing in the 1970 Census, PHC(E)-5, 1973.

[139] U. S. Bureau of the Census. 1970 Census of Population and Housing Evaluation and Research Program: the Medicare Record Check: an Evaluation of the Coverage of Persons 65 Years of Age and Over in the 1970 Census, PHC(E)-7, 1973.

[140] U. S. Bureau of the Census. 1970 Census of Population and Housing Evaluation and Research Program: the CPS-Census Match, PHC(E)-11, 1974.

[141] U. S. Public Health Service. Statistics of the United States, 1950, vol. 1, 1954, pp. 108-112.

[142] U. S. Public Health Service. Weight at birth and its effect on survival of the newborn in the United States, early 1950, Vital Statistics -- Special Reports, vol. 39, no. 1, 1954.

[143] U. S. Public Health Service. Reporting of hospitalization in the Health Interview Survey, Public Health Service Publication no. 584-D4. Vital and Health Statistics, series 2, no. 6, 1965.

[144] U. S. Public Health Service. Comparison of hospitalization reporting in three survey procedures, Vital and Health Statistics, series 2, No. 8, 1965.

[145] U. S. Public Health Service. Interview response on health insurance compared with insurance records, Vital and Health Statistics, series 2, no. 18, 1966.

[146] U. S. Public Health Service. A study of infant mortality from linked records, Vital and Health Statistics, series 20, nos. 12, 13, and 14, 1972-1973.

[147] U. S. Social Security Administration. Workers covered under Social Security: 1963 administrative information and pilot link results compared, Studies from Interagency Data Linkages, report no. 3 (In preparation).

[148] U. S. Social Security Administration. Individual income taxpayers: 1963 administrative information and pilot link results compared, Studies from Interagency Data Linkages, report no. 5 (In preparation).

[149] U. S. Social Security Administration. Report on Policies and Procedures for Establishing Initial Entitlement to RSDI Benefits, series 1, nos. 21-24, 1974.

[150] U. S. Social Security Administration. The 1% Sample Longitudinal Employee-Employer Data File, 1971 (Processed).

[151] Wells, B. Optimum Matching Rules, Univ. of North Carolina, 1974.

DISCUSSION

Roger Herriot, Census Bureau

The work discussed in the three papers given at this session represents the foundation of the 1973 Match Study. Taken together, the papers provide a clear and concise blueprint of the construction of that foundation and of its strength. I think that the authors -- and Fritz Scheuren who directed the work -- should be commended for the fine job they did. Although more remains to be done, soon we expect to have a file on which about 90 percent of the persons 14 years or older in the March 1973 Current Population Survey will have social security numbers that have a high probability of being correct. Furthermore, since some of the remaining 10 percent of the CPS sample have not received a number, we estimate that we will have SSN's for about 92 percent or more of the persons in the Study who actually have had a number issued to them.

Foundations have a tendency to get covered over as people begin to build upon them. Much that might be learned, which would be of use to future studies, can, of course, be lost because of this. Every effort has been made in the 1973 Match Study, as the papers at this session attest, to examine and reflect on the methods used with a view towards improving similar studies in the future. A number of suggestions along these lines can be found in the papers. Some of them would require major changes in the CPS that may be too expensive to implement, such as asking for father's name, mother's maiden name, and place of birth, etc. Other changes could be made at almost no cost, such as having the interviewers attempt to get the middle names or initials of all household members. Procedural changes at Social Security might also be made in future matching efforts. For example, as was pointed out in the second paper, the Study would have reaped significant benefits if there had been time to "re-search" for the missing SSN's of individuals for whom no number was found after just one search. Hopefully, in future studies, multiple searches will be possible, perhaps using somewhat more automated techniques.

In the remainder of my remarks, let me re-emphasize and expand on some of the points already made about the general nature and purpose of the 1973 Match Project. I would like to begin by stating categorically that this work is not being undertaken in anticipation of the development of any large-scale data bank. In fact, I believe that matched <u>sample</u> data sets, such as this one, constitute the best defense against the development of such a data bank for all persons. The reason is that such data, for a relatively small sample of persons, can fulfill virtually all of the statistical needs, without running the risks for abuse inherent in a population data bank.

STUDY GOALS

As has been said, the 1973 Match Study is a joint statistical project between the Social Security Administration and the Bureau of the Census. It involves an exact match for persons in the March 1973 CPS -- of their survey data, their Social Security earnings and benefits data, and a small amount of data from Federal income tax returns. The purposes of this project are as follows:

1. To use the administrative data to determine the correctness of the survey data and to ascertain any biases in the Bureau's statistics. For example, among the particular projects being planned are:

   (a) evaluation of the CPS reporting of income with respect to both recipiency and amounts;

   (b) evaluation of procedures used to allocate missing income amounts; and

   (c) evaluation of the weighting and control procedures used to adjust for noninterviews.

2. To use the administrative data to correct misreporting of the survey information and to augment the survey with data which the respondents were not asked to provide, such as work experience or income for a number of years. These data can then be used in many ways; for example,

   (a) as inputs to models which simulate the effect of proposed new tax-transfer programs or changes in existing programs;

   (b) to develop corrected income size distributions and to prepare estimates of after-tax income distributions; and

   (c) to test and evaluate various statistical matching procedures. This, we feel, is particularly important considering the increasing amount of statistical matching which is being undertaken.

CONFIDENTIALITY

The three-way data linkage required that special operating procedures be instituted at Social Security and at the Bureau of the Census to insure the confidentiality of the shared information--in particular, to ascertain that the linked data was used only for statistical purposes and not administrative ones. The laws and regulations under which the involved agencies operate impose very definite restrictions on such exchanges.

Census Requirements.--Information derived from the Bureau of the Census' Current Population Survey is governed by policies and procedures established under Title 13 of the U.S. Code. This title requires that information about identifiable individuals remain under the direct control of employees of the Census Bureau at all times. On rare occasions, to better achieve its statistical goals (such as in this linkage project), the Census Bureau swears in, as its own temporary employees, a small group of employees of other agencies. In this instance, those Social Security Administration employees directly involved in the linkage, about 15 or so, were hired and sworn in as Census employees without compensation. These few individuals--technically employees of both agencies at once--have been legally given the means for access to both Census Bureau and SSA data, so that the linkage could be performed. Both "regular" and "special" employees of the Census Bureau are, of course, sworn to uphold the confidentiality of all census information and are subject to criminal penalties should they fail to do so.

Social Security Administration Requirements.-- Information derived from Social Security Administration files is governed by Title II of the Social Security Act and the regulations established under that Act (specifically, Regulation No. 1, Sections 401 and 422). To release to the Census Bureau SSA earnings and benefit information for identifiable individuals, a special Commissioner's decision had to be obtained. This decision, dated June 28, 1973, was made subject to the following conditions:

1. No SSA information was to be given to the Census Bureau for any CPS respondent who refused to give his social security number to the Census interviewer.

2. All SSA data given to the Census Bureau were to continue to have the protected treatment required by the Social Security laws and regulations. Furthermore, the data were to be subjected to Census' own confidentiality restrictions as imposed under Title 13.

3. After linkage, all individual identifications must be removed or scrambled in the resultant file.

Internal Revenue Service Requirements.--The Internal Revenue Service (IRS), under an executive order (promulgated under IR Code Section 6103) provided a magnetic tape file of abstracts of 1972 individual income tax returns to the Census Bureau for statistical purposes. Subsequently, IRS agreed to permit the Census Bureau to match a very limited amount of this data to CPS and SSA information, subject to the following provisions:

1. that individually identifiable IRS data continue to be subject to Title 13 and the various IRS confidentiality restrictions (specifically, IRS Code Section 7213); and

2. that after matching and the removal of individual identifiers, IRS is to have veto power over any data item on subsequent match files to be prepared for SSA, if IRS believes that the inclusion of the data item could possibly result in disclosure.

As with Census data, unauthorized disclosure of Social Security or Internal Revenue Service information is a punishable offense which can result in fines or imprisonment or both.

A number of operational implications flow from the above confidentiality requirements. Some examples of the steps taken were:

1. when competing confidentiality regulations existed, the strictest provisions were followed;

2. project computer tapes, whether they contained linked data or not, were stored in locked facilities when not in use; and

3. all Census confidential information was in the custody of Census regular or special employees at all times.

As an added precaution, just limited abstracts of the basic CPS and SSA data were used in the searching and matching work done at Social Security.

LONGITUDINAL COMPARISONS AMONG SURVEY STATISTICS WITH SPECIAL REFERENCE TO THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Edward C. Bryant and Morris H. Hansen, Westat, Inc.

1. Introduction

The purpose of this paper is to discuss comparison of survey statistics across time, where the objective is to measure changes in knowledge, skills, and attitudes. While the problem may be encountered in a variety of settings, we are concerned primarily with measuring "progress" in the context of the title of the project, "National Assessment of Educational Progress" (NAEP).

That project administers "exercises" to a sample of 9-, 13-, and 17-year-old children and to young adults (26 to 35 years of age). The exercises are usually scored as correct or incorrect, but some are scored into multiple categories. The aggregative measure for an individual exercise is a "p-value," which represents the percentage of persons whose answers are correct, or whose answers fall into (let's say) a given attitudinal response category. The p-values are not aggregated into test scores for an individual person although, as we shall see, some aggregation across exercises for groups of people is implied in summary measures used for generalizations about progress (or lack of it) over time.

After the first administration, some of the exercises were "released," i.e., made public along with their p-values. The others were not disclosed and were retained for use in future administrations to measure change in knowledge, skill, or attitude (by comparing p-values). There is concern that released exercises might be picked up and "taught to" or used in teaching, and consequently not be acceptable for measuring change. In the following sections we discuss procedures and problems in obtaining comparability of exercises and scoring and in increasing the sensitivity of comparisons across time by adjustment for background factors.

2. Comparability of Exercises

For approximately the half of the exercises given during the first administration that were not released, the difference between the p-values at the second administration and the p-values at the first administration is a measure of educational progress. Such measures are, of course, subject to sampling error. They may also be subject to bias:

a. if the scoring standards change,

b. if changes in method of administration affect the p-values,

c. if there is a failure in security of unreleased exercises, and

d. if the exercises themselves are related to temporal issues.

We will discuss changes in scoring standards in the next section. An important objective of administration procedures is to minimize the effect of the method of administration on exercise outcome. Changes in procedures, either to improve them or to cut costs, will have the inevitable effect of reducing comparability. Whether it is feasible to introduce changes by splitting the sample into halves, one half receiving the old procedure and the other the new, depends upon cost and the availability of this feature in the design.

The security of unreleased exercises probably is not of major concern as long as NAEP can retain its isolation from political and funding issues. NAEP has sometimes been criticized because the earth doesn't tremble when its outcomes are reported. This probably fosters comparability across time.

As an example of the relationship of an exercise to temporal things, consider the question: "What is the name of the President of the United States?" Here, the question may remain the same, but the p-value of this exercise may be expected to change through the office-holding period and may be substantially different for a President who is much in the news than for a President who is less visible.

We turn now to the question of release of exercises after the second round. Some alternatives are:

<u>Alternative 1</u>. Retain the unreleased exercises from the first round indefinitely, using them at each round to measure progress.

This procedure would seem to produce the maximum comparability, but has some obvious shortcomings. First, it puts additional stress on the matter of security and, in case of any doubts about security, confounds the measures of progress with possible breaches in security. Second, there will be substantial interest in identifying the exercises that show extreme gains or extreme losses in p-values, but this cannot be done under this plan. One can, of course, identify the objectives with which the extreme gains or losses are associated, but this disclosure may not be satisfactory to the educational community. We believe that, as a practical matter, it would not be feasible to apply this procedure -- an effort to take this course is likely to prove exceedingly difficult if not impossible to follow.

<u>Alternative 2</u>. Release, at the second round, all exercises that were unreleased after the first round, retaining as unreleased the new exercises at the second round.

This plan retains all exercises for two rounds with half of them "expiring" and therefore being released at each round. Thus, a link with the previous round is always available with half of the exercises, and comparison with earlier rounds can be made by chaining the links. Implied in this approach is the ability to aggregate exercises or to pair them in some meaningful way.

Suppose one pairs an unreleased exercise at time t with a released exercise, where the pairing is made within common objectives and similar difficulty. We will refer to the set of possible exercises that are eligible to be paired as an exercise family. Then it is possible to measure "progress" with respect to a common objective for an exercise and all of its "ancestors" as follows.

Consider an exercise e that is administered at time t-1 and time t, and released after it is administered at time t, and let $p_{(t-1)ue}$ and p_{tre} be the p-values for the same exercise on the two dates, the subscript u denoting "unreleased" and r denoting "released." For convenience, we will let e represent an exercise used at times t and t-1 and all of its predecessors in the chain, which are presumed to be samples of the same exercise family. Then, a comparison of

$$\tilde{p}_{te} = p_{ore} \frac{p_{1re}}{p_{0ue}} \cdots \frac{p_{(t-1)re}}{p_{(t-2)ue}} \cdot \frac{p_{tre}}{p_{(t-1)ue}} \qquad (1)$$

with p_{ore} indicating gain (or loss) over time, subject, or course, to a sampling error. However, if time-linked p-values in a given subject matter show an upward movement in their distribution, measured perhaps by the median of the linked p-values, one might have some confidence that gain or progress was being demonstrated. If we assume that the exercises used in the chain are independently sampled from an exercise family, and that the persons to whom the exercises are administered are independently selected samples, the relative sampling error of r, the ratio of the time-linked p-value of released exercises at time t ($\tilde{p}_{tre}$ to p_{ore}) is approximately

$$\sigma_r^2 = 2tR^2V^2(1 - \rho)$$

where

R = the expected value of r ;

t = the number of chained time comparisons, that is, the number of ratios of the type $r_t = p_{tre}/p_{(t-1)ue}$ in estimate (1);

$V^2 = V_b^2 + \frac{V_w^2}{t}$ is here assumed to be approximately constant over time;

V_b^2 = the relvariance between the expected p-values for the exercises in the exercise family;

V_w^2 = the average relvariance among students within exercises for the exercise family; and

$\rho = (V_{b12})/V^2$ with V_{b12} equal to the relcovariance of expected p-values for an exercise in two adjacent years.

The definition of these terms might be made clearer by the following illustration for two adjacent years, and with the simplifying assumption that the students taking the exercise are a simple random sample of eligible students. The observed p-value for exercise e at time t is p_{te} and its expected value is P_{te} .

Exercise (e) Within the Family	Expected p-value for Individual Exercises in the Family in Year	
	1	2
1	P_{11}	P_{21}
2	P_{12}	P_{22}
3	P_{13}	P_{23}
4	.	.
.	.	.
.	.	.
M	P_{1M}	P_{2M}

(We illustrate M as finite, but it may be regarded as indefinitely large.)

The average relvariances and relcovariances above are further defined as follows:

Average:

$$P_{t.} = \sum_{e}^{M} P_{te}/M$$

Relvariance between expected exercise p-values (for simplicity, assumed to be approximately equal over time):*

$$V_b^2 = \sum_{e}^{M} (P_{te} - P_{t.})^2/MP_{t.}^2$$

Relvariance within exercises (for simplicity assumed to be approximately equal over time):*

$$V_w^2 = \sum_{e}^{M} P_{te}Q_{te}/MP_{t.}^2$$

Relcovariance of expected p-values:

$$V_{b12} = \frac{\sum_{e}^{M} (P_{(t-1)e} - P_{(t-1).})(P_{te} - P_{t.})}{MP_{(t-1).}P_{t.}}$$

* Actually, if $P_{2.}$ is substantially greater or less than $P_{1.}$, it may be unreasonable to assume that the V_b^2 and V_w^2 are equal, but the amount of such variation in V_w^2 is not important for our present purposes.

Estimation of variances would call for use of two or more exercises from an exercise family at each time.

Alternative 3. Create a large pool of exercises, stratified by objective and (possibly) difficulty, and develop a rotation plan for adding (and releasing) a scheduled proportion of exercises at each round.

This approach differs from Alternative 2, above, among other reasons, because it involves creating a substantial pool of exercises and doing stratified random selection from the pool. In Alternative 2, we simply assumed that presumably comparable exercises are identified and chained in subsequent tests, without necessarily creating a family of exercises in advance and using an explicit random procedure for selection of exercises. We believe Alternative 3 (or a modification of procedures along these lines) has important advantages.

There is clearly an assumption of aggregation of exercises in this approach. That is, one assumes that there is a parameter representing change in knowledge, skills, or attitudes in a given subject matter, which can be estimated by average exercise scores. The estimate would have two components. The first component would be comprised of the difference in estimated average p-values of identical exercises given at two dates, presumably using weighted averages. The second component would consist of differences in average p-values of exercises that were different at the two dates, but that were drawn from the same exercise pool. Thus, an overall estimate of gain (or loss) could be expressed as

$$\bar{d} = z\bar{d}_c + (1 - z)\bar{d}_u \qquad (2)$$

where z is chosen so as to minimize the variance of $\bar{d}$, $(0 \leq z \leq 1)$,

$$\bar{d}_c = \sum_j w_j (P_{ctj} - P_{c(t-1)j}) \;, \qquad (3)$$

P_{ctj} denotes the p-value for the j^{th} common exercise at time t , the w_j are assigned weights, and

$$\bar{d}_u = \frac{\sum_j w'_j P_{utj}}{\sum_j w'_j} - \frac{\sum_j w''_j P_{u(t-1)j}}{\sum_j w''_j} \qquad (4)$$

where P_{utj} is the p-value of the j^{th} uncommon exercises at time t .

The variance and covariance estimates could be developed along the lines discussed above.

3. Comparability of Scoring

As we have pointed out before [1, 3], it is not a simple matter to maintain comparability of scoring across time for subjective exercises such as writing exercises or performance exercises -- playing of musical instruments, singing, etc. One way to obtain comparability is to have exercises given earlier rescored by persons who are scoring current performance. Then, assuming proper control can be maintained on other factors, one can obtain a set of p-values on exercises common to both administrations that are comparable. It may be necessary to photocopy writing exercises, for example, to get materials of equal readability.

4. Adjustments to Increase Comparability Across Time

Inherent in the adjustment process is the effort to define reasonably homogeneous subgroups of the population whose performance or outcome at a subsequent time can be compared with that of a similar group at an earlier time, such that if changes in average performance have occurred, it is reasonable to infer that the change reflects some real changes in performance and not simply changes in the composition of the subgroup. Such a subgroup might be "13-year-old Southern rural black males, neither parent completed high school." We assume that the characteristics that define the group (age, sex, geographic region, urbanization, education of parents, and race) have relatively stable definitions over time and that the classification of an individual as either in the group or not in the group is substantially error-free. Of the characteristics listed in the example, only degree of urbanization and education of parents are subject to change over time, and that change is likely to be slow enough for a five-year period that one need not be greatly concerned about such change on the comparability of classification.

The fact that one can identify similar groups over time is quite important. There are other variables, however, such as community and school variables, occupation of parents, items in the home, and other indicators of socioeconomic status (SES) that have been found to be useful in "adjusting" educational outcomes in order to make comparisons among population subgroups at a given date [3]. These variables typically do not remain stable over time -- a $10,000 income in 1974 does not represent the same thing as a $10,000 income in 1970, owning a color TV set does not have the same meaning in 1974 as it had in 1970, and so on. The question we address here is whether such background factors can be used successfully in increasing comparability of outcome measures across time.

The comparison problem can be explained with reference to the cells of the following table:

	Group 1	Group 2	Both Groups
Time 1	P_a	P_b	P_{a+b}
Time 2	P_c	P_d	P_{c+d}

in which the appropriate p-value is shown in the table.

During any one administration of the tests, one is interested in comparing one group with another, let's say, in comparing P_c with P_d.

The fact that NAEP usually compares one group against total U.S. performance, which includes the group of interest (i.e., P_c with P_{c+d}), tends to reduce the magnitude of the differences, but does not change the fundamental nature of the comparison. In NAEP, one is also interested in comparing performance of a defined group at time t with performance at time t-1, e.g., P_c with P_a, or P_d with P_b. These constitute measures of group gain and, of course, one is also interested in measuring overall gain, e.g., comparing P_{c+d} with P_{a+b}. Finally, one is interested in comparing gains by groups, such as $P_c - P_a$ with $P_d - P_b$.

Most of the literature has concerned itself with reducing bias in comparisons at one point in time, such as P_a with P_b or P_c with P_d. Adding the time dimension introduces some complexities, as we shall see.

We consider first the comparison of group means without adjustments for other variables. Let $\bar{y}_{gt}$ denote the mean outcome score (in National Assessment terms, a p-value) for group g at time t. Then, assuming that problems in comparability of exercises and in comparability of scoring have been solved (see above), the difference between the group means

$$d_g = \bar{y}_{gt} - \bar{y}_{g(t-1)} \tag{5}$$

is a meaningful measure of the gain in achievement for group g if the group is reasonably comparable at both dates, as discussed earlier.

An unadjusted aggregate measure of change for all groups combined is

$$\bar{d} = \bar{y}_t - \bar{y}_{(t-1)} \tag{6}$$

where

$$\bar{y}_t = \sum_g w_{gt}\bar{y}_{gt} \tag{7}$$

and the weights w_{gt} are the appropriate population or sampling weights at time t.

A problem with this comparison is that, even though there are no changes in the individual group averages $\bar{y}_g$, the $\bar{d}$ may show a significant change simply because the proportion of the population in the various groups is changed, that is, because w_{gt} is not equal to $w_{g(t-1)}$.

A simple and widely used adjustment procedure is to adopt a common set of weights, w_g, to apply to the group means at each period of time [2]. Thus, an adjusted measure of change and adjusted means are given by

$$\bar{d}' = \Sigma\, w_g\bar{y}_{gt} - \Sigma\, w_g\bar{y}_{g(t-1)} = \Sigma\, w_g\bar{d}_g . \tag{8}$$

The common weights to be used may be the sampling or population weights at time t or time (t-1), some average of them, or may be chosen from some external source. Their choice is somewhat arbitrary, but they should be chosen in a rational way for the purpose of the comparison.

We are now ready to consider adjustment for background factors that are multivalued and scaled, or continuous variables, and that may provide unstable background measures over time. For purposes of the exposition, we will assume that there is a simple linear relationship between the outcome y and a background measure x, where it is to be understood that x may be a composite measure constructed by regression or other methods and the result of various linearizing or normalizing transformations. The linear relationship is expressed as

$$y'_i = \mu + \beta x_i + \varepsilon_i \tag{9}$$

where, it is assumed, ε_i are random residuals with mean zero and, at least for the present, we assume x_i is measured without error.

It is convenient to talk about various cases that may arise, based upon assumptions concerning the stability over time of the background variable x and whether or not the regression is common to all groups.

Case 1. The distribution of x has not changed between time (t-1) and time t; in particular, $\bar{X}_{(t-1)} = \bar{X}_t$ where these are the true means at the two dates and

$$\sigma_{X_1} = \sigma_{X_2} .$$

There is common regression across all groups, i.e., $\beta_g = \beta$ and this common regression has not changed over time.

This is a simple case. One can simply adjust the differences in group means at the two times according to the methodology used in regression estimation applied to sample data. Note that

$$\bar{y}'_{gt} = \bar{y}_{gt} + b(\bar{X} - \bar{x}_{gt}) \tag{10}$$

An estimate of the adjusted difference between outcomes for group g between the two dates is as follows:

$$d'_g = \bar{y}_{gt} - \bar{y}_{g(t-1)} + b(\bar{x}_{g(t-1)} - \bar{x}_{gt}) \tag{11}$$

Note that the parameter $\bar{X}$ need not be known since it subtracts out. The adjustment clearly has the effect of reducing sampling error. An aggregate estimate of the difference across all groups is provided by

$$d' = \sum_g w_g d'_g \tag{12}$$

where w_g is chosen as before.

Clearly, one is faced with a dilemma if he is not quite sure that $\bar{X}_{(t-1)} = \bar{X}_t$, $\sigma_{X_1} = \sigma_{X_2}$, and that the regression coefficients are approximately equal. He can test the hypotheses of equality, but then he becomes involved in the interpretation of sequential tests of hypotheses, and the case is no longer a simple one.

Before proceeding, it may be worth noting that one can adjust for auxiliary variables either by "adjustment by subclassification," as described by Equations (5) and (8), or by regression methods. The method of subclassification is algebraically equivalent to regression when dummy variables (1 or 0) are assigned for each subclass. Cochran [2] has shown that by breaking up "continuous" x variables into classes, one can adjust for major portions of the bias in group comparisons. For monotonic relationships between x and y, his analytical results suggest that one can remove from 64% to 92% of the bias in $\bar{y}$ by using from two or six classes of x. The method is particularly good when one is uncertain of the relationship between x and y. This may be particularly important in NAEP adjustments since most outcome measures are dichotomous.

It may also be worth noting that, for purposes of adjustment by subclassification, one can tolerate relatively small average frequencies in the adjustment classes since the increased sampling error of small classes is offset by the decreased weight given to each class. (It is only necessary to ensure that an unusually small class does not get a relatively large weight.)

In any case, assuming a regression adjustment for auxiliary variables, whether continuous or not, is simply a convenience in the presentation.

It should also be observed here that a shift in the distribution of the auxiliary variable x can sometimes be adjusted for by a deflator that is external to the survey itself. An obvious example is use of the Consumer Price Index to deflate income. There may also be other deflators that are not commonly used, such as Census estimates of the proportion of persons in specified age groups who have completed high school. It is conceivable that such a deflator could be used to adjust for educational level of parents when t and $t-1$ are widely separated. Also, it is possible that one should concentrate on finding measures of SES that remain relatively stable over time (such as educational attainment of parents) rather than more volatile measures (such as items in the home).

Case 2. The distribution of x has not changed from time 1 to time 2; there is a separate regression within each group that has not changed.

One only needs to replace b by b_g in expressions (10) and (11). There are no further complications.

Case 3. The distribution of x has changed; there is a separate regression within each group that remains constant on the normalized value of x.

Suppose the x variable is income, which can be presumed to change over time. However, it may be that the regression of y on $(x - \bar{X}/\sigma_x)$ will remain constant over time. If so, one can adjust each group mean at time t by

$$d'_g = \bar{y}_{gt} + b_g(\bar{x} - \bar{x}_{gt})/S_x \quad (13)$$

where S_x is the sample estimate of σ_x and the adjusted gain in outcome can be expressed as

$$d'_g = \bar{y}_{gt} - \bar{y}_{g(t-1)} + b_g(\bar{x}_{g(t-1)} - \bar{x}_{gt})/S_x \quad (14)$$

In this case, as with the earlier cases, there is the problem of determining whether the regression coefficients have remained fixed and, if they have, how they should be estimated.

What one accomplishes by this adjustment perhaps should be discussed further. Let us suppose that a particular group had an average SES measure at time 1 that fell at the 37th percentile of the national SES distribution. At time 2, their average SES measure may have moved up to (say) the 39th percentile, or (say) down to the 33rd. The adjustment represented by Equation (13) adjusts the average of the outcome measure upward or downward accordingly. Thus, it presumes that the changes in the SES measure are a result of changes in the measurement process, and that equivalent percentile ranks at the two dates identify equivalent SES groups for adjustment purposes.

The interpretation of the adjusted gain is important. It seems evident that, if one's interest lies in evaluating the educational process (both in and out of school), one might very well make such an adjustment because it tends to free the estimate of gain from the gain in the SES measure. However, if one is interested in using educational outcome as a measure of social gain, then it seems inappropriate to make such adjustments. This same principle holds, of course, in all of the adjustments discussed here.

Case 4. The distribution of x has changed; there is a separate regression for each group that has changed and cannot be stabilized by normalization.

This case does not appear to lend itself to adjustment, although some gains might be achieved by assuming one of the simpler models if departures from those assumptions are minor.

There are, of course, other cases, but the ones discussed above appear to be of most interest to NAEP.

5. Additional Comments on Data Adjustment

Much of the interest in adjustment of survey data stems from the desire to infer cause from observed effects in nonexperimental situations. In many social science evaluative studies, it is impossible, within a political system that recognizes rights of individuals, to experiment with human beings, and often it is unwise or terribly expensive to do so in other cases. Also, even though experimentation might be feasible, the time required for the experiment to run its course may be so great that retrospective surveys are employed. In such cases, one generally tries to accomplish a partitioning of the variation in the outcomes into portions "due to" various characteristics of the observational units, their environments, or the processes to which they have been exposed, or to compare sets of outcomes after such partitioning. (The words "due to" are not to be interpreted as implying cause and effect.)

Sometimes the partitioning of the variation in outcomes is the key analytical result, and a statement such as "Fifty percent of the variation in outcome is accounted for by Factor X" will lead to the conclusion that Factor X needs to be modified through intervention of some kind. Note that cause and effect cannot be inferred from the mathematical statement, but are implied by the decision to intervene. This is a typical exercise in retrospective surveys and has been discussed ably by Dorn [4].

In cases where the analytical result is a comparison of outcomes of two or more groups, focus is usually on controlling bias, i.e., by statistical adjustment for confounding variables. Cochran and Rubin [5] examined, under an assumed linear model, some of the common procedures that have been used, including linear regression adjustment and several matching procedures. Not surprisingly, linear regression adjustment proved to be superior over matching when the linear model with parallel regressions was used and declined in relative merit with respect to "category-matching" with departures from the linear, parallel regression model. Their concept of category-matching is essentially equivalent to our unadjusted and adjusted comparison of groups represented by Equations (5) and (8) above. Their category-matching followed by regression adjustment corresponds closely to procedures we have discussed under Cases 1 through 3.

McKinlay [6] investigated methods for removing bias where the outcome is dichotomous (generally the situation with NAEP data) and the covariate is continuous. Methods investigated were pair-matching and stratification of the covariate. A Monte Carlo analysis of these procedures showed that, for the simulation models studied, pair-matching did not appear to be more effective than stratification. The group comparisons we have discussed in the previous section are quite similar to the concept of stratification on auxiliary variables.

Earlier work, as well as a number of recent studies, have been well summarized by McKinlay [7], and we will not attempt to discuss that work here.

References

1. Morris H. Hansen and Edward C. Bryant, "National Assessment Design Implications," December 1972 meetings of the American Association for the Advancement of Science, Washington, D.C.

2. W.G. Cochran, "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," Biometrics, 24 (June 1968).

3. Edward C. Bryant, Ezra Glaser, Morris H. Hansen, and Arthur Kirsch, "Associations Between Educational Outcomes and Background Variables: A Review of Selected Literature," under contract to Education Commission of the States, 300 Lincoln Tower, Denver, Colorado, 1974 (Monograph of the National Assessment of Educational Progress).

4. Harold F. Dorn, "Philosophy of Inferences from Retrospective Studies," American Journal of Public Health (June 1953), 677-683.

5. William G. Cochran and Donald B. Rubin, "Controlling Bias in Observational Studies: A Review," available from Donald B. Rubin, Educational Testing Service, Princeton, New Jersey 08540.

6. Sonja M. McKinlay, "Removing Bias Due to a Continuous Covariate from a Dichotomous Response in Pair-Matched and Stratified Samples," Department of Mathematics, Boston University, Boston, Massachusetts.

7. Sonja M. McKinlay, "The Design and Analysis of the Observational Study -- A Review," Harvard University Medical School, Boston Massachusetts (in preparation).

ADJUSTING EDUCATIONAL SURVEY DATA

Robert C. Larson and Donald T. Searls

National Assessment of Educational Progress
The Education Commission of the States

1. ASSESSING EDUCATIONAL ACHIEVEMENT

When the United States Office of Education was founded in 1867, one charge set before its commissioner was to determine the nation's progress in education. Almost 100 years later, the National Assessment of Educational Progress began charting educational change under the guidance of Dr. Ralph Tyler and Dr. Francis Keppel. The project has now grown to where five years of field assessment have already been completed.

National Assessment is a project of the Education Commission of the States and was established to measure educational achievement. The project's goal is to provide reliable information describing what young Americans (at the ages of nine, thirteen, seventeen and twenty-six to thirty-five) know and can do. Specifically, the assessment is designed (1) to obtain census-like data on the knowledge, skills, and attitudes at regular intervals and (2) to measure the growth or decline in educational attainments.

Presently we have completed a first assessment in Citizenship, Reading, Literature, Music, Social Studies, Mathematics and Career and Occupational Development, and have completed second assessments in Science and Writing. The tasks included in each assessment have been judged important by representative panels of scholars, laypersons and educators and represent things that should be taught in American schools.

The reporting categories National Assessment uses were selected because they reflect groups where differences in achievement occur in the population. The four age levels essentially mark the end of primary, intermediate, secondary and post-secondary education. For each age we report results for groups defined by region, sex, color, level of parents' education and size-and-type of community. Of these, region and sex groups have traditionally shown large differences in educational attainments; schools are thought to vary with the size and type of community (STOC) they serve; and color and level of parental education (PED) are believed to differentiate socio-economic and home and family environments.

An estimate of the percent of people who can perform a task or group of tasks is the basic measure of educational achievement used by National Assessment. The difference between the percentage of a group and the percentage of people in the nation at that age who can perform the task is called the "group effect."

Observed group effects estimate achievement levels of subpopulations such as Southeast or Blacks as they exist in our country. Interpretations of observed group percentages, however, can be misleading in several ways. The fact that a group's relative performance is labeled as a Northeast or Southeast regional "effect" does not mean that differences in these effects occur solely because the respondents live in the Northeast or Southeast. For example, a large fraction of respondents in the Northeast live in large cities while a larger fraction of respondents in the Southeast live in rural areas. Consequently, size and type of community effects may be masquerading as part of an observed regional effect. Similarly, persons whose parents went beyond high school are more frequent in affluent communities than in the country as a whole and persons whose parents had no high school are more frequent in extreme rural communities. In this case parental education effects may be masquerading as size-and-type of community effects.

Confusion about group effects due to masquerading arise when the mixture of characteristics among groups are unbalanced. Since most of the groups are in fact distributed disproportionately in the national population, our weighted probability sample automatically ensures this imbalance in our percentage estimates.

Demands are continually placed on education to find solutions to difficult problems and provide a better education for everyone. At the same time, the pace of changing views and emerging opinion about the nature and the solutions to problems facing education far outdistances the aggregation of supporting data. Unfortunately, this pressure increases the tendency to go beyond the capability of observational data and decreases resistance to making inferences about causes which are at best uncertainties.

Much has been written about the differences between observation and experimentation and the danger of inferring cause from observational studies. Not much has been said, however, about the need to combine skilled observation with incisive analysis in order to generate suggestions for experimental studies. In his lecture on social experimentation, Frederick Mosteller (3) pointed out that "...we need both mechanisms, one to generate suggestions that might lead to improvement, the other to eliminate most suggestions as ineffective."

Analysis and reanalysis of observational data cannot dispel uncertainties about causes, but they can help us gain new perspectives, new hypotheses and perhaps a better understanding of the data. For example, a statistical adjustment which balances the distribution of group characteristics may lead to a better understanding of differences among group effects. This is the basic purpose for performing a "balanced analysis" of National Assessment data.

2. THE BALANCING PROCEDURE

Sample survey data like the kind National Assessment has been collecting over the last five years is frequently adjusted in the sense of forcing sample frequencies or sample ratios to agree with population figures that are known from other sources (See Deming (1)). This is a desirable procedure because presumably the adjusted sample represents the population better and sampling variability of the adjusted frequencies is reduced to some extent.

As a result of this adjustment, the disproportionate distributions are not greatly altered, merely refined. For our purpose, however, we need to go beyond this. It is easier to think about a single marginal group effect if the distribution of the other factors in that group are represented in the same proportions as in the whole age population. Then a direct comparison of any two marginal groups is unconfounded with differing mixtures of these other variables.

Many analyses or data adjustments are possible and specific concerns often direct choices for adjusting data. John Tukey (5) specified a data adjustment procedure which defined marginal effects to be non-sensitive to simple disproportionate distributions of other variables. For reasons which favor reader/listener understandability rather than statistical efficiency Tukey chose to fit an additive linear model to the margins. This led to the "conditions for balance" where the observed number of successes equals the fitted (balanced) number of successes for each marginal group. These conditions for balance were first written by Tukey as follows:

$$(2.1)\quad \Sigma n_{ijk\ell m}(P_{nat} + \Delta P_i + \Delta P_j + \Delta P_k + \Delta P_\ell + \Delta P_m) = \Sigma C_{ijk\ell m}$$

The sum in (2.1) is taken in turn over all indices except one thereby generating a set of 21 equations. Each equation corresponds to a group denoted by one value of the indices $i = 1, \ldots, 4$; $j = 1,2$; $k = 1, \ldots, 7$; $\ell = 1,2,3$; and $m = 1, \ldots, 5$ belonging to Region, sex, STOC, color and PED respectively, and where:

- P_{nat} is the overall national percent correct for the age group;
- ΔP is the balanced group effect corresponding to each value of the indices;
- $N_{ijk\ell m}$ is the weighted number of observations in each 5-way cell; and
- $C_{ijk\ell m}$ is the weighted number of correct responses in each 5-way cell.

A solution to the 21 simultaneous equations gives a set of fitted group effects. These effects were designed so that when added together and with the national percentage, and when multiplied by the actual number of cases they would give a fitted number of successes equal to the observed number of successes. This set of equations is, however, not of full rank and cannot yield unique balanced ΔP's directly. The number of linearly independent equations is 16 but can be increased by appropriately replacing 5 of the 21 equations with the following usual side conditions imposed when an additive linear model is fitted to a multi-way crossed classification

$$\sum_i n_{i\cdots\cdot}\Delta P_i = \sum_j n_{\cdot j\cdots}\Delta P_j = \sum_k n_{\cdot\cdot k\cdot\cdot}\Delta P_k = \sum_\ell n_{\cdots\ell\cdot}\Delta P_\ell = \sum_m n_{\cdots\cdot m}\Delta P_m = 0$$

of the data where the dot "·" notation denotes the sum over the replaced subscript.

A solution of the independent set of equations results in a unique set of balanced group ΔP's. Exhibit 1.1 provides a simple example of the balancing equations and the computation of fitted group effects.

Tukey's balancing equations can be shown to be algebraically equivalent to the usual set of normal equations resulting from minimizing the error sum of squares for a five factor additive linear model

$$(2.2) \quad Y_{ijk\ell m} = \hat{P}_{nat} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\theta}_\ell + \hat{\phi}_m + e_{ijk\ell mr}$$

where $Y_{ijk\ell m}$ is the weighted response for the r-th person in the ijkℓmr-th cell, and α, β, γ, θ, and φ correspond to the ΔP's with the same subscript and $e_{ijk\ell mr}$ is random error associated with the r-th person. The solution to the normal equations provides simple least squares estimates of the group effects. If the variability of the percentages in the five-way cell combinations happen to be proportional to the reciprocal of the corresponding cell weights, then these group effects are minimum variance estimates.

Since the original conditions for balance are preserved in the algebraic equivalence, balanced group effects are also least squares estimates.

3. EXAMINING THE RESULTS OF BALANCING

An understanding of balanced group effects can be guided by thinking of a "conceptual" balanced population where the mixture of characteristics in each group is the same as the mixture in the age population. For example, in 1972 approximately 43% of all thirteen-year-olds in the nation had at least one parent who was educated beyond high school. In the Southeast, however, only 29% had at least one parent educated beyond high school. In the "conceptual" thirteen-year-old balanced population, the Southeast region would have effectively the same proportion of children with at least one parent having a post-high school education as the nation.

Consider the result of balancing. If persons having parents with post-high school education do well and are more frequent in one region than in others then we would expect that the balanced effect of that region to be less than its unadjusted effect. In contrast, if persons having parents with no high school education do poorly and are more frequent in another region, then we would expect that balanced effect to appear better than the unadjusted effect. If the magnitude of a group effect is generally reduced by the balancing adjustment, one might conclude that the factor named by the group label itself is not what "accounts" for the observed differences as much as the unbalanced representation of these other variables.

The proportion of an observed group effect that can be attributed to imbalance of other variables can be shown by considering two of the normal equations. First consider the normal equation obtained from (2.2) by summing over all subscripts:

$$(3.1) \quad n_{.....}\hat{P}_{nat} + \sum_i n_{i....}\hat{\alpha} + \sum_j n_{.j...}\hat{\beta}_j + \sum_k n_{..k..}\hat{\gamma}_k + \sum_\ell n_{...\ell.}\hat{\theta}_\ell + \sum_m n_{....m}\hat{\phi}_m = \sum_{ijk\ell m} n_{ijk\ell m} P_{ijk\ell m}.$$

Note that the right-hand side is equal to the observed weighted number of successes for an age group. Dividing both sizes by $n_{.....}$

$$(3.2) \quad \hat{P}_{nat} + \sum_i \frac{n_{i....}}{n_{.....}}\hat{\alpha}_i + \sum_j \frac{n_{j....}}{n_{.....}}\hat{\beta}_j + \sum_k \frac{n_{..k..}}{n_{.....}}\hat{\gamma}_k + \sum_\ell \frac{n_{...\ell.}}{n_{.....}}\hat{\theta}_\ell + \sum_m \frac{n_{....m}}{n_{.....}}\hat{\phi}_m = \frac{\text{wted \# successes}}{\text{wted \# cases}}.$$

Since terms 2-6 of the left hand side are all equal to zero, then $\hat{P}_{nat}$ equals the ratio of successes to cases as expected. Note that the ratio of group weights to total weight in these five terms are the marginal group proportions in the observed populations. Now consider a normal equation corresponding to the summation over all subscripts except one, say i, then

$$(3.3) \quad n_{i....}\hat{P}_{nat} + n_{i....}\hat{\alpha}_i + \sum_j n_{ij...}\hat{\beta}_j + \sum_k n_{i\cdot k..}\hat{\gamma}_k + \sum_\ell n_{i..\ell.}\hat{\theta}_\ell + \sum_m n_{i...m}\hat{\phi}_m = \sum_{jk\ell m} n_{ijk\ell m} P_{ijk\ell m}.$$

The right hand side is equal to the weighted number of successes in group i, and $n_{i....}$ is the weighted number of cases in group i. Dividing both sides by $n_{i....}$ and subtracting $\hat{P}_{nat}$ gives

$$(3.4) \quad \hat{\alpha}_i + \sum_j \frac{n_{ij...}}{n_{i....}}\hat{\beta}_j + \sum_k \frac{n_{i\cdot k..}}{n_{i...}}\hat{\gamma}_k + \sum_\ell \frac{n_{i..\ell.}}{n_{i....}}\hat{\theta}_\ell + \sum_m \frac{n_{i...m}}{n_{i....}}\hat{\phi}_m = \frac{\text{wted \# successes group i}}{\text{wted \# cases group i}} - P_{nat}.$$

The right hand side is the observed

group i effect, and $\hat{\alpha}_i$ is the balanced group i effect. In a balanced population the marginal proportions of the other groups in group i are equal to the corresponding proportions in the population as shown in equation (3.5). That is, the proportion in (3.4) are equal to their corresponding proportions in (3.2) as shown below:

$$(3.5) \quad \frac{n_{ij\cdots}}{n_{i\cdots\cdot}} = \frac{n_{\cdot j\cdots}}{n_{\cdots\cdot}},$$

$$\frac{n_{i\cdot k\cdot\cdot}}{n_{i\cdots\cdot}} = \frac{n_{\cdot\cdot k\cdot\cdot}}{n_{i\cdots\cdot}}, \quad \frac{n_{i\cdot\cdot\ell\cdot}}{n_{\cdots\cdot\cdot}} = \frac{n_{i\cdots m}}{n_{i\cdots\cdot}}$$

$$\text{and} \quad \frac{n_{i\cdots m}}{n_{i\cdots\cdot}} = \frac{n_{\cdots\cdot m}}{n_{\cdots\cdot\cdot}}.$$

If the proportions of groups in subpopulation i are the same as the proportions of groups in the total population, then terms 2-4 of the left hand side in (3.4) are zero and the balanced group i effect $\hat{\alpha}_i$ equals the observed group i effect. To the extent that subpopulation and population proportions in (3.5) differ, the balanced group effect $\hat{\alpha}_i$ will differ from the observed group i effect.

Note that the observed group i effect is equal to the balanced group i effect plus four terms. Each term estimates how much of the balanced effects of one variable are transmitted to the observed group effect i through imbalance of that factor. For example,consider the imbalance of the STOC groups in the Southeast group. If i denotes the Southeast region and k sums over the size-and-type of community variable (STOC), then $\sum_k \frac{n_{i\cdot k\cdot\cdot}}{n_{i\cdots\cdot}} \hat{\gamma}_k$ is the portion of the balanced effects of STOC that is transmitted through the imbalance of the distribution of STOC groups in the Southeast when compared to the distribution in the nation as a whole. Exhibit (3.1) provides an example showing transfers of balanced effects when different variables are included in the balancing equations.

4. LIMITATIONS ON THE INTERPRETATION OF BALANCED EFFECTS

There are several kinds of limitations on the interpretation of balanced results. The group names National Assessment uses for data analysis are labels standing not only for the factor indicated by its name but also for a variety of other factors National Assessment did not (or could not) measure--factors associated or correlated with the named factor. Like observed results, balanced group effects do not show what is caused by the labelled factor. They show only the part of the unadjusted effect that can be conveniently named and attached to a group for bookkeeping purposes. They provide a means of comparing groups of individuals, free of confounding due to various mixtures of other groups.

Balancing of National Assessment data is limited to the five variables assessed. Some important factors may be partially represented in our factors and others not represented at all. Factors may exist which have smaller "proxy" bundles of other factors. If other factors had been included in the balancing equations then the balanced effects and the portion of the balanced effects transmitted though imbalance would be different. The difficulty in interpreting balanced group effects is the same as those encountered in interpreting regression coefficients--that is, the meaning of a balanced group effect depends heavily on what other factors are balanced at the same time. Thus, a balanced group effect represents the influence of unnamed background variables that are still associated with the balanced group name after considering the other known variables included in the balancing set.

Our balancing procedure utilizes an additive model which emphasizes balancing of marginal group effects and ignores balancing on interactions or effects of combinations of groups. For example, the fraction of rural Blacks living in the Southeast is greater than the fraction of rural Blacks living in the Northeast. If rural Blacks living in the Southeast do poorly compared to all rural Blacks, then we would expect the balanced Southeast region to do poorly. Though problems of confounding also exist from disproportionality of combinations of groups, one hopes they exist to a lesser extent. Similar disproportionate representation exists for the other two-, three-, and four-way group interaction effects.

5. SUMMARY AND CONCLUSIONS

The observed group effects from our surveys are facts describing the extent and level of educational performance for subpopulations in our country. Insightful analyses and data adjustments can be helpful in understanding these facts but they cannot change them.

We have shown that balancing is a combination of data adjustment and marginal main effect analysis. Typically, a five dimension data adjustment forces observed sample sizes to known marginal population totals leaving existing disproportionate marginal distributions within a single group pretty near the same. The balanced fit was shown to be equivalent to a least-squares fit of an additive linear model. A unique solution of the corresponding normal equations yields balanced group effects consistent with what one would obtain if the distribution of marginal proportions in each group were the same as the distribution of the marginal groups in the population.

Although a statistically more sophisticated data adjustment could have been employed, we can begin to see how simple adjustments help in removing effects of masquerading and the portion of a balanced group effect that may be transferred to the observed group effect due to imbalance of other characteristics. The greatest limitation is that other important background factors which were not included may still be masquerading as balanced group effects.

For National Assessment (4) the balancing of marginal baseline data is just the beginning. As we obtain repeated measures over time, it will be helpful to know if the proxy bundles maintain their relative importance. As characteristics of the population change, which is expected in our ever changing society, it could be that the "factors" we measure by these variables are changing too. In the future it may be even more important to have well-measured variables and well-conceived data adjustments if we are to go beyond the observed data and obtain some guidance about the mechanisms involving the complex set of factors affecting education. Though the problems in education are far reaching and urgent, one must guard continually against overstating conclusions about the results of sample surveys and data adjustments. The full value of adjusting data depends on careful, clear documentation of the limited steps we have taken, what we have learned, and what we still don't know.

REFERENCES

(1) Deming, W. Edwards, Statistical Adjustment of Data, 2nd publication, New York: Dover Publications Inc., 1964.

(2) Larson, Robert; Wayne Martin; Todd Rogers; Donald Searls; Susan Sherman; and David Wright, A Look at the Analysis of National Assessment Data presented by J. Stanley Ahmann in Frontiers of Educational Measurement and Information Systems-1973, ed. William C. Coffman (Boston: Houghton Mifflin Company, 1973), pp. 89-111.

(3) Mosteller, Frederick, "Social Experimentation," Fifth S.S. Wilks Memorial Lecture, Princeton University, Princeton, NJ., 1973.

(4) National Assessment of Educational Progress, Report 7: Science Group Results B 1970, Washington, D.C.: (U.S. Government Printing Office), 1973.

(5) Tukey, J.W. "Technique for Analysis of Groups". A personal memo to National Assessment staff and the Analysis Advisory Committee, 1970.

EXHIBIT 11, AN EXAMPLE OF THE PROCEDURE FOR OBTAINING A BALANCED FIT.

Consider the simple example of two variables, variable A having 3 levels and variable B having 2 levels. The layout for the number of cases and number of successes is shown below. Note that the number of successes in each cell are to be fitted to the marginals and are left blank.

Number of Cases

		B 1	B 2	
A	1	100	100	200
	2	50	150	200
	3	0	200	200
		150	450	600

Number of Successes

		B 1	B 2	
A	1			80
	2			90
	3			100
		80	190	270

The representation of the fitted number of successes in each cell is shown below where marginal effects are denoted by the letters and the national percentage is 45%.

		B 1	B 2
A	1	$100\ (45\% + a_1 + b_1)$	$100(45\% + a_1 + b_2)$
	2	$50(45\% + a_2 + b_1)$	$150(45\% + a_2 + b_2)$
	3	$0(45\% + a_3 + b_1)$	$200(45\% + a_3 + b_2)$

The balancing equations are formed by combining cell representations of successes and equating to each marginal number of successes in turn.

Cell Combinations	Balancing Equations
$a_1b_1 + a_1b_2$	(1) $100(45\% + a_1 + b_1) + 100(+5\% + a_1 + b_2) = 80$
$a_2b_1 + a_2b_2$	(2) $50(45\% + a_2 + b_1) + 150(45\% + a_2 + b_2) = 90$
$a_3b_1 + a_3b_2$	(3) $0(45\% + a_3 + b_1) + 200(45\% + a_3 + b_2) = 100$
$a_1b_1 + a_2b_1 + a_3b_1$	(4) $100(45\% + a_1 + b_1) + 50(45\% + a_2 + b_1) + 0(45\% + a_3 + b_1) = 80$
$a_1b_2 + a_2b_2 + a_3b_2$	(5) $100(45\% + a_1 + b_2) + 150(45\% + a_2 + b_2) + 200(45\% + a_3 + b_1) = 190$
	Usual Side Conditions
	(6) $200\ a_1 + 200\ a_2 + 200\ a_3 = 0$
	(7) $150\ b_1 + 450\ b_2 = 0$

Only 3 of the set of 5 balancing equations are linearly independent. One way to obtain full rank is to replace equations (3) and (5) by equations (6) and (7). The unique solution for the fitted balanced effects is shown below.

$a_1 = -10\%$ $\quad b_1 = 15\%$
$a_2 = 0\%$ $\quad b_2 = -5\%$
$a_3 = 10\%$

EXHIBIT 3.1a. EXAMPLES FROM THE YEAR 01 SCIENCE ASSESSMENT ILLUSTRATING MEDIAN TRANSFERS OF BALANCED GROUP EFFECTS TO AN UNADJUSTED SOUTHEAST GROUP EFFECT FOR SEVERAL BALANCING COMBINATIONS.

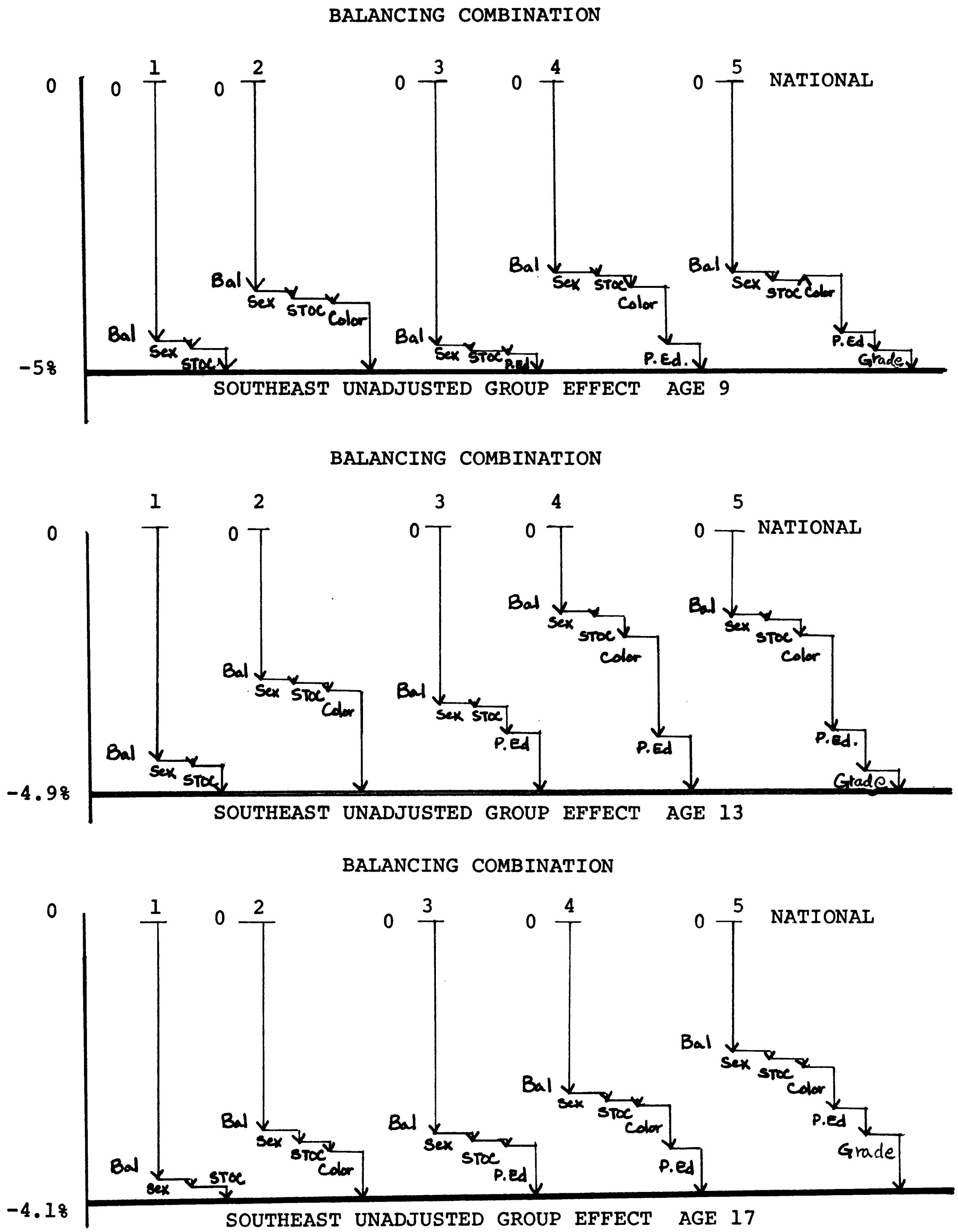
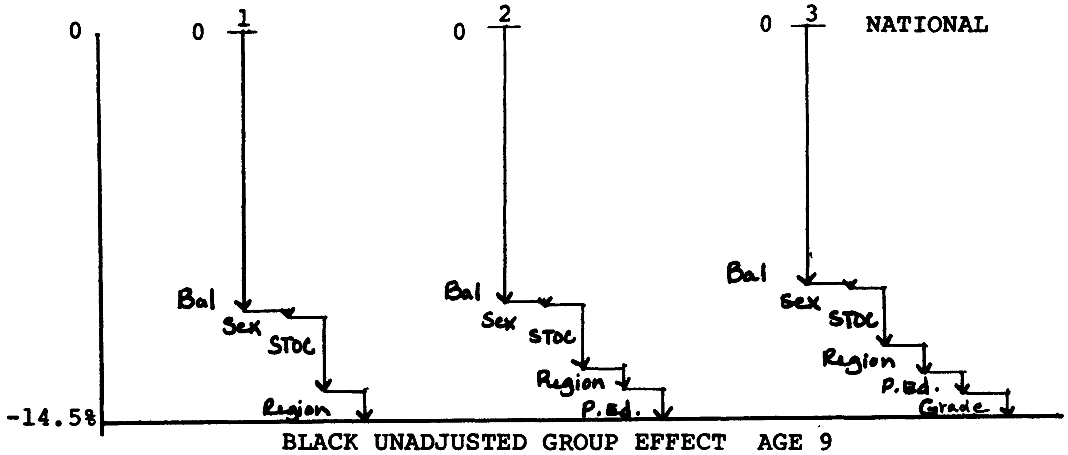


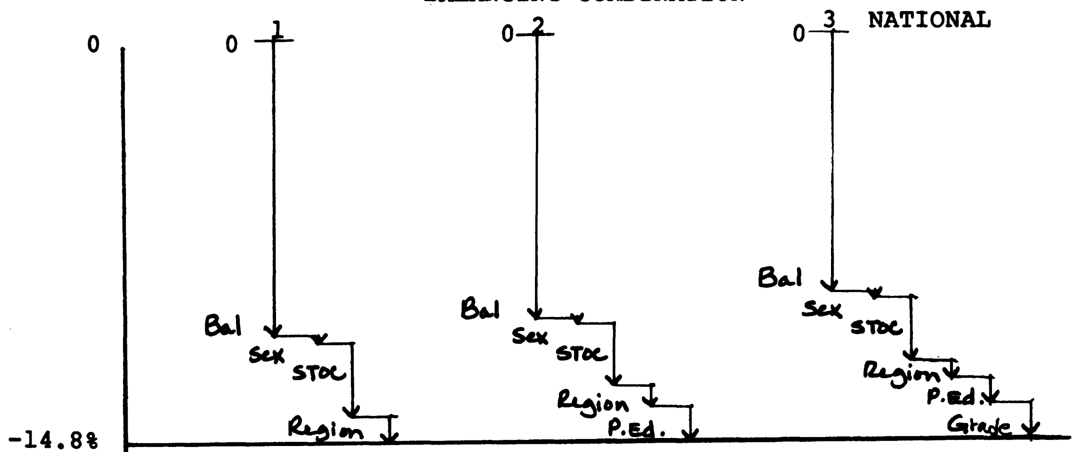
EXHIBIT 3.1b. EXAMPLES FROM THE YEAR 01 SCIENCE ASSESSMENT ILLUSTRATING MEDIAN TRANSFERS OF BALANCED GROUP EFFECTS TO AN UNADJUSTED BLACK GROUP EFFECT FOR SEVERAL BALANCING COMBINATIONS.

BALANCING COMBINATION



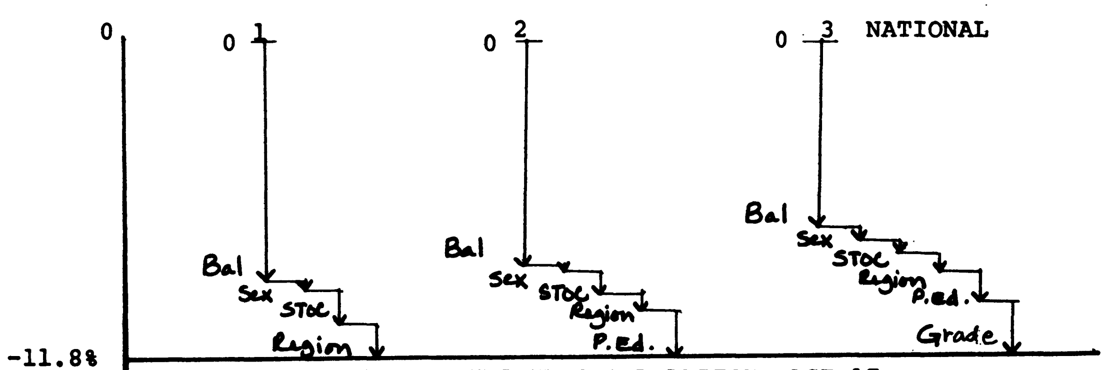
BLACK UNADJUSTED GROUP EFFECT AGE 9

BALANCING COMBINATION



BLACK UNADJUSTED GROUP EFFECT AGE 13

BALANCING COMBINATION



BLACK UNADJUSTED GROUP EFFECT AGE 17

THE USE OF MONETARY INCENTIVES IN NATIONAL ASSESSMENT HOUSEHOLD SURVEYS

James R. Chromy and D. G. Horvitz, Research Triangle Institute

BACKGROUND

The National Assessment of Educational Progress survey of young adults employs a rather unique monetary incentive procedure to increase the overall response rate and to decrease the cost of data collection. The procedure was adopted as a result of a special quality check study conducted after obtaining poor results in the first year of the assessment program [1].

Monetary incentives or other respondent remunerations are not extensively employed in survey practice, particularly in the case of studies under government sponsorship. Their greatest potential appears to be in surveys that place unusual demands upon the respondent or require continued cooperation over extended periods of time.

Pearl [2] suggests experimental testing of alternative incentive procedures as part of the development of methodology for maintaining cooperation in consumer expenditure surveys.

In a recent paper, Ferber and Sudman [3] review a number of observational and experimental studies which utilized monetary incentives or gifts to respondents. Their review shows that beneficial, neutral, and detrimental results have been reported in different instances. Factors they recognize as affecting the potential success of incentives include the amount of compensation; the required period of cooperation (e.g., single interview, record keeping, or panel participation); the auspices under which the survey is conducted; the socioeconomic status of the respondents; and the nature of the information sought. They conclude that there does not appear to be a need for compensation in one time interviews, but that there is a rationale for using monetary incentives to maintain cooperation in record keeping or panel studies.

In another recent paper, Cannell and Henson [4] consider the effects of incentives on the accuracy of the data obtained as well as on the level of cooperation. They recognize the determination of an appropriate payment as a research effort in itself, since overpayment may in fact have negative effects.

One of the most favorable effects of remuneration was experience in the 1971 Health and Nutrition Examination Survey (HANES) as reported by Miller, Kennedy, and Bryant [5]. In this study, persons age 25-74 were asked to submit to a physical examination. In a designed experimental study, a $10 incentive plan produced a response rate of .82 compared with .70 in the no incentive control group. As a result, the procedure was adopted for subsequent stands in the survey.

The National Assessment experience tends to parallel that of the HANES study. An experimental test of alternative incentive procedures was employed after obtaining poor results in an initial survey. As a result, an incentive plan has been adopted and employed in all subsequent surveys. This paper reports on the initial experimental study and on results obtained since that time.

THE YEAR 01 SURVEY

A national household sample survey was conducted in the summer of 1969 to obtain benchmark estimates of the level of knowledge of young adults in the areas of citizenship, science, and writing. Young adults between the ages of 26 and 35 were asked to respond to a selected set of National Assessment Exercises. Ten different packages or forms were used. Each eligible individual located by a household screening process was asked to respond to only one of these packages (one hour or less of the respondent's time was required to complete one package); therefore, 10 respondents were required to complete the entire set of 10 packages.

The essential features of the Year 01 survey and of subsequent household surveys of young adults are shown in table 1. More detailed descriptions of the sample and survey designs are given in sources listed in the reference list [6-8].

The Year 01 sampling plan is summarized in table 2. The sample was selected in stages. The first stage or primary sample consisted of 208 primary sampling units (PSUs). Fifty-two PSUs were selected in each of the 4 regions--Northeast, Southeast, Central, and West. Further stratification within regions was based on type of community and income level. Low-income areas were oversampled to attempt to achieve a NAEP objective of effectively measuring the low socioeconomic status (SES) population. The PSUs used for the Year 01 household survey of young adults were also used as the primary sample for the selection of 9-year-olds, 13-year-olds, and 17-year-olds enrolled in school; the secondary sampling frames for the 3 in-school age classes were constructed from school lists.

The second stage or secondary sample consisted of clusters of housing units (or households) which were screened for eligible young adults. Oversampling of low-income areas within the PSU was again employed to attempt to increase the sample size for the low SES population. All eligible young adults identified in the screening process were asked to complete a single package of exercises requiring less than an hour of their time.

The response to the Year 01 young adult survey was extremely disappointing. As shown in tables 3 and 4, 77.4 percent of the occupied housing units were successfully screened for young adults, and 57.4 percent of the eligible young adults agreed to participate by completing a package of exercises. The overall response rate, considering both household screening and package completion, was 44.4 percent.

A number of factors were believed to be responsible for the poor Year 01 response. The staff conducting the field work had been employed primarily for use in the school assessment, and

TABLE 1 - OVERVIEW OF THE NATIONAL ASSESSMENT YOUNG ADULT SAMPLE SURVEYS

Item	Year 01	Year 02	Year 03	Year 04	Year 05
Subject areas assessed	Citizenship Science Writing	Literature Reading	Music Social Studies	Mathematics Science	Writing, COD (Career and Occupational Development)
Age eligibility based on data of birth	July 1933- June 1943	April 1935- March 1945	April 1936- March 1946	January 1937- December 1946	January 1938- December 1974
Data collection period Dates:	June-August 1969	February- August 1971	January- June 1972	October 1972- May 1973	April- August 1974
Length (approximate)	3 months	7 months	6 months	8 months	5 months
Data collection instruments: Background questionnaire (BQ)	Attached to each package	Short separate questionnaire	Short separate questionnaire	Short separate questionnaire	Expanded separate questionnaire
Number of packages	10	6	8	8	3
Use of stimulus tapes	None	Nonpaced for Literature	Nonpaced	Nonpaced	Nonpaced
Concurrent assessment of out-of-school 17-year-olds	Yes	Yes	Yes	Yes	No
Incentive offered to adult respondents.	None. Each respondent responded to one package only.	None for 1 pkg. $10 for 2 pkgs. $15 for 3 pkgs. $20 for 4 pkgs.	None for 1 pkg. $10 for 2 pkgs. $15 for 3 pkgs. $20 for 4 pkgs.	$5 for 1 pkg. $10 for 2 pkgs. $15 for 3 pkgs. $20 for 4 pkgs.	$10 for BQ and 1 pkg., $15 for BQ and 2 pkgs. $20 for BQ and 3 pkgs.
Distinct primary sampling unit areas	171	47	94	94	100
Number of area segments	2,087	520	936	1,059	970
Field staff:					
Regional supervisors	4	-	2	2	2
Field supervisors	27	5	12	11	10
Interviewers	624 (approx.)	57	110	113	120
Interviewer training	3 days	3 days	5 days	4 days	4 days

Source: Taken in part from Vern Acherman, Operations Documentation Report: Adult Assessment (Years 01-04), National Assessment of Educational Progress, March 1974.

most had no experience in household interviewing. The period allotted for completion of the survey was too short to allow for establishment of good control procedures with adequate time for corrective action. The magnitude of the household screening operation made it difficult to manage effectively over a short period of time. The nature of the assessment exercise packages may have had an adverse effect on voluntary cooperation. As a result of this poor response experience, plans were developed to conduct a quality check study. Specific objectives of the study were:

TABLE 2 - YEAR 01 PLANNED SAMPLE COUNTS

Item	Per area segment	Per PSU	Entire sample
Primary sampling units	--	1	280
Area segments	1	10	2,080
Housing units	33	330	69,000
Persons screened	108.7	1,087	226,096
Eligible adults (26-35)	12.5	125	26,000
Responding adults (approximate 80% response rate)	10	100	20,800

TABLE 3 - YEAR 01 HOUSEHOLD SCREENING EXPERIENCE

Item	Number	Percent of occupied housing units
Sample housing units listed	64,506	
Less		
Vacant	3,602	
Usual residence elsewhere	864	
Other unoccupied	775	
Occupied housing units	59,265	100.0
Not at home after 4 calls	4,667	7.9
Temporarily away	1,255	2.1
Refusal	5,478	9.2
Illness	124	0.2
No reliable respondent	184	0.3
No English spoken	357	0.6
Incomplete or missing questionnaire	460	0.8
Refused access by ordinance, apartment manager, etc.	700	1.2
Other	191	0.3
Respondents	45,849	77.4

TABLE 4 - YEAR 01 PACKAGE ADMINISTRATION EXPERIENCE

Item	Number	Percent of eligible young adults
Total eligible by age definition	14,676	
Less		
Physically or mentally handicapped	98	
English not understood	203	
Eligible young adults	14,375	100.0
Temporarily away	431	3.0
Refusal	3,834	26.7
Appointment not kept	1,001	7.0
Misclassified by interviewer	295	2.1
Unresolved	393	2.7
Other	164	1.1
Respondents	8,257	57.4

(1) To determine the extent to which coverage might have been improved through the use of better field staff and better methods;
(2) To determine the effect of monetary incentives on participation by eligible respondents;
(3) To determine the magnitude of the bias in estimates derived from the initial survey data;
(4) To collect the data needed to compute estimates adjusted for any bias due to nonresponse.

SAMPLE DESIGN

Since this study had mixed objectives, a sample design adaptive to these several objectives was required.

An implicit assumption was made that improved procedures could only improve response; as a result, it was decided to meet objective 1 by conducting the study in a subsample of the same area segments that were used in the initial study. Persons who had participated in this initial study were identified and not asked to take any additional exercise packages. The field evaluation was therefore aimed at the nonrespondent portion of the population from the initial study. This approach also served to meet objectives 3 and 4 most effectively. The experimental part of the study involving the effects of alternative monetary incentive procedures under this plan was also applied only to persons who had not participated (voluntarily for no monetary incentive) in the initial study. It was assumed that a monetary incentive would have no adverse effect upon the cooperation of those persons who had cooperated previously without the incentive.

A probability subsample of the area segments used in the initial study was selected to allow the development of bias adjustment procedures, if necessary (objective 4), which would be applicable to all area segments in the initial sample.

The sampling frame for the quality check study consisted of the PSUs and the area segments (secondary sampling units) which were included in the initial household survey PSUs.

The quality check sample was drawn in two stages. The first stage units were PSUs from the initial study or clusters of these PSUs within a compact area. Second stage units were the area segments in the initial study. In most cases, a sample of 4 segments was selected in each sampled cluster of PSUs for the incentives experiment.

The metropolitan areas of New York-Newark, Philadelphia, Chicago, Los Angeles-Long Beach, and San Francisco were treated somewhat differently than the remaining clusters. The PSUs selected from each of these areas were considered as a single cluster and included in the quality check sample with probability one; first-stage units selected in this are said to be self-representing. The number of quality check segments selected from each area was not restricted to 4, but to a multiple of 4, depending upon the size of the area. These areas coincide with some of the worst experiences in obtaining response in the initial household survey.

The remaining areas in the sampling frame were first stratified by region into 4 regions: (1) Northeast; (2) Southeast; (3) Central; and (4) West. Within each region further stratification was carried out based on type of community (TOC) strata used in the initial sample of PSUs. These TOC strata are:

(1) Central cities with a 1960 population of 180,000 or more;
(2) The remainder of the Standard Metropolitan Statistical Area (SMSA) for each of the TOC 1 cities;
(3) Remaining metropolitan counties belonging to an SMSA and counties with a city of 25,000 or more;
(4) All areas not included in TOC 1, 2, and 3.

These strata were collapsed into 2 new strata consisting of a combined TOC 1 and TOC 2 stratum and a combined TOC 3 and TOC 4 stratum within each region.

The allocation of the quality check sample to these strata is shown in table 5. The overall segment sampling rate is shown in the last column. This rate was computed as the ratio of quality check segments to total segments available in the sampling frame (10 times the number of PSUs).

Thirteen clusters or areas were drawn, one from each of the strata shown in table 5. The sample areas were selected with probability proportional to estimated size. The size estimate used was an approximation of the number of potential nonrespondents in each cluster based on an expectation of 100 adults per PSU less some preliminary tabulations of packages administered.

Within each sample area, 4 segments were selected for the quality check study. (Two expections to this rule are area 1, New York City-Newark, and area 2, Los Angeles County, which had 8 and 12 quality check segments,

TABLE 5 - ALLOCATION OF QUALITY CHECK SAMPLE TO THE ORIGINAL SAMPLE STRATA

Description		Sampling Frame No. of PSUs	Sampling Frame No. of clust.	QC Sample No. of clust.	QC Sample No. of QC segs.	Overall segment sampling rate
Self-representing areas						
NYC-Newark		11	1	1	8	.075
Philadelphia		5	1	1	4	.080
Chicago		5	1	1	4	.080
San Francisco		8	1	1	4	.050
Los Angeles		12	1	1	12	.100
Subtotal		41	5	5	32	.078
Remainder						
Region	TOC					
NE	1 & 2	15	8	1	4	.027
	3 & 4	21	17	1	4	.019
SE	1 & 2	14	7	1	4	.029
	3 & 4	38	34	1	4	.010
Central	1 & 2	19	12	1	4	.021
	3 & 4	28	25	1	4	.014
West	1 & 2	11	7	1	4	.036
	3 & 4	21	21	1	4	.019
Subtotal		167	131	8	32	.019
Total		208	136	13	64	.031

respectively.) Additional data on the number of eligible respondents by segment and the number of adult packages administered by segment were examined for the sample areas. An estimated number of potential additional respondents was computed for each segment by adding those who refused to take a package and an estimate of eligible adults in the housing units not screened. Stratification of the segments within an area was then carried out by classifying the segments into the following six strata:

(1) The 10 percent of the segments having the largest number of potential additional respondents per segment.

(2) The next 20 percent;

(3) The next 30 percent;

(4) The remaining segments that had at least one potential respondent or one housing unit which had not been screened;

(5) Segments with zero potential additional respondents;

(6) Segments not received or lacking screening questionnaires as of the date of selecting the sample.

One segment from each of the first four strata (2 per stratum in area 1 and 3 per stratum in area 5) was selected at random for the quality check sample. An additional segment was selected from the fifth stratum in each cluster. Field procedures for these "zero" segments were limited to a check of the listing and a check of the screening in a subsample of those "zero" segments which contained some housing units.

The few segments in stratum 6 were omitted from the study completely. Table 6 shows the stratum size in number of segments by area. The range of estimated potential additional respondents for the first 4 strata is given in the right half of table 6. A pre-survey estimate of the potential additional respondents for the quality check study is given in the bottom line of table 6 by secondary stratum. The total over the 4 strata is 596 potential additional respondents.

EXPERIMENTAL DESIGN

The purpose of the experimental part of the quality check study was to test 4 different monetary incentive procedures in conjunction with first-quality field procedures. The 4 procedures were:

(A) No monetary incentive;

(B) Five dollar incentive to take a package;

(C) Ten dollar incentive to take a package;

(D) A variable incentive procedure.

The variable incentive procedure assumed a different tradeoff between time and money for different persons. Adult respondents were asked to take one package for no reimbursement, 2 packages for 10 dollars, 3 packages for fifteen dollars or 4 packages for twenty dollars.

If an 80% response could not be obtained using procedure A with 4 callbacks, 2 additional callbacks along with a monetary incentive were used to bring the response rate up as high as possible. For purposes of determining the effectiveness of procedure A, the response rates were calculated on the basis of packages and interviews completed prior to offering the incentive.

All 4 procedures were assigned in each area or cluster. To balance the administration of different incentive procedures across the secondary strata, the procedures were assigned to the quality check segments by random selection of 4 x 4 Latin Square experimental designs. The particular assignments made are shown in table 7.

Since not all 10 packages could be administered in every segment, different package start numbers were assigned to each of the segments within an area. The random assignment scheme used produced an approximate balance in the number of different packages administered within area, within Latin Square, and within the entire sample.

FIELD PROCEDURES

A number of intensive field interviewing procedures were employed or attempted for this study in addition to the use of monetary incentives to obtain cooperation from eligible adults.

Newspaper releases were prepared for local newspapers and for distribution by the interviewers. Although newspapers did not publish the releases, they were used effectively by the interviewers. Police departments and Better Business Bureaus were advised of the study.

Only experienced, highly qualified interviewers were used. A total of 27 interviewers, 26 females and 1 male, were recruited. A male escort was also used in one inner city area. Interviewers were trained and closely supervised by 6 team leaders. The team

TABLE 6 - SECONDARY STRATIFICATION BY AREA

Area	Sample* segments per stratum	Number of segments per stratum 1	2	3	4	5	6	Range of potential additional respondents per segment** 1	2	3	4
1	2	10	20	30	31	7	2	19-26	14-19	6-14	0.4-6
2	1	5	10	15	18	2	0	18-55	10-18	6-10	0.4-5
3	1	5	10	14	15	6	0	15-20	8-13	5-8	0.4-5
4	1	8	16	24	27	1	4	19-32	13-19	6-12	0.4-6
5	3	10	20	30	34	4	2	17-33	10-17	6-10	0.4-6
6	1	3	6	9	11	0	1	16-25	9-15	6-8	1-5
7	1	1	2	3	3	1	0	10	7-8	4-5	2-4
8	1	2	4	6	7	1	0	10-12	5-10	3-5	0.4-2
9	1	1	2	3	3	1	0	7	6-7	4-6	0.4-3
10	1	2	4	6	7	1	0	20-29	9-17	6-9	1-6
11	1	1	2	2	3	2	0	6	2-3	1-2	0.4-1
12	1	2	4	6	6	2	0	6-22	5-6	1-5	0.4-1
13	1	1	2	3	4	0	0	5	3-4	1-2	0.4-1
Total------------------------		51	102	151	169	28	9				
Midrange of potential additional respondents-------------------------								297	158	102	39

*Applies to first 4 strata.
**The corresponding value for stratum 5 is always 0. No information was available for stratum 6.

leaders were either central staff or senior field supervisory personnel.

The listings of housing units in sample segments were checked for completeness and proper designation of sampling housing units. Housing units omitted in the initial listings were added and special procedures were applied to determine if these housing units were to be included in the sample.

The household screening questionnaire was used to obtain a roster of all household members. Eligible adults were then identified according to birth date. Up to four calls were made to contact households for household screening. Neighbors were contacted after the first call to determine, if possible, the best time for a followup call. If neighbors, did not specify a time, the second call was made between 6:30 and 9:00 p.m.; a letter of explanation indicating that further efforts would be made to contact household members was left under the door. If no household member could be contacted after four calls, a neighbor questionnaire was completed if possible.

All refusals at the household screening level were referred to the team leaders for possible followup. Table 8 shows that almost a third of the initial refusals were converted

TABLE 7 - ASSIGNMENTS OF INCENTIVE PROCEDURES TO QUALITY CHECK SEGMENTS

Latin square number	Area	Secondary Stratification: Stratum 1 Sample number	Stratum 1 Incentive* procedures	Stratum 2 Sample number	Stratum 2 Incentive procedure	Stratum 3 Sample number	Stratum 3 Incentive procedure	Stratum 4 Sample number	Stratum 4 Incentive procedure
1	1	111	C	121	B	131	A	141	D
	1	112	A	122	D	132	C	142	B
	2	211	B	221	C	231	D	241	A
	3	311	D	321	A	331	B	341	C
2	4	411	D	421	B	431	A	441	C
	5	511	B	521	D	531	C	541	A
	5	512	A	522	C	532	B	542	D
	5	513	C	523	A	533	D	543	B
3	6	611	B	621	D	631	C	641	A
	7	711	A	721	C	731	D	741	B
	8	811	C	821	B	831	A	841	D
	9	911	D	921	A	931	B	941	C
4	10	1011	A	1021	C	1031	D	1041	B
	11	1111	B	1121	D	1131	C	1141	A
	12	1211	D	1221	A	1231	B	1241	C
	13	1311	C	1321	B	1331	A	1341	D

*Incentive procedures are indicated as: A-No incentive; B-$5 to complete a single package; C-10 to complete a single package; and D-The variable incentive procedure which offers no incentive for 1 package, $10 for 2 packages, $15 for 3 packages, and $20 for 4 packages.

TABLE 8 - FOLLOWUP PROCEDURES FOR HOUSEHOLD SCREENING REFUSALS

Procedure	Times attempted	Times successful	Percent successful
Telephone follow-up by team leader	5	0	0
Personal follow-up by second interviewer	21	9	42.9
Personal follow-up by team leader	35	10	28.6
Total*	61	19	31.1
Total personal followup only	56	19	33.9

*Refusals initially occurred in 64 households; due to timing or other circumstances, not all of these were followed up.

TABLE 9 - FOLLOWUP PROCEDURES FOR PACKAGE REFUSALS

Procedure	Times attempted	Times successful	Percent successful
Telephone followup by team leader	21	0	0
Personal follow-up by second interviewer	43	8	18.6
Personal follow-up by team leader	44	6	13.6
Total*	108	14	13.0
Total personal followup only	87	14	16.1

*Refusals initially occurred in 111 cases.

to respondents as a result of these followup procedures.

If monetary incentives were assigned to a segment, the interviewer was instructed to make the following statement as part of the introduction to the screening questionnaire:

"Anyone in your home who is eligible to participate in this study will be paid for his time."

A list of respondents from the initial study was provided to each interviewer and after verification with the respondent, those who had previously participated were not eligible to participate in the quality check study.

If package administration with the eligible young adults could not be completed during the screening interviews, an appointment was made for a return visit. Interviewers were instructed to make up to three calls to complete package administration. Team leaders were informed of all refusals at the package administration level. Table 9 shows the action taken and the results achieved. As in the case of the screening followup procedures, telephone procedures were not effective. Of 87 initial refusals contacted personally by a different interviewer or a team leader, 14 were converted.

Since the study was designed to evaluate bias in the estimates as well as the relative effectiveness of alternate monetary incentive procedures, it was necessary to obtain high response in all areas including the no incentive segments. To accommodate these two study requirements, the no incentive segments were first treated according to the experimental design. All response information was recorded at that point and if the overall response (screening rate time package completion rate) was less than 80 percent, further attempts were to be made to achieve, at least, an 80 percent response. These further attempts were to employ the $10 per package incentive (monetary incentive procedure C). Table 10 shows that in these particular areas, over half of the refusals were converted by the use of the monetary incentive.

EXPERIMENTAL COMPARISONS

The results of the experimental study are shown in tables 11 through 14. Data from one sampling unit was not useable because the interviewer inadvertently applied the wrong incentive procedure. The data for incentive procedure A, (no incentive) excludes responses obtained after invoking incentive procedure C to obtain response rates of at least 80 percent.

Table 15 shows the response rates achieved using each of the 4 incentive procedures. To determine the significance of any observed differences in response rates among the 4 incentive plans, a linear model was fitted to the segment data. The model adjusted for area (13 areas in the sample) and for the secondary strata. Three dependent variables were analyzed:

(1) The ratio of initial survey respondents plus quality check survey respondents to quality check eligibles plus initial survey respondents;
(2) The ratio of quality check eligibles plus initial survey respondents to housing units screened;
(3) The ratio of housing units screened to total occupied housing units.

The variability of each of the dependent variables can be expected to depend upon the value of the denominator of the ratio. For the purposes of fitting the model, the variances were assumed to be inversely proportional to the denominators of the dependent variable ratios.

Three contrasts were considered to be of special interest and were estimated for each of the dependent variables. These three contrasts were:

(1) Incentive versus no incentive [(B+C+D)/3 - A];
(2) Variable incentive versus fixed incentives [D - (B+C)/2];
(3) $10 incentive versus $5 incentive.

TABLE 10 - EFFECTS OF FOLLOWUP WITH INCENTIVE IN NO INCENTIVE SEGMENTS

Sampling unit number	Screening: Initial refusals	Screening: Successful followup	Pkg. Admin.: Initial refusals	Pkg. Admin.: Successful followups
321	0	0	1	1
541	0	0	2	1
921	3	1	10	5
1011	1	0	9	3
1221	3	0	5	2
1331	6	6	7	5
Total	13	7	34	17

It appears appropriate to test the hypothesis that each of these contrasts is zero against the one-sided alternative that it is greater than zero. Table 16 shows the estimated values of these contrasts and the value of the t-statistic for each contrast. Only the contrast which compared package response rates for the incentive procedures and the no incentive plan was shown to be statistically significant.

COST COMPARISONS

In light of the small differences in response rates observed among the the 3 incentive plans, one basis for selecting a plan for future surveys was the cost per unit of data obtained. Cost comparisons at the time of the study clearly favored plan D.

An updated variable field cost model based on subsequent experience with plan D is shown in table 17. This model reflects Year 03 response experience and current costs. The cost per package and cost per respondent based on the table 17 model are shown in table 18. The cost per package for plan D is less than half that obtained for any of the other plans. Cost per respondent is, of course, highest for plan D.

TABLE 11 - STUDY RESULTS FOR PROCEDURE A

Sampling unit number	Quality check survey: Occupied HUs: Sample	Occupied HUs: Screened*	Elig. adults: Total**	Elig. adults: Respdnts	Initial survey respdnts
112	(Data not useable)				
131	37	32	22	11	3
241	21	20	8	8	0
321	41	36	4	3	4
431	38	37	12	7	14
512	44	36	9	8	0
523	37	34	21	21	1
541	14	14	6	4	0
641	35	33	8	2	3
711	59	54	10	10	1
831	32	32	6	4	1
921	39	36	17	7	4
1011	46	45	15	6	4
1141	19	19	0	0	2
1221	28	22	7	2	7
1331	21	15	9	2	2
Total	511	465	154	95	46

*Excludes 7 additional housing units screening using incentive procedure C and shown in table 10.

**Includes eligible respondents identified in 7 additional housing units shown in table 10.

TABLE 12 - STUDY RESULTS FOR PROCEDURE B

Sampling unit number	Quality check survey: Occupied HUs: Sample	Occupied HUs: Screened	Elig. adults: Total	Elig. adults: Respdnts	Initial survey respdnts
121	47	36	10	6	2
142	20	11	9	5	1
211	56	53	22	20	7
331	28	25	2	1	2
421	53	51	21	18	3
511	41	31	19	7	1
532	27	22	9	6	1
543	23	23	4	1	0
611	23	23	16	12	1
741	36	34	12	12	0
821	25	25	15	14	10
931	40	38	5	5	1
1041	14	13	3	2	2
1111	40	38	7	6	5
1231	26	24	3	3	3
1321	30	30	4	1	14
Total	529	477	161	119	53

ADDITIONAL EXPERIENCE

Plan D was adopted in the surveys conducted in Years 02 through 05. The package response experience for Years 01 through 04 are shown in table 19. The Year 05 experience has not yet been tabulated.

TABLE 13 - STUDY RESULTS FOR PROCEDURE C

Sampling unit number	Quality check survey: Occupied HUs: Sample	Occupied HUs: Screened	Elig. adults: Total	Elig. adults: Respdnts	Initial survey respdnts
111	65	49	19	14	1
132	21	21	10	4	0
221	27	25	9	6	5
341	19	18	5	4	5
441	14	14	4	4	1
513	34	23	4	4	0
522	31	29	8	6	8
531	20	19	9	5	0
631	47	44	13	10	4
721	56	50	6	6	5
811	48	45	13	12	12
941	20	19	3	2	4
1021	31	26	7	6	4
1131	20	17	3	3	4
1241	32	31	2	1	1
1311	70	68	6	6	16
Total	555	498	121	93	70

TABLE 14 - STUDY RESULTS FOR PROCEDURE D

Sampling unit number	Quality check survey: Occupied HUs Sample	Screened	Elig. adults Total	Respdnts	Initial survey respdnts
122	44	35	24	18	4
141	21	18	3	2	2
231	34	32	2	1	1
311	58	45	12	9	10
411	56	49	16	10	2
521	49	49	3	3	3
533	50	47	18	15	6
542	20	20	1	0	1
621	24	22	18	18	0
731	44	43	3	3	0
841	8	8	1	1	5
911	44	44	8	4	12
1031	26	25	7	5	4
1121	19	19	3	3	4
1211	17	17	6	3	2
1341	29	29	1	0	4
Total	543	502	126	95	60

TABLE 15 - INCENTIVE PROCEDURE COMPARISONS

Incentive procedure	Percent HUs screened	Percent completing packages: Quality check eligibles only	Including initial survey respdnts*
A	91.0	61.7	70.5
B	90.2	73.9	80.4
C	89.7	76.8	85.3
D	92.4	75.4	83.3

*Initial survey respondents are added to both the numerator and denominator used in computing package completion rates for the quality check survey.

TABLE 16 - EXPERIMENTAL COMPARISONS

Estimated contrasts	Value	t-statistics
Respondents/eligibles		
(B+C+D)/3 - A	.1034	2.097*
D - (B+C)/2	.0467	0.869
C - B	.0749	1.281
Eligibles/HUs screened		
(B+C+D)/3 - A	-.0397	-0.660
D - (B+C)/2	-.0271	-0.433
C - B	-.0780	-1.065
HUs screened/OHUs		
(B+C+D)/3 - A	.0059	0.251
D - (B+C)/2	.0290	1.187
C - B	.0021	0.073

TABLE 17 - COST MODEL FOR SCREENING AND PACKAGE ADMINISTRATION*

Paramater	Assumed value
Calls required to complete screening at 99.7 percent of occupied housing units (OHUs)	1.809
Eligible young adults per OHU	.338
Calls per eligible in addition to screening calls to obtain cooperation and administer at least one package	.920
Additional calls per respondent taking 2 or more packages (Plan D only)	.118
Response rate	R
Proportion of respondents willing to complete	
1 package	.019
2 packages	.022
3 packages	.007
4 packages	.952
Average packages administered per respondent (Plan D)	3.892
Average incentive cost per package (Plan D)	$4.91
Other unit costs:	
Package administration (1 hour)	$3.50
Cost per call (.571 hours + 8.57 miles)	$3.11

*Based in part on NAEP Year 03 experience.

As shown in table 19, the response rates obtained in the experimental study compare favorably with results obtained for plan D in the full scale surveys.

A large majority of respondents have chosen to take all 4 packages. In Year 04, the policy for incentive payment was modified to include a $5 incentive for one package. As a result, the number of respondents opting to take only one package increased. A small increase in the percent agreeing to take any package at all was also experienced.

TABLE 18 - HYPOTHETICAL COST PER PACKAGE BY INCENTIVE PLAN

Plan	Assumed response rate (R)	Cost per package	Cost per respondent
A	.70	$31.16	$31.16
B	.80	$32.69	$32.69
C	.80	$37.69	$37.69
D	.80	$14.72	$57.29

TABLE 19 - ADDITIONAL PACKAGE RESPONSE EXPERIENCE

	Year 01		Year 02		Year 03		Year 04	
	Number	Percent of total	Number	Percent of total	Number	Percent of total	Number	Percent of total
Refusals	3,834	26.7	517	20.0	892	15.4	733	14.7
Other/not at home	2,28	15.9	125	4.8	68	1.2	57	1.1
Respondents:	8,257	57.4	1,938	75.2	4,822	83.4	4,211	84.2
1 package	8,257	57.4	40	1.6	93	1.6	168	3.4
2 packages	--	--	72	2.8	108	1.9	96	1.9
3 packages	--	--	58	2.3	33	.6	36	.7
4 packages	--	--	1,768	68.5	4,588	79.3	3,911	78.2
Total eligible	14,375	100.0	2,580	100.0	5,782	100.0	5,001	100.0
Total packages	8,257	--	7,072	--	18,760	--	16,112	--
Packages/eligible	.57	--	2.74	--	3.24	--	3.22	--
Packages per respondent	1.00	--	3.65	--	3.89	--	3.83	--

REFERENCES

[1] D. G. Horvitz, J. R. Chromy, and D. A. King, Field Evaluation of the National Assessment Household Survey, Project SU-497 Preliminary Report, Research Triangle Institute, 1970.

[2] Robert B. Pearl, Methodology of Consumer Expenditure Surveys, U. S. Bureau of Census Working Paper No. 27, Washington, D. C.: Government Printing Office, 1968.

[3] Robert Ferber and Seymour Sudman. "Effects of Compensation in Consumer Expenditure Surveys," Annals of Economic and Social Measurement, March 3, 1974, pp. 319-331.

[4] Charles F. Cannell and Ramon Henson. "Incentives, Motives, and Response Bias." Annals of Economic and Social Measurement, March 2, 1974, pp. 307-317.

[5] Henry W. Miller, Jacqueline Kennedy, and E. Earl Bryant, "A Study of the Effect of Remuneration upon Response in a Health and Nutrition Examination Survey," Proceedings of the Social Statistics Section of the American Statistical Association, 1972, pp. 370-375.

[6] J. R. Chromy and D. G. Horvitz, Appendix C: Structure of Sampling and Weighting, Report 1, 1969-1970 Science: National Results and Illustrations of Group Comparisons, Washington, D. C.: Government Printing Office, 1970.

[7] National Assessment of Educational Progress, Report 02-GIY, Reading and Literature: General Information Yearbook, Washington, D. C.: Government Printing Office, 1972.

[8] R. Paul Moore, James R. Chromy, and W. Todd Rogers, The National Assessment Approach to Sampling, Denver, Colorado: National Assessment of Educational Progress, 1974.

THE NATIONAL ASSESSMENT NO-SHOW STUDY: AN EXAMINATION OF NONRESPONSE BIAS[1/]

William D. Kalsbeek, Ralph E. Folsom, Jr., Anne Clemmer
Research Triangle Institute

1. INTRODUCTION

In any sample survey, nonresponse is a serious problem. Nonrespondents may either be difficult to locate or be reluctant to participate once they are located. Cochran and Kish [1,4] enumerate several types of nonrespondents who are difficult to locate, such as young marrieds who do not have small children, the employed females, the higher socioeconomic classes, and the lowest socioeconomic classes; nonrespondents who are sometimes reluctant to participate include the old and the widowed. There are certain individuals who will refuse to be interviewed no matter how skillful the interviewer, who are incapacitated, or who are away from the interview site for the duration of the field-work; these people compose the "hard core" [1] nonrespondents and represent a source of bias in virtually every sample survey.

Several methods have been proposed for reducing the number of nonrespondents. These methods include improved methodological techniques with emphasis on increased cooperation rates among respondents, repeated callings or mailings to nonrespondents, subsampling of nonrespondents, and use of a more expensive callback procedure such as personal interviews [2,3,6], imputation or editing of missing data [7], or a technique of weighting the data inversely proportional to its frequency of occurrence called the Politz Scheme [5]. None are completely satisfactory.

As Kish and Cochran [1,4] have noted, it has become standard practice for better sample surveys to report the size and sources of nonresponse. Nonresponse is a problem which must be handled in different ways for each survey. The National Assessment of Educational Progress[2/] (NAEP) has conducted a nonresponse study on a subsample of 17-year-olds who were enrolled in school. The subsample was composed of 17-year-olds who were intially selected for the National Assessment School sample but who failed to appear for exercise administration. Data on absenteeism and course grades were collected for respondents (original or regular National Assessment respondents) and nonrespondents (No-Shows). These data are termed auxiliary data (appendix A). In addition, data were also collected from No-Shows in an attempt to determine their reasons for failing to appear during the regular Assessment. These data are termed supplementary data, (appendix B).

Attempts were made, over a 3-week period, to contact selected individuals in school. A subsample of those individuals who had not been contacted in school was selected for an out-of-school follow up. Using the data from both in-school and out-of-school nonrespondents, as well as from the original National Assessment, respondents, a model was developed to estimate the nonresponse bias in exercises package performance.

2. SELECTING THE SAMPLE OF NONRESPONDENTS

The National Assessment sample is a three-stage national probability sample which is composed of a school sample, termed the in-school sample, and a household sample, termed the out-of-school sample. This particular nonresponse study involves only the in-school portion of the National Assessment sample.

An in-school sample has been drawn annually since the inception of National Assessment in 1969. The first year of National Assessment was termed Year 01. The nonrespondent sample was selected as a subsample of the Year 04 sample, i.e., 1972-73 school year. In order to understand the method by which the nonrespondent sample was selected, it is necessary to briefly explain the Year 04 sampling design.

The Year 04 Primary Sampling Units (PSU's) were composed of counties or groups of contiguous counties. The primary units were stratified by region, size of community, and socioeconomic characteristics before selection.

Primary units were selected using probabilities proportional to the population of the sampling unit. A total of 118 units were selected by this procedure. The secondary sampling units consisted of public and private schools within selected PSU's. Stratification of the secondary units by income characteristics and size of school took place before selection. Schools were selected using probabilities proportional to the estimated number of eligibles in each school. The tertiary sampling units were students who were enrolled in sample schools, who met certain age requirements, and who were not ineligible for any other reason.[3/] Although the in-school assessment is interested in student respondents from three different ages, 9-year-olds, 13-year-olds, and 17-year-olds, this particular study was limited to the 17-year-old subpopulation since it would appear that the nonresponse problem is most serious in that age class.

Students selected for National Assessment may be administered exercises on a group basis or be given individual interviews. Response rates for the three in-school age classes are recorded by type of administration in table 1. These data are from Year 03 of National Assessment and indicate that the response rates for individual packages tended to be slightly higher than those for group packages. This response difference might be explained by the fact that NAEP places particular importance on the administration of individual exercises. One purpose of the No-Show study is to determine the reasons for this relatively low rate of response among 17-year-olds.

Table 1. Student response rates by age class and package administration mode*

Description	Planned sample size	Sample students assessed	Response rate
9-year-olds			
Group packages	10,368	9,102	87.8
Individual packages	6,480	5,745	88.7
13-year-olds			
Group packages	18,144	15,489	85.4
Individual packages	6,480	5,629	86.9
17-year-olds			
Group packages	23,328	17,229	73.9
Individual packages	6,480	4,842	74.7

*Data recorded from Year 03 of National Assessment.

During any assessment year, two or three different subject matter areas are assessed. The subject matter areas for Year 04 were mathematics and science. Exercises are grouped together into packages. Every package in Year 04 contained a mixture of mathematics and science exercises. From the set of Year 04 packages for 17-year-olds, three group-administered packages and one individually administered package were arbitrarily selected. The group packages were numbered as 01, 03, 09; the individual package was numbered 13. These packages were designated as No-Show packages, and it is by means of these packages that the nonresponse bias in the NAEP-reporting of student performance is assessed.

As was mentioned earlier, the nonrespondent sample was a subsample of the Year 04 sample. The Year 04 PSU's were subdivided into two heterogeneous clusters. The clusters were constructed so as to be well balanced with respect to region, size of community, and socioeconomic characteristics. One cluster was then randomly selected for the No-Show study using equal probabilities. The No-Show primary sample was composed of 57 PSU's.

Eligible schools in the No-Show secondary sample consisted of all 17-year-old sample schools in No-Show PSU's in which at least one of the No-Show packages had been administered during regular National Assessment. Within selected schools, students eligible for the nonresponse study were selected for a particular No-Show package on a matched-sample basis. That is, all students who were originally selected for a group-administered package but who had not appeared for assessment were eligible for any of the No-Show group packages administered in the school during 17-year-old assessmwnt. Similarly, any student who was selected for an individually administered package but who had failed to appear for assessment was eligible for the No-Show individual package, provided that the same package had been administered in the school during 17-year-old assessment. This matched sampling procedure was adopted so that the analysis of differences between respondents and nonrespondents could be made on a within-school basis. Eligible students were selected for specific No-Show packages using cyclic systematic sampling. A subsample of 2,771 students was selected from an original 7,725 17-year-old nonrespondents in Year 04.

Attempts were made to contact selected individuals in school over a 3-week period following the regular assessment. Of the 2,771 students selected for the in-school portion of the study, 34 were determined to be ineligible; a total of 1,990 students out of the 2,737 who were eligible and selected were assessed; thus, the response rate for the in-school portion of the nonresponse study was 72.7 percent. At the end of the 3-week period, the names and addresses of all individuals who had not been contacted were requested from the schools. Several schools refused to release this type of information; however, names and addresses were obtained for 598 of the 747 eligible in-school nonrespondents. A systematic subsample of 130 of No-Show study nonrespondents was selected for the out-of-school portion of the No-Show study. During the out-of-school phase of the study, selected individuals were encouraged to take all four No-Show packages and were given an incentive payment of five dollars for each package which they completed. Ten of the individuals selected for the out-of-school portion of this study were determined to be ineligible. The total number of out-of-school respondents was 102; thus, a response rate of 85 percent was achieved during the out-of-school portion of the No-Show study.

3. NOTATION

The notation given below and used in section 4.1 is defined by PSU (i.e., conditional upon selection of the first-stage PSU's). Furthermore, formulas are developed specifically to subject matter exercises within the package since the subsequent analytic development is similar in each case. A symbol is intended to define an entity, while the attached subscript serves to determine its applicability. A block symbol refers to a random variable, and a script symbol refers to a parameter. Finally, let an upper case script symbol refer to the population of all units and let the corresponding lower case script symbol refer to an estimate of the parameters associated with a sample of these units. Specifically, we define

Y = number of exercises answered correctly,

P (p) = population proportion of eligible NAEP participants,

$\bar{Y}$ ($\bar{y}$) = mean number of exercises answered correctly,

E (e) = number of eligible students.

The first-position subscript (α) associated with the above symbols refers to the total population (o), regular assessment respondents (1), and nonrespondents or No-Shows (2). Population totals F_α and C_α refer to the quantities

$$F_\alpha \equiv \sum_{j\varepsilon\Omega} E_{oj} \; P_{1j} \; \bar{Y}_{\alpha j} \; ; \; (\alpha=1,2)$$

$$C_\alpha \equiv \sum_{j\varepsilon\Omega} E_{oj} \; P_{2j} \; \bar{Y}_{\alpha j} \; ; \; (\alpha=1,2)$$

which are estimated by f_α and c_α respectively. These quantities will be combined to assess the magnitude of nonresponse bias in NAEP regular assessment statistics.

The following symbols are used in the preceding and subsequent formulation:

h = pseudo stratum,

i = PSU within pseudo-stratum,

j = school,

k = student within school,

m = number of eligible sample students taking a package,

w = package sample nonresponse adjusted weight (i.e., inverse of the probability of selection into the study),

Ω = set of all eligible schools,

ω = sample set of elibible schools,

\+ = summation over all possible subscript values.

4. METHODOLOGY

4.1 First-Order PSU Estimators

First, note that the "true" value of $\bar{Y}$ is

$$\bar{Y}_o = \frac{\sum_{j\varepsilon\Omega} E_{oj} \; \bar{Y}_{oj}}{E_o} = \frac{\sum_{j\varepsilon\Omega} E_{oj} \; (P_{1j} \; \bar{Y}_{1j} + P_{2j} \; \bar{Y}_{2j})}{E_o} \; ;$$

$$E_o = \sum_{j=\Omega} E_{oj} .$$

If one lets

$$\bar{y}_1 = \frac{\sum_{j\varepsilon\omega_1} \sum_{k=1}^{m_{1j}} w_{1jk} \; Y_{1jk}}{\sum_{j\varepsilon\omega_1} \sum_{k=1}^{m_{1j}} w_{1jk}} ,$$

then the expectation of the estimator $\bar{y}_1$ is

$$E(\bar{y}_1) = \frac{\sum_{j\varepsilon\Omega} E_{oj} \; \bar{Y}_{1j}}{E_o} .$$

Computation of the w_{1jk}, the regular assessment weights adjusted for total nonresponse, is documented elsewhere [8].

Thus,

$$\text{Bias}(\bar{y}_1) = E(\bar{y}_1) - \bar{Y}_o \qquad (4.1.1)$$

$$= \frac{C_1 - C_2}{E_o} ,$$

since $P_{2j} = 1 - P_{1j}$ by definition

Similarly,

$$\text{Rel-Bias}(\bar{y}_1) = \frac{\text{Bias}(\bar{y}_1)}{\bar{Y}_o} \qquad (4.1.2)$$

$$= \frac{C_1 - C_2}{F_1 - F_2} .$$

Ratio-type estimators are used to estimate values associated with equations (4.1.1) and (4.1.2)

$$\text{bias}(\bar{y}_1) = \frac{c_1 - c_2}{e_o} \equiv \frac{\Upsilon}{\Delta} \qquad (4.1.3)$$

$$\text{rel-bias}(\bar{y}_1) = \frac{c_1 - c_2}{f_1 + f_2} \equiv \frac{\Upsilon}{\xi} \qquad (4.1.4)$$

where

$$e_o = \sum_{j\varepsilon\omega_1} \sum_{k=1}^{m_{1j}} w_{1jk}$$

$$c_1 = \sum_{j\varepsilon\omega_1} p_{2j} \sum_{k=1}^{m_{1j}} w_{1jk} \; Y_{1jk}$$

$$f_1 = \sum_{j\varepsilon\omega_1} p_{1j} \sum_{k=1}^{m_{1j}} w_{1jk} \; Y_{1jk}$$

$$c_2 \equiv f_2 = \sum_{j\varepsilon\omega_2} \sum_{k=1}^{m_{2j}} w_{2jk} \; Y_{2jk} .$$

The parameters P_{1j} and P_{2j} are estimated from school response rates during regular assessment. The estimate for a No-Show study group package is found as the response rate to all group packages given in that school. Similarly, the No-Show study individual package response rate is obtained from the response rate to all individual packages given in the school. Recall that the w_{2jk} weights denote the reciprocals of No-Show selection probabilities adjusted for No-Show nonresponse.

The preceding statistics yield bias estimates involving in-school regular assessment respondents and all No-Show respondents. Another set of meaningful bias estimates involves in-school regular assessment respondents and in-school No-Show respondents. The definition changes indicated by the (*) were motivated by the attempt to form a matched school bias estimator based exclusively on in-school No-Shows. The set of schools ω_1^* is the subset of regular assessment ω_1 schools which provided in-school No-Show responses for the particular package in question. The deleted schools either had no cooperating in-school No-Show respondents for the package, or were subsampled out at the No-Show package assignment stage to control the package yield per PSU. The regular assessement respondent for the set of ω_1^* schools with in-school No-Show responses for the package were inflated to account for the deleted schools, hence the adjusted w_{1jk}^* weights.

Regarding the components of equations (4.1.3) and (4.1.4),

$$\text{bias}^*(\bar{y}_1) = \frac{c_1^* - c_2^*}{e_o^*} \equiv \frac{\Upsilon^*}{\Delta^*} \qquad (4.1.5)$$

and

$$\text{rel-bias}^*(\bar{y}_1) = \frac{c_1^* - c_2^*}{f_1^* + f_2^*} \equiv \frac{\Upsilon^*}{\xi^*} \qquad (4.1.6)$$

where

$$e_o^* \equiv \sum_{j\varepsilon\omega_1^*} \sum_{k=1}^{m_{1j}} w_{1jk}^*$$

$$c_1^* \equiv \sum_{j\varepsilon\omega_1^*} p_{2j} \sum_{k=1}^{m_{1j}} w_{1jk}^* \; Y_{1jk}^*$$

$$f_1^* \equiv \sum_{j\varepsilon\omega_1^*} p_{1j} \sum_{k=1}^{m_{1j}} w_{1jk}^* \; Y_{1jk}$$

$$c_2^* \equiv \delta_2^* = \sum_{j\varepsilon\omega_1^*} \frac{p_{2j}\ m_{1j}}{m_{2j}^*} \sum_{k=1}^{m_{2j}^*} w_{1jk}^*\ Y_{2jk}^* .$$

With m_{2j}^* denoting the number of in-school No-Show responses from school $j\varepsilon\omega_1^*$, the definition of the set of schools ω_1^* assures that $m_{2j}^* > 0$. Since the adjusted weights $w_{1jk}^* = w_{1j}^* E_{oj} / m_{1j}$ with w_{1j}^* denoting the adjusted school by package weight, one can recast c_1^* and c_2^* as follows:

$$c_1^* = \sum_{j\varepsilon\omega_1^*} w_{1j}^*\ p_{2j}\ E_{oj}\ \bar{y}_{1j}^* ,$$

$$c_2^* = \sum_{j\varepsilon\omega_1^*} w_{1j}^*\ p_{2j}\ E_{oj}\ \bar{y}_{2j}^* .$$

The numerator of equations (4.1.5) and (4.1.6) is therefore

$$(c_1^* - c_2^*) = \sum_{j\varepsilon\omega_1^*} w_{1j}^*\ p_{2j}\ E_{oj}\ (\bar{y}_{1j}^* - \bar{y}_{2j}^*). \quad (4.1.7)$$

4.2 Overall First-Order Estimates

To facilitate the ensuing discussion, attach subscripts to γ, Δ, and ξ of (4.1.3) and (4.1.4) and γ^*, Δ^*, and ξ^* of (4.1.5) and (4.1.6) (i.e., subscripts "hi" to indicate PSU-i within pseudo stratum-h). Using these quantities, one obtains the overall estimate involving all No-Shows as

$$\text{bias}\ (\bar{y}_1) = \gamma_{++} / \Delta_{++} \quad (4.2.1)$$

$$\text{rel-bias}\ (\bar{y}_1) = \gamma_{++} / \xi_{++} \quad (4.2.2)$$

and involving only in-school No-Shows as

$$\text{bias}^*\ (\bar{y}_1) = \gamma_{++}^* / \Delta_{++}^* \quad (4.2.3)$$

$$\text{rel-bias}^*\ (\bar{y}_1) = \gamma_{++}^* / \xi_{++}^* \quad (4.2.4)$$

4.3 Second-Order Estimators

The second-order estimators of variance for expressions (4.2.1) through (4.2.4) are based upon a form of the "jackknife" technique introduced by Quenouille [9] and advanced for interval estimation by Tukey [10]. The exact form used here was presented by Frankel [11]. The procedure is presented for estimates involving all No-Shows, although the procedure for estimates involving only in-school No-Shows is similar.

First, these definitions are given:

$$\beta_{h1} = \frac{2\gamma_{++}}{\Delta_{++}} - \left[\frac{\gamma_{++} + \gamma_{h1} - \gamma_{h2}}{\Delta_{++} + \Delta_{h2} - \Delta_{h2}}\right]$$

$$\beta_{h2} = \frac{2\gamma_{++}}{\Delta_{++}} - \left[\frac{\gamma_{++} + \gamma_{h2} - \gamma_{h1}}{\Delta_{++} + \Delta_{h2} - \Delta_{h1}}\right]$$

$$\zeta_{h1} = \frac{2\gamma_{++}}{\xi_{++}} - \left[\frac{\gamma_{++} + \gamma_{h1} - \gamma_{h2}}{\xi_{++} + \xi_{h1} - \xi_{h2}}\right]$$

$$\zeta_{h2} = \frac{2\gamma_{++}}{\xi_{++}} - \left[\frac{\gamma_{++} + \gamma_{h2} - \gamma_{h1}}{\xi_{++} + \xi_{h2} - \xi_{h1}}\right].$$

Since the 57 PSU's make up a half-sample of NAEP regular assessment PSU's, the desirable condition of having two PSU selections per stratum does not hold. Instead, pseudo strata were formed by sequentially pairing the No-Show PSU's according to region and size. Since the number of PSU's is odd, one pseudo stratum (h^o) was assigned three PSU's The associated jackknife estimators of variance are

$$\text{var}\{\text{bias }(\bar{y}_1)\} = 1/4 \sum_{\substack{h=1\\ h\neq h^o}}^{H} \left[\beta_{h1} - \beta_{h2}\right]^2 + 1/8 \sum_{i=1}^{2} \sum_{j=i+1}^{3} \left[\beta_{h^o i} - \beta_{h^o j}\right]^2$$

and

$$\text{var}\{\text{rel-bias }(\bar{y}_1)\} = 1/4 \sum_{\substack{h=1\\ h\neq h^o}}^{H} \left[\zeta_{h1} - \zeta_{h2}\right]^2 + 1/8 \sum_{i=1}^{2} \sum_{j=i+1}^{3} \left[\zeta_{h^o i} - \zeta_{h^o j}\right]^2 .$$

To assess the significance of the bias and rel-bias estimates, one might be willing to assume that

$$T = \frac{\text{bias }(\bar{y}_1)}{\left[\text{var}\{\text{bias }(\bar{y}_1)\}\right]^{1/2}}$$

$$T' = \frac{\text{rel-bias }(\bar{y}_1)}{\left[\text{var}\{\text{rel-bias }(\bar{y}_1)\}\right]^{1/2}}$$

are distributed as a "Student's" t-statistic with 29 degrees of freedom. Under this assumption, significance with a Type I error of 0.05 is indicated when $|T| \geq 2.045$ or $|T'| \geq 2.045$.

5. RESULTS AND DISCUSSION

Results obtained by applying the methodology of section 4 to group packages, 01, 03, and 09 and individual package 13 are now presented and discussed. Sample sizes associated with these findings are found in table 2. The actual results are presented in table 3. Weighted estimates of the mean correct response for subject matter exercises within a package are found in table 4.

In many instances, comparisons involving all No-Shows are statistically significant subject to the assumptions made previously. Since most biases and rel-biases are positive, it would appear that the regular assessment respondents are somewhat better students than No-Shows. These results parallel the results of appendix A in which regular assessment and No-Show data are compared with respect to other auxiliary measures. These data indicate that regular assessment respondents tend to miss school less frequently, get better course grades, take more courses, and have more academic-oriented goals. The figures of table 4 indicate a similar type of performance differential between regular assessment and No-Show respondents.

As one might anticipate from these preliminary findings, most of the bias values in table 3 are positive. The only negative results are small and could be attributable to sampling variation.

With the group packages, the magnitude of bias is reduced and generally not statistically significant when only in-school No-Shows are involved. The figures of table 4 imply similar results when one notices the intermediate performance of the in-school No-Shows relative to regular assessment and out-of-school No-Show respondents. This indicates that in-school No-Shows may be more similar to the regular assessment respondents and have been absent from the regular assessment for different reasons than the out-of-school No-Shows.

The results of table B.1 in appendix B indicate that the majority of in-school No-Shows were not absent from school on the day in which they were originally scheduled to be assessed. Furthermore, a moderate proportion of those who said they were not absent from school indicated that they had other school-oriented commitments.

The results for individual package 13 do not conform to the patterns established with the group packages. The magnitude of the biases associated with comparisons involving only in-school No-Shows are somewhat greater than comparisons involving all No-Shows. Both science exercise biases are small and not significant while significance is indicated with the latter explanation for these findings is not clear; however, it should be noted that the content and administration of this individual package 13 differ from the group package. Group packages are administered to a much greater proportion of participating students and are essentially self-administered. Individual packages on the other hand, require continual interaction by exercise administrators and contain several exercises which involve opinion, probing, and branching among parts of the exercise.

Table 2. No-Show study sample sizes

Package	Regular Assessment Respondents	No-Show Respondents		
	Total	In-School	Out-of-School	Total
01	1148	522	100	622
03	1209	536	100	636
09	1144	492	98	590
13	1086	439	99	538

Table 3. Bias estimates and estimates of precision

Package	Number of Exercises	Bias (B) or Rel-Bias (RB)	All No-Shows		In-School No-Shows Only	
			Estimate	Variance	Estimate	Variance
MATHEMATICS:						
01	16	B	0.3385*	$(0.0780)^2$	0.0721	$(0.0472)^2$
		RB	0.0374*	$(0.0089)^2$	0.0078	$(0.0051)^2$
03	19	B	0.5545*	$(0.1012)^2$	0.1771*	$(0.0670)^2$
		RB	0.0554*	$(0.0106)^2$	0.0172*	$(0.0065)^2$
09	19	B	0.3316	$(0.2166)^2$	0.0148	$(0.1278)^2$
		RB	0.0360	$(0.0243)^2$	0.0016	$(0.0136)^2$
13	22	B	0.1387	$(0.1650)^2$	0.2284*	$(0.0856)^2$
		RB	0.0113	$(0.0136)^2$	0.0191*	$(0.0072)^2$
SCIENCE:						
01	23	B	0.3801*	$(0.0935)^2$	0.1445	$(0.0508)^2$
		RB	0.0327*	$(0.0082)^2$	0.0124	$(0.0044)^2$
03	13	B	0.2510*	$(0.0570)^2$	0.0674	$(0.0488)^2$
		RB	0.0425*	$(0.0102)^2$	0.0111	$(0.0081)^2$
09	28	B	0.3874	$(0.4517)^2$	-0.0345	$(0.1832)^2$
		RB	0.0210	$(0.0249)^2$	-0.0018	$(0.0097)^2$
13	7	B	0.0001	$(0.0384)^2$	0.0120	$(0.0271)^2$
		RB	0.0000+	$(0.0122)^2$	0.0038	$(0.0087)^2$

*Probable significance with a Type I error of 0.05

Table 4. Weighted estimates of mean number correct responses*

Package	Regular Assessment Respondents	No-Show Respondents		
	Total	In-School	Out-of-School	Total
MATHEMATICS:				
01	9.4	8.8	5.7	8.1
03	10.6	9.8	6.4	9.0
09	9.5	9.1	5.8	8.4
13	12.4	11.0	8.1	10.3
SCIENCE:				
01	12.0	11.3	8.9	10.7
03	6.2	5.9	4.6	5.6
09	18.9	18.5	15.7	17.9
13	3.2	3.0	2.3	2.8

*
$$\bar{y}_\alpha = \frac{\sum_{j \epsilon \omega_\alpha} \sum_{k=1}^{m_{\alpha j}} w_{\alpha j k} Y_{\alpha j k}}{\sum_{j \epsilon \omega_\alpha} \sum_{k=1}^{m_{\alpha j}} w_{\alpha j k}}$$

Appendix A

Absentee and course grade data were collected for certain in-school respondents and nonrespondents. These data were termed auxiliary data. The data were collected for all in-school respondents who were selected for either the first group administered No-Show package (i.e., package 01) or the individually administered No-Show package (i.e., package 13). Auxiliary data were also collected for all nonrespondents who were selected for the in-school portion of the study and who were assigned either the first group administered No-Show package or the individually administered No-Show package.

Appendix B

A special supplementary questionnaire was developed for the nonresponse study to ascertain the reason a selected student failed to appear for assessment. Separate questionnaires were developed for the in-school and out-of-school portions of the study. These supplementary data were collected from all individuals who were selected for either the in-school or out-of-school portions of the nonresponse study and who participated in the study.

Table A.1. Auxiliary comparative frequency data by percent days absent through February 1973

Percent Days Absent	Selected Respondents: Absolute	Selected Respondents: Adjusted Relative (Percent)	No-Show Respondents: Absolute	No-Show Respondents: Adjusted Relative (Percent)
0-10	1362	81.3	674	61.2
11-20	253	15.1	254	23.1
21-30	39	2.3	89	8.1
31-40	14	0.8	34	3.1
41-50	5	0.3	18	1.6
51-60	0	0.0	23	2.1
61-70	2	0.1	4	0.4
71-80	0	0.0	3	0.3
81-90	0	0.0	2	0.2
91-100	0	0.0	0	0.0
NA	78	-	223	-
Total	1753	100.0	1324	100.0
Mean	5.9		11.3	
Median	4.2		7.3	

Table A.2. Auxiliary comparative frequency data by individual grade point average for the most recent available reporting period ending prior to March 1, 1973

Grade Point Average*	Selected Respondents: Absolute	Selected Respondents: Adjusted Relative (Percent)	No-Show Respondents: Absolute	No-Show Respondents: Adjusted Relative (Percent)
0.0-0.5	10	0.6	73	6.6
0.6-1.0	35	2.1	62	5.6
1.1-1.5	119	7.1	125	11.3
1.6-2.0	198	11.8	164	14.8
2.1-2.5	315	18.8	233	21.1
2.6-3.0	339	20.2	159	14.4
3.1-3.5	387	23.1	191	17.3
3.6-4.0	275	16.4	99	9.0
NA	75	-	218	-
Total	1753	100.0	1324	100.0
Mean	2.60		2.15	
Median	2.66		2.22	

*Based upon a four-point system: A = 4, B = 3, C = 2, D = 1, F = 0; involves only those courses for which a letter grade was given.

Table A.3. Auxiliary comparative frequency data by number of courses taken in the most recent available reporting period ending prior to March 1, 1973

Number of Courses	Selected Respondents Absolute	Selected Respondents Relative (Percent)	No-Show Respondents Absolute	No-Show Respondents Relative (Percent)
0	0	0.0	56	4.2
1	74	4.2	159	12.0
2	21	1.2	13	1.0
3	43	2.5	54	4.1
4	178	10.2	152	11.5
5	536	30.6	342	25.8
6	598	34.1	364	27.5
7	181	10.3	112	8.5
8	122	7.0	72	5.4
Total	1753	100.0	1324	100.0
Mean	5.40		4.69	
Median	5.54		5.17	

Table A.4. Auxiliary comparative frequency data by curriculum type

Curriculum	Selected Respondents Absolute	Selected Respondents Adjusted Relative (Percent)	No-Show Respondents Absolute	No-Show Respondents Adjusted Relative (Percent)
College Preparatory - Academic	935	55.6	446	39.5
Vocational - Occupational	329	19.6	307	27.2
Other	258	15.3	199	17.6
Cannot Determine	159	9.5	177	15.7
Left Blank	72	-	195	-
Total	1753	100.0	1324	100.0

Table B.1. In-school No-Show supplementary frequency data by whether the respondent was absent from school on the scheduled participation date

Were you absent from school the day in which you were scheduled to participate?	Absolute	Adjusted Relative (Percent)
Yes	715	36.1
No	1032	52.0
Cannot Remember	236	11.9
Refused	0	-
Blank	6	-
Total	1989	100.0

Table B.2. In-school No-Show supplementary frequency data by reason for nonparticipation in original assessment

If you were not absent, why then did you not participate?	Absolute	Relative (Percent)	Adjusted Relative (Percent)
Was late and the session had begun	126	6.3	17.5
Had extracurricular activities	63	3.2	8.7
Had an examination or important class	95	4.8	13.2
Was notified but forgot	79	4.0	10.9
Went home sick	18	0.9	2.5
Was late for school	17	0.9	2.4
Had work study commitments	45	2.3	6.2
Had job commitments	31	1.6	4.3
Cannot remember	32	1.6	4.4
Other	216	10.9	29.9
Refused	0	0.0	0.0
Blank	1267	63.7	-
Total	1989	100.0	100.0

REFERENCES

[1] Cochran, W. G., Sampling Techniques, New York: John Wiley and Sons, Inc., 1963.

[2] El-Badry, M. A., "A Sampling Procedure for Mailed Questionnaires," Journal of the American Statistical Association, 51(1956), 209-27.

[3] Hansen, M. H., and Hurwitz, W. N., "The Problem of Nonresponse in Sample Surveys," Journal of the American Statistical Association, 41(1946), 517-29.

[4] Kish, Leslie, Survey Sampling, New York: John Wiley and Sons, Inc., 1965.

[5] Politz, A. N., and Simmons, W. R., "An Attempt to Get the 'Not at Homes' into the Sample Without Callbacks," Journal of the American Statistical Association, 44(1949), 9-31.

[6] Srinath, K. P., "Multiphase Sampling in Nonresponse Problems," Journal of the American Statistical Association, 66(1971), 583-86.

[7] Taeuber, C., and Hansen, M. H., "A Preliminary Evaluation of the 1960 Censuses of Population and Housing," Proceedings of the Social Statistics Section, American Statistical Association, 1963, 56-73.

[8] Clemmer, A., "Formulation and Summary for Age 17 Weights NAEP Year 04 In-School Sample," Working Paper No. 796, September 1973.

[9] Quenouille, M. H., "Approximate Tests of Correlation in Time Series," Journal of the Royal Statistical Association, Series B, 11(1949), 68-84.

[10] Tukey, J. W., "Bias and Confidence in Not-Quite Large Samples" (Abstract), Annals of Mathematical Statistics, 29(1958), 614.

[11] Frankel, M., "An Empirical Investigation of Some Properties of Multivariate Statistical Estimates from Complex Samples," Institute for Social Research Monograph, Ann Arbor, Michigan, 1971.

Footnotes

1/ The work upon which this publication is based was performed pursuant to a Contract with the Education Commision of the States, utilizing funds from the U.S. Office of Education, Department of Health, Education and Welfare, Contract No. OEC-0-74-0506. However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education or the Education Commission of the States, and no official endorsement by the U.S. Office of Education or the Education Commission of the States should be inferred.

2/ A project of the Education Commission of the States (ECS).

3/ Individuals who are emotionally or mentally retarded, functionally disabled, non-English speaking, or nonreaders, are excluded from the NAEP sample.

THE NEW SUPPLEMENTAL SECURITY INCOME PROGRAM

Richard Bell, Division of Supplemental Security Studies
Office of Research and Statistics, Social Security Administration

On January 1, 1974, the supplemental security income program (SSI) became the vehicle for providing monthly cash benefit payments to the aged, blind, and disabled with inadequate income. 1/ This new program replaced the former Federal grants to States for old-age assistance (OAA), aid to the blind (AB), and aid to the permanently and totally disabled (APTD) in the 50 States and the District of Columbia.

The SSI program establishes uniform, nationwide basic payment standards and eligibility requirements. It is administered by the Social Security Administration with benefits being paid from U.S. Treasury general funds. In addition, there is provision for State supplementation of the basic Federal payments. During June 1974, nearly 3.6 million persons received federally administered SSI benefits totalling $448 million. Of this amount, $333 million represented Federal SSI payments and $115 million State supplementary payments.

A. DESCRIPTION OF THE PROGRAM

Benefits and eligibility requirements

The basic monthly Federal benefit amount is $146 for an eligible individual and $219 2/ for an eligible couple, living alone and without any income. However, if the individual or couple is living in another's household and receiving support and maintenance there, the standard payment is reduced by one-third. SSI beneficiaries in public or private institutions, who receive more than 50 percent of the cost of their care from the Medicaid program under Title XIX of the Social Security Act, may receive a maximum payment of $25 per month. Eligible beneficiaries in private institutions whose care is not met from Medicaid funds may receive the standard payment amount of $146. Inmates of public institutions for whom Medicaid is not the main source of support are ineligible for SSI benefits.

There is a requirement that the States maintain income at the December 1973 level for an individual converted from OAA, AB, or APTD to the SSI program. Consequently, his December 1973 State public assistance payment plus other income is compared with his basic Federal SSI payment plus current income. If the latter sum is smaller than the former, the difference must be made up in a "mandatory" State supplementary payment.

Moreover, "essential persons" are also considered. These are persons (a) whose needs were taken into account for December 1973 in determining the needs of an eligible individual for State assistance under the State plan in June 1973, (b) who live in the same household as the individual, and (c) who are not eligible in their own right or as a spouse for SSI payments. Thus, the only beneficiaries who could receive higher benefits because of the presence of an essential person are those whose State assistance payments in December 1973 took the needs of such persons into account. The increase may be as much as $73 for each essential person.

In addition to the "mandatory" payment level maintenance mentioned above, a State may choose to further supplement the Federal SSI payment of beneficiaries irrespective of prior eligibility for OAA, AB, or APTD. The extent of coverage under this "optional" State supplementation varies from State to State. Some States provide additional benefits for all persons who qualify for basic SSI payments, other States limit supplementation to certain groups, such as the blind or those in domiciliary care facilities.

Income and resource provisions

There are different kinds of income exclusions. First, on a monthly basis, $20 of income is disregarded, whether earned--wages, salaries, or self-employment income--or unearned income, e.g., social security benefits, other governmental or private pensions, veteran's compensation or workman's compensation. In addition, $65 monthly of earned income and one-half of the remainder is also disregarded. Further, there are also a number of special exclusions including the following: certain earnings of a student beneficiary, infrequent or irregular income, income necessary to fulfilling an approved self-support plan of a blind or disabled beneficiary, work expenses of blind beneficiaries, and one-third of support payments made by an absent parent.

To be eligible for benefits, the value of a person's resources may not exceed $2,250 if the individual has a spouse with whom he is living; otherwise, their value cannot exceed $1,500. However, there are also a number of resource exclusions. A home, automobile, household goods, and personal effects of reasonable value are all excluded. If the face value of all insurance policies on any one person is less than $1,500, they will not be taken into account.

Administration of State supplementation programs

The mandatory and optional State supplementation program mentioned above can be administered, at the State's discretion, either by the State itself or by the Federal government. In the latter case the State reimburses the amount of the payments made, while the administrative costs are borne by the Federal government.

For those States choosing Federal administration of their supplementary payments, there is the advantage of the following "hold harmless" provision. To the extent that the State payments do not, on the average, exceed its "adjusted payment level", the State's financial liability for the amount of supplementary payments under the program is limited to the total expenditures for cash assistance to the aged, blind, and disabled in calendar year 1972; the Federal government pays the rest of the cost of such State supplementary payments.

A State's adjusted payment level is the average of payments which individuals, with no other income, received in January 1972. The adjusted payment level may also include the bonus value of food stamps. Where payments were less than State standards, there is also provision for adjustment. 3/

B. SUPPLEMENTAL SECURITY INCOME PROGRAM STATISTICS

Summary of national data

Approximately 3.4 million persons received federally administered supplemental security income (SSI) benefits in May 1974. 4/ Of these 2.0 million were awarded benefits on the basis of age, 1.3 million because of disability, and 72,000 on the basis of blindness (table 1). Of those eligible for benefits, 2.9 million had been eligible for State public assistance payments in December 1973 (table 2). The average monthly amounts for the Federal SSI and the federally administered State supplementary payments for aged, blind, and disabled individuals and couples converted from the former State programs were considerably higher than the corresponding amounts for new SSI beneficiaries. The above relationship indicates that the SSI program is reaching a population with higher income than did old-age assistance, aid to the blind, and aid to the permanently and totally disabled (table 4).

For both adult individuals and couples the majority had Federal payments only (table 3). However, a substantial number received both Federal and State payments, whereas a relatively small number received State supplementation only. The average monthly Federal payments were considerably higher than the State payments to individuals, but this margin was small for couples. This difference partially follows from the fact that couples more frequently than individuals have social security benefits which are deducted from the Federal SSI payments (80 and 66 percent for aged couples and individuals, respectively). Moreover, in some cases the previous State assistance programs treated couples more favorably, relative to individuals, than does the Federal SSI program. Finally, California, which accounts for 30 percent of all couples with federally administered State supplementation, provides optional monthly payments of $95 to individuals without income and living in their own households; and $230 to couples.

Impact of social security benefits

About 50 percent of the persons with federally administered SSI payments also received social security benefits (table 6). Although approximately two-thirds of aged SSI individuals had social security payments, only about one-third of blind and one-fourth of disabled adult individuals received such payments. Only 6.8 percent of the children under the program also received social security benefits.

The national average monthly Federal SSI payments for adult individuals were $95.37, $135.22, and $136.97 for the aged, blind, and disabled, respectively (table 4). The relatively greater amounts for the blind and disabled reflect the fact that there is a larger proportion of aged with social security benefits than for the blind and disabled.

In fact, the amount of social security payments accounts for almost all the countable income used in determining the SSI benefits. Theoretically if social security benefits were the only source of income, for a given subset of Federal SSI individual beneficiaries, the average Federal SSI monthly benefit, B, would be computed as follows:

$$B = \frac{140I_1 + 93.33I_2 + 25I_3}{I_1 + I_2 + I_3} - (S - 20)\ P,$$

where,

I_1 = number of individuals living in their own household,

I_2 = number of individuals living in another's household,

I_3 = number of individuals living in institutions covered by Medicaid,

P = proportion of individuals with Federal SSI payments in concurrent receipt of social security benefits,

S = average social security benefit of individuals with Federal SSI benefits having social security benefits.

Analogously, for couples we have:

$$B = \frac{210I_1 + 140I_2 + 50I_3}{I_1 + I_2 + I_3} - (S - 20)\ P,$$

where all the variables now refer to numbers of and averages among couples.

Deviations between the tabulated average monthly Federal SSI amount and the number computed by the above formula measure the effects of earned and unearned income (other than social security) upon the reduction of SSI benefits. In particular, the small differences between these tabulated and computed amounts given below indicate the impact of social security benefits and the limited effect of earned and other unearned income upon the determination of the overall SSI average payment amounts.

Value (based on March 1974 data)	Aged		Disabled	
	Individuals	Couples	Individuals	Couples
Computed value	$81.53	$106.84	$112.42	$148.37
Tabulated value	81.29	100.61	113.26	145.10

The slight excess of the computed value over the tabulated value 5/ reflects the contribution of earned and other unearned income to the total income of SSI beneficiaries.

The relatively small amount of earned income and of unearned income among SSI beneficiaries receiving OAA, AB, and APTD in December 1973 should be noted (tables 7 and 9). This further points to the importance of social security benefits to the low income population receiving supplemental security income.

C. APPLICANTS UNDER THE SSI PROGRAM

Although payments under the supplemental security income program began in January 1974, applications for benefits were taken by Social Security Administration district offices beginning July 1973.

Applications during the make-ready period (July-December 1973)

During the 6-month period July-December 1973, 507,400 applications for benefits under the SSI program were filed with Social Security Administration district offices. Of these, 325,600 came from persons over 65 years of age, and 181,800 from applicants under 65 claiming blindness or disability.

It should be noted that a significant number of the SSI applicants also filed for the former State programs during the make-ready period. Some of these individuals were awarded benefits under the State programs and were subsequently converted to the SSI program on January 1, 1974 along with all the other OAA, AB, and APTD beneficiaries. The number of such duplicate applicants was approximately 200,000.

Applications during January-June 1974

During the first 6 months of the SSI program, 1,314,000 people filed for benefits, 544,000 individuals were aged applicants; the remaining 770,000 claimed blindness or disability (table 11).

As of early July, one million applications had been fully processed, resulting in payment awards in about two-thirds of the cases and a finding of ineligibility in about one-third. Of the half-million cases in process, the majority are disability cases that require medical determination and review in State agencies.

D. THE ORS STATISTICAL PROGRAM FOR SSI

The Division of Supplemental Security Studies in the Office of Research and Statistics has the responsibility for collecting and analyzing data on the characteristics of participants in the new program and the amounts they receive. In addition, information will be developed on the dynamics of change in beneficiary circumstances. The Division is also responsible for providing estimates of program participants and costs for budgetary purposes, legislative planning, SSA workload determinations, and research into measures of income adequacy as they affect the low income group.

Some of the Division's activities are described below:

The Survey of Low Income Aged and Disabled (SLIAD) is a two-stage panel study providing data to assess the effects of SSI on its target population. Information concerning income, assets, debts, household expenditures, health, mobility, employment, social interaction, adequacy of and attitudes towards welfare assistance was obtained by means of personal interviews with 18,000 aged, blind, and disabled during the last three months of 1973. The same sample respondents will be asked identical questions during the last three months of 1974. By providing descriptive profiles of beneficiaries prior and subsequent to the initiation of SSI, analyses will be developed on the extent to which SSI benefits secure standards of living consonant with program intent.

SLIAD consists of four independent samples. To allow a comparison of SSI with the former State public assistance programs, two samples were selected, the first from a frame consisting of persons with OAA, and the second, those with AB and APTD. Further, to gauge the impact of SSI upon the general population, two additional samples were taken from the universe of low income persons, not necessarily receiving State assistance payments.

For the low income general population sample, the frame consists of aged or disabled individuals in the CPS sample with incomes of less than $6,500 if married or less than $5,000 if living alone. The sample was selected by the Bureau of the Census under standard CPS procedures.

The Longitudinal Supplementary Security Income Sample (LSSI), housing 550,000 cases selected from the SSI applicant universe, is designed to provide statistical data based on SSI program records for in-depth analyses of selected areas.

In particular, data will be gathered to reflect changes in incomes and their effects on benefit levels, residence changes; and for disabled persons--primary and secondary diagnosis, occupation and education.

The LSSI design is a stratified sample, where each State-category (aged, blind, or disabled) group, based on initial application, represents a stratum. Digital sampling, based on digits of the social security numbers is used for selection of cases within the stratum. The sample size, 550,000, was determined so that the maximum sampling variability for the proportion of cases in a given State possessing a particular attribute would be no greater than ±2 percent at the 95 percent confidence level.

Program statistics

Statistics relevant to beneficiaries with federally administered payments (Federal SSI or federally administered State supplementation) are being obtained from the Supplemental Security Record 6/ on a universe basis. Similar information concerning State administered State supplementation is being collected separately from the States with these programs.

1/ The SSI program--Title XVI of the Social Security Act---was established under the 1972 Social Security Amendments.

2/ Prior to July 1, 1974, these amounts were $140 and $210 respectively.

3/ For a detailed account of the basic provisions of the SSI program see James C. Callison, "Early Experience Under the Supplemental Security Income Program", Social Security Bulletin, June 1974, pp. 3-12.

4/ Since May is the latest month for which comprehensive data are available at the time of this writing, most of the figures in this section are based on experience during that month.

5/ The payments for essential persons corresponding to each category is not available at this time. However, allocating the total number, 117,200 of essential persons to each category and assuming a maximum payment of $70.00 to them, we can adjust upward the computed values for aged and disabled individuals by $3.95 and $1.71, respectively. This correction can explain, moreover, the excess of the tabulated value over the computed value for disabled individuals.

6/ The tape file containing basic data on income, living arrangements, and other characteristics necessary to establish and maintain the eligibility and benefit amounts for all individuals under the SSI program.

TABLE 1.--Supplemental security income for the aged, blind, and disabled: Number of persons receiving federally administered payments and total amount, 1974

Period	Number of persons 1/ 2/				Benefits (in thousands) 1/		
	Total	Aged	Blind	Disabled	Total	Federal SSI	State supplementation 3/
1974							
January.......	3,215,632	1,865,109	72,390	1,278,133	$365,149	$260,159	$104,989
February......	3,236,167	1,882,690	72,584	1,280,893	405,050	311,686	93,364
March.........	3,333,017	1,944,966	74,048	1,314,003	416,841	313,171	103,670
April.........	3,334,493	1,953,693	73,189	1,307,611	392,122	290,029	102,092
May...........	3,425,891	2,016,177	71,833	1,337,881	410,240	307,322	102,918
June..........	3,583,894	2,093,301	72,883	1,417,710	448,142	332,776	115,366

1/ Excludes emergency advance payments made by Social Security Administration district offices. Figures have not been adjusted for returned checks and refunds of overpayments.

2/ Does not include persons receiving only State supplementation under State-administered programs.

3/ Does not include payments for State supplementation under State-administered programs.

TABLE 2.--Supplemental security income for the aged, blind, and disabled: Number of persons receiving federally administered payments by type of payment and reason for eligibility, May 1974 1/

Type of payment	Total	Aged	Blind	Disabled
	All persons			
Total....................................	3,425,900	2,016,200	71,800	1,337,900
Federal SSI payments...........................	3,135,100	1,807,400	65,600	1,262,100
Federal SSI payments only....................	1,988,800	1,259,700	38,600	690,500
Federal SSI and State supplementation.........	1,146,200	547,700	26,900	571,600
State supplementation..........................	1,437,100	756,500	33,200	647,400
State supplementation only...................	290,800	208,800	6,200	75,800
	Converted from State programs			
Total....................................	2,927,900	1,664,700	71,000	1,192,200
Federal SSI payments...........................	2,698,000	1,506,600	64,800	1,126,600
Federal SSI payments only....................	1,645,700	1,024,400	38,000	583,300
Federal SSI and State supplementation.........	1,052,300	482,200	26,800	543,300
State supplementation..........................	1,282,000	640,300	32,900	608,800
State supplementation only...................	229,800	158,000	6,200	65,600
	Not converted from State programs			
Total....................................	498,100	351,500	900	145,700
Federal SSI payments...........................	437,000	300,700	800	135,500
Federal SSI payments only....................	342,900	235,200	600	107,100
Federal SSI and State supplementation.........	94,100	65,500	200	28,400
State supplementation..........................	155,100	116,300	200	38,600
State supplementation only...................	61,100	50,800	100	10,200

1/ Number of persons partly estimated.

TABLE 3.--Supplemental security income: Number of adult units and children receiving federally administered payments by type of payment, reason for eligibility and conversion status, May 1974 1/

Type of payment	Adult units						Children	
	Aged		Blind		Disabled			
	Individual	Couple	Individual	Couple	Individual	Couple	Blind	Disabled
	Total							
Total	1,717,900	150,700	64,000	2,800	1,263,200	23,200	1,900	18,800
Federal SSI payments	1,543,700	133,300	58,400	2,500	1,193,900	20,700	1,900	18,600
Federal SSI payments only	1,056,300	102,700	34,000	1,500	643,700	12,600	1,300	14,800
Federal SSI and State supplementation	487,400	30,600	24,400	1,000	550,200	8,000	600	3,800
State supplementation	661,500	48,100	30,000	1,300	619,500	10,600	600	4,000
State supplementation only	174,200	17,500	5,600	300	69,300	2,600	2/	200
	Converted from State payments							
Total	1,421,400	122,300	63,200	2,800	1,139,700	20,500	1,900	4,700
Federal SSI payments	1,288,500	109,700	57,700	2,500	1,079,700	18,200	1,900	4,600
Federal SSI payments only	861,700	81,600	33,400	1,500	555,100	10,400	1,200	3,000
Federal SSI and State supplementation	426,800	28,100	24,200	1,000	524,600	7,800	600	1,700
State supplementation	559,700	40,700	29,800	1,300	584,600	10,100	600	1,800
State supplementation only	132,900	12,600	5,500	300	60,000	2,300	2/	100
	Not converted from State programs							
Total	296,500	28,400	800	2/	123,500	2,700	100	14,100
Federal SSI payments	255,200	23,600	700	2/	114,200	2,500	100	14,000
Federal SSI payments only	194,700	21,000	600	2/	88,600	2,200	2/	11,900
Federal SSI and State supplementation	60,600	2,500	200	2/	25,600	200	2/	2,100
State supplementation	101,800	7,400	200	2/	34,900	500	2/	2,200
State supplementation only	41,200	4,900	100	2/	9,300	200	2/	100

1/ Number of adult units and children partly estimated.

2/ Less than 50.

TABLE 4 .--Supplemental security income: Average monthly amount payable June 1, to adult units and children eligible for federally administered payments, May 1974, by type of payment and reason for eligibility

Type of payment	Adult units						Children	
	Aged		Blind		Disabled			
	Individual	Couple	Individual	Couple	Individual	Couple	Blind	Disabled
	Total							
Total	$ 95.37	$116.80	$135.22	$203.50	$136.97	$181.00	$147.48	$ 92.07
Federal SSI payments	78.95	98.22	106.90	150.35	110.60	142.57	117.95	79.00
Federal SSI payments only	82.17	101.87	113.54	167.08	106.79	146.35	116.34	67.20
Federal SSI and State supplementation	136.45	162.89	174.82	261.65	180.38	244.31	214.93	191.49
State supplementation	63.42	93.94	80.40	149.06	66.14	118.90	93.52	65.98
State supplementation only	60.48	123.77	94.42	191.32	72.63	153.91	1/	63.31
	Converted from State programs							
Total	104.47	129.20	135.90	204.11	143.18	191.59	148.28	141.07
Federal SSI payments	86.03	107.87	107.32	150.78	114.78	149.13	118.19	118.36
Federal SSI payments only	90.97	114.29	114.23	167.89	112.99	157.51	116.63	114.79
Federal SSI and State supplementation	143.41	165.41	175.14	261.61	182.64	246.16	215.66	191.36
State supplementation	67.24	97.53	80.63	149.14	67.16	120.48	94.32	68.06
State supplementation only	66.90	145.07	94.99	191.91	77.57	161.54	1/	91.11
	Not converted from State programs							
Total	51.78	63.52	80.00	1/	79.65	101.78	120.69	75.60
Federal SSI payments	43.21	53.33	71.53	1/	71.10	94.81	107.14	65.91
Federal SSI payments only	43.23	53.69	70.63	1/	68.02	94.33	1/	55.29
Federal SSI and State supplementation	87.45	135.09	122.45	1/	134.01	185.95	1/	191.92
State supplementation	42.45	74.13	47.17	1/	49.21	86.24	1/	64.18
State supplementation only	39.76	68.58	47.02	1/	40.74	85.71	1/	38.81

1/ Averages based on 50 or fewer individuals or couples not shown.

TABLE 5.--Supplemental security income for the aged, blind, and disabled: Number and percentage distribution of persons receiving federally administered payments by reason for eligibility and living arrangements, May 1974

Living arrangement	Total	Adults			Children	
		Aged	Blind	Disabled	Blind	Disabled
Total number................	3,425,900	2,016,200	69,900	1,319,100	1,900	18,800
Total percent...............	100.0	100.0	100.0	100.0	100.0	100.0
Own household....................	87.3	89.1	88.9	85.1	78.1	47.0
Another's household..............	8.0	6.9	7.8	9.6	11.6	10.8
Institutional care covered by Medicaid (Title XIX)............	4.7	4.0	3.3	5.3	10.2	42.2

TABLE 6.--Supplemental security income for the aged, blind, and disabled: Number and percent of persons, adult units, and children receiving federally administered payments May 1974, in concurrent receipt of social security benefits; average monthly amount of social security benefit

Classification	Total number	With social security benefits		Average monthly social security benefit
		Number 1/	Percent of total	
All persons..................	3,425,900	1,740,400	50.8	$112.55
Adult units				
Aged individual..............	1,717,900	1,132,100	65.9	114.54
Aged couple..................	150,700	120,800	80.1	172.13
Blind individual.............	64,000	23,000	35.9	119.52
Blind couple.................	2,800	1,300	46.7	167.41
Disabled individual..........	1,263,200	323,400	25.6	120.03
Disabled couple..............	23,200	12,200	52.8	173.48
Children				
Blind........................	1,900	200	10.9	80.82
Disabled.....................	18,800	1,100	6.1	68.23

1/ Partly estimated.

TABLE 7.--Supplemental security income for the aged, blind, and disabled: Number and percent of employed persons eligible for Federal SSI payments previously receiving OAA, AB, and APTD in December 1973; and average monthly earnings

Reason for eligibility	Number employed 1/	Percent with employment	Average monthly earnings
Total.............	55,300	2.1	$ 54.74
Aged 65 and over........	32,000	2.0	37.61
Disabled under 65.......	20,300	1.9	59.79
Blind under 65..........	3,000	8.1	176.57

1/ Partly estimated.

TABLE 8 --Supplemental security income for the aged, blind, and disabled: Distribution of monthly earnings of employed persons eligible for Federal SSI payments previously receiving OAA, AB, and APTD, December 1973

Earnings	Distribution
Total number....................	55,300
Total percent..................	100.0
Less than $20........................	34.5
$21-39..........................	21.1
40-59..........................	15.3
60-79..........................	9.2
80-99..........................	5.7
100-149..........................	6.9
150-199..........................	3.0
200-249..........................	1.7
250-299..........................	1.2
300 and over.....................	1.3

TABLE 9.--Supplemental security income for the aged, blind, and disabled: Number and percent of persons eligible for Federal SSI payments previously receiving OAA, AB, and APTD in December 1973 with unearned income other than Social Security, and average monthly unearned income

Reason for eligibility	Unearned income
	Number with unearned income
Total 1/...............................	166,700
Age 65 and over..........................	113,100
Disabled under 65........................	52,300
Blind under 65..........................	1,200
	Percent with unearned income
Total..................................	5.6
Aged 65 and over........................	6.3
Disabled under 65.......................	4.3
Blind under 65..........................	1.7
	Average monthly unearned income
Total..................................	$52.04

1/ Partly estimated.

TABLE 10.--Supplemental security income for the aged, blind, and disabled: Number and percentage distribution of persons under State programs of OAA, AB, and APTD by race and sex, December 1973

Race and sex	Total	Aged 65 and over	Under 65	
			Blind	Disabled
Total number.........	3,135,700	1,876,500	49,200	1,210,100
	Race			
Race reported.....	2,574,300	1,551,300	41,500	981,500
Total percent........	100.0	100.0	100.0	100.0
White.............	69.6	71.3	64.9	67.2
Negro.............	30.4	28.7	35.1	32.8
	Sex			
Sex reported......	2,944,200	1,765,000	46,400	1,132,800
Total percent........	100.0	100.0	100.0	100.0
Men..............	36.0	30.3	49.6	44.3
Women............	64.0	69.7	50.4	55.7

1/ Data on races other than white or Negro not available.

TABLE 11.--Supplemental security income for the aged, blind, and disabled: Number of persons applying for federally administered payments January-June 1974

Year	Total	Aged	Blind or disabled
Total.............	1,314,260	543,903	770,357
January.........	307,740	148,796	158,944
February........	248,365	108,077	140,288
March...........	220,721	89,868	130,853
April...........	178,166	67,582	110,584
May.............	206,079	75,133	130,946
June............	153,189	54,447	98,742

TRENDS AND COMPOSITION OF THE LOW-INCOME POPULATION

Renée Miller and Arno I. Winard, U.S. Bureau of the Census

This paper presents a summary of the changes in the composition of the low-income population since 1959, the first year for which such data were available, and discusses the limitations of the data used to quantify poverty.

Since May 1968, when the report entitled "The Extent of Poverty in the United States: 1959 to 1966" was published, the Bureau of the Census has been issuing annual reports on the number and characteristics of the poor based on data collected from the March Supplement of the Current Population Survey (CPS). The definition of poverty used is based on an index developed about 10 years ago by Mollie Orshansky at the Social Security Administration in response to the need, generated by the "War on Poverty," to quantify and describe the poor. This index was revised by a Federal Interagency Committee in 1969, and a directive from the Bureau of the Budget made the Census Bureau series on poverty the official statistical standard. (1)

The low-income concept is a statistical yardstick based on the Department of Agriculture's 1961 economy food plan and reflects the different consumption requirements of families in relation to their size and composition, sex and age of the family head, and farm-nonfarm residence. It was determined from the Department of Agriculture's 1955 survey of food consumption that families of three or more persons spend approximately one-third of their income on food; the poverty level for these families was therefore set at three times the cost of the economy food plan. For smaller families and persons living alone, the cost of the economy food plan was multiplied by factors that were slightly higher in order to compensate for the relatively larger fixed expenses of these smaller households. In all, there are 124 different thresholds which are updated annually to reflect changes in the cost of living as measured by the Consumer Price Index (see table 7). These thresholds are applied annually to the data from the March Current Population Survey, thus providing a consistent series of data for the past decade and a half.

The limitation of applying one statistical standard based on food alone was recognized from the beginning. In a January 1965 article entitled "Counting the Poor: Another Look at the Poverty Profile," Mollie Orshansky stated that there would probably never be one definition of adequacy that is accepted by all persons and that could be applied across the board to all persons without the danger of misclassification. She continued to explain the rationale for the poverty definition by writing, "... and if it is not possible to state unequivocally 'how much is enough,' it should be possible to assert with confidence how much, on an average, is too little." (2) The poverty threshold in 1973 for a nonfarm family of four was $4,540; it was $2,973 in 1959. More detailed thresholds for selected years are shown in table 1.

The poverty index is presently used as an absolute rather than a relative measure and some persons criticize the use of such a definition to determine changes in poverty. For example, the relationship between the poverty threshold for a nonfarm family of four and the median family income has changed considerably since 1959. In that year, the poverty threshold was over one-half (55 percent) of the overall median income; by 1973, it was about 38 percent of the median income. The ratio of the poverty threshold to the median income for each of the years is shown in table 2. There are other limitations of the definition that will be discussed in later sections such as the lack of adjustment for geographic variations in the cost of living, the fact that the poverty definition is based only on the cost of the food budget, and that it only includes money income received during the previous year and does not take account of assets.

We are left with the question of what a statistical yardstick that is admittedly imperfect at measuring a socioeconomic concept such as poverty can show. Increasing the poverty thresholds would no doubt increase the total number of poor persons but would not change the fact that certain groups are more likely to be poor than others. Families headed by women, especially those that are black, are so much more likely to be poor than families headed by men, especially those that are white, that raising the thresholds would not alter this relationship. The following table shows the proportion of families below the poverty level, below 125 percent of the poverty level, and below one-half of the median family income cross-classified by sex and race of head.

Table A. Percent of Families Below Selected Low-Income Levels in 1973

	Percent below the poverty level	Percent below 125% of the poverty level	Percent below one-half of median income
FAMILIES WITH MALE HEAD			
White	4.6	7.7	13.8
Black	15.4	22.7	27.8
FAMILIES WITH FEMALE HEAD			
White	24.6	31.7	45.7
Black	52.7	63.3	68.5

Source: Current Population Survey, Bureau of the Census.

Perhaps we do not really know what the "true" number of poor is, since we do not know the answer to the question of what poverty is. (Perhaps one is poor if one considers oneself to be so.) The trend line showing the number of poor persons from 1959 to 1973 is meaningful as a measure of well-being even if the absolute level is actually higher or lower. It should be stressed at this time that the poverty statistics discussed in this paper are based on a sample survey and are, therefore, subject to sampling variability. All statements of comparison using CPS data are made at the 95-percent confidence level, and estimates of standard errors are included in text table B and table 3. It should also be noted that, as in all field surveys of income, these figures are subject to errors of response and nonreporting.

Changes in the Low-Income Population: 1959 to 1973

The fact that the number of persons below the poverty level declined considerably since 1959 (when there were 39.5 million poor persons) has been pointed out many times in our reports; however, the decline has neither been steady nor universal for all groups. During the period from 1959 to 1964, there were several years in which the number of poor did not change significantly; overall, the number of poor declined by only 3 million during this 5-year period. (3) A sharp decrease in poverty was observed during the 1964-69 period when the number of poor declined by 12 million. In recent years, the number of poor has fluctuated, so that by 1973 the level was only about 1 million below that for 1969, providing considerable contrast to the preceding period. In 1973, there were about 23 million persons below the low-income level, comprising 11 percent of the total U.S. population. As figure 1 demonstrates, the trend line for the number of persons below 125 percent of the low-income level is similar to the poverty trend line. In 1973, there were 33 million persons below 125 percent of the low-income level, which was $5,675 for a nonfarm family of four.

Unfortunately, the Current Population Survey was not designed to measure the impact of Federal programs that aid the poor. It is, therefore, not possible to determine the role that these

Table B. Number of Persons Below the Low-Income Level and Below 125 Percent of the Low-Income Level: 1959 to 1973

(Numbers underlying figure 1)

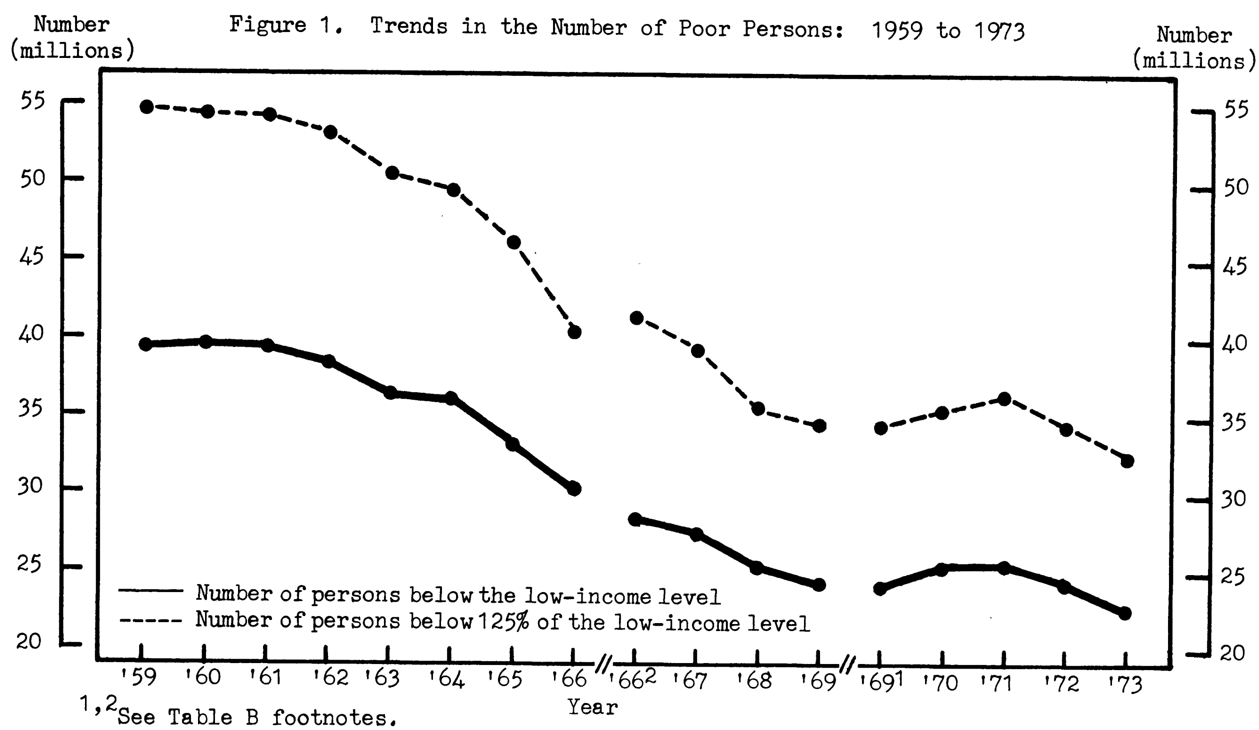
Year	Persons below low-income level (thous.)	Standard error (thous.)	Persons below 125% of low-income level (thous.)	Standard error (thous.)
1973	22,973	345	32,828	395
1972	24,460	353	34,653	403
1971	25,559	360	36,501	410
1970	25,420	359	35,624	406
1969[1]	24,147	352	34,665	403
1969	24,289	352	34,875	404
1968	25,389	359	35,905	408
1967	27,769	371	39,206	419
1966[2]	28,510	375	41,267	426
1966	30,424	384	40,617	424
1965	33,185	556	46,163	616
1964	36,055	571	49,819	629
1963	36,436	574	50,778	631
1962	38,625	584	53,119	637
1961	39,628	589	54,280	640
1960	39,851	591	54,560	641
1959	39,490	588	54,942	643

[1]Beginning with the March 1970 CPS, data based on 1970 census population controls; therefore, not strictly comparable with data for earlier years.

[2]Beginning with the March 1967 CPS, data based on revised methodology for processing income data.

Source: Current Population Survey, Bureau of the Census.

Figure 1. Trends in the Number of Poor Persons: 1959 to 1973



[1,2]See Table B footnotes.

programs played in reducing the number of poor between 1964 and 1969 as compared with the role of general economic growth during this period. There is, of course, the additional problem of the two factors being interrelated and the difficulty of separating the effect of each.

The Aged Poor

Substantial increases in Social Security benefits have been enacted since 1970, and it is possible to observe their impact on the aged poor since nearly all persons 65 years old and over receive such benefits. The number of low-income aged persons has declined steadily from 4.7 million in 1970 to 3.4 million in 1973, even though the total number of aged persons has increased during that period, thereby reducing the proportion of the aged population that was poor from 25 percent in 1970 to 16 percent in 1973 (see table 3). However, aged persons still have a higher poverty rate than those under 65 years. In addition, while the number of poor persons 65 years and over has been declining steadily since 1970, the number of aged persons between 100 and 125 percent of the poverty level has increased during the last three years, suggesting a slight but not very substantial upgrading of income levels.

Overall, aged persons comprise a larger proportion of the white population than of the black population; therefore, increases in Social Security benefits would tend to affect whites more than blacks. Moreover, the original Social Security Act did not cover agricultural or domestic workers, and many aged blacks were formerly employed in these job categories. Although the Social Security Act was later amended to cover these workers, aged blacks have not had as much time as whites to build up equity in the Social Security fund. (4) It is, therefore, not surprising that the change in the number of low-income aged black persons between 1970 and 1973 did not parallel the sharp decrease observed for low-income aged whites and was not large enough to be statistically significant.

Families Headed by Females

Families with a female head constitute another group that has a high risk of being poor. The number of such families with incomes below the low-income level has actually increased since 1959, and the proportion of the low-income population that they represent has about doubled (see table 4). By 1973, families headed by women comprised about two-thirds of all low-income black families. The first poverty report published by the Census Bureau summarized the situation of black families as follows: "The impression is widespread that nonwhite families are poor because of the absence of a male breadwinner and the presence of too many children. The statistics lend some support to these impressions, but the more important finding is that nonwhite families are far more likely than white families to be in poverty whatever their composition. Within each sex of head and size of family group, the incidence of poverty among nonwhite families far exceeds that of white families." (5) Although this statement was made in 1968 based on data for the year 1966, table C illustrates similar findings for the year 1973.

Table C. Percent of Families Below the Low-Income Level in 1973 by Size of Family and Sex and Race of Head

	Male head		Female head	
Size of family	White	Black	White	Black
Total	4.6	15.4	24.5	52.7
2 and 3 persons	4.4	12.6	20.5	42.5
4 and 5 persons	3.8	12.5	34.0	59.0
6 persons or more	8.1	27.0	51.3	71.1

Source: Current Population Survey, Bureau of the Census.

The fact that families headed by women, regardless of race, were more likely to be poor in 1973 than those headed by men is also illustrated in table C. In other words, at this time, the likelihood of a family being poor is lowest if the family is headed by a white male (5 percent of such families were poor in 1973); families headed by black males had the next lowest chance of being poor, followed by families headed by white females, and finally by families headed by black females (the likelihood of these families being poor was about 53 percent in 1973). This relationship, however, has not always been true for other years. From 1959 until 1967, the poverty rate for male-headed families of Negro and other races was about equal to or greater than that for white families headed by females. By 1968, the poverty rate for these male-headed families had declined substantially and was lower than that for white female-headed families (see table 5). The decline in the poverty rate for black male-headed families was related in part to the increase between 1959 and 1968 in the proportion of wives who worked year-round full time. Although progress has been made in reducing the poverty rate for black male-headed families, it should be noted that the poverty rate for such a family in 1973 was about the same as that for a white male-headed family in 1959.

The divorce rate continued to increase during the mid-1960's, contributing to the growth of female-headed families in the total population, but this provides only a partial explanation of the increase of poor families with a female head. Another explanation is suggested by a recent study on public welfare in 100 local areas prepared for the Joint Economic Committee of Congress. While this study does not show how welfare actually influences changes in family composition, it does indicate what a hypothetical family can gain by splitting up. As the study points out, there is "little doubt that welfare does establish large incentives for low-income families to break up, or to never form in the first place." (6)

The effect of changes in family status on poverty status can be better seen when using longitudinal rather than cross-sectional data. The CPS was not designed to provide longitudinal data, but it is possible to make comparisons for about one-half of the sample because of the sample rotation scheme. A household is interviewed for 4 months in a row, then dropped from the survey for 8 months, and then interviewed for another 4 months before being permanently dropped; thus, approximately 50 percent of the households are common over a two-year period. A family was considered matched if at least one member of the household was the same as in the previous year; the matched member did not have to be the head. If the entire household moved, it was not included in the matched file.

A very limited study using the matched file of families from the March 1971 and 1972 Current Population Surveys shows that 4 out of every 10 female heads of families who became poor in the income year 1971 had been "married, spouse present" in 1970. These families had been classified as nonpoor male-headed families in 1970. The remaining 60 percent of the female-headed families that became poor in 1971 had been classified as nonpoor female-headed families during the previous year. In addition to the fact that families headed by women were more likely to fall below the poverty level than those headed by men, they were also more likely to remain poor. About three out of four families headed by women that were poor in 1970, were also poor in 1971, compared to three out of five families headed by men.

Historically, female heads of families have been less likely to work year-round full time than male heads, mainly because of their family responsibilities. The number of men who have not worked because of family responsibilities has been insignificant. Obtaining work has been an important means of moving out of poverty. Since 1959, the number of low-income families headed by workers has been decreasing faster than the number of such families headed by nonworkers (see table 6). While it appears that families headed by women remain poor because the head does not work, it should be noted that employment does not always provide sufficient income to bring a family over the low-income thresholds. The poverty rate in 1973 for black mothers who were heads of families or subfamilies and worked year-round full time was still higher than the base poverty rate for all families.

Black female heads of families are not the only persons who work and remain within the low-income classification, however. About half of all low-income persons 14 to 64 years old worked at some time during 1973, and about one-fourth of these workers were employed year-round full time. (It should be noted here that the classification of persons by poverty status is based on the total money income of the family of which they are a member rather than on the person's own income. For unrelated individuals, poverty status is determined by their own total income.) The basic minimum hourly wage was $1.60 in 1973, but it would have required an hourly wage of $2.27 to bring a nonfarm family of four with one year-round full time worker above the poverty line in that year. The minimum wage was increased to $2.00 an hour as of May 1974, and it will rise to $2.10 an hour in 1975 and to $2.30 an hour in 1976. (7)

For low-income female-headed families, public assistance is an important source of income. The majority of low-income black female-headed families reported receiving public assistance in 1972, and for one-half of those receiving it, public assistance was their only source of income. (8) Except for two States (Alaska and Hawaii), these payments alone were not enough to bring a family of four above the poverty line.

Students

Another change observed in the composition of the low-income population since 1959 includes the increase in the proportion of poor persons who are unrelated individuals (persons who live alone or with nonrelatives). Unrelated individuals have increased as a proportion of the total population as well, especially in the 14 to 24 and 25 to 34 year age groups. The majority of unrelated individuals 14 to 24 years of age are college students who are living in their own households or with nonrelatives in a household.

The economic status of students with incomes below the poverty level cannot be compared with that of other low-income persons because students are voluntary nonworkers; that is, they opt to continue their education rather than seek employment. They are most likely living on savings or occasional allowances from their families which are not counted as income in the CPS. Students can be considered "transient poor" in the sense that they are making an investment and are likely to have lifetime earnings well above average once they graduate from college.

Limitations of Money Income as a Measure of Economic Status

This raises the question concerning the adequacy of money income received during the previous year as a measure of economic well-being. Data on consumer income collected in the CPS cover money income received before payments for personal income taxes and deductions for Social Security, union dues, Medicare, etc. and is the sum of the following sources: money wages or salary, net income from nonfarm self-employment, net income from farm self-employment, Social Security, public assistance, and other unearned income such as dividends, interest, rent, private pensions, alimony, etc. An analysis of families and unrelated individuals reporting annual incomes under $500 in 1971 suggests that, for self-employed persons having low family incomes, cash income alone may not be the most appropriate measure of classification by poverty status. In view of the fact that $500 is approximately the same as the cost of the economy food plan for the group with the lowest poverty cutoffs (female unrelated individuals), it appears unrealistic to assume that a family could live on such income alone. Families in which a member reported a loss which brought the total family income below $500 are a very small proportion of the total population, but the question is raised whether total money income is an adequate measure of economic well-being in such cases.

Nonmoney transfer payments and money receipts from capital gains are not included in the income concept. Some analysts are concerned that the number of poor is overstated because such items as food stamps and subsidized housing are not taken into consideration. It should be noted that in-kind payments are received by families throughout the income distribution. For example, farm families often receive nonmoney income in the form of rent-free housing and goods produced and consumed on the farm. Nonmoney transfers received by nonfarm residents often take the form of business expense accounts and full or partial payments by business for retirement programs, medical and educational expenses, etc. Nevertheless, these transfer payments are more likely to be received by the nonpoor than the poor. Although adding nonmoney transfer payments to the income of the nonpoor would not change the number of poor by the present definition of poverty, as Mollie Orshansky suggested, if it were possible to translate in-kind payments into money income for all families and if money received from sources such as sale of stocks, bonds, and inheritances were included in the income concept, the upper end of the income distribution might benefit more than the lower end. As she stated in her testimony before the House Committee on Education and Labor, "The full effect of properly incorporating these (nonmoney transfers) into the income distribution might be to skew it even more than now with a resultant upping of the poverty line." (9)

Data from the June 1973 CPS indicate that only about one-half (53 percent) of all poor families received food stamps. (10) Furthermore, only a small proportion of the poor live in subsidized housing units. There have been discussions of imputing specific monetary values for both subsidized housing and food stamps to all poor, but in view of the above facts, the result would be quite inaccurate and misleading.

Another limitation of using income as a measure of economic status is the tendency for respondents in household surveys to underreport their income. Underreporting is especially pronounced for income sources that are not derived from earnings, such as Social Security, public assistance, and net income from interest, dividends, rentals, etc. Overall, income earned from wages or salary is much better reported than other sources of income and is nearly equal to independently derived benchmark estimates. By contrast, recent CPS income data on Social Security and public assistance payments to beneficiaries have averaged approximately 84 and 73 percent, respectively, of their benchmark estimates.

The original intent of the poverty definition was never to classify specific individual families below or above the poverty line. The definition was not developed for administrative use in any specific program; it was to be a statistical yardstick. The problem arises when this concept, in the absence of any other measure, is used to determine eligibility in programs to aid the poor. In April 1973, the Interagency Committee on Income Distribution and Poverty was requested by the Statistical Policy Division of the Office of Management and Budget to conduct a thorough review of the statistics on income and poverty. Three subcommittees, one focusing on updating the poverty thresholds, another on measuring cash income, and the third on measuring noncash income, met throughout the summer of 1973 and came up with a number of recommendations. (11) One of the conclusions reached by the Subcommittee on Updating the Poverty Threshold was that the development of separate measures for eligibility in programs should receive high priority.

It should also be noted that the income concept is limited in scope by the available data. The Current Population Survey, which is the only source of detailed income data available on a regular basis, does not provide data on assets; therefore, including them in the income concept would not be feasible. With the exception of food stamps, questions on in-kind benefits are not presently asked and even data on food stamps are not collected in the CPS at the same time as information on income and other socioeconomic characteristics of the population is gathered. Proposals have been advanced to obtain information in the CPS on the number of persons living in publicly owned or subsidized rented housing units. Such data, together with the results of the studies now underway at the Census Bureau on the characteristics of food stamp recipients, should shed further light on the impact of nonmoney transfers on the low-income population.

Limitations of the Poverty Concept

The poverty definition itself is based on food, in part because an accepted standard is available of what a minimal, but adequate, diet should comprise--the economy food plan. There are no comparable standards for housing, transportation, or medical care. The Subcommittee on Updating the Poverty Threshold recommended that, ideally, the poverty thresholds should be based on a standard which takes all of the above factors into account. An expenditure survey for families of varying income levels would be necessary to determine what proportion of income is spent on these items.

The poverty thresholds are now updated each year to reflect changes in the cost of living as reflected in the Consumer Price Index (CPI). They were originally updated based on changes in the average per capita cost of the food in the economy food plan, but during the 1960's the cost of the economy food plan did not increase as much as the overall cost of living (see table D). Therefore, changes in the poverty thresholds did not parallel changes in the overall cost of living. As a result of deliberations of a Federal Interagency Committee in 1969, it was decided that annual adjustments in the low-income thresholds would be based on changes in the CPI rather than on changes in the cost of food included in the economy food plan. The year 1963 was retained as the base year (the nonfarm thresholds for this year remained the same). Since the change in the CPI between 1959 and 1963 was greater than the change in the original thresholds based on the cost of the economy food plan, the 1959 thresholds were lowered, resulting in a reduction of 260,000 poor nonfarm families in 1959 as compared with the number originally published. For the years following 1963, the thresholds were increased to reflect the changes in the CPI, so that the number of poor nonfarm families in 1967 was 210,000 more than the number originally published. (12) Both changes cited were less than one standard error and are therefore not statistically significant.

As the figures in table D indicate, changes in the food component of the CPI in most years did not differ much from the CPI for all items. However, between 1972 and 1973 the cost of food, for the first time, increased considerably more than the total CPI (14.5 percent compared with 6.2 percent). Should this difference continue, the overall CPI may not be entirely reflective of the cost of living for the poor since the poor spend a higher proportion of their income on food than the total population, and therefore it may be desirable to revise the method of updating the thresholds annually to account for trends in food and nonfood components of the CPI separately. Furthermore, it may be preferable to establish a special cost of living index for the poor. This index could be based on the changes in the cost of such items as food, housing, medical care, and transportation, weighted by the proportion of income a poor family has to spend for them. This, too, would necessitate the undertaking of special consumer expenditure surveys.

Table D. Average Annual Percent Change in the Cost of the Economy Food Plan and in the Consumer Price Index (All Items and Food): 1959 to 1973

Year	Economy food plan	Consumer Price Index	
		All items	Food
1959-60	2.2	1.6	1.0
1960-61	-2.2	1.0	1.3
1961-62	0.0	1.1	0.9
1962-63	2.2	1.2	1.4
1963-64	0.0	1.3	1.3
1964-65	2.2	1.7	2.2
1965-66	4.2	2.9	5.0
1966-67	0.0	2.9	0.9
1967-68	(NA)	4.2	3.6
1968-69	(NA)	5.4	5.1
1969-70	(NA)	5.9	5.5
1970-71	(NA)	4.3	3.0
1971-72	(NA)	3.3	4.3
1972-73	(NA)	6.2	14.5

NA Comparable data not available.

Source: Economic Report of the President, February 1974, and U.S. Department of Agriculture.

Another question that is often raised concerning the measurement of poverty is whether such a measure should be absolute or relative. The Poverty Subcommittee favored retaining an absolute measure of poverty rather than changing to a relative measure. Supporters of a poverty definition that is tied to the median income claim that by using the present absolute measure, we will eventually eliminate poverty by definition. They are referring to the fact that as median family income increases after adjustment for increases in prices, the poverty threshold moves farther away from it. (This relationship is illustrated in table 2.) However, median family income in constant dollars does not necessarily increase each year. It remained the same between 1969 and 1970 and again between 1970 and 1971. During those periods the ratio of the poverty threshold to the median family income remained about the same. Those who prefer a relative measure assume that real median income will increase over time, and therefore such a measure would result in a continuously higher minimum standard of living for the poor.

The Poverty Subcommittee recognized that nutritional standards and consumption patterns change over time and therefore recommended that the relationships contained in the poverty series be updated every 10 years, while revising data for the previous 10 years so that comparisons over time could be made. As stated earlier, the poverty thresholds were based originally on the economy food plan, and the assumption was that, on the average, families of three or more persons spend one-third of their incomes on food. The poverty thresholds could be revised using the same base, if, for example, it were determined that the low-cost food plan (which is 25 percent higher than the economy food plan) was a better measure of a minimum standard of nutrition and that families of three or more persons spend one-fourth of their incomes on food.

Because of the limited data available which are suitable for making place-to-place comparisons of costs of living, the subcommittee decided that the introduction of regional poverty indexes or indexes based on size of metropolitan area was not practical at this time. In addition, Bureau of Labor Statistics studies indicate that geographical variations in the cost of living at the poverty level would not be significant. However, as the Poverty Subcommittee points out, variations in the availability of services such as transportation and medical care would make a difference in a family's ability to manage on a small income.

Conclusion

Earlier in this paper, the point was made that although the definition of poverty has certain limitations, the definition is still useful as a measure of well-being since it reflects changes in the economy as measured by independent sources along with changes resulting from the implementation of programs to help the poor. Social Security benefits have been increasing since 1970, and the data reflect this change for the aged poor. The mid 1960's are generally accepted as a time of economic growth by any standard, and the decrease of 12 million poor persons between 1964 and 1969 mirrors this. It is particularly important that the data continue to reflect the general state of the economy, because these data are, in turn, used as measures of our general economic well-being.

Since the publication of our first poverty report, the Census Bureau has increased significantly the number of cross-tabulations it publishes. The intent of the poverty definition was to quantify and describe the poor, and in our publications we have done that. Each year we state how many poor there are and who they are. Having identified the poor with a concept developed in 1964, it now appears that it is time to improve this definition and to try to understand just why a family is poor and remains so. To measure chronic poverty would require longitudinal studies which would follow the same individuals over a number of years. As James Morgan points out in his summary analysis of a five-year longitudinal study, "Observations of family well-being over time also permit a more satisfactory definition of low-income or poverty population. If all families with a low-income in a single year could be observed over several years, it would be found that some are only transitory members of a poverty population while the remainder are its permanent members." (13) Another method of expanding the data base would be to link administrative records of other agencies with households in the CPS sample. Finally, it would be desirable to supplement the information now collected in the March CPS with questions that would directly measure the impact of government programs on the poor.

Table 1. Nonfarm Weighted Average Thresholds at the Low-Income Level by Size of Family: 1959 and 1967 through 1973

Size of family unit	1973	1972	1971	1970	1969	1968	1967	1959
1 person	$2,247	$2,109	$2,040	$1,954	$1,840	$1,748	$1,675	$1,467
Under 65 years	2,307	2,168	2,098	2,010	1,893	1,797	1,722	1,503
65 years and over	2,130	2,005	1,940	1,861	1,757	1,667	1,600	1,397
2 persons	2,895	2,724	2,633	2,525	2,383	2,262	2,168	1,894
Head under 65 years	2,984	2,808	2,716	2,604	2,458	2,333	2,238	1,952
Head 65 years and over	2,688	2,530	2,448	2,348	2,215	2,102	2,017	1,761
3 persons	3,548	3,339	3,229	3,099	2,924	2,774	2,661	2,324
4 persons	4,540	4,275	4,137	3,968	3,743	3,553	3,410	2,973
5 persons	5,358	5,044	4,880	4,680	4,415	4,188	4,019	3,506
6 persons	6,028	5,673	5,489	5,260	4,958	4,706	4,516	3,944
7 or more persons	7,435	6,983	6,751	6,468	6,101	5,789	5,550	4,849

NOTE: Farm thresholds may be obtained by multiplying the nonfarm thresholds by 0.85.

Source: Current Population Survey, Bureau of the Census.

Table 2. Poverty Threshold for a Nonfarm Family of Four as a Percent of Median Family Income: 1959 to 1973

Year	Median income (current dollars)	Poverty threshold for a nonfarm family of four	Poverty threshold as a percent of median income
1973	$12,051	$4,540	37.7
1972	11,116	4,275	38.5
1971	10,285	4,137	40.2
1970	9,867	3,968	40.2
1969	9,433	3,743	39.7
1968	8,632	3,553	41.2
1967	7,933	3,410	43.0
1966	7,500	3,317	44.2
1965	6,957	3,223	46.3
1964	6,569	3,169	48.2
1963	6,249	3,128	50.1
1962	5,956	3,089	51.9
1961	5,737	3,054	53.2
1960	5,620	3,022	53.8
1959	5,417	2,973	54.9

Source: Current Population Survey, Bureau of the Census.

Table 3. Persons 65 Years Old and Over by Low-Income Status: 1959 and 1966 to 1973
(Numbers in thousands. Persons as of March of the following year)

Year	Total	Below low-income level	Standard error	Percent below low-income level	Between 100 and 125 percent of low-income level	Standard error
All races						
1973	20,602	3,354	142	16.3	2,168	114
1972	20,117	3,738	149	18.6	1,992	110
1971	19,827	4,273	160	21.6	2,001	110
1970	19,254	4,709	167	24.5	1,820	105
1969	18,899	4,787	169	25.3	1,860	106
1968	18,549	4,632	166	25.0	(NA)	(NA)
1967	18,245	5,388	178	29.5	(NA)	(NA)
1966	17,931	5,114	173	28.5	(NA)	(NA)
1959 1/	15,557	5,481	108	35.2	(NA)	(NA)
White						
1973	18,754	2,698	127	14.4	1,865	106
1972	18,340	3,072	136	16.8	1,766	103
1971	18,087	3,605	147	19.9	1,745	103
1970	17,684	3,984	154	22.5	1,642	100
1969	17,370	4,052	155	23.3	1,687	101
1968	17,062	3,939	153	23.1	(NA)	(NA)
1967	16,791	4,646	166	27.7	(NA)	(NA)
1966	16,514	4,357	161	26.4	(NA)	(NA)
1959 1/	14,344	4,744	101	33.1	(NA)	(NA)
Negro						
1973	1,672	620	57	37.1	279	39
1972	1,603	640	58	39.9	201	33
1971	1,584	623	58	39.3	252	37
1970	1,422	683	60	48.0	171	31
1969	1,373	689	60	50.2	154	29
1968	1,374	655	59	47.7	(NA)	(NA)
1967	1,341	715	62	53.3	(NA)	(NA)
1966	1,311	720	62	54.9	(NA)	(NA)
1959 1/	1,138	711	37	62.5	(NA)	(NA)

NA Not available.

1/ Data for 1959 are based on a 1-in-1,000 sample of the 1960 census because such figures are not available from the March 1960 Current Population Survey.

Source: Current Population Survey, Bureau of the Census.

Table 4. Families With a Female Head as a Percent of All Families by Low-Income Status and Race of Head: 1959 to 1973

(Numbers in thousands. Families as of March of the following year)

Year	All income levels				Below low-income level			
	All races	White	Negro and other races		All races	White	Negro and other races	
			Total	Negro			Total	Negro
FAMILIES WITH FEMALE HEAD								
1973	6,804	4,853	1,951	1,849	2,193	1,190	1,002	974
1972	6,607	4,762	1,935	1,822	2,158	1,135	1,023	972
1971	6,191	4,489	1,702	1,642	2,100	1,191	908	879
1970	6,002	4,408	1,594	1,535	1,951	1,102	851	834
1969 1/	5,581	4,186	1,395	1,349	1,827	1,069	758	737
1968	5,441	4,055	1,387	1,327	1,755	1,021	734	706
1967	5,333	4,008	1,324	1,272	1,774	1,037	737	716
1966 2/	5,203	4,034	1,170	1,138	1,721	1,036	685	674
1965	4,992	3,860	1,132	(NA)	1,916	1,196	720	(NA)
1964	5,006	3,881	1,126	(NA)	1,822	1,125	697	(NA)
1963	4,882	3,797	1,085	(NA)	1,972	1,191	781	(NA)
1962	4,741	3,626	1,115	(NA)	2,034	1,230	804	(NA)
1961	4,642	3,606	1,036	(NA)	1,954	1,208	746	(NA)
1960	4,611	3,676	936	(NA)	1,955	1,252	703	(NA)
1959	4,493	3,544	949	(NA)	1,916	1,233	683	(NA)
FAMILIES WITH FEMALE HEAD AS A PERCENT OF ALL FAMILIES								
1973	12.4	9.9	31.8	34.0	45.4	37.0	62.3	63.8
1972	12.2	9.6	32.8	34.6	42.5	33.0	62.6	63.6
1971	11.6	9.4	30.1	31.8	39.6	31.8	58.5	59.2
1970	11.5	9.5	28.9	30.5	37.1	29.7	54.8	56.3
1969 1/	10.9	9.1	26.7	28.3	36.5	29.9	52.9	54.0
1968	10.8	8.9	27.3	28.6	34.8	28.2	51.3	51.7
1967	10.7	8.9	26.4	27.7	31.3	25.6	45.7	46.0
1966 2/	10.6	9.1	23.6	25.0	29.8	25.2	40.8	41.6
1965	10.3	8.9	23.7	(NA)	28.5	24.8	38.0	(NA)
1964	10.5	9.0	23.7	(NA)	25.4	21.4	36.6	(NA)
1963	10.3	8.9	22.7	(NA)	26.1	21.8	37.4	(NA)
1962	10.1	8.5	24.4	(NA)	25.2	20.9	36.7	(NA)
1961	10.0	8.6	23.2	(NA)	23.3	19.5	34.1	(NA)
1960	10.1	8.9	21.6	(NA)	23.7	20.5	33.0	(NA)
1959	10.0	8.7	22.4	(NA)	23.0	19.9	32.0	(NA)

NA Not available.

1/ Beginning with the March 1970 CPS, data based on 1970 census population controls; therefore, not strictly comparable with data for earlier years.

2/ Beginning with the March 1967 CPS, data based on revised methodology for processing income data.

Source: Current Population Survey, Bureau of the Census.

Table 5. Percent of Families Below the Low-Income Level by Sex and Race of Head: 1959 to 1973

Year	Families with male head			Families with female head		
	White	Negro and other races		White	Negro and other races	
		Total	Negro		Total	Negro
1973	4.6	14.5	15.4	24.5	51.4	52.7
1972	5.3	15.4	16.2	24.3	52.9	53.3
1971	5.9	16.3	17.2	26.5	53.3	53.5
1970	6.2	17.9	18.6	25.0	53.4	54.3
1969 1/	6.0	17.3	17.9	25.7	53.1	53.3
1968	6.3	18.9	19.9	25.2	52.9	53.2
1967	7.4	23.7	25.3	25.9	55.7	56.3
1966 2/	7.7	26.2	27.6	25.7	58.5	59.2
1965	9.2	32.3	(NA)	31.0	63.6	(NA)
1964	10.5	33.2	(NA)	29.0	61.9	(NA)
1963	11.0	35.5	(NA)	31.4	72.0	(NA)
1962	12.0	40.2	(NA)	33.9	72.1	(NA)
1961	13.1	42.1	(NA)	33.5	72.0	(NA)
1960	13.0	41.9	(NA)	34.0	75.1	(NA)
1959	13.3	44.2	(NA)	34.8	72.0	(NA)

NA Not available.

1/ Beginning with the March 1970 CPS, data based on 1970 census population controls; therefore, not strictly comparable with data for earlier years.

2/ Beginning with the March 1967 CPS, data based on revised methodology for processing income data.

Source: Current Population Survey, Bureau of the Census.

Table 6. Work Experience of Family Heads Below the Low-Income Level by Sex: 1973 and 1959

Work experience of head	All families		Male head		Female head	
	1973	1959	1973	1959	1973	1959
Total (thousands)	4,828	8,320	2,635	6,404	2,193	1,916
Percent	100.0	100.0	100.0	100.0	100.0	100.0
Worked	51.4	67.5	61.8	74.9	38.9	42.9
50 to 52 weeks	21.7	36.6	31.1	42.8	10.5	15.8
Full time	18.2	31.5	27.4	37.6	7.2	10.9
1 to 49 weeks	29.7	31.0	30.7	32.1	28.4	27.1
Did not work	48.3	30.5	37.6	22.5	61.1	57.1
Reason for not working:						
Ill or disabled	15.5	9.5	19.8	10.8	10.4	5.4
Keeping house	20.7	10.9	-	-	45.7	47.5
Going to school	1.6	0.1	1.6	0.1	1.7	0.4
Unable to find work	1.8	1.2	1.6	1.0	2.0	1.5
Other	8.6	8.7	14.6	10.6	1.3	2.3
In Armed Forces	0.3	1.9	0.6	2.5	-	-

- Represents zero.

Source: Current Population Survey, Bureau of the Census.

Table 7. Income Thresholds at the Poverty Level in 1973 by Sex of Head, Size of Family, and Number of Related Children Under 18 Years Old, by Farm-Nonfarm Residence

Size of family unit	Number of related children under 18 years old						
	None	1	2	3	4	5	6 or more
NONFARM							
Male Head							
1 person (unrelated indiv.):							
Under 65 years	$2,396	--	--	--	--	--	--
65 years and over	2,153	--	--	--	--	--	--
2 persons:							
Head under 65 years	2,996	$3,356	--	--	--	--	--
Head 65 years and over	2,690	3,356	--	--	--	--	--
3 persons	3,488	3,601	$3,806	--	--	--	--
4 persons	4,598	4,666	4,505	$4,733	--	--	--
5 persons	5,549	5,616	5,436	5,299	$5,413	--	--
6 persons	6,365	6,386	6,251	6,115	5,934	$6,025	--
7 or more persons	8,016	8,085	7,926	7,790	7,610	7,337	$7,270
Female Head							
1 person (unrelated indiv.):							
Under 65 years	$2,217	--	--	--	--	--	--
65 years and over	2,125	--	--	--	--	--	--
2 persons:							
Head under 65 years	2,768	$3,022	--	--	--	--	--
Head 65 years and over	2,656	3,022	--	--	--	--	--
3 persons	3,375	3,215	$3,556	--	--	--	--
4 persons	4,415	4,574	4,553	$4,505	--	--	--
5 persons	5,299	5,459	5,436	5,391	$5,209	--	--
6 persons	6,183	6,296	6,251	6,205	6,002	$5,819	--
7 or more persons	7,767	7,881	7,858	7,790	7,587	7,429	$7,066
FARM							
Male Head							
1 person (unrelated indiv.):							
Under 65 years	$2,036	--	--	--	--	--	--
65 years and over	1,830	--	--	--	--	--	--
2 persons:							
Head under 65 years	2,546	$2,852	--	--	--	--	--
Head 65 years and over	2,286	2,852	--	--	--	--	--
3 persons	2,965	3,061	$3,235	--	--	--	--
4 persons	3,909	3,967	3,829	$4,023	--	--	--
5 persons	4,717	4,774	4,620	4,504	$4,601	--	--
6 persons	5,410	5,428	5,314	5,198	5,044	$5,121	--
7 or more persons	6,815	6,873	6,738	6,622	6,469	6,237	$6,180
Female Head							
1 person (unrelated indiv.):							
Under 65 years	$1,884	--	--	--	--	--	--
65 years and over	1,806	--	--	--	--	--	--
2 persons:							
Head under 65 years	2,353	$2,569	--	--	--	--	--
Head 65 years and over	2,258	2,569	--	--	--	--	--
3 persons	2,868	2,733	$3,023	--	--	--	--
4 persons	3,754	3,887	3,870	$3,829	--	--	--
5 persons	4,504	4,640	4,620	4,582	$4,428	--	--
6 persons	5,256	5,352	5,314	5,275	5,102	$4,947	--
7 or more persons	6,601	6,700	6,680	6,622	6,449	6,314	$6,006

-- Not applicable.
Source: Current Population Survey, Bureau of the Census.

REFERENCES

(1) U.S. Bureau of the Budget, Office of Statistical Policy, "Definition of Poverty for Statistical Purposes," Statistical Reporter, No. 70-3, September 1969.

(2) Orshansky, Mollie. "Counting the Poor: Another Look at the Poverty Profile," Social Security Bulletin, January 1965, p. 3.

(3) U.S. Department of Commerce, Bureau of the Census, Current Population Reports, Series P-60, No. 94, "Characteristics of the Low-Income Population: 1973" (Advance report).

(4) Lindsay, Inabel B. "The Multiple Hazards of Age and Race: The Situation of Aged Blacks in the United States," a Preliminary Survey for the Special Committee on Aging, United States Senate, September 1971, p. 2.

(5) U.S. Department of Commerce, Bureau of the Census, Current Population Reports, Series P-60, No. 54, "The Extent of Poverty in the United States: 1959 to 1966," p. 6.

(6) U.S. Congress, Joint Economic Committee, Subcommittee on Fiscal Policy, "Welfare in the '70's: A National Study of Benefits Available in 100 Local Areas," Studies in Public Welfare, Paper No. 15, July 1974, pp. 6-7.

(7) Elder, Peyton. "The 1974 Amendments to the Federal Minimum Wage Law," Monthly Labor Review, July 1974, Vol. 97, No. 7, pp. 33-37.

(8) U.S. Department of Commerce, Bureau of the Census, Current Population Reports, Series P-60, No. 91, "Characteristics of the Low-Income Population: 1972," table 42.

(9) Orshansky, Mollie. Testimony before the House Committee on Education and Labor, November 15, 1973, p. 64.

(10) Coder, John F. "Results of a Survey on Household Participation in the Food Stamp Program: Data from the June 1973 Current Population Survey," paper presented at the Annual Meeting of the American Statistical Association, St. Louis, Missouri, August 1974, p. 14.

(11) Mahoney, Bette S. "Review of Poverty and Income Distribution Statistics," Statistical Reporter, Statistical Policy Division, Office of Management and Budget, January 1974.

(12) U.S. Department of Commerce, Bureau of the Census, Current Population Reports, Series P-23, No. 28, "Revision in Poverty Statistics, 1959 to 1968," p. 2.

(13) Morgan, James N., Dickinson, K., Dickinson, J., Benus, J., and Duncan, G. Five Thousand American Families--Patterns of Economic Progress, Vol. I, Institute for Social Research, The University of Michigan, 1974, p. 11.

DIRECTIONS IN AFDC STATISTICAL RESEARCH

Mitsuo Ono
Social and Rehabilitation Service*

I. Introduction

Since 1937, the Aid to Families with Dependent Children (AFDC) program has provided public assistance to needy families with children who are deprived of parental support or care. In addition, beginning May 1961, States could extend the AFDC program to include not only foster home care for these children but also assist intact and needy families with children whose fathers were inadequately or temporarily unemployed and ineligible for unemployment insurance benefits. 1/ Currently, 25 jurisdictions have approved plans operating under the unemployed father segment of the AFDC program.

As of February 1974, 10.9 million persons or 5.2 percent of the total civilian population in the nation were covered under the AFDC program. Included were 7.9 million children and 3.0 million adults in about 3.2 million families (or 5.8 percent of the total number of families in the nation). The program currently covers about 97 AFDC children for every 1000 under 21 years of age in the Nation (the AFDC child recipient rate).

In February, 1974, 461,000 recipients (including 277,000 children) within 101,000 (mostly double parent) families were in the unemployed father segment of the AFDC program. The remaining segment covered essentially single parent families, mostly families headed by a woman. In January, 1973, a special study found that about two million families of the 2.9 million AFDC families (or 79 percent of the total) were families headed by a woman.

Four States (New York, California, Illinois and Pennsylvania) with about 30 percent of the nation's population had 36 percent of the total AFDC recipient caseload.

AFDC assistance groups not only receive cash maintenance payments but some are also eligible to receive benefits such as food stamps, Medicaid, housing allowances and social services. The amount and type of assistance received depends on States' payments plans and on how much assistance units are able to obtain necessities on their own. Means-tested assistance payments make up the gap between income received and the minimal living cost standards established by States (including adjustments which differed by States.) 2/

In February 1974, for the nation as a whole, an average AFDC case unit (approximately a 4-person family) received about $200 monthly of cash assistance or an annualized amount of about $2400. This cash assistance was roughly about 53 percent of the low income threshhold for a 4-person family. However, due to different payment standards among States, average monthly payments per unit ranged from $339 in New York to $51 in Mississippi.

AFDC cash income maintenance payments for February 1974 amounted to 638 million dollars. In calendar year 1973, total AFDC income maintenance payments totaled 7.2 billion dollars, approximately 35 percent of all Federally aided assistance program expenditures of 20.5 billion dollars (including AFDC, the adult programs and emergency assistance).

In February 1974, Federally aided Medicaid payments (in the form of vendor payments) amounted to 842 million dollars. In calendar year 1973, total Medicaid payments totaled 9.8 billion dollars. Overall, the Federal Government paid about 54 percent (income maintenance) and 52 percent (Medicaid) of total assistance costs with the balance paid by State and local jurisdictions.

The establishment of timely and comprehensive statistical reporting channels from State welfare offices to Federal regional offices and to NCSS is critical for the proper planning and administration of Federally aided but State operated public assistance programs, such as AFDC.

The main purpose of this preliminary paper, divided into three parts, is to outline better information needs for policymaking purposes. The first part highlights changes in AFDC caseload and AFDC family characteristics occurring in past years. The second part presents some of the efforts being made in NCSS to improve the current AFDC statistical information systems.

While the first and second parts are descriptive, the third part is exploratory. The third part points out needs for further research to develop better ways to collect household survey data. It notes that for certain types of information data collection methods fail to produce relevant information because of the lack of a theoretical framework and related data collection techniques, especially in household surveys covering low income families. It further suggests that the human capital approach be used in developing a conceptual framework for use in conducting household surveys.

II. Changes in AFDC Caseload and AFDC Characteristics

A. Macro Changes

Since the beginning of the program in 1936, AFDC caseload has been on an upward trend in response to many complex interrelated variables. As of December 1936, with 23 States included, the program covered 161,600 AFDC families and 546,200 recipients (403,980 children). As of December

1973, with 54 jurisdictions included, (50 States Puerto Rico, District of Columbia, Guam, Virgin Islands) there were 3,155,500 AFDC families and 10,814,300 recipients (including 7,811,700 children). Between 1936 and 1940, the average annual percentage rate of increase in the number of families and recipients were both about 21 percent per annum primarily because of new States coming into the program. However, between 1941 and 1973, the annual rate of increase averaged out to 6.7 percent for families and 6.9 percent for recipients. The ratio of AFDC families to all families in the Nation (the AFDC family rate) and the ratio of AFDC recipients to the resident population in the Nation (the AFDC recipient rate) increased at an average annual rate of 6.5 percent and 5.4 percent, respectively, during this same period. These data indicate that although AFDC families increased proportionately with the increase in the number of families in the Nation, the AFDC recipient population increased at a slower rate than the overall population growth from 1941 to 1973.

A full analysis of the reasons for the upward trend in AFDC caseload over the past 37 years is outside the scope of this present paper. Only an outline of the more important factors involved is presented. For those who wish to pursue this problem further, publications on this general subject are available. 3/

The growth of AFDC caseload during the past 37 years has been in response to many complex factors. Simplified, they relate to demographic changes, e.g., child population increase, mobility and migration; economic, e.g., rising standards of living, unemployment, etc.; sociological, e.g., increasing teenage-age marriages, more broken homes, urbanization, etc.; and the most important-administrative, judicial and legislative program changes, including amendments in the Social Security Act, several important court decisions, and changes in State and Federal regulations resulting in expanded coverage, improvement of standards, etc. Clearly, this expansion in AFDC caseload did not occur at an uniform rate but at variable rates depending on how various factors affected the creation of new eligibles, how fast they were converted to program recipients, and on how fast AFDC recipient units were leaving the system.

Beginning in 1936 and continuing through 1940, the high average annual percentage increase (21 percent per annum) in the number of families and recipients between 1936 and 1940 is attributed primarily to new States entering into the program. Between 1936 and 1941, 21 additional States joined this program. By April, 1941, 44 States had adopted the AFDC program. Not included were several States with about 10 percent of the total civilian population in the Nation.

From December 1941 to December 1945, the continual rise in AFDC caseload was broken due to World War II. During this period, the average annual percentage decline was 8.5 percent and 7.5 percent, respectively, for families and recipients. However, after this temporary contraction, the AFDC caseload again rose. From 1946 to 1950, the average annual percentage increase in AFDC families and recipients were 17 percent and 19 percent, respectively. The AFDC recipient rate showed an average annual rate of increase of about 15 percent. Major reasons for this upswing were increases in marriages and marital breakups, expansion in the child population, economic problems, and program changes. In August 1950, the caseload reached a peak of 655.8 thousand families and 2.24 million recipients. As of December, 1950, 36 children per 1000 population under 18 years of age in the Nation were aided in the AFDC program.

This expansion in AFDC caseload was temporarily stopped during the period of the Korean conflict. From 1951 to 1953, the average annual percentage reduction in the number of AFDC families and recipients were 5.6 percent and 2.5 percent, respectively. As of December 1953, there were 547.3 thousand AFDC families and 1.94 million recipients on the rolls. This represented a net loss of 108 thousand AFDC families in three years. The child recipient rate was 27 per 1000, significantly below the high of 36 per 1000 in December of 1950.

From the end of 1953 to the beginning of the 1957-58 recession, the overall AFDC caseload moved steadily upward again. Between December 1953 and December 1956, the average annual percentage increases in AFDC families and recipients were 4.0 percent and 5.4 percent, respectively. The AFDC recipient rate grew at an annual average of 3.4 percent. The child recipient rate was 29 per 1000 as of June 1956, still below the high of 36 per 1000 in December 1950. However, from 1956 to 1958, the average annual rate of increase accelerated to 11 percent for families and 12 percent for recipients, probably affected by the 1957-58 recession. Although the average annual rate of increase in AFDC caseload leveled off for a short period from 1959 to 1960 (4.3 percent for recipients) it started to advance again during the 1960-61 period. This amounted to an average of 11 percent for recipients between December 1960 and December 1962. Two main reasons were involved: (1) the 1960-61 recession and (2) effective May 1961, the AFDC program was broadened to include the unemployed father segment. There was a net increase of 112,000 AFDC families and a net increase of 493,000 recipients between December 1960 and 1961.

From 1962 through 1966, the volume of AFDC families and recipients moved forward only gradually, averaging an annual increase of 4.9 percent and 5.3 percent, respectively. Both the AFDC family rate and the AFDC recipient rate increased at an annual average rate of 4.0 percent; in contrast, the child recipient rate creeped up from 42 per 1000 in June 1962 to 48 per 1000 in June 1966, reflecting a higher share of the Nation's children under 18 years within the AFDC program.

Beginning in August 1966, the longest rise in AFDC caseload began and continued upward until

April 1972. From August 1966 to March 1970, the average annual rate of increase of the 12 months moving average (to eliminate seasonal variations) in the number of AFDC families was 16 percent per annum; from April 1970 to December 1970, it accelerated to a historical peak of 33 percent per annum. From January 1971 to March 1972, the expansion slowed down but still advanced at a very high rate of 22 percent per annum. Between December 1966 to December 1971, the AFDC family caseload continued to increase at an average annual rate of 18 percent while the AFDC recipient caseload rose at 21 percent per annum. The comparable measures for the AFDC family rate and the AFDC recipient rate were 19 percent and 16 percent, respectively.

In December 1966, the AFDC child recipient rate was 48, but in December of 1971, this rate was approximately 83 children per 1000 population under 18 years of age, a near doubling within 5 years.

As of December 1971, the AFDC caseload consisted of 2.9 million families and 10.7 million recipients (including 7.7 million children). Five years before (December 1966) the caseload covered about 1.1 million families and 4.4 million recipients (including 3.3 million children).

The dramatic rise in AFDC caseload during this period can be attributed to several factors. The more important ones were the 1967 amendments to the Social Security Act which required States to reprice their cost standards for basic needs to reflect price changes and required States to disregard specified amounts of earnings in determining the assistance amount that a family with earnings would receive. Additional families became eligible for assistance under the first amendment while cases that normally would had been closed continued to be eligible for reduced assistance payments under the second. In addition, more units entered into the AFDC program as they became aware of their eligibility as publicized under the "War on Poverty" program. Also, many organizations working in this program area caused program regulations to be revised which made possible more people to be eligible under the AFDC program. In addition, internal administrative changes caused more AFDC eligible families to receive benefits.

However, beginning with April 1972, the average annual percentage rate of increase of the 12 months moving average began to decline. From April 1972 to August 1973, the average annual rate of increase of the 12 months moving average for recipients dropped about a third (7 percent) of the rate for the January 1971-August 1973 period (22 percent). From September 1973 to February 1974, the average annual rate of increase in the moving average dropped significantly to a level of 1.2 percent. The rate for 1970 to 1971 was 10.3 percent, from 1970 to 1972, 7.1 percent, and from 1970 to 1973, 3.8 percent. Declines in the average rates of expansion were equally significant for families, 14.3 percent, 10.6 percent and 7.3 percent, respectively. For the AFDC family rate, the rates of increase were 12.9 percent, 8.8 percent and 5.5 percent while for the AFDC recipient rate the comparable rates were 9.0 percent, 5.9 percent, and 2.8 percent, respectively.

This decline in the rate of increase in the dramatic upward movement of AFDC caseload can be attributed to many interrelated factors among which are administrative (e.g., tightening of administrative procedures in screening eligibles), legislative (e.g., shift of the disabled to APTD and subsequently to SSI prior to July 1973), demographic (e.g., probable saturation of female headed families eligible for AFDC) and economic (e.g., budgetary pressures on governmental units). The various reasons responsible for the rise and decline of the AFDC caseload are to be analyzed in more detail.

With respect to seasonal variations in the AFDC caseload, there is a definite seasonality pattern in the caseload for the unemployed father segment. Overall, the national caseload tends to increase about 17 percent during the late winter and early spring period and tends to decrease about 14 percent during the late summer and early fall period. The AFDC-UF segment caseload hits its highest peak during January-February and decline to its lowest level during the summer months. These variations are presumably associated with seasonal farm and service-type work, involving the high risk AFDC population.

In summary, this overview noted that AFDC rolls have not expanded on a steady course during the past 37 years but have grown with different periods of rise and decline, depending how different caseload variables interacting with each other. More detailed investigations of reasons for fluctuations in AFDC caseload will be a subject for future papers.

B. Micro Changes

Characteristics of AFDC families have also changed historically, many in line with national trends. Data depicting these changes were obtained from National surveys of AFDC families conducted recurrently since 1948 by NCSS in cooperation with State welfare offices. These changes in characteristics reflect changes in administrative, demographic, economic and sociological variables. The statistical information compiled in these surveys is limited to those which can be readily obtained by caseworkers from case schedules kept in State welfare offices. All information is handled under very strict procedures to protect the confidentiality of individuals. Since more detailed statistical information is available elsewhere, only key changes in characteristics are highlighted. 4/

In line with National trends, AFDC families are becoming more urbanized (an increase of about 31 percent between 1953 and 1973). The proportion of Black families in the National caseload has increased, from 22 percent in 1942 to 40 percent in 1961 and 46 percent in 1973. The average number of child recipients declined in AFDC families from 3.2 in 1961 to 2.6 in 1973.

Accordingly, the average size of an AFDC family has dropped from an average of 4.2 persons in 1961 to 3.6 persons per family in 1973. Moreover, the median age of AFDC mothers has declined from about 35 in 1961 to about 30 years in 1973.

In 1961, the median time on assistance rolls of those on the caseload at the time of the survey for a family was 2.1 years. In 1971, the comparable figure was 1.6 years. This drop was heavily influenced by the large number of AFDC families newly coming into the program during the 1966-71 caseload expansion period noted previously. In 1973, however, the median time on assistance had risen to two years. This rise was due to a combination of factors including a slowdown in the caseload expansion which resulted in a smaller proportion of families which were on AFDC rolls for shorter periods and also a result of the income retention provision of the 1967 amendment.

The percentage of illegitimate children in the AFDC program has risen. 5/ In 1948 about 11 percent of all children in the program were illegitimate. This percentage rose to 24 percent in 1961, and to 33 percent in 1973. The January 1973 study showed that about 46 percent of all AFDC families had at least one illegitimate child.

Also in line with national trends, there are more AFDC families with marital breakups. In 1961, about 22 percent of all AFDC families had no fathers because of marital breakup. In 1973, this rate was approximately 45 percent, twice the percentage of 1961.

Approximately 18 in every 100 mothers was employed or looking for work in 1961. However, by 1973, this rate had risen to 29 in every 100. AFDC mothers working full-time also increased. In 1961, one in every 20 mothers had worked full time. In 1973, this rate was 1 in 10.

The average educational level of workers has also risen from 9 years of schooling in 1961 to 11 years of schooling in 1973.

Overall, these data indicate that as compared with the past, AFDC families have better chances for getting off AFDC rolls, i.e., relatively smaller sized AFDC families with breadwinners having more work experience and higher educational levels than before. However, further reduction also requires that on the demand side, more work opportunities be made available for them, and on the supply side, barriers for employment outside of the home be reduced (e.g., availability of child care facilities).

III. Ongoing NCSS Statistical Research 6/

In view of the need for more relevant, timely and accurate statistical information on SRS program recipients, eligibles, and the high risk AFDC population, a demonstration project to be implemented in a single State is in progress to (1) provide information on unduplicated counts of SRS program participants, (2) tie in assistance from multiple programs to individuals and families, (3) identify factors which lead to programs participation among eligible and (4) provide detailed characteristics of recipients and services received by these recipients.

This redesign of the recipient statistical reporting system brings into focus data of the "in-scope" population who are already on the rolls of welfare agencies. Information will also be sought about the high risk AFDC population by means of a proposed longitudinal sample survey of low income households. A small pilot study is presently being conducted by the Census Bureau covering families with an income level of less than 1.5 the low income threshold. This survey will attempt to learn more about the AFDC eligible population, with or without public assistance payments, and the low income population not eligible for the AFDC program. In this test study, data on household compositional changes will also be collected. The results of this study should provide valuable information on problems involved in collecting such data in future household surveys covering low income families.

Furthermore, NCSS is publishing results of the AFDC study conducted in January 1973, of which some finding were presented in Part II. This is a continuation of a series of such studies conducted periodically since the program originated. Surveys were made in 1948 (June), 1953 (November), 1956 (January, February, March, or April), 1961 (December), 1967 (November or December), 1969 (May), and 1971 (January). Although no evaluation has been made to determine the accuracy of study findings, plans are being made to conduct such analysis in the future.

In addition to the above, action is being taken to expedite the timely reporting of key statistical data from State welfare offices by telephonic reporting to NCSS (via Regional Offices) in advance of the regular reporting requirements. Also to speed up the processing and publication of statistical data collected by NCSS, efforts are being made to computerize the production of NCSS publications.

The above work will materially improve the current statistical information system within NCSS and will result in better statistical data on AFDC caseloads and payments for use in the administration of SRS programs.

IV. Directions in AFDC Statistical Research

It goes without saying that decision-makers make choices under conditions of uncertainty. The statistician's main task is to help reduce this uncertainty by providing the proper information so that the decision-maker will make better decisions. Ideally, this information should be relevant, accurate, timely and understandable. However, working under imperfect conditions, the statistician sorts out the more important requirements and attempts to meet them as best as possible under the constraints of time and resources. In this effort, tradeoffs among alternatives are made depending on how information is to be used, costs of obtaining the desired information and the penalties of making wrong decisions. For example, under certain circumstances, timeliness may be more important than rigor and in others, accuracy supersedes, regardless of costs.

In making these tradeoffs, implicitly or explicitly, the statistician applies a crude information effectiveness index. In this regard, this writer believes that many statistical studies on household activities, especially on low income households could have been materially improved by the use of some analytical framework in compiling statistical data. Typically, much of the information asked in surveys is not related. This writer believes that these surveys have a low information effectiveness index because of the lack of an appropriate conceptual model and the use of deficient data collection techniques. Under these circumstances, the decision-maker could receive misleading quantitative information.

Framework of Analysis

In trying to understand a complex social phenomena involving many interrelated variables, e.g., demographic, social, economic, psychological, political and administrative, decision makers need to formulate a framework by which these interrelationships can be analyzed. It is suggested initially that simplified social accounts be used which correlate transactions between and among the three sectors in the domestic economy - government, business and households.

This is shown in the form of circular flow accounts:

Sales (Income)	Purchases (Expenditures)
Private and Household Enterprises (Production)	
Net sales of goods and services to enterprises, government and to households	Net purchases of goods/services Labor services purchased from households Net Government payments Residual
Households (Production and Consumption)	
Earnings received from labor services sold to enterprises, government and to households Transfers receipts Other receipts	Net purchase of goods/services Transfer payments Net government taxes Residual
Government (Production)	
Government Receipts	Net purchase of goods/services Labor services purchased from households Net transfer payments Residual

In this scheme, households purchase goods and services (from business and household enterprises, government and households) for consumption purposes. The value of consumption expenditures, other things equal, depends on the amount of aggregate income received, which in turn, is divided into earnings received from labor and capital services, and transfer payments from governmental or private sources. The value of individual earnings is a product of earnings received per time unit and the amount of time unit worked. For public assistance households, e.g., those in AFDC, means-tested transfer payments are cash assistance receipts and income-in-kind or vendor payments for housing, social and medical services.

Services provided to eligible public assistance households are paid by Government in the form of vendor payments or direct compensations. Within the household sector, given fixed levels of household consumption expenditures, earnings and transfer receipts are substitutes since as one source increases, the other decreases and vice versa. Under these circumstances, transfer payments are substitutes on the income side; but on the expenditures side, they are complements.

A key problem is to examine how rates of substitution between these income sources, given a fixed consumption level, are derived by households under varying conditions and circumstances.

This framework helps to examine the following statistical conceptual issues in conducting household surveys, especially among low income households:

(a) Need for a labor utilization approach instead of the presently used labor force approach to collect work experience data;

(b) Need for a human capital conceptual approach in designing household surveys;

(c) Need for integrated household surveys, and,

(d) Need for integration of information on transactions between and among the three sectors - government, business, and households.

Need for A Labor Utilization Approach

For the typical household, aggregate income consists mostly of earnings. However, for the public assistance households, income consists of mostly transfer payments, whether in the form of cash payments or income-in-kind. Many studies on the employment potential of AFDC families have concluded that breadwinners of these families are employable and are willing to work to get off public assistance rolls if employment barriers, such as need for training, inadequate child care facilities, illness, etc., could be somehow reduced. 7/

The size of individual earnings is a product of the volume of paid work measured in time units, and the corresponding average time unit earnings, which depends on the occupational activity performed. For aggregate household earnings, its size depends on (1) number of earners in the household and (2) the size of individual earnings.

Thus, the overall level of aggregate household expenditures and income which affects the assistance level, depends upon the amount of earnings received which in turn is related to the amount and type of paid work performed. Study findings show that work activity being performed by AFDC mothers was typically sporadic employment-either full or part time. 8/

The currently used labor force approach used for measuring employment activity in AFDC households is not fully adequate to produce information which reflects the fluid and dynamic work situation typically found among these households. A basic assumption of the labor force approach is that labor agents work at and seek jobs - related to established work activities involving regular work, regular pay, and rather standardized working schedules. In AFDC households, it is not unusual to find working age members engaged in low productive work in marginal services activities while seeking regular (and more productive work elsewhere). For others, they temporarily apply for assistance until suitable work is found. They move in and out of different work (and nonwork) activities as circumstances change. Under these conditions, classifications such as "inside and outside of the labor force" based on a week's time, become meaningless and very difficult to interpret for policy-making.

Instead of the labor force approach this writer suggests that a labor utilization approach may be more suitable for measurement purposes. This proposal employs a flow accounting concept rather than the stock accounting concept used in the labor force approach. 9/

The labor utilization approach attempts to analyze the sources and uses of the labor input flow in the same way that the economic accounts attempt to measure the value-added flow in productive activity. In this approach, labor agents are classified in different categories based upon how their streams of labor energy are utilized. In contrast, the labor force approach primarily attempts to measure the number of people in the economically active population who were either employed or not employed at certain points of time as shown by the labor force participation rate and the unemployment rate. The mechanics of getting meaningful information under the labor utilization approach in household surveys involve the development of a time disposition questionnaire which allows for the recording of irregular hours worked by labor agents in primary and second work activities during the month. In view of the typical irregularity of work activities performed by persons in AFDC households, the use of a time disposition questionnaire provides the opportunity to record the variability not only in the hours worked but also in the types of work performed. For recording purpose, the survey period can be divided into (1) the survey reference period and (2) the survey accounting period. Labor agents are classified based on the information compiled in the survey reference period. More detailed information is collected for those labor agents which have very irregular working conditions, e.g., migrant laborers, in the supplementary survey accounting period. This allows more flexibility in collecting survey information, especially for the dynamic work situation typically found among AFDC households.

Need for A Human Capital Conceptual Approach in Designing Household Surveys

Once the formulation of the labor utilization approach is accepted the applicability of the human capital approach in designing household surveys logically follows. 10/

The basic idea of this approach is that persons or labor agents represent human capital with embodied investment flows such as inputs of health care, education, training, etc. 11/ They are also carriers of work energy that can be utilized in both production and non-production activities. In combination with other factor agents of production, labor agents produce goods and services (and receive earnings) for current consumption or further investment.

A good example of this relationship is that between caloric consumption, work activity and value added. 12/

A basic problem is to investigate the flow process by which this work energy is created and eventually utilized in different renumerative and non-renumerative end-use activities. This type of analysis should materially add to the better understanding of needs of low income households for policymaking purposes.

Need for Integrated Household Surveys

In line with the above, requirements exist to analyze the socio-economic behavior of public assistance units in a more comprehensive fashion using integrated household surveys. It is unfortunate that much of this interdependent information on human and social capital (investment) is collected separately in ad hoc surveys, e.g., a sample survey on the labor force, a sample survey on income and expenditures, etc. This problem can be attributed partly to the lack of a comprehensive analytical framework which can be used to integrate different types of household data into a socio-economic theoretical system.

As soon as we accept the need for the use of the human and social capital approach in designing household surveys, the use of integrated household surveys becomes more critical.

Need for More Information Analyzing Transactions Between and Among Sectors of the Economy

The discussion up to now relates to data needs to analyze more effectively the dynamic behavior of the household sector. However, for policy-making purposes, data are also required which show linkages between and among the selling (supply) and buying (demand) components, e.g., the provision of social and medical services to AFDC recipients. For this, a marketing matrix can be developed. Thus, a social service statistical information system can be developed in the same way that retail consumer sales by different types of outlets can be related to purchase of goods and services by different types of households.

This matrix can be used to trace types of social services received by different groups of public assistance recipients from various vendor groups under different SRS program areas. Work is now ongoing to develop the basic groundwork by which such information systems can be implemented.

Summary

This preliminary note is a beginning of a series of more detailed papers presenting statistical research efforts currently being conducted at NCSS. It indicates the need for more analysis on the sources and structure of growth in AFDC caseload. Also, it notes that more work is needed in developing new conceptual approaches to compile statistical data covering potential and current AFDC households in household surveys. Finally, it calls for more investigation to improve the statistical measurements of transactions dealing with public assistance payments between and among the three sectors of the economy-business, government and households.

Footnotes

* National Center for Social Statistics, Office of Information Science (OIS). The statistical assistance of Mr. Stanley Nachimson, National Center for Social Statistics (NCSS) is gratefully acknowlledged. Other staff members were also consulted in preparing this preliminary note.

1/ See Handbook of Public Income Transfer Programs, Paper No. 2 of Studies in Public Welfare, by Irene Cox, Joint Economic Committee Print, U. S. Government Printing Office, Washington, D. C., 1972, p. 136.

2/ See the following: "Families Receiving AFDC: What Do They Have to Live On" by Gerald Kahn and Ellen Perkins, Welfare In Review, October 1964; "Characteristics Assóciated with Receipt or Non-Receipt of Financial Aid from Welfare Agencies" by Robert Stone and Fredric Schlamp, Welfare In Review, July 1965. See also Studies in Public Welfare, various papers.

3/ See the following: "AFDC in Review"; Ellen Perkins, Welfare in Review, November 1963; "Trend in AFDC Recipients, 1961-65"; John Lynch, Welfare in Review, May 1967; Decline in Number of AFDC Orphans, 1935-1966", David Eppley, Welfare in Review, Sept.-Oct. 1968. NCSS Report (A-2) for March 1970; NCSS Report (H-4); "Trends in AFDC," NCSS Report (H-1), October 1968. Published analyses by various State agencies, e.g., the Illinois Department of Public Aid. Studies in Public Welfare series prepared by the Jöint Economic Committee. "Have the Poor Been Regulated" Toward a Multivariate Understanding of Welfare Growth" by Eugene Durman, Social Services Review, September 1973, pp. 339-359. Aid to Dependent Children by Winifred Bell, Columbia University, 1965.

4/ See "The AFDC Family in the 1960's" by David Eppley, Welfare in Review, Sept.-Oct. 1970; "A Study of Children in AFDC Families," by Perry Levinson, Welfare in Review, March-April, 1969; "Changes in AFDC, 1969-71" by Betty Burnside, Welfare in Review, March-April 1972. See also various NCSS publications published under the title Findings of AFDC Studies.

5/ See "The Incidence of Illegitimacy in the United States" by Arthur Campbell and James Cowhig, Welfare in Review, May 1967. See also "The Role of the Unmarried Father" by Mignon Sauber, Welfare in Review, November 1966.

6/ This covers the research work currently being conducted only in NCSS and does not include the other research work being completed in other organizational elements within SRS.

7/ See "The Employment Potential of AFDC Mothers" by Genevieve Carter, Welfare in Review, July-August 1968; "How Employable are AFDC Women?" by Perry Levinson, Welfare in Review, July-August 1970; "MDTA Training and Employability" by Edward Prescott, et.al. Welfare in Review, January-February 1971; "The Employment Potential of AFDC Mothers" by Betty Burnside, Welfare in Review, July-August 1971. For studies relating employability and public assistance see "The Employability of AFDC Mothers and Fathers" by Martin Warren and Sheldon Berkowitz, Welfare in Review, July-August 1969; Measuring the Potential of AFDC Families for Economic Independence" by Harold Morse, Welfare in Review, November-December 1968; "Predicting Use of Public Assistance" by Oliver Moles, Welfare in Review, November-December 1961.

8/ See "Patterns of Work and Welfare in AFDC" by Mildred Rein and Barbara Wishnov, Welfare in Review, November-December 1971. See also "Families on Welfare in New York City," Welfare in Review, March-April 1968.

9/ This concept involves a flow/stock model where a person is envisoned as a carrier of human capital. There is a stock (an inventory of potential human energy) and a flow (the use of this labor energy as measured in time periods). The human population at a point of time then represents a collective of persons holding an inventory of potential labor energy to be used for some end use.

10/ According to T. W. Schultz, investments in human capital can be classified under (1) schooling and higher education, (2) pre-school training and learning, (3) preschool learning activities, (4) migration, (5) health, (6) information and (7) investment in children (population). See his "Human Capital: Policy Issues and Research Opportunities" in Human Resources, Fiftieth Anniversary Colloquim, Part VI, National Bureau of Economic Research, New York, 1972.

11/ See "Food Consumption, Nutrition and Economic Development in Asian Countries" by H. T. Oshima, Economic Development and Cultural Change, July 1967, pp. 385-397.

12/ For useful studies, see The Economics of Human Resources by H. Correa, North Holland Publishing Co., Amsterdam, 1963; Human Capital Formation and Manpower Development, edited by R. A. Wysha, The Free Press, N. Y., 1971; Journal of Political Economy, Supplement on "Investment in Human Beings," Vol. LXX, No. 5, Part 2, Oct. 1962; "The Increasing Economic Value of Human Time" by T. W. Schultz. The American Journal of Agricultural Economics, December 1972 pp. 843-850.

DISCUSSION

W. Michael Mahoney, Social Security Administration

The three papers presented here are closely related. One is about the Aid to Families of Dependent Children (AFDC), which provides benefits to poor female headed families. Another is about the Supplemental Security Income (SSI) program, which provides benefits to poor aged, blind and disabled persons. And one is about the poor population as a whole. Unfortunately, despite the wealth of interesting and related data in each of them, one cannot move readily from one to the other. The first two papers rely on "program" data, i.e. data generated by the programs themselves. Since each of the two programs uses different eligibility criteria and different definitions of eligible units, the data they generate are not directly comparable. The third paper relies on data collected by the Bureau of the Census in the regular March income supplement to the Current Population Survey (CPS). The Bureau of the Census uses definitions of families and households which do not match the programs' definitions of eligible units. Moreover, it reports family income on a calendar year basis whereas program eligibility is determined for much shorter periods of time. Finally, the CPS provides only limited information on asset holdings which are often an important determinant of program eligibility.

These differences have been with us for a long time. Why do I stress them now? Because in the last six years increasingly sophisticated models have relied on the CPS data to simulate public assistance and other income transfer programs. These models have been the basis for dozens of studies of program impact and have been the source of most cost and caseload estimates produced for the many new welfare programs proposed during these years. The first real test of the ability of these models to produce accurate estimates is far from inspiring.

Recent experience with the SSI program, which was implemented in January 1974 and which thus far has far fewer recipients than first estimated, suggests that one or more of the following are wrong:

(a) the techniques and inferences used to conform the CPS data to program concepts
(b) assumptions about participation rates
(c) the program data
(d) the CPS data

Quite likely there are some biases in the way the various simulation models manipulate the data. But it is an act of faith to assume that any remaining differences are attributable to program participation rather than the data themselves. I prefer to remain skeptical and to assume that all four sources of error are equally likely.

Income transfer programs have become the largest component of Federal expenditures. It is essential that there be adequate data and procedures for exploring these programs and the impact of possible changes in them. Efforts to provide and improve this capability in the Office of the Secretary of Health, Education and Welfare, in the Office of Research and Statistics of the Social Security Administration, at the Urban Institute, Mathematica and other private firms are to be commended. The pilot study to which Mitsuo Ono refers in his paper is another important step in the right direction. It seems likely, however, that a new income survey which would replace the March supplement will be necessary.

There have of course already been a number of discussions about such a new survey, what it should look like, and the problems that are likely to be associated with it. I would put high on the list of problems for any future discussions the difficulties that will be faced in gathering data that conform to program usage.

I have a few comments about each of the papers.

Renee Miller and Arno Winard provide us with a concise and useful description of the poverty population and how it has changed over time. But in attributing some of the decline in the number of working poor to the impact of minimum wage legislation they provide an interpretation that most economists would surely dispute. If minimum wage legislation has the affect that economic theory predicts, the reduction in the number of working poor is not a result of their greater earning power but because fewer of them work! In fact, the very high unemployment rates among out-of-school youth and others with marginal skills may be a result of minimum wages. If these rates were lower, poverty rates might also be lower.

Miller and Winard point out another trend: the increasing number of female headed families. Citing the evidence of a recent publication of the Fiscal Subcommittee of the Joint Economic Committee that welfare benefits available to female headed families are generally greater than they are for male headed families, Miller and Winard suggest that financial incentives may account for the trend. It is a possibility I readily concede. But we must guard against the possibility of assuming that because an identifiable and measurable incentive exists, behavior is necessarily strongly influenced thereby. For example, it has long been assumed that the financial disincentives in most welfare programs effectively discouraged recipients from working. Recent evidence raises some doubt about this assumption. Where incentives exist, it is appropriate to learn whether they affect behavior but we must be cautious when we attempt to explain the behavior of others.

In his paper, Richard Bell provides us with a straightforward description of the new SSI program. But his suggestion that social security benefits are an important source of income for SSI recipients is misleading. Persons who receive both supplemental security and social security benefits are only twenty dollars a month better off than those who receive only supplemental security benefits. It is on the total expenditures in the SSI programs that the social security payments have their impact. These expenditures are substantially less than they otherwise would be.

I will note without comment that Bell's favorable assement of the Social Security Administration effort to identify and enroll eligible SSI recipients is not shared by a number of spokespen for the elderly poor.

Mitsuo Ono's paper about the AFDC program is a refreshing change from previous papers by employees of the Social and Rehabilitative Services agency which administers the AFDC program. Many (though not all) of those papers displayed an almost perverse unwillingness to to look at and think logically about facts. I don't know how many times Social and Rehabilitative Services publications pointed out that that the median length of time on the rolls was declining. Ono points out correctly that when the rolls are increasing of course the median time on declines.

Ono does slip up once or twice, however. He used differences in average welfare payments to illustrate the differing generosity among the states. However, average welfare benefits are a function of: a) the income distribution within the individual states; b) the formula the state uses to determine the payment to an eligible family; c) the administrative practices of the state; as well as d) the amount the state will pay to a family with no other income. The last of these, while not perfect, is a better indicator of generosity than average payment.

Ono observes that the proportion of families receiving AFDC where the father has deserted or never married the mother has been steadily growing. One reason for this which is seldom noted is that the number of widowed mothers requiring assistance has been steadily declining. This is in part a result of the maturation of the social security program and its increased benefit levels.

Ono is far more optimistic about the prospect of reducing the welfare rolls than I. While I agree with him that a greater proportion of them are employable than before, I doubt that many will be able to earn enough to reduce their welfare entitlement to zero.

I was asked by our Chairlady to introduce a "policy perspective" into our discussion of these papers. Ono's comment that timeliness is often as important as accuracy is a good turning point for me to do that. I could not agree with him more. Far too often statisticians, economists, and program analysts have sought a degree of perfection that simply was not worth the cost in time and resources. There is never a time when accuracy is important regardless of cost. This is a possibility that Ono admits but I suspect that his point his really that we should be pragmatic about our work. I certainly agree.

Pragmaticism is not the same as having undue concern about the inferences for policy that people may draw from our work. Because most of us are concerned about social welfare, we are constantly tempted to worry about this. <u>It is a temptation devoutly to be resisted</u>. It inevitably leads us to bias our work towards those policy outcomes we prefer. Our job is to help our employers make informed decisions not to lead them down the paths we wish them to follow.

Let me give an example.

Miller and Winard discuss the fact that non-cash income is not included in the CPS income measures. They point out that a number of people have concluded from this that the number of poor is overstated. They then quote Mollie Orshansky to the effect that the non-poor may receive as many or more non-cash benefits than the poor. Miss Orshansky is certainly right. Our subsidized presence at this meeting is one example of the non-cash benefits we non-poor receive.

But, if we are using an absolute measure of poverty, what is the point?[1] The amounts of non-cash transfer to the poor have increased rapidly and dramatically. We should not let our preference that the poor be still less poor, or that their poverty be relieved by cash rather than in-kind obscure that important point.

1/ Miss Orshansky has pointed out that if the income distribution reflected in-kind as well as cash and the poverty measure were reconstructed using her original methodology, the counted number of poor might actually increase. As constructed, the poverty measure is not precisely "absolute". But the measure is used as though it were absolute. In any event, the poor are less poor with their in-kind benefits than they would be without them.

ON SAMPFORD'S PROCEDURE OF UNEQUAL PROBABILITY SAMPLING WITHOUT REPLACEMENT

C. Asok and B. V. Sukhatme
Iowa State University, Ames, Iowa

1. Introduction: When sampling is done with unequal probabilities and without replacement for estimating the population total Y of a characteristic y defined over a population of size N, Horvitz and Thompson (1952) have proposed an unbiased estimator,

$$\hat{Y} = \sum_{1}^{n} \frac{y_i}{\pi_i} , \qquad (1)$$

with variance

$$V(\hat{Y}) = \sum^{N} \frac{Y_i^2}{\pi_i} + \sum_{i}^{N} \sum_{j(\neq i)}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} Y_i Y_j - Y^2 \qquad (2)$$

where n is the sample size, π_i is the probability of including the i-th unit in the sample and π_{ij} is the probability for the i-th and j-th units to be both in the sample. When information on an auxiliary characteristic x, assuming the value X_i on the i-th unit, is available for all the units where Y_i is approximately proportional to X_i, considerable reduction in the variance can be achieved by making $\pi_i \propto X_i$. Such a scheme must obviously satisfy the condition

$$\pi = n p_i , \qquad (3)$$

where $p_i = X_i/X$, X being the sum of all the X_i's. Condition (3) obviously puts a restriction on the X_i's viz., $\max_i X_i \leq \frac{X}{n}$ which is not a severe one. Among the schemes that satisfy condition (3) and are applicable for general sample size $n \geq 2$, we will consider here the Goodman and Kish (G and K) (1950) procedure and the Sampford's (1967) procedure. The procedure of Sampford, which is a generalization of the Durbin's (1967) procedure for sample size 2, is described as follows:

Assuming without loss of generality that $np_i < 1$ for all i, define

$$\lambda_i = p_i/(1-np_i) \qquad (4)$$

Let $L_0 = 1$ (5)

and $L_m = \sum_{S(m)} \lambda_{i_1} \lambda_{i_2} \ldots \lambda_{i_m}$, $(1 \leq m \leq N)$ (6)

where S(m) denotes a set of m different units $i_1, i_2, \ldots, i_m$ and the summation in (6) is over all such possible sets drawn from the population which are $\binom{N}{m}$ in number. The procedure then consists of selecting the particular sample S(n), consisting of units $i_1, i_2, \ldots, i_n$ with probability

$$P\{S(n)\} = nK_n \lambda_{i_1} \lambda_{i_2} \ldots \lambda_{i_n} (1 - \sum_{u=1}^{n} p_{i_u}) \qquad (7)$$

where

$$K_n = (\sum_{t=1}^{n} t L_{n-t}/n^t)^{-1} \qquad (8)$$

Since the evaluation of the expression in (7) for all the $\binom{N}{m}$ sets is out of question in practice, Sampford has suggested two alternative ways of achieving the probabilities P{S(n)} as given in (7). Method (i) is to select the units without replacement, with the probabilities evaluated at each drawing according to the rule described and illustrated by Sampford in his article. Method (ii) is to select n units with replacement, the first drawing being made with probabilities p_i and all subsequent ones with probabilities proportional to $p_i/(1-np_i)$ and rejecting completely any sample that doesn't contain n distinct units and to start afresh. In practice method (ii) could be convenient because a sample can be discarded as soon as a duplicate unit is drawn. However for small samples one may take as a guide line in the relative preference of methods (i) and (ii), the value of the expected number of samples that must be drawn to obtain an acceptable sample which is given by $K_n \cdot (\Sigma\lambda_t)^{n-1}/(n-1)!$. For this scheme of sampling Sampford has shown that π_i is given by (3) and

$$\pi_{ij} = K_n \cdot \lambda_i \lambda_j \phi_{ij} \qquad (9)$$

where $\phi_{ij} = n \cdot \sum_{\substack{S(n-2) \\ i,j \notin S}} \lambda_{\ell_1} \lambda_{\ell_2} \ldots \lambda_{\ell_{n-2}} \cdot$

$$\{1 - (p_i + p_j) - \sum_{u=1}^{n-2} p_{\ell_u} \} \qquad (10)$$

It has also been shown by Sampford that the condition $\pi_i\pi_j - \pi_{ij} > 0$, is satisfied which ensures the nonnegativity of the Yates and Grundy variance estimator.

Even though the exact expression for π_{ij} of the Sampford's procedure is available, the computations become quite cumbersome particularly for N and/or n large. Since the simplicity of computations is one of the factors to be considered in choosing a sampling procedure, it will be of advantage if reliable approximate expressions for π_{ij} are derived because one may prefer to use the approximate expressions that would be quite satisfactory and easy for numerical evaluation. Also since the procedures of Goodman and Kish, and Sampford are two competitive schemes, it will be worth while if we could compare the efficiencies of the two schemes. Thus it would be realistic for comparison purposes to derive the approximate expressions for π_{ij} and hence the variance for the Sampford's procedure using the asymptotic approach of Hartley and Rao (1962). In order to evaluate the variance expression of the Horvitz-Thompson (H. T.) estimator under the Sampford's procedure, we will first evaluate π_{ij} correct to $O(N^{-4})$ under the assumptions of Hartley and Rao (1962) viz., n is small relative to N and p_i is of $O(N^{-1})$.

2. Evaluation of the approximate expression for π_{ij} of the Sampford's procedure: Since $np_i < 1$, from (4) we get by expanding in Taylor's series and retaining terms up to $O(N^{-4})$ only,

$$\lambda_i\lambda_j = p_ip_j \{1 + n(p_i+p_j) + n^2(p_i^2+p_j^2+p_ip_j)\} \quad (11)$$

Since the leading term in $\lambda_i\lambda_j$ above is of $O(N^{-2})$, in order to evaluate π_{ij} in (9) correct to $O(N^{-4})$ only, it would be sufficient to evaluate K_n and ϕ_{ij} each correct to $O(N^{-2})$. For evaluating K_n and ϕ_{ij} we need the following lemma.

Lemma 1: Let $\ell_1, \ell_2 \ldots \ell_m$ be the units drawn in that order when a simple random sample without replacement of size m is drawn from a population of N units. Let $p_1, p_2 \ldots p_N$ be such that each p_i is of $O(N^{-1})$ and $\sum_1^N p_i$ is not necessarily equal to one. Then for $m \geq 3$ where m is small relative to N, the following relations are true correct to $O(N^{-2})$:

$$N_{(m)} \cdot E(p_{\ell_1} p_{\ell_2} \cdots p_{\ell_m}) = (\Sigma p_t)^m - \binom{m}{2}(\Sigma p_t)^{m-2} \cdot \Sigma p_t^2 + 2 \cdot \binom{m}{3} \cdot (\Sigma p_t)^{m-3} \cdot \Sigma p_t^3 + 3 \cdot \binom{m}{4} \cdot (\Sigma p_t)^{m-4} \cdot (\Sigma p_t^2)^2 \quad (12)$$

$$N_{(m)} \cdot E(p_{\ell_1}^2 p_{\ell_2} p_{\ell_3} \cdots p_{\ell_m}) = (\Sigma p_t)^{m-1} \cdot \Sigma p_t^2 - (m-1)(\Sigma p_t)^{m-2} \cdot \Sigma p_t^3 - \binom{m-1}{2} \cdot (\Sigma p_t)^{m-3} \cdot (\Sigma p_t^2)^2 \quad (13)$$

$$N_{(m)} \cdot E(p_{\ell_1}^3 p_{\ell_2} p_{\ell_3} \cdots p_{\ell_m}) = (\Sigma p_t)^{m-1} \cdot \Sigma p_t^3 \quad (14)$$

and

$$N_{(m)} \cdot E(p_{\ell_1}^2 p_{\ell_2}^2 p_{\ell_3} p_{\ell_4} \cdots p_{\ell_m}) = (\Sigma p_t)^{m-2} \cdot (\Sigma p_t^2)^2 \quad (15)$$

wherein $N_{(m)} = N(N-1)\ldots(N-m+1)$ and $\binom{\mu}{v}$ is to be taken as zero if $\mu < v$.

Proof is by induction which is straightforward and hence is omitted.

Remark: Even though the proof of the lemma assumes that $m \geq 3$, the relations (12)-(15) are true for m = 0, 1, and 2 also which can be easily verified.

L_m of (6) can be written as

$$L_m = \binom{N}{m} \cdot E[\lambda_{\ell_1} \lambda_{\ell_2} \cdots \lambda_{\ell_m}] \quad (16)$$

where E denotes the expectation over the scheme described in Lemma 1.
Substituting the value of λ_{ℓ_i} from (4) and by expanding in Taylor's series we get from (16)

$$L_m = \frac{N_{(m)}}{m!} \cdot E[\prod_{i=1}^m p_{\ell_i} \{1 + np_{\ell_i} + n^2 p_{\ell_i}^2 + \ldots\}]. \quad (17)$$

Now, it can be easily seen that for any set of positive integers $\alpha_1, \alpha_2, \ldots, \alpha_m$; the contribution of $N_{(m)} \cdot E[p_{\ell_1}^{\alpha_1} p_{\ell_2}^{\alpha_2} \cdots p_{\ell_m}^{\alpha_m}]$ to L_m, correct to $O(N^{-2})$, would be zero if $\sum_1^m \alpha_i > (m+2)$. Further from the basic properties of simple random sampling it is also known that $E[p_{\ell_1}^{\alpha_1} p_{\ell_2}^{\alpha_2} \cdots p_{\ell_m}^{\alpha_m}]$ is the same for all the permutations of $(\alpha_1, \alpha_2, \ldots, \alpha_m)$. Hence from (17) it follows that the expression for L_m that could contribute to $O(N^{-2})$ is

$$L_m = \frac{N_{(m)}}{m!}[E(p_{\ell_1} p_{\ell_2} \cdots p_{\ell_m}) + nm \cdot E(p_{\ell_1}^2 p_{\ell_2} p_{\ell_3} \cdots p_{\ell_m}) + \frac{n^2 m(m-1)}{2} \cdot E(p_{\ell_1}^2 p_{\ell_2}^2 p_{\ell_3} p_{\ell_4} \cdots p_{\ell_m})]$$

Substituting from (12)-(15) we get for $m \geq 3$,

$$L_m = \frac{1}{m!}[1 + \{\binom{m}{1}n - \binom{m}{2}\} \cdot \Sigma p_t^2 + \{\binom{m}{1}n^2 - 2\cdot\binom{m}{2}n + 2\binom{m}{3}\} \cdot \Sigma p_t^3 + \{\binom{m}{2}n^2 - 3\cdot\binom{m}{3}n + 3\cdot\binom{m}{4}\} \cdot (\Sigma p_t^2)^2] \tag{18}$$

It can be verified that (18) in fact holds for $m = 0, 1,$ and 2 also with the convention that $\binom{\mu}{v} = 0$ if $\mu < v$.

Theorem 1: For $n \geq 5$, the expression for $\frac{1}{K_n}$ correct to $O(N^{-2})$ is

$$\frac{1}{K_n} = \frac{1}{(n-1)!} + \frac{n}{2(n-2)!}\Sigma p_t^2 + \frac{n(n+1)}{3(n-2)!}\Sigma p_t^3 + \frac{n(n+1)(n-2)}{8(n-2)!}(\Sigma p_t^2)^2 \tag{19}$$

Proof: From (8) we get by using the transformation $s = n-t$,

$$\frac{1}{K_n} = \sum_{s=0}^{n-1}(n-s)\cdot L_s/n^{n-s} = \frac{L_0}{n^{n-1}} + \frac{(n-1)\cdot L_1}{n^{n-1}} + \frac{(n-2)\cdot L_2}{n^{n-2}} + G, \tag{20}$$

where $G = \sum_{s=3}^{n-1}(n-s)\cdot L_s/n^{n-s}$. Let $T_s = n^s/s!$, and for any nonnegative integers $0 \leq \ell \leq m$, let

$$I_{(\ell,m)} = \sum_{s=\ell}^{m} T_s$$

and

$$J_{(\ell,m)} = \sum_{s=\ell}^{m} s \cdot T_s$$

Then by using the relations

$$\sum_{s=\ell}^{m}(n-s)\cdot T_{s-\alpha} = (n-\alpha)\cdot I_{(\ell-\alpha,\, m-\alpha)} - J_{(\ell-\alpha,\, m-\alpha)}$$

for any $\alpha \leq \ell \leq m$;

$$J_{(\ell,m)} = n\cdot I_{(\ell-1,\, m-1)};$$

and

$$I_{(\ell+1,\, m+1)} - I_{(\ell,m)} = T_{m+1} - T_\ell,$$

we get by substituting the value of L_s from (18) into (20),

$$\frac{1}{K_n} = \frac{1}{(n-1)!} + \frac{n}{2(n-2)!}\Sigma p_t^2 + \frac{n(n+1)}{3(n-2)!}\Sigma p_t^3 + \frac{n(n+1)(n-2)}{8(n-2)!}(\Sigma p_t^2)^2 \qquad \text{Q.E.D.}$$

Remark: Direct evaluation of $\frac{1}{K_n}$ from (8) for $n = 2, 3,$ and 4 shows that (19) in fact holds for $n \geq 2$. From (19) we get for $n \geq 2$

$$K_n = \{(n-1)!\}[1 - \frac{n(n-1)}{2}\Sigma p_t^2 - \frac{n(n-1)(n+1)}{3}\Sigma p_t^3 + \frac{n(n-1)(n^2-n+2)}{8}(\Sigma p_t^2)^2] \tag{21}$$

correct to $O(N^{-2})$.

As the expression for ϕ_{ij} in (10) is not meaningful to consider for $n=2$, we derive the approximate expression for ϕ_{ij} assuming $n \geq 3$. Expression (10) can alternatively be written as

$$\phi_{ij} = n\cdot\binom{N-2}{n-2}\cdot E'[\lambda_{\ell_1}\lambda_{\ell_2}\cdots\lambda_{\ell_{n-2}} \{1 - (p_i + p_j) - \sum_{u=1}^{n-2} p_{\ell_u}\}] \tag{22}$$

where E' denotes the expectation taken over the scheme of selecting (n-2) units from the population excluding the i-th and j-th units with simple random sampling without replacement. Using the results of Lemma 1 with suitable modifications we get, from (22), for $n \geq 3$,

$$\phi_{ij} = \frac{n}{(n-2)!}[1 + \{\frac{(n-2)(n+1)}{2}\cdot\Sigma p_t^2 - (n-1)(p_i + p_j)\}$$
$$+ \{(n-1)(n-2)p_ip_j - (n-2)(p_i^2 + p_j^2)$$
$$- \frac{(n-2)(n^2-3)}{2}(p_i + p_j)\Sigma p_t^2$$
$$+ \frac{(n-2)(n^2+2n+3)}{3}\Sigma p_t^3$$
$$+ \frac{(n-2)(n-3)(n^2+3n+4)}{8}\cdot(\Sigma p_t^2)^2\}], \quad (23)$$

correct to $O(N^{-2})$.

Sampford's procedure is a generalization, for sample size $n \geq 2$, of the Durbin's (1967) procedure which is for sample size 2. The expression for π_{ij} of the Durbin's procedure is

$$\pi_{ij} = K_2 p_i p_j (\frac{1}{1-2p_i} + \frac{1}{1-2p_j})$$

Substituting the value of K_2 from (21) in the above, we get after retaining terms to $O(N^{-4})$ only,

$$\pi_{ij} = 2p_ip_j[1 + \{(p_i+p_j) - \Sigma p_t^2\} + \{2(p_i^2 + p_j^2) - 2\Sigma p_t^3$$
$$- (p_i + p_j)\Sigma p_t^2 + (\Sigma p_t^2)^2\}] \quad (24)$$

Substituting from (11), (21), and (23) into (9), we get for $n \geq 3$,

$$\pi_{ij} = n(n-1)p_ip_j[1 + \{(p_i+p_j) - \Sigma p_t^2\}$$
$$+ \{2(p_i^2+p_j^2) - 2\Sigma p_t^3 - (n-2)p_ip_j$$
$$+ (n-3)(p_i+p_j)\Sigma p_t^2 - (n-3)(\Sigma p_t^2)^2\}] \quad (25)$$

correct to $O(N^{-4})$.

Observation of (24) shows that (25) is in fact true for $n \geq 2$.

3. <u>Comparison of the variances of the corresponding H. T. estimators for the Sampford's procedure and the Goodman and Kish procedure</u>: The expression for π_{ij} of the Goodman and Kish proeedure correct to $O(N^{-4})$, as derived by Hartley and Rao (1962)(5.15 of p. 369), can be written in the modified form as

$$\pi_{ij} = n(n-1)p_ip_j[1 + \{(p_i+p_j) - \Sigma p_t^2\}$$
$$+ \{2(p_i^2 + p_j^2) - 2\Sigma p_t^3 + 2p_ip_j - 3(p_i+p_j)$$
$$\cdot \Sigma p_t^2 + 3(\Sigma p_t^2)^2\}] \quad (26)$$

Theorem 2: Given any unequal probability sampling scheme for selecting a sample of size n whose π_i and π_{ij} are given by

$$\pi_i = np_i \quad (27)$$

and
$$\pi_{ij} = n(n-1)p_ip_j[1 + \{(p_i+p_j) - \Sigma p_t^2\}$$
$$+ \{2(p_i^2 + p_j^2) - 2\Sigma p_t^3 + a_n\cdot p_ip_j$$
$$- (a_n + 1)(p_i+p_j)\Sigma p_t^2$$
$$+ (a_n+1)(\Sigma p_t^2)^2\}], \quad (28)$$

correct to $O(N^{-4})$, where a_n is some constant that does not depend on p_t's but may depend on n, the variance expression correct to $O(N^0)$ of the corresponding H. T. estimator is given by

$$V(\hat{Y}_{H.T.}) = \frac{1}{n}[\Sigma p_iz_i^2 - (n-1)\Sigma p_i^2z_i^2]$$
$$- \frac{(n-1)}{n}\cdot[2\Sigma p_i^3z_i^2 - \Sigma p_i^2\cdot\Sigma p_i^2z_i^2$$
$$- a_n\cdot(\Sigma p_i^2 z_i)^2], \quad (29)$$

where
$$z_i = \frac{Y_i}{p_i} - Y. \quad (30)$$

Proof: Substituting the values of π_i and π_{ij} from (27) and (28) in

$$V(\hat{Y}_{H.T.}) = \Sigma\frac{Y_i^2}{\pi_i} + \sum_i\sum_{j(\neq i)}\frac{\pi_{ij}}{\pi_i\pi_j}Y_iY_j - Y^2,$$

simplifying and retaining terms to $O(N^0)$ we get the expression in (29).

Q.E.D.

From (25) and (26) it can be observed that condition (28) of Theorem 2 is satisfied for both the procedures of Sampford as well as Goodman and Kish, the values of a_n being $-(n-2)$ and 2 respectively. Since condition (27) is known to be satisfied for both the procedures, we get from (29) that:
Variance of the H.T. estimator for the Sampford's procedure correct to $O(N^0)$ is

$$V(\hat{Y}_{H.T.})_{Samp} = \frac{1}{n}[\Sigma p_i z_i^2 - (n-1)\Sigma p_i^2 z_i^2] - \frac{(n-1)}{n}\cdot[2\Sigma p_i^3 z_i^2 - \Sigma p_i^2 \cdot \Sigma p_i^2 z_i^2 + (n-2)\cdot(\Sigma p_i^2 z_i)^2] \quad (31)$$

and the variance of the H.T. estimator for the Goodman and Kish procedure correct to $O(N^0)$ is

$$V(\hat{Y}_{H.T.})_{G \text{ and } K} = \frac{1}{n}[\Sigma p_i z_i^2 - (n-1)\Sigma p_i^2 z_i^2] - \frac{(n-1)}{n}\cdot[2\Sigma p_i^3 z_i^2 - \Sigma p_i^2 \cdot \Sigma p_i^2 z_i^2 - 2(\Sigma p_i^2 z_i)^2] \quad (32)$$

From (31) and (32) it follows that when the variance is considered to $O(N^1)$ only,

$$V(\hat{Y}_{H.T.})_{Samp} = V(\hat{Y}_{H.T.})_{G \text{ and } K} = \frac{1}{n}[\Sigma p_i z_i^2 - (n-1)\Sigma p_i^2 z_i^2]; \quad (33)$$

and when the variance is considered to $O(N^0)$,

$$V(\hat{Y}_{H.T.})_{G \text{ and } K} - V(\hat{Y}_{H.T.})_{Samp} = (n-1)\cdot(\Sigma p_i^2 z_i)^2 \geq 0 \quad (34)$$

Further percentage gain in efficiency of the Sampford's procedure over the Goodman and Kish procedure is

$$E = \frac{(n-1)\cdot(\Sigma p_i^2 z_i)^2}{V(\hat{Y}_{H.T.})_{G \text{ and } K}} \times 100 \quad (35)$$

E will be an increasing function of the sample size since the numerator increases and the denominator decreases as the sample size increases. Thus from (33), (34) and (35) it can be concluded that

Theorem 3: when the variance is considered to $O(N^1)$, the H.T. estimators corresponding to the Sampford's procedure and the Goodman and Kish procedure are equally efficient; and when the variance is considered to $O(N^0)$, H.T. estimator corresponding to the Sampford's procedure is always more efficient than the H.T. estimator corresponding to the Goodman and Kish procedure and the percentage gain in efficiency will be larger for larger sample sizes.

Thus, this result is a generalization of the result due to Rao (1965) wherein he compared the H.T. estimators corresponding to the Durbin's procedure and the Goodman and Kish procedure for sample size 2.

4. Numerical Illustration: The data relates to that of 35 Scottish farms, considered by Sampford (1962). In order to have an idea as to how good the approximate expressions for π_{ij} are in a given situation, the π_{ij} are calculated for the above data by using both the exact as well as the approximate expressions for samples of size 3. The variance also is evaluated using both the sets of π_{ij}. The set of probabilities π_{1j} $(j = 2, 3, \ldots, 35)$ are shown along with the approximate π_{1j} $(j = 2, 3, \ldots, 35)$ in Table 1. Variance calculated using the exact π_{ij} is $V(\hat{Y}) = 68319$, whereas the variance calculated using the approximate π_{ij} is $V(\hat{Y}) = 68341$, which suggests that in many practical situations the approximate expressions for π_{ij} given in (25) will serve the purpose quite adequately. In Table 2 are presented the variances computed to various orders for both the procedures of Sampford as well as Goodman and Kish when samples of size 4 are considered. The value computed to $O(N^2)$ represents the true variance of the customary estimator in the varying probability with replacement procedure. Values of the successive approximations suggest that the convergence is quite satisfactory even though the population size, $N = 35$, is much smaller than the sizes usually encountered in practice. For larger sample sizes, the relative difference between the two variances correct to $O(N^0)$ is however expected to be much higher than it is in this case.

Acknowledgement. This work was supported in part under OEC contract No. 0-73-6640 by the U.S. office of Education, Department of Health, Education and Welfare.

References

Durbin, J. 1967. Design of multi-stage surveys for the estimation of sampling errors. Applied Statistics 16: 152-164.

Goodman, R. and Kish, L. 1950. Controlled selection-A technique in probability sampling. Journal of the American Statistical Association 45: 350-372.

Hartley, H.O. and Rao, J.N.K. 1962. Sampling with unequal probabilities and without replacement. Annals of Mathematical Statistics 33: 350-374.

Horvitz, D.G. and Thompson, D.J. 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47: 663-685.

Rao, J.N.K. 1965. On two simple schemes of unequal probability sampling without replacement. Journal of the Indian Statistical Association 3: 173-180.

Sampford, M.R. 1962. An introduction to sampling theory. London: Oliver and Boyd, Ltd.

Sampford, M.R. 1967. On sampling without replacement with unequal probabilities of selection. Biometrika 54: 499-513.

Table 1.
Exact π_{1j}'s and the approximate π_{1j}'s of the Sampford's procedure with n = 3.

Unit No. j	Exact π_{1j}	Approximate π_{1j}	Unit No. j	Exact π_{1j}	Approximate π_{1j}
2	.000439	.000439	19	.001249	.001250
3	.000456	.000457	20	.001249	.001250
4	.000510	.000510	21	.001396	.001397
5	.000527	.000528	22	.001396	.001397
6	.000527	.000528	23	.001712	.001713
7	.000545	.000546	24	.001787	.001788
8	.000572	.000572	25	.001891	.001891
9	.000572	.000572	26	.002185	.002185
10	.000599	.000599	27	.002512	.002512
11	.000625	.000626	28	.002765	.002765
12	.000652	.000653	29	.002794	.002794
13	.000688	.000688	30	.002873	.002873
14	.000796	.000796	31	.003001	.003001
15	.000804	.000805	32	.003061	.003060
16	.000813	.000814	33	.003321	.003319
17	.000850	.000850	34	.003870	.003866
18	.000976	.000977	35	.004077	.004072

Table 2.
Approximations to $V(\hat{Y}_{H.T.})$

Order of approximation	Sampford's procedure	Goodman and Kish procedure
$O(N^2)$	55852	55852
$O(N^1)$	49321	49321
$O(N^0)$	48952	48979

EMPIRICAL DISTRIBUTIONS OF BALANCED HALF-SAMPLE REPLICATION VARIANCE ESTIMATOR AND LINEARIZATION VARIANCE ESTIMATOR

Judy A. Bean, University of Iowa

1. Introduction

The majority of sample surveys of any magnitude are complex multi-stage probability surveys. This type of survey may include 1) unequal probabilities of selection of the different sampling units in the population; 2) stratification of the units; 3) two or more stages of clustering; and 4) nonlinear estimation. The use of these complex surveys creates problems that cannot be solved by classical statistical theory. One of the problems is the estimation of the variance of the parameter estimator.

The solution three of the major survey organizations, the United States Bureau of the Census, the National Center for Health Statistics and Survey Research Center, University of Michigan, use is to calculate variance estimates from the sample data by one of the following general variance estimators: 1) balanced half-sample replication method; 2) linearization method; and 3) jackknife method. Because the mathematics is intractible, the properties of the estimators have not been derived.

Recently, Bean (1, 2) and Frankel (3) have investigated the estimators by Monte Carlo sampling from a completely specified universe. Simulation studies from an empirical distribution permit comparison of the methods under conditions of known population values, complete response from the sampling units, and no observational errors. In the present study, the empirical behavior of the balanced half-sample replication variance estimator and the linearization variance estimator was examined in terms of their histograms generated from Monte Carlo sampling of empirical data.

2. Methodology

2.1 Balanced Half-Sample Replication Method and Linearization Method.

Before describing the sample design of this study, the main features of the two variance estimators will be outlined.

Suppose we have a population of individuals grouped in clusters of households which are themselves grouped into larger clusters called primary sampling units (PSU's); then these PSU's are classified into L strata. Consider a survey whose design is two primary sampling units selected from each of the L strata with subsampling within each of the chosen PSU's. An estimate X' of the population parameter is computed from the X'_{hi}'s where X'_{hi} is the sample estimate for the i^{th} PSU in the h^{th} stratum.

To obtain an estimate of variance for X' by the replication procedure, a half-sample is created by randomly selecting one of the two PSU's from each of the L strata. Another estimate, X'_j, of the same population parameter is then made, utilizing only the data from the one half-sample. The quantity $(X'_j - X')^2$ is an estimate of variance of X'; however, this variance estimate has itself a high variance. To produce a less variable variance estimator, k such half-samples are drawn with the mean of the squared differences being the variance estimate.

$$\hat{\sigma}_{X'}^{2} = \frac{1}{k} \sum_{j=1}^{k} (X'_j - X')^2$$

The linearization method is derived from the theorem by Keyfitz (4) which states that the variance of a sum of two independent estimates of a parameter equals the expected value of the square of the difference between them. This theorem can be extended to estimates produced when the design is two PSU's selected from each of the L strata, assuming selections among the strata are independent. The method consists of estimating the stratum totals and then linearly combining them to yield an estimate of variance.

$$X' = \sum_{h=1}^{L} (X'_{h1} + X'_{h2})$$

$$\hat{\sigma}_{X'}^{2} = \sum_{h=1}^{L} (X'_{h1} - X'_{h2})^2$$

2.2 Universe.

The universe for the study consisted of morbidity data collected by the United States Health Survey (HIS) in 131,575 civilian noninstitutional individuals. A description of the survey has been published by the National Center for Health Statistics (6).

The sample design in HIS is a stratified two stage cluster sample with the first stage units (PSU's) being counties of the United States and the second units (ultimate sampling units) being clusters of 6 households within the counties. The 357 original PSU's in HIS were regrouped to form 148 PSU's; the ultimate sampling units containing less than 3 individuals were combined with similar units.

2.3 Sample Design.

The design used included unequal probabilities of selection, stratification and clustering. The 149 PSU's were classified into 19 strata, eight containing only one primary sampling unit. The first stage of sampling consisted of the independent selection of two PSU's drawn with replacement from eleven strata with probability proportional to size. The other eight strata entered the sample with a probability of one. In the second stage of sampling, the ultimate sampling units in selected PSU's were randomly subsampled with replacement.

In order to observe the behavior of the variance estimates for different sample sizes, three sample sizes were used (smaller = design I, intermediate = design II, largest = design III). Since the design called for the selection of two primary sampling units from 11 of the strata with the remaining 8 strata automatically being in the sample, the subsampling rate applied within each PSU was varied to achieve the different sample sizes. For each design 900 samples were independently drawn with a total of 2700 samples being selected.

2.4 Variables and Estimators.

In order to study the behavior of the variance estimates for different distributions, the following variables were selected: 1) family income; 2) number of restricted activity days; 3) number of physician and dental visits; 4) number of days spent in short-stay hospitals; and 5) whether or not the person has seen a physician within a 12-month period. The average of variables 1 through 4 above per person per year were estimated as was the proportion of the population seeing a physician.

For each one of the 2700 samples drawn, a ratio post-stratified estimate of each of the five characteristics was computed. The post-stratification was performed for 24 ethnic-sex-age classes (white and non-white, male and female, ages 0-4, 5-14, 15-24, 25-44, 45-64 and 65+).

$$R' = \frac{\sum_{a=1}^{A} \frac{x'_a}{y'_a} y_a}{y}$$

where

R' = the final post-stratified ratio estimate of the x^{th} characteristic,

x'_a = the simple inflated estimate of the x^{th} characteristic for the a^{th} ethnic-sex-age class,

y'_a = the simple inflated estimate of the population in the a^{th} ethnic-sex-age class,

and

y_a = the known population in the a^{th} ethnic-sex-age class.

2.5 Estimators for the Variance of a Ratio Sample Statistic.

a. Replication Method. There are three different ways of estimating the variance of a statistic by using the balanced half-sample replication method; these are described in McCarthy (5). Variances were computed by all three versions but since one of them is the average of the other two, only it will be discussed. The variance estimator is

$$\hat{\sigma}^2(R') = \tfrac{1}{2}\left[\sum_{j=1}^{k} \frac{(R'_j - R')^2}{k} + \sum_{j=1}^{k} \frac{(R^*_j - R')^2}{k}\right]$$

where

k = the number of half-samples,

R' = the final post-stratified ratio estimate,

R'_j = the post-stratified ratio estimate secured from the j^{th} half-sample,

and

R^*_j = the post-stratified ratio estimate secured from the complement half-sample (the PSU's not in the half-sample).

The full orthogonal balance pattern was used to select the PSU's that form the repeated half and complement half-samples. The pattern is presented in Bean (2). For each R' computed, an estimate of variance was calculated using the above equation.

b. Linearization Method.

The estimator R' can be written as

$$R' = \frac{\dfrac{\sum_{h=1}^{L}(x'_{ah1} + x'_{ah2})}{\sum_{h=1}^{L}(y'_{ah1} + y'_{ah2})} y_a}{y} +$$

where

x'_{ahi} = the inflated estimate of the x^{th} characteristic for class a of the i^{th} PSU in the h^{th} stratum.

and

y'_{ahi} = the inflated estimate of the population in class a of the i^{th} PSU in the h^{th} stratum.

Then the linearization variance estimator of R' is

$$\hat{\sigma}^2(R') = \frac{\sum_{h=1}^{L}\left[(x'_{h1} - x'_{h2}) - \sum_{a=1}^{A} \frac{x''_a}{y_a}(y'_{ah1} - y'_{ah2})\right]^2}{y^2}$$

where

$$x''_a = \sum_{a=1}^{A} \frac{x'_a}{y'_a} y_a$$

For each of the 8 stratum consisting of only one PSU, two pseudo-PSU's were created from the sampled segments. The details are given in Bean (2). Again for each R' estimate, a variance estimate was produced from the above equation.

3. Summary of Empirical Work

Because of the volume of the primary estimates and variance estimates generated, only the results for the largest sample size, design III, are presented. In Table 1 can be seen the average of the 900 statistics calculated along with the mean variance estimate of the replication method and of the linearization method. The two means are very similar for all the variables studied.

Tables 2 and 3 present the frequency and the cumulative percent for the variables restricted activity days and proportion seeing a physician. The results are nearly the same.

Figures 1 through 5 show the histograms of the variance estimates produced by the two techniques. As can be seen, the distributions are very much alike between methods. Notice the distributions are skewed with bunching below the mean and a tail above; this skewness decreased as the sample size increased. Bean (2) compared the estimates in terms of their means, sampling variances and biases. Each estimator has a small bias.

Because the balanced half-sample replication method and the linearization method provide similar results, either may be used to calculate variance estimates for data gathered in a complex multistage probability sample survey. The validity of applying either variance estimator method to this type of survey has been reported on in detail elsewhere [Bean (1,2) and Frankel(3)]. The studies showed that the variance estimates produced by the methods are satisfactory.

ACKNOWLEDGMENT

The author wishes to thank the National Center for Health Statistics for providing the data and to thank Mr. Walt R. Simmons for help and encouragement received in doing the research.

REFERENCES

1. Bean, Judy A. "Behavior of Replication and Linearization Variance Estimators for Complex Multistage Probability Samples." University of Texas, 1973. (unpublished dissertation).

2. Bean, Judy A. "Distribution and Properties of Variance Estimators for Complex Multistage

Probability Samples." Vital and Health Statistics. (In press).

3. Frankel, Martin R. Inference to Survey Samples. An Empirical Investigation. Ann Arbor: University of Michigan, 1971.

4. Keyfitz, Nathan. "Estimates of Sampling Variance where Two Units are Selected from each Stratum." Journal of the American Statistical Association, 52, (1957), 503-510.

5. McCarthy, Philip J. "Replication: An Approach to the Analysis of Data from Sample Surveys." Vital and Health Statistics. PHS Publication No. 1000, Series 2, No. 14. Washington, D. C.: Government Printing Office, 1966.

6. National Health Survey. "The Statistical Design of the Health Household - Interview." Health Statistics. PHS Publication No. 584-A2. Washington, D. C.: Government Printing Office, 1958.

TABLE 3. The frequency and cumulative per cent of 900 variance estimates as calculated by the balanced half-sample replication variance estimator and by the linearization variance estimator for the variable proportion seeing a physician within the last 12 months for Design III.

Variance Estimate (var. times $x10^{-4}$)	Replication Frequency	Replication Cumulative Per Cent	Linearization Frequency	Linearization Cumulative Per Cent
0.00-0.10	0	0.00	0	0.00
0.11-0.20	1	0.11	2	0.22
0.21-0.30	9	1.11	10	1.33
0.31-0.40	35	5.00	43	6.11
0.41-0.50	75	13.33	79	14.89
0.51-0.60	110	25.56	114	27.56
0.61-0.70	141	41.22	147	43.89
0.71-0.80	123	54.89	124	57.67
0.81-0.90	99	65.89	93	68.00
0.91-1.00	69	73.56	69	75.67
1.01-1.10	73	81.67	53	81.56
1.11-1.20	42	86.33	50	87.11
1.21-1.30	33	90.00	37	91.22
1.31-1.40	39	94.33	31	94.67
1.41-1.50	13	95.78	15	96.33
1.51-1.60	11	97.00	8	97.22
1.61-1.70	8	97.89	6	97.89
1.71-1.80	7	98.67	9	98.89
1.81-1.90	4	99.11	4	99.33
1.91-2.00	7	99.89	2	99.56
2.01-2.10	1	100.00	2	99.78
2.11-2.20	0	100.00	2	100.00

TABLE 1. A summary of the values obtained by repeated sampling for Design III.

Value	Family Income	Restricted Activity Days	Physician and Dental Visits	Short-Stay Hospital Days	Proportion Seeing Physician
Average Sample Estimate - R'	8392.80	14.6595	4.6548	1.0597	0.6840
Average Variance Estimate .. Replication	26554.80	0.9206	0.0178	7.0842×10^{-3}	8.3010×10^{-5}
Average Variance Estimate .. Linearization	26174.70	0.8915	0.0175	6.7380×10^{-3}	8.1135×10^{-5}

TABLE 2. The frequency and cumulative per cent of 900 variance estimates as calculated by the balanced half-sample replication variance estimator and by the linearization variance estimator for the variable average number of restricted activity days per person per year for Design III.

Variance Estimate	Replication Frequency	Replication Cumulative Per Cent	Linearization Frequency	Linearization Cumulative Per Cent
0.00-0.20	0	0.00	0	0.00
0.21-0.40	37	4.11	48	5.33
0.41-0.60	131	18.67	136	20.44
0.61-0.80	225	43.67	225	45.44
0.80-1.00	190	64.79	207	68.44
1.01-1.20	141	80.44	131	83.00
1.21-1.40	80	89.33	64	90.11
1.41-1.60	46	94.44	52	95.89
1.61-1.80	26	97.33	20	98.11
1.81-2.00	14	98.89	10	99.22
2.01-2.20	6	99.56	4	99.67
2.21-2.40	1	99.67	1	99.78
2.41-2.60	2	99.89	1	99.89
2.61-2.80	0	99.89	0	99.89
2.81-3.00	1	100.00	1	100.00

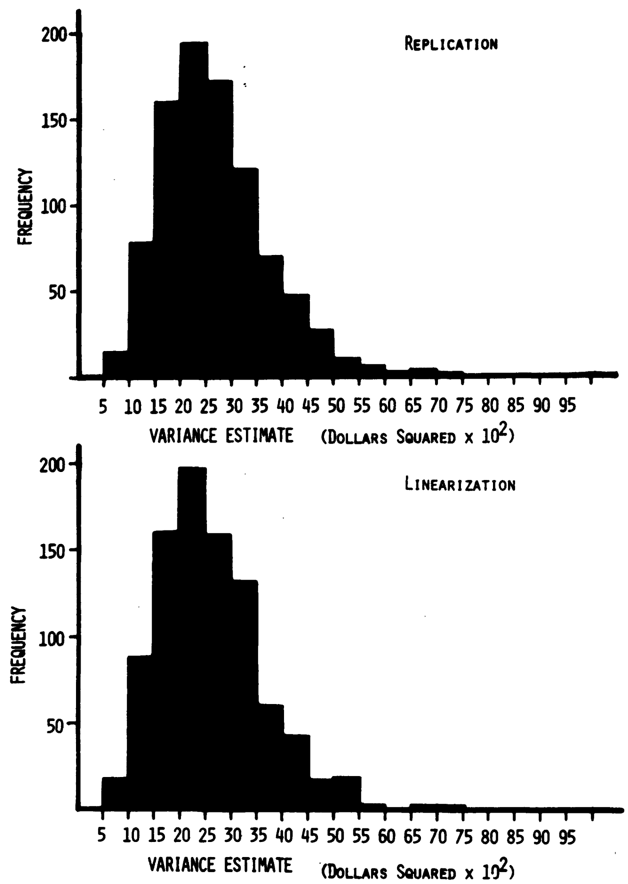


FIGURE 1: HISTOGRAMS OF 900 VARIANCE ESTIMATES AS CALCULATED BY THE BALANCED HALF-SAMPLE REPLICATION VARIANCE ESTIMATOR AND BY THE LINEARIZATION VARIANCE ESTIMATOR FOR THE POPULATION ESTIMATE AVERAGE INCOME PER PERSON FOR DESIGN III.

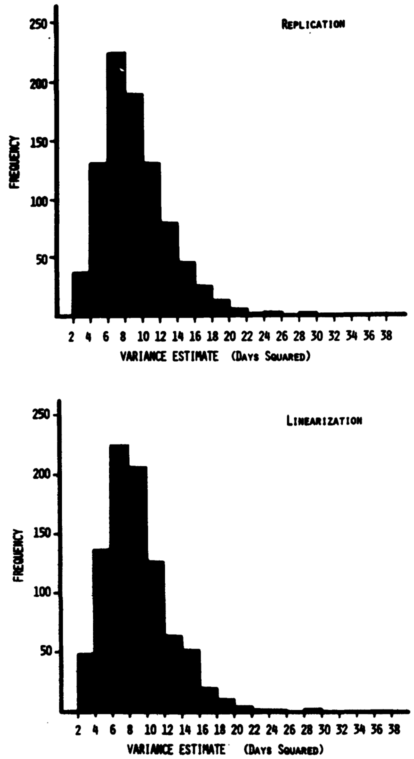


FIGURE 2: Histograms of 900 variance estimates as calculated by the balanced half-sample replication variance estimator and by the linearization variance estimator for the population estimate the average number of restricted activity days per person per year for Design III.

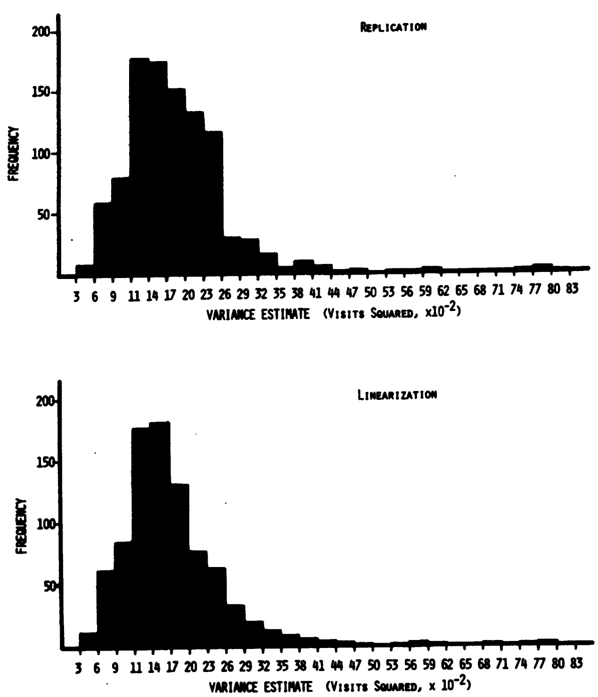


FIGURE 3: Histograms of 900 variance estimates as calculated by the balanced half-sample replication variance estimator and by the linearization variance estimator for the population estimate average number of visits per person per year for Design III.

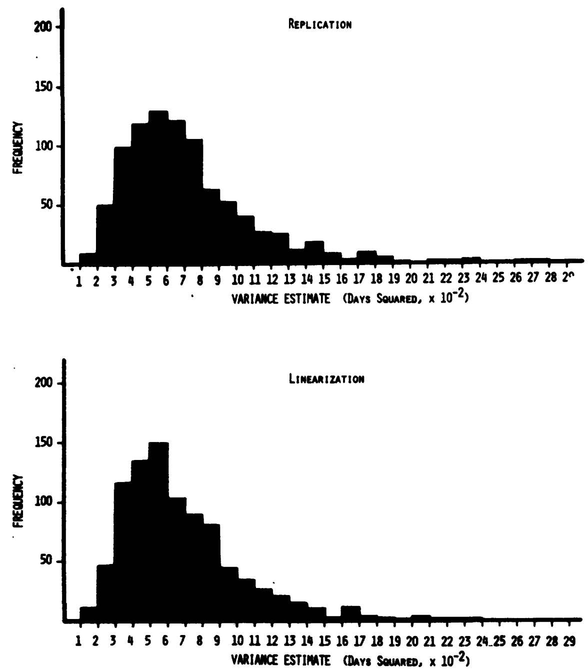


FIGURE 4: Histograms of 900 variance estimates as calculated by the balanced half-sample replication variance estimator and by the linearization variance estimator for the population estimate average number of short-stay hospital days per person per year for Design III.

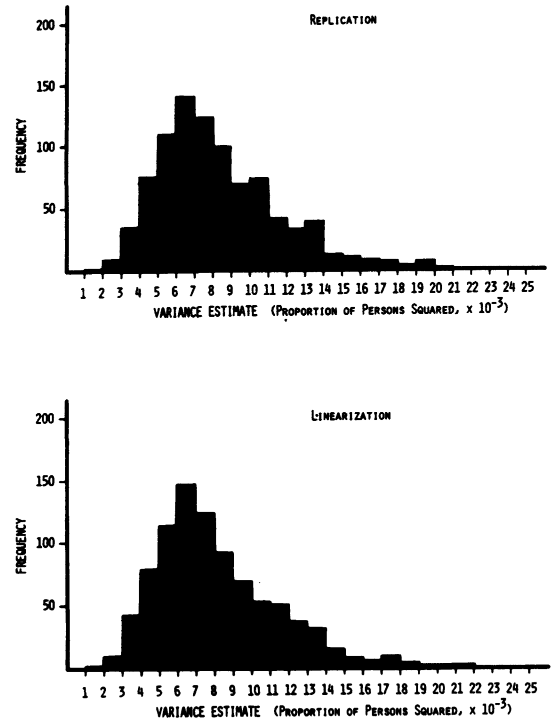


FIGURE 5: Histograms of 900 variance estimates as calculated by the balanced half-sample replication variance estimator and by the linearization variance estimator for the population estimate proportion of population seeing a physician in last 12 months for Design III.

MAXIMIZING ADDITIVITY OF SCALE VALUES

Purnell H. Benson, Rutgers University Graduate School of Business

In research to optimize product acceptability from the standpoint of consumer satisfaction, it is sometimes necessary for the measurements of preference to possess the property of additivity. The additivity is a necessary condition of optimization analysis if preferences for articles are separately measured under different conditions and the summation of these preferences maximized.

The problem of additivity of scale values is illustrated by figure 1. In this diagram OA, OB, OC and OD represent measurements of the degree of consumer desire for stimuli A, B, C and D which we will say are apples, bananas, cherries and dates. In the situation where the consumer is indifferent to whether he has apples and bananas, or has cherries, assuming these articles are independently consumed, measurement of desire for cherries should equal the sum of the measurements of desires for apples and for bananas. That is, OC should equal OA plus OB. In the diagram it is apparent that this relationship does not hold, since OC is smaller in length than the sum of the distances OA plus OB. Similarly, if the consumer is indifferent to whether he has bananas and cherries or has dates, OB plus OC should equal OD.

However, a transformation of the form $S' = \emptyset(S)$ which shifts the origin and alters the relationships of proportionality among the scale values, represented by the curve, changes the scale values into O'A', O'B', O'C' and O'D' for which the relationship of additivity does hold for the relationships of indifference specified. A scale form for whose values the relationship of additivity holds reflects the natural way in the personality system of the consumer whereby separate desires combine to determine choice.

Correction of Relationships of Proportionality Among Scale Values

L. L. Thurstone and L. V. Jones [2] have published a procedure for achieving additivity which they use to correct scale values for the zero point of the scale. In their procedure they use stretching factors expected from the error variance in comparing single-single, single-double or double-double stimuli. The zero point is then taken so as to minimize errors in applying the relationship of addition to stimuli scaled separately and in doublets.

In the diagram origin O" is located so as to minimize errors in applying the relationships of addition O"A + O"B = O"C and O"B + O"C = O"D when the same interval OO" is subtracted from all of the scale values. It is apparent from the diagram that this is not the full correction which is possible. From an algebraic standpoint it is a correction only for the first term in a series of powers to define new scale values S' from the initial scale values S in order to minimize errors in applying the relationship of addition to stimuli for which desires are equivalent. Beyond correcting for the zero point and using three stretching factors for adjusting deviates from comparing single-single, single-double and double-double stimuli, a functional transformation can be applied to improve additivity of scale values further.

We seek a transformation of scale values S into new scale values S' which maximize their additivity, that is, which minimize the error variance in applying the relationship of additivity. The expression to be minimized is

$$\frac{\sum_{m} [S_{ij}' - (S_i'+S_j')]^2 / m}{[\sum_{m+n} S_{ij,i,j}'^2 - (\sum_{m+n} S_{ij,i,j}')^2/(m+n)]/(m+n)} \qquad (1)$$

The numerator represents the error variance in applying the relationship of addition to the $\underline{m}$ pairs of scale values S_i' and S_j' for single stimuli in order to equal scale values S_{ij}' for double stimuli. This is divided by the scale variance of the $\underline{n}$ scale values for single stimuli and the $\underline{m}$ scale values for double stimuli. The denominator is needed in order to consider error variance in relation to the numerical spread in the scale values to which the relationship of addition is applied. An alternative procedure is to minimize the numerator subject to the constraint that the denominator is constant.

We seek constants in a functional transformation $S' = \emptyset(S)$ which will minimize expression (1). The scale values to be transformed are the original unweighted paired comparison scale values before applying stretching factors to the deviates. A more highly desired single stimulus with multiple uses is to be given the same mathematical treatment as a multiple stimulus. In either case the size of the discriminal error in making comparisons of stimuli would be similarly affected, and a single functional transformation without use of stretching factors seems sufficient.

Maximizing Additivity of Scale Values From the Thurstone-Jones Data

A linear-cubic function was selected for accomplishing the transformation of initial scale values into ones which are additive, the function having the following form.

$$S' = (S + a) + b\,(S + a)^3 . \qquad (2)$$

More than two power terms were not introduced into the analysis since there is question of whether the data are adequate for more detailed treatment. A cubic, rather than a squared, term was used since this produced a symmetric transformation for positive or negative scale values. The way in which the constant $\underline{a}$ is employed gives a mathematically meaningful adjustment for the zero point in the initial scale values distinct from altering the proportional relationships of scale values to each other. An additional constant would be required as coeffi-

TABLE 1

TRANSFORMATIONS OF PAIRED COMPARISON SCALE VALUES FROM THURSTONE-JONES DATA

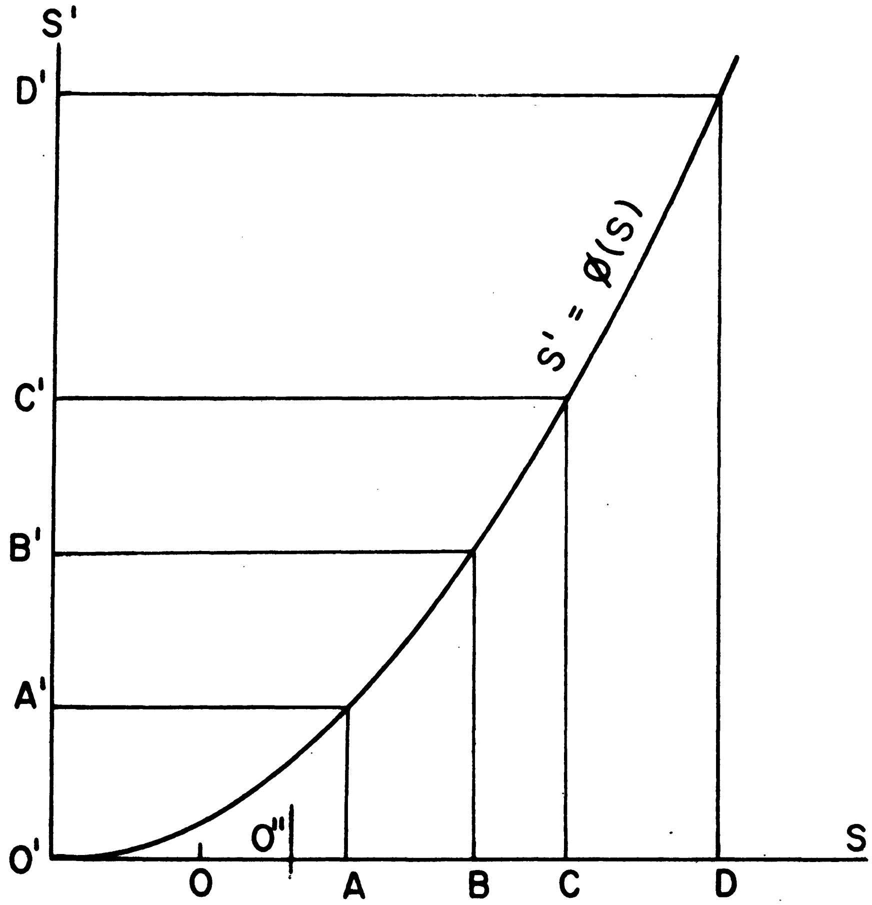
Stimulus	Treatment of Scale Values			
	(I) No corrections applied to scale values	(II) Corrected for zero point, normal deviates unweighted	(III) Corrected by weighting normal deviates and altering zero point	(IV) Scale Values transformed to maximize additivity
A. Briefcase	-1.04	.30	.62	.69
B. Dictionary	-1.06	.28	.61	.66
C. Record Player	.39	1.73	2.81	2.97
D. Desk lamp	- .63	.71	1.26	1.19
E. Pen & pencil set	- .34	1.00	1.68	1.60
A. & B.	- .42	.92	1.48	1.48
A. & C.	.76	2.10	3.72	3.89
A. & D.	- .14	1.20	2.02	1.92
A. & E.	.06	1.40	2.40	2.28
B. & C.	.73	2.07	3.65	3.81
B. & D.	- .27	1.07	1.76	1.71
B. & E.	- .09	1.25	2.12	2.01
C. & D.	.94	2.28	4.07	4.41
C. & E.	.93	2.27	4.10	4.38
D. & E	.18	1.52	2.63	2.52

TABLE 2

PRECISION OF TRANSFORMATIONS OF PAIRED COMPARISON SCALE VALUES

Statistical measure	Treatment of Scale Values: (I) No corrections applied to scale values	(II) Corrected for zero point, normal deviates unweighted	(III) Corrected by weighting normal deviates and altering zero point	(IV) Scale Values transformed to maximize additivity
Standard error in adding scale values, $SE[S_{ij} - (S_i + S_j)]$	1.357	.212	.228	.191
Standard error of scale values, $SE(S_{ij,i,j})$	.783	.123	.132	.110
Standard deviation of scale values, $SD(S_{ij,i,j})$	.635	.635	1.120	1.217
Precision of scale values, $SE(S_{ij,i,j})$ / SD	1.23	.19	.12	.09

FIGURE 1



cient for the linear, as well as the cubic, term if the procedure were followed of minimizing the numerator of (1) under the constraint of constant variance in the scale values. This alternative procedure increases somewhat the complexity of the non-linear system of equations to be solved.

Substituting (2) in (1) and minimizing with respect to a and b gives the two stationary equations whose solution defines a and b.

The original unweighted paired comparison scale values, not given by Thurstone and Jones, were computed from the unweighted normal deviates in their table 2 [2, p. 464]. These scale values, recorded in column I of table 1 of the present paper, were adjusted before furtheranalysis by starting the lowest value at 0 to simplify computation of powers of values. From these adjusted scale values were computed a long series of terms up to the 6th power required for solving the stationery equations. The details of the algebra are not reproduced here. A simpler route of calculation may be feasible than that which was followed here. After substituting these figures in the two partial derivative equations, the solution of the equations was obtained uniquely yielding $a = .64$ and $b = .096$. Since the initial values in column I of the table were increased by 1.06 before computation of the transformation, the value for a applied to these values becomes 1.70.

Relative Precision of Scale Forms

Table 1 records several sets of scale values computed from the Thurstone-Jones data by alternative procedures. These scale forms show the effect of successive types of correction. Column I shows paired comparison values computed from unweighted normal deviates. Column II shows these same values together. Column III records scale values computed by Thurstone and Jones after weighting normal deviates with three stretching factors and adjusting for the natural zero point. The normal deviates were weighted by them to allow for the expected increase in the discriminal dispersion in comparing multiple stimuli. In column IV are scale values computed from the unweighted values in column I by the transformation

$$S' = (S + 1.70) + .096(S + 1.70)^3 . \qquad (5)$$

Table 2 gives the standard errors resulting in adding values, the standard errors estimated for the values themselves, the standard deviations of values, and the relative precisions of the sets of scale values in table 1. The error variance in applying the relationship of addition to scale values is equal to the sum of the variances of the errors of the three values in the relationship of addition. The standard error of individual scale values is then obtained by dividing the standard error in adding values by $\gamma 3$. The precision of the scale can then be defined as the ratio of the standard error of scale values divided by the standard deviation of the scale values.

As evidenced by their additivity, the precisions of the four scales listed in table 2 are 1.23, .19, .12, and .09. The unweighted scale values unadjusted for a natural zero point afford the poorest values for satisfying the criterion of additivity. Scale values adjusted for their natural zero point by the addition-of-stimuli-procedure of Thurstone and Jones but without adjustment by stretching factors show marked improvement in precision. The precision is sharpened by use of stretching factors.

The improvement in precision is greater if scale values are transformed to maximize additivity. The F level for addition of the cubic term is 4.91. With 7 and 1 degrees of freedom, this falls between a 10% level of significance, 3.59, and a 5% level of significance, 5.59.

Further Comments

A theoretical point at issue is whether a scale metric based upon the standard deviation of the discriminal dispersion is sufficient to schieve additivity of scale values after correction for the natural zero point and adjustments for comparing single-single, single-double and double-double stimuli have been made. The analysis given here indicates that improved correction is obtained by a continuous functional transformation to maximize additivity.

The conclusion that the raw discriminal dispersion does not produce scale values whose proportionality after zero point correction assures additivity of scale values is not unexpected. In scaling sensory intensities rather than intensities of desire, the addition of scale values does not conform to the addition of physical stimulus intensities together. Instead, equal increments in physical stimulus produce diminishing increments in the sensory stimulus experienced.

Consideration of consumer purchasing indicates a similar result. The standard error in choosing between two similar articles worth $101 and $102 is much larger than in choosing between articles worth $1 and $2 to the individual judge. A previous paper by the author and John H. Platten, Jr., indicated that scale values based upon the discriminal dispersion of preference judgments require a Fechner-type correction in order for the relationship of additivity to be satisfied [1].

References

[1] Benson, Purnell H., and Platten, John H., Jr., "Preference Measurement by the Methods of Successive Intervals and Monetary Estimates," Journal of Applied Psychology, Vol. 40 (1956), pp. 412-414.

[2] Thurstone, Louis L. and Jones, Lyle V., "The Rational Origin for Measuring Subjective Values," Journal American Statistical Association, Vol. 52 (1957), pp. 458-471.

TRANSFORMING CATEGORY SCALE DATA TO EQUAL INTERVALS FOR A HEALTH INDEX[1]

W.R. Blischke, University of Southern California
J.W. Bush and R.M. Kaplan, University of California, San Diego

The persistent need for a comprehensive social indicator of health for planning and evaluation has led health services researchers, despite the well-known difficulties [Sullivan 1966; Torrance 1973], to again attack the problems of creating a health status index [Berg 1973; Goldsmith 1973]. Bush and his colleagues have undertaken a series of studies to define the problems more precisely and to propose approaches that avoid some of the earlier criticisms [Fanshel and Bush 1970; Bush et al., 1972; Patrick et al., 1973a, b; Berry and Bush 1974; Chen, et al. 1975]. The present proposal resolves another problem in index construction--the creation of an equal interval measure of social preference for states and levels of function. Before presenting the problem in detail, however, the next section will outline the general conceptual framework for defining health status.

THE HEALTH INDEX

The social construct "health" is composed of two distinct components: Level of Well-being and prognosis. Level of Well-being refers to the measured social preference or weight assigned to a person's level of functioning at some point in time. Prognoses are the probabilities of transition to other levels of functioning at subsequent times. Treating these components as analytically distinct separates two frequently confused aspects of health status and permits the separate measurement and quantitative expression of the two variables.

Prognoses or transitional probabilities are matters for empirical determination in follow-up studies of different patient and population groups. Thus, no precise statement of health status can be made for an individual or a group without knowledge of the expected transitions among the function levels over time. Since function level and prognoses vary independently for different individuals and populations, we shall reserve the term "health" for some joint or composite expression of current Level of Well-being and prognosis.

The present report concerns the value dimension of health--the preference or Level of Well-being that society assigns to levels of function on a continuum from death (0.0) to optimum function (1.0). When these values have been measured, health status can be expressed statistically as the expected value (product) of the preferences associated with the levels of function and the probabilities of transition among the levels over a defined standard life [Bush, et al. 1972; Chen, et al. 1975], as follows:

$$Q = \sum_{j=1}^{30} W_j Y_j$$

where j is the index for the function levels [$j = 1, 2, \ldots, 30$],
Q is the quality-adjusted or weighted life expectancy,
W_j are the function weights or measured social preferences associated with each function level, and
Y_j are the expected durations in each function level computed from the transition probabilities [Bush, et al. 1971].

Measuring both prognoses and preferences requires operational definitions of the function levels [Patrick, et al. 1973a]. Abstracts of several hundred medical case descriptions revealed the spectrum of disturbances that diseases and disabilities can cause in role performance. Several well-known survey instruments provide items that span the range of disturbances in function status. Three ordinal rankings--Mobility, Physical Activity and Social Activity--organize the items into mutually exclusive and collectively exhaustive scales. Omitting the rare or impossible, at least 30 combinations of the scale steps exist that can be referred to as Function Levels [Table 3]. Available survey instruments will classify individuals into one and only one of the Function Levels [Bush, et al. 1974].

An independent set of 42 symptom/problem complexes comprise the specific disturbances that cause dysfunction. To compute the Levels of Well-being for the index, human judges must rate a series of cases each comprised of a function level plus a symptom/problem complex. Because subjects must rate a large number of cases to adequately sample the function status domain, a simple and efficient method is necessary for laboratory and survey research. Previous research indicated that category rating, a partition method in which subjective differences between stimuli are assessed via a numbered category, is more reliable and gives values equivalent to methods that generate ratio scales and imply social choice [Patrick, et al. 1973b; Kaplan and Bush 1974]. Health index construction requires that a large number of case descriptions be rated by a method that is simple enough for household interview surveys. Thus, complex or time-consuming methods such as Von Neumann-Morgenstern or paired comparisons are impractical for field use.

Patrick, Bush and Chen [1973a] describe in detail the experiment that provided the data set for the present analysis. Thirty-one panelists rated each of 400 case descriptions (items), selected randomly to represent the function status domain, on a fifteen-point category scale. Each standardized case description included an age group, three scale steps composing a function level, and one of 42 symptom/problem complexes. A total of 12,000 observations were available for analysis.

THE MEASUREMENT PROBLEM

Although unnecessary for most statistical hypothesis testing, even exponents of "weak" measurement models agree that estimating "true" scale locations and computing expected values requires measurements with interval properties [Baker, et al. 1966]. Although a significant body of literature contends that complex case descriptions can be rated on interval scales [Anderson 1974; Stone 1970], Stevens and others have argued strongly that category data do not possess metric (interval) properties [1966]. Since category scaling is so useful in the field, more extensive tests and procedures to transform the data to assure equal intervals became desirable.

Thurstone originally developed the method of successive intervals to obtain interval measures from ordered category data. Based on many of the same assumptions as paired comparisons, successive intervals can be considered an extension of Fechner's method of constant stimuli. Among the existing computational procedures, the most common are graphical [Jones and Thurstone 1955], least squares [Gulliksen 1954; Tucker 1964], and maximum likelihood estimation [Schonemann and Tucker 1967; Ramsey 1973]. Maximum likelihood is the most efficient procedure for estimating the item parameters and the interval widths, but the equations are highly nonlinear and numerical techniques are required for their solution. Therefore we used Edwards' least squares method [1956]. The method that we propose for estimating the widths of the end intervals and for data transformation apply no matter how the interval widths are estimated.

ASSUMPTIONS OF SUCCESSIVE INTERVALS

If N items are to be scored on an integer scale from 1 to n, the method of successive intervals assumes 1) that an unknown and unobservable psychological continuum underlies each subject's scoring, 2) that the underlying continuous random process that determines a score for say, the ith item scored by a particular subject, follows a normal probability distribution with mean μ_i and variance σ_i^2, 3) that the recorded score is the nearest integer to the score-value resulting from the continuous process, with category 1 representing any number less than 1.5, category 2 any number in the interval (1.5, 2.5), ... and category n any number in the interval (n-.5, ∞) and 4) that the normal distribution with parameters μ_i and σ_i^2 determines the probabilities assigned to the intervals (-∞, 1.5), (1.5, 2.5), and so on.

Subsequently we shall modify the implicit assumption that both the end intervals are conceptually infinite, effectively replacing the normal by a truncated distribution. Comparing and aggregating items to compute function level values requires estimating μ_i (the "scale value") for each item. The new procedure given below adjusts for inequality in the intervals before it estimates the μ_i's. When items have values near the scale extremes, successive intervals assigns substantial probability to the region beyond the end of the measurement scale.

Constriction of the scale skews the distribution of observed integer scores. This curtailing of the distribution of stimuli near either extreme is a well-known property of category scales [Torgerson 1958; p. 74]. At the upper extreme, for example, we would expect to observe mostly scores of n with some (n-1)'s and perhaps a few (n-2)'s. This skewness occurs even though the underlying preference continuum is normal.

ESTIMATING WIDTHS OF INTERIOR INTERVALS

Since normality assumes infinitely wide end categories, all successive interval methods provide width estimates for interior intervals only, that is, for the intervals (1.5, 2.5), (2.5, 3.5), ..., (n-1.5, n-.5), but not for the intervals (-∞, 1.5) and (n-.5, ∞). The basic data used for estimating are the proportion of responses (about 30 in this study) that are 1's, 2's, ..., n's for each of the N items.

We shall denote these proportions P_{ij}, where i identifies the item and j the response category or score. Each such proportion is an estimate of the probability assigned to score j by the normal distribution associated with item i. This probability depends upon μ_i and σ_i, the parameters of the normal distribution associated with the ith item. Thus, if the parameters for each item are known, the widths of the intervals can be calculated in units of the σ_i. For example, about 2/3 of the observations should be within σ_i of μ_i, and so forth.

By standardizing the normal distribution associated with each item, successive intervals estimates the interval widths without explicitly estimating the μ_i and σ_i. A table of the cumulative standard normal distribution gives the standard normal deviate, z_{ij}, corresponding to each cumulative sum, $P_{ij} = \sum_{\alpha=1}^{j} p_{i\alpha}$. The difference between successive z_{ij}'s is then an estimate, in standard units, of the width, w_{ij}, of the jth category for the ith item. The overall estimate of this interval width is then calculated as an average across all items of the w_{ij}, namely

$$W_j = \frac{1}{N}\sum_{i=1}^{N} w_{ij}.$$

The procedure omits all P_{ij}'s less than .02 or greater than .98 since the proportions in the tails are poorly estimated and slight sample fluctuations can disproportionately influence the final result. The z_{ij}'s are recorded only for the remaining P_{ij}'s and the final results are means of the corresponding remaining w_{ij}'s. The number of estimates for the final computation depends on the number of non-zero w_{ij}'s. As noted above, the procedure estimates only the n-2 interior category widths, 2, 3, ..., n-1, that is, the intervals from 1.5 to 2.5, 2.5 to 3.5, and so forth.

To estimate the category widths in the Health Index study, computations are required for j = 1,

..., 15 and i = 1, ..., 400, all with roughly equal variances. As examples the w_{ij}'s for five illustrative items were caluclated and these values are displayed in Table 1A. In the analysis the end P_{ij}'s are dropped and the interval width estimates (w_{ij}) are calculated as differences between successive z_{ij}'s. Since the normal distribution sets $Z_{i,15}$ to infinity, only 14z_{ij}'s exist for each i. Thus only 13 differences, w_{ij}, are available to estimate n-2 = 13 interval widths. An average over all nonzero w_{ij}'s (of the total N = 400) produces the estimate of category widths. The number of nonzero w_{ij}'s and the estimates for the thirteen intervals are given in Table 1B. As expected, the interval widths are quite similar in the middle of the scale but increase toward both ends [Guilford 1954; Torgerson 1958].

TABLE 1: COMPUTATION OF MEAN CATEGORY WIDTHS USING SUCCESSIVE INTERVALS ANALYSIS, SHOWING FIVE EXAMPLES

	CATEGORY (INTERVAL) NUMBER (j)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Item (i)				A. Interval Width Estimates (w_{ij})											
1		.000	.000	.000	.000	.000	.000	.000	.000	.331	.000	.387	.671	1.021	
2		.000	.000	.000	.331	.218	.311	.340	.189	.088	.332	.412	.088	1.389	
3		.000	.000	.000	.331	.000	.529	.237	.200	.180	.251	.243	.631	.548	
4		.000	.000	.549	.000	.311	.437	.180	.251	.081	.162	.251	.493	.654	
5		.759	.000	.169	.218	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		B. Mean Estimates of Interior Category Widths (400 Items)													
Number of Nonzero w_{ij}'s		59	130	233	286	323	344	356	349	334	293	204	126	44	
Estimated Width (in Z units)		.582	.492	.413	.436	.392	.379	.384	.353	.367	.409	.433	.521	.858	

TESTING FOR EQUALITY OF INTERVAL WIDTHS

Since each w_{ij} serves as an estimate of an interval width, these values are used to form a simple *F*-test, based on a within-and-among groups analysis of variance, with the hypothesis that average interval widths are equal. Although the analysis could be formulated as a two-way cross classification with intervals as "treatments" and items as "replicates," the procedure "adjusts" for item differences by transforming to the standard normal distribution, so we performed the analysis as a nested design.

Although the assumptions of normality and independence are violated, a very large or a very small *F*-value will give some information regarding the equality of interval widths. If the *F*-value is large the hypothesis of equality should be rejected as usual; if it is very small, the hypothesis should not be rejected. The test is logical, and at least some gross judgments are possible even though the violations destroy any ability to make exact probability statements. The approximate ANOVA test yielded a large *F*-value [*F* (12/3068) = 34.64] which provides evidence for rejecting equality.

We do not compare this *F* with a tabulated value because of lack of independence, which invalidates even non-parametric procedures. Since lack of independence effectively reduces the degrees of freedom for experimental error, it is sometimes possible, depending on the correlation structure, to adjust degrees of freedom downward. That analysis did not seem necessary since the observed *F* would be significant under nearly any such reduction. Even if reduced by a factor of five, for example, the *F* would be 6.93 with 12 and 614 degrees of freedom. The tabulated value with 12 and 120 degrees of freedom, at the .0005 level is 3.22. We shall proceed under the assumption that the interval widths are unequal.

ESTIMATING THE WIDTHS OF END INTERVALS

The most popular procedure for estimating scale values from category data is to multiply the rank order of each category by the frequency of responses for that category and to obtain a mean for these products. A second possibility is to calculate the median of the frequency distribution. Both procedures require the category widths to be equal.

The categories near either end of the scale most consistently violate the assumption, especially when they are unbounded. Even with finite end categories, the response processes outlined previously would produce intervals wider at the end than in the interior. The usual successive intervals analysis requires an alternative method, however, to extrapolate this process to the end categories.

Using the category number 2 through n-1 as the predictor variables, a polynomial regression will estimate the widths of the end intervals from the previously estimated widths of the interior intervals. Since the widths are smallest toward the center of the range and increase toward either end, the distribution will require at least a second degree polynomial. To allow for possible asymmetry, we increased the degree of the polynomial to four.

Using the category numbers and the widths given in Table 1B as the predictor and response variables, a regression analysis was performed which compared a model of a given degree with the model which is one degree lower. This analysis suggested that at least a cubic model was required. All $\underline{F}$'s were significant at the 1% level ($\underline{F}_{.99}$ = 11.3 with 1/8 df.), but we cannot interpret the results strictly because the assumptions are not met. On the other hand, the plot of the observed and predicted interval widths given in Figure 1 shows that the fit is quite good. Accepting the fourth degree model as adequate, the estimated polynomial is

$$Y = 1.1362 - .4268X + .0932X^2 - .0089X^3 + .0003X^4,$$

where X is the predictor variable (j), and
Y is the response variable (w_{ij}).

This equation estimates widths for the end categories (X = 1 and X = 15) as .794 and 1.227, respectively.

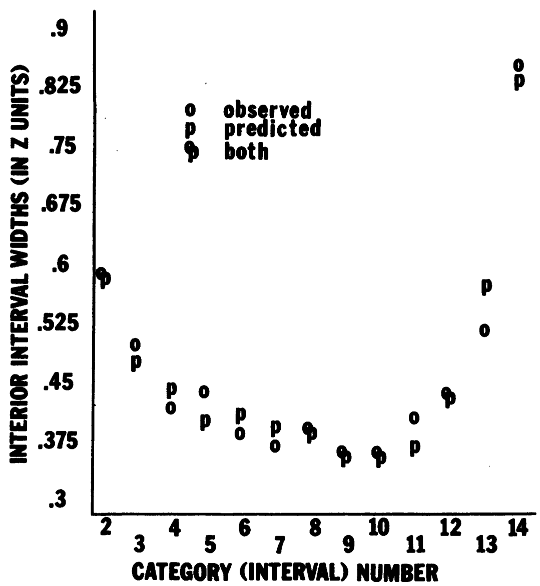


Figure 1. Observed (calculated) mean interval widths (w_{ij}) compared to widths predicted using a fourth degree polynomial fit by category number.

TRANSFORMING FUNCTION LEVEL PREFERENCES

For health index research, it is convenient to transform the data to a scale of 0 (for death) to 1.0 (for well), since data taken originally on different scales are directly comparable after transformation. If subjects are instructed that "death" is at the lower bound of the lowest category, the end-points of the scale will be located at 1-.5=.5 and n+.5. The proposed transformation therefore maps .5 to 0 and n+.5 to 1. An equal interval scale then maps 1 to (1/2n), 2 to (1/2n + 1/n), and so forth. Values that correct for the unequal interval widths will replace these values.

Broadening the middle intervals and shortening the end ones compensates for the unequal intervals. Making the new intervals proportional to the reciprocals of the original provides an approach simpler than other methods. The sum of the reciprocals taken as the proportionality constant again transforms the values to the total scale interval (0,1). The transformed widths are given by

$$w_j' = \frac{\frac{1}{w_j}}{\sum_{i=1}^{n}\left(\frac{1}{w_i}\right)}$$

where w_j is the estimated and w_j' is the transformed width, all in z units.

To use these results for transforming item values, the values to be assigned to the integers 1,..., n must be determined. Since the endpoint of the original scale was .5, the distance to 1 will be one-half the first interval width, or $.5w_1'$. Similarly, the distance from 1 to 2 will be $.5(w_1' + w_2')$. Thus the integer 2 is transformed to $.5w_1' + .5(w_2') = w_1' + .5w_2'$, and so forth. The results are then used to estimate scale values for each item by calculating means or medians in the usual way.

Table 2 gives w_j's, their reciprocals, the transformed w_j's, and the transformation set for mapping the original category numbers. The transformation maps 1 into 0.0199, 2 into 0.0668, etc.

TABLE 2: VALUES FOR TRANSFORMED CATEGORY DATA

Category (j)	w_j	$1/w_j$	w_j'	Transformed value
1	.794	1.259	.0397	.0199
2	.582	1.718	.0541	.0668
3	.492	2.033	.0641	.1259
4	.413	2.421	.0763	.1961
5	.436	2.294	.0723	.2704
6	.392	2.551	.0804	.3467
7	.379	2.639	.0832	.4285
8	.384	2.604	.0821	.5112
9	.353	2.833	.0893	.5969
10	.367	2.725	.0859	.6845
11	.409	2.445	.0771	.7660
12	.433	2.309	.0728	.8409
13	.521	1.919	.0605	.9076
14	.858	1.166	.0367	.9562
15	1.227	0.815	.0257	.9874
TOTAL		31.731	1.0002	

Table 3 displays the mean value of items in each Function Level, averaged over age groups and symptom/problem complexes, before and after the successive intervals transformation. The transformation moves the mean preference for the lower function levels closer to 0.0, and for higher function levels closer to 1.0, spreading the values more evenly across the 0 to 1 scale. Thus, the transformation has the desired effect--moving estimates away from the middle of the scale.

TABLE 3: FUNCTION LEVEL MEANS BEFORE AND AFTER TRANSFORMATION

Function Level Number(j)	Mobility (Step)	Physical Activity (Step)	Social Activity (Step)	Before	After
L 30	Travelled Freely (5) (No Symptom/Problem Complex)	Walked freely (4)	Performed major and other activities (5)	1.000	1.000
L 29	Travelled freely (5) (Symptom/Problem Complex Present)	Walked freely (4)	Performed major and other activities (5)	0.804	0.848
L 28	Travelled freely (5)	Walked freely (4)	Performed major but limited in other activities (4)	0.690	0.738
L 27	Travelled freely (5)	Walked freely (4)	Performed major activity with limitations (3)	0.694	0.744
L 26	Travelled freely (5)	Walked freely (4)	Did not perform major but performed self-care activities (2)	0.646	0.688
L 25	Travelled with difficulty (4)	Walked freely (4)	Performed major but limited in other activities (4)	0.516	0.537
L 24	Travelled with difficulty (4)	Walked freely (4)	Performed major activity with limitations (3)	0.536	0.561
L 23	Travelled with difficulty (4)	Walked freely (4)	Did not perform major but performed self-care activities (2)	0.495	0.512
L 22	Travelled with difficulty (4)	Walked with limitations (3)	Performed major but limited in other activities (4)	0.519	0.538
L 21	Travelled with difficulty (4)	Walked with limitations (3)	Performed major activity with limitations (3)	0.522	0.542
L 20	Travelled with difficulty (4)	Walked with limitations (3)	Did not perform major but performed self-care activities (2)	0.469	0.479
L 19	Travelled with difficulty (4)	Moved independently in wheelchair (2)	Performed major activity with limitations (3)	0.503	0.520
L 18	Travelled with difficulty (4)	Moved independently in wheelchair (2)	Did not perform major but performed self-care activities (2)	0.457	0.465
L 17	In house (3)	Walked freely (4)	Did not perform major but performed self-care activities (2)	0.594	0.628
L 16	In house (3)	Walked freely (4)	Required assistance with self-care activities (1)	0.505	0.522
L 15	In house (3)	Walked with limitations (3)	Did not perform major but performed self-care activities (2)	0.519	0.538
L 14	In house (3)	Walked with limitations (3)	Required assistance with self-care activities (1)	0.436	0.439
L 13	In house (3)	Moved independently in wheelchair (2)	Did not perform major but performed self-care activities (2)	0.491	0.504
L 12	In house (3)	Moved independently in wheelchair (2)	Required assistance with self-care activities (1)	0.444	0.448
L 11	In house (3)	In bed or chair (1)	Did not perform major but performed self-care activities (2)	0.534	0.555
L 10	In house (3)	In bed or chair (1)	Required assistance with self-care activities (1)	0.436	0.439
L 9	In hospital (2)	Walked freely (4)	Did not perform major but performed self-care activities (2)	0.528	0.548
L 8	In hospital (2)	Walked freely (4)	Required assistance with self-care activities (1)	0.440	0.443
L 7	In hospital (2)	Walked with limitations (3)	Did not perform major but performed self-care activities (2)	0.440	0.442
L 6	In hospital (2)	Walked with limitations (3)	Required assistance with self-care activities (1)	0.388	0.381
L 5	In hospital (2)	Moved independently in wheelchair (2)	Did not perform major but performed self-care activities (2)	0.445	0.449
L 4	In hospital (2)	Moved independently in wheelchair (2)	Required assistance with self-care activities (2)	0.397	0.392
L 3	In hospital (2)	In bed or chair (1)	Did not perform major but performed self-care activities (2)	0.428	0.428
L 2	In hospital (2)	In bed or chair (1)	Required assistance with self-care activities (1)	0.342	0.333
L 1	In special unit (1)	In bed or chair (1)	Required assistance with self-care activities (1)	0.267	0.248
L 0	Death (0)	Death (0)	Death (0)	0.000	0.000

CONCLUSION

A FORTRAN program, written to perform the calculations of the above analysis, will: a) estimate interval widths by Edwards' [1956] procedure for large data sets, b) test for equality of average widths, c) fit a fourth degree polynomial (using a regression routine that includes plots and tests of fit), d) estimate the end intervals, e) give the values for transforming to compensate for unequal intervals (optimally on the 0-1 scale), f) calculate item means and medians of the transformed data, and g) provide punched-card output of the transformed data as an option. The programs, which use standard BMD analysis of variance and regression routines are available from the authors.[2]

The procedures and program described above may be of benefit to investigators using category scaling in a wide variety of other research applications.

---

[1]Research supported by Grant No. 7-R18-HS00702 from the National Center for Health Services Research.

[2]The authors gratefully acknowledge the assistance of Pete Mundle who wrote the computer programs and Donald Patrick who collected the original data.

REFERENCES

Anderson, N.H. Information Integration Theory: A Brief Survey, CHIP 24, 1972, in press in Contemporary Developments in Mathematical Psychology, Vol. 2, D.H. Krantz, R.C. Atkinson, R.D. Tuce, and P. Suppes (Eds.), W.H. Freeman & Co., 1974.

Baker, B.O., Hardyck, C.F. and Petrinovich, L.F. Weak Measurements vs. Strong Statistics: An Empirical Critique of S.S. Stevens' Proscriptions on Statistics, Educational and Psychological Measurement, 26, 1966.

Berg, R. (Ed.), Health Status Indexes, Chicago: Hospital Research and Educational Trust, 1973.

Berry, C.C. and Bush, J.W. Maintaining Health in a Defined Population Using a Health Status Index, presented before the Pacific Division, American Association for the Advancement of Science, Irvine, California, June 1974.

Bush, J.W., Chen, M.M. and Zaremba, J. Estimating Health Program Outcomes Using a Markov Equilibrium Analysis of Disease Development, American J. of Public Health, 61(12):2362-2375, 1971.

Bush, J.W., Kaplan, R.M., Berry, C.C. and Blischke, W.R. Design of a Survey to Assess the Properties of a Health Status Index, 1974, in process.

Bush, J.W., Robinson, J.D. and Chen, M.M. A General Computer Simulation Model for Disease Histories and Changes in Health Status, in R. Yoder, (Ed.), Proceedings of the San Diego Biomedical Symposium, pp. 169-175, 1972.

Chen, M.M., Bush, J.W. and Patrick, D.L. Social Indicators for Health Planning and Policy Analysis, Policy Sciences, in press, 1975.

Dixon, W.J. (Ed.) Biomedical Computer Programs, Berkeley: University of California Press, 1971.

Edwards, A.L. Techniques of Attitude Scale Construction, New York: Appleton-Century-Crofts, Inc. 1956.

Fanshel, S. and Bush, J.W. A Health Status Index and Its Application to Health Services Outcomes, Operations Research, 18(6):1021-1066, 1970.

Goldsmith, S.B. A Reevaluation of Health Status Indicators. Health Services Reports, 88(10): 937-941, 1973.

Guilford, J.P. Psychometric Methods, New York: McGraw-Hill, 1954.

Gulliksen, H.A. A Least Squares Solution for Successive Intervals Assuming Unequal Standard Deviations, Psychometrika, 19:117-140, 1954.

Jones, L.V. and Thurstone, L.L. The Psychophysics of Semantics: An Experimental Investigation, J. of Applied Psychology, 39:31-36, 1955.

Kaplan, R.M. and Bush, J.W. A Multitrait Multimethod Study of Value Ratings for a Health Status Index, in process, 1974.

Patrick, D.L., Bush, J.W. and Chen, M.M. Toward an Operational Definition of Health, J. of Health and Social Behavior, 14(1):6-23, 1973a.

Patrick, D.L., Bush, J.W. and Chen, M.M. Measuring Levels of Well-Being for a Health Status Index, Health Services Research 8,3:228-245, Fall, 1973b.

Ramsey, J.O. The Effect of Number of Categories in Rating Scales on Precision of Estimation of Scale Values, Psychometrika, 38(4):513-532, 1973.

Schonemann, P.H. and Tucker, L.R. A Maximum Likelihood Solution for the Method of Successive Intervals Allowing for Unequal Stimulus Dispersions, Psychometrika, 32(4):403-417, 1967.

Stevens, S.S. A Metric for the Social Consensus. Science, 151:530-541, 1966.

Stone, L.A. Magnitude Estimation and Numerical Category Scale Evaluations of Category Scale Adjectival Stimuli on Three Clinical Judgmental Continua, J. of Clinical Psychology 26(1):24-27, 1970.

Sullivan, D.F. Conceptual Problems in Developing an Index of Health, Washington, D.C.: HEW, National Center for Health Statistics, Public Health Service Publication No. 1000, Series 2, No. 17, 1966.

Torgerson, W.S. Theory and Methods of Scaling, New York: John Wiley, 1958.

Torrance, G.W. Health Index and Utility Models: Some Thorny Issues. Health Services Research, 8(1):12-14, 1973.

Tucker, L.R. A Maximum Likelihood Solution for Paired Comparisons Scaling by Thurstone's Case V. Technical Report, Urbana, Illinois: University of Illinois, 1964.

A NEW APPROACH TO ANALYZING QUASI-EXPERIMENTAL DATA: VALUE-ADDED ANALYSIS

Anthony S. Bryk, Huron Institute - Herbert I. Weisberg, Contract Research Corporation

I. Introduction

In a true educational experiment, the experimental units (e.g., individuals, classes, schools, pre-school centers) are randomly assigned to the different treatments, or programs, under study. As the number of units per treatment group increases, the pre-experimental mean differences among the treatment groups on any background variable tend to become small. As a result of this "natural equation of groups," differences among the post-treatment outcomes can, with reasonable assurance, be attributed to treatment effects.

For a variety of political, practical, and ethical reasons, however, randomized experiments are rarely feasible in educational settings. Recently, (Cohen, 1973) even the desirability of such experiments has been called into question. These practical and theoretical considerations have led to the implementation of what Campbell and Stanley (1963) have referred to as quasi-experiments. In these designs, the experimental units are not randomly assigned to treatments. As a result, the pre-treatment equation of groups is not assured. Observed differences among the post-treatment outcomes are attributable to pre-treatment differences in addition to the effects of the treatments. We must now resort to statistical techniques to adjust away the relevant pre-experiment differences among the treatment groups. Because of the assumptions required by these techniques and certain artifacts involved in the estimation of effects, interpretation of the results of these analyses requires extreme care. Some authors (Lord, 1967; Campbell and Erlebacher, 1970) are quite pessimistic about the ability of quasi-experiments to yield any useful and valid inferences.

In this paper, we introduce the value-added strategy as an alternative approach to the analysis of data from educational quasi-experiments. This technique was developed in response to shortcomings inherent in currently existing adjustment strategies, such as analysis of covariance, matching, standardization (Wiley, 1970), gain scores, and analysis of residuals.

II. An Alternative Strategy - Value-Added Analysis

The basic idea of the value-added analysis is to estimate for each subject in each experimental program the post-test score he would have obtained had he not been in any experimental program. Comparing this artificially-constructed post-test score with the actual pre- and post-test scores, we can estimate how much of his growth is the result of "natural" maturation, and how much is the value-added by the program in which he was enrolled. Smith (1973) used this approach to estimate overall effects of Head Start. We have refined his method and provided more theoretical underpinning.

A. Theory

Let Y_{ij} and Y'_{ij} represent the observed pre- and post-test scores on some measure for individual i in treatment, or program, group j. Let T_{ij} and T'_{ij} be the corresponding true scores. We assume the classical test theory model that

$$Y_{ij} = T_{ij} + e_{ij} \qquad (1)$$

$$Y'_{ij} = T'_{ij} + e'_{ij} \text{ for } i=1\ldots n_j,\ j=1\ldots J \qquad (2)$$

where $E(e_{ij}|T_{ij})=E(e'_{ij}|T'_{ij})=E(e_{ij}e'_{ij})=0$ and $\mathrm{Var}\,(e_{ij}|T_{ij}) = \mathrm{Var}\,(e'_{ij}|T'_{ij}) = \sigma^2_e$.

Let us define:

a_{ij}=age of individual i in group j at pre-test

a'_{ij}=age of individual i in group j at post-test

M_{ij}=component of true scores, T_{ij} and T'_{ij}, representable as a linear function of measurable covariates other than age

U_{ij}=nonmeasurable component of the true scores, T_{ij} and T'_{ij}, which is independent of both age and other measurable covariates.

For any specific individual i in group j, a_{ij}, a'_{ij}, M_{ij}, and U_{ij} are considered fixed. They do vary, however, over subjects in the population of individuals from which the sample for treatment group j was chosen. Let us define:

$$E(U_{ij}) = U_j \text{ and } \mathrm{Var}\,(U_{ij}) = \sigma^2_U \qquad (3)$$

Further, since U_{ij} is independent of age and the other measurable covariates, it follows that

$$\mathrm{Cov}(U_{ij},a_{ij}) = \mathrm{Cov}(U_{ij},a'_{ij}) = \mathrm{Cov}(U_{ij},M_{ij})=0 \qquad (4)$$

For the simplest case, we assume that growth in the domain under study is a known linear function of age, measurable covariates, and nonmeasurable covariates. We can, then, represent the true score for subject i in group j at the pre-test as:

$$T_{ij} = \mu + \beta a_{ij} + M_{ij} + U_{ij} \qquad (5)$$

Further, if there is no intervention between the pre-test and post-test - i.e., if only natural maturation is operational - then we can represent the post-test score as:

$$T'_{ij} = \mu + \beta a'_{ij} + M_{ij} + U_{ij} \qquad (6)$$

Let us define a variable Δ_{ij} where there is no intervention between pre- and post-test:

$$\Delta_{ij} = T'_{ij} - T_{ij} \qquad (7)$$

Clearly, Δ_{ij} represents the growth increment for subject i in group j which is attributable solely to natural maturation. Specifically, in terms of the growth model represented by equations (5) and (6),

$$\Delta_{ij} = \beta\,(a'_{ij} - a_{ij}) \qquad (8)$$

In general, we could make different assumptions about the nature of growth in the domain under study than were made for our simple case. For example, one could hypothesize a growth model involving interaction terms or non-linear terms in a_{ij}. In this case, we would replace the mathematical model of equation (5) with a more complex model. As a result, the expression for Δ_{ij} would also become more complex. The basic derivation presented here, however, would remain unchanged.

Let us now examine the alternative situation where an intervention has occurred between the pre- and post-tests. If we assume that the effect of the intervention is to increment uniformly the growth of all children in program j, then

$$T'_{ij} = \mu + \beta a'_{ij} + M_{ij} + U_{ij} + V_j \qquad (9)$$

where V_j represents the incremental effect of experimental program j, or what we have termed the value-added by program j. From equations (5) and (9), we see that

$$V_j = (T'_{ij} - T_{ij}) - \Delta_{ij} \qquad (10)$$

Thus, the value-added by program j, V_j, can be interpreted as the true difference (gain or change) score between pre- and post-test adjusted for the growth increment which is expected on the basis of natural maturation. Alternatively, $T_{ij} + \Delta_{ij}$ may be interpreted as the post-test score predicted solely on the basis of natural maturation. From this point of view, $T'_{ij} - (T_{ij} + \Delta_{ij})$ is a "residual score;" i.e., the observed post-test minus a predicted post-test.

B. Application to the Educational Quasi-Experiment

Our goal in the value-added analysis is to estimate for each program a value-added, $\hat{V}_j$, which is a measure of the absolute program effect over and above what one might expect on the basis of natural maturation. Our model assumes that the true pre- and post-test scores can be represented as a function of age, measured covariates, and unmeasured covariates. The model also assumes that the function is known. We now consider what can be done when each of two important assumptions (true scores known; growth model known) in this idealized situation is lifted. This will lead us to a way of implementing the value-added approach in practice.

1. True Scores Known: First Approach

The first problem that we encounter in applying this model is that the true scores T_{ij} and T'_{ij} are rarely known. The first and obvious approach is to substitute the observed pre- and post-test scores, Y_{ij} and Y'_{ij}, in equations (5) and (9). For the pre-test, this yields

$$Y_{ij} = \mu + \beta a_{ij} + M_{ij} + V_{ij} + e_{ij} \qquad (11)$$

Similarly, for the post-test,

$$Y'_{ij} = \mu + \beta a'_{ij} + M_{ij} + U_{ij} + V_j + e'_{ij} \qquad (12)$$

It follows from these last two equations that

$$Y'_{ij} - (Y_{ij} + \Delta_{ij}) = V_j + e'_{ij} - e_{ij} \qquad (13)$$

we may now define, V_{ij1}, an individual value-added for subject i in program j, where

$$V_{ij1} = V_j + e'_{ij} - e_{ij} \qquad (14)$$

For any randomly selected individual i in treatment group j,

$$E\,(V_{ij1}) = V_j \text{ and } Var\,(V_{ij1}) = 2\sigma^2_e \qquad (15)$$

Thus, each subject provides an unbiased estimate of V_j with variance $2\sigma^2_e$.

We can obtain a more efficient estimate of V_j by pooling these individual estimates to obtain an average across the n_j subjects in treatment group j. This approach yields

$$V_{\cdot j1} = \sum_{i=1}^{n_j} \frac{V_{ij1}}{n_j} \qquad (16)$$

where, for any treatment group j,

$$E(V_{\cdot j1}) = V_j \text{ and } Var\,(V_{\cdot j1}) = \frac{2\sigma^2_e}{n_j} \qquad (17)$$

Like our individual subject estimate, our pooled estimate, $V_{\cdot ij1}$, is unbiased, but the variance is now reduced to $2\sigma^2_e/n_j$. Thus, the direct substitution of the observed pre- and post-test scores for the unknown true scores provides us with one approach for estimating the incremental effect of our experimental program.

2. True Scores Unknown: Second Approach

While the above approach is intuitively appealing, we might ask whether we can obtain a better estimate of V_j by some alternative approach. In particular, since the observed pre- and post-tests are measured with error, perhaps the substitution of an estimated or predicted true score might yield a more efficient estimate. Several approaches for estimating true scores have been reviewed by Cronbach and Furby (1970). Their "complete estimator" has intuitive appeal, and it merits further investigation. This estimator is quite complex, however, and its usefulness in the value-added approach is an unresolved question which is under further investigation.

A simple and natural spin-off of this estimated true score approach would be the use of a predicted pre-test score, $\hat{Y}_{ij}$, generated from the observable components of equation (5). Simply,

$$\hat{Y}_{ij} = \mu + \beta a_{ij} + M_{ij} \quad (18)$$

$\hat{Y}_{ij}$ may be viewed as an alternative estimate for T_{ij}. In terms of equation (5),

$$T_{ij} = \hat{Y}_{ij} + U_{ij} \quad (19)$$

As a second approach to dealing with the unknown true scores, we substitute the observed post-test score Y'_{ij} and the predicted pre-test score $\hat{Y}_{ij*}$ into equation (10) for the unknown true scores. This yields,

$$Y'_{ij} - (\hat{Y}_{ij} + \Delta_{ij}) = V_j + U_{ij} + e'_{ij} \quad (20)$$

We now have a second individual subject estimator, V_{ij2}, of the value-added for subject i in program j, where

$$V_{ij2} = V_j + U_{ij} + e'_{ij} \quad (21)$$

For any randomly selected individual i in treatment group j,

$$E(V_{ij2})=V_j+U_j \text{ and } Var(V_{ij2})=\sigma^2_U + \sigma^2_e \quad (22)$$

Thus, each subject provides a biased estimate of V_j (bias = U_j) with variance $\sigma^2_U + \sigma^2_e$.

Pooling these individual estimates across the n_j subjects in treatment group j yields,

$$V_{\cdot j2} = \sum_{i=1}^{n_j} \frac{V_{ij2}}{n_j} \quad (23)$$

where, for any treatment group j,

$$E(V_{\cdot j2})=V_j+U_j \text{ and } Var(V_{\cdot j2})=\frac{\sigma^2_U + \sigma^2_e}{n_j} \quad (24)$$

Thus, the pooled estimate, $V_{\cdot j2}$, is also biased, but it has reduced variance as compared to the individual subject estimates.

3. <u>Comparing the Two Estimates</u>

If we compare equations (17) and (24), we see that neither estimate of V_j, $V_{\cdot j1}$, or $V_{\cdot j2}$ is clearly superior. $V_{\cdot j1}$ is unbiased, while $V_{\cdot j2}$ is in general biased, except when $U_j=0$; j=1...J. On the other hand, $V_{\cdot j1}$ will have larger variance if $\sigma^2_e > \sigma^2_U$. The <u>mean square error</u> (MSE), which combines both of these criteria, is a useful measure of the accuracy of an estimator. The MSE equals the sum of the variance and the bias squared. For our two estimators,

$$MSE\ (V_{j1}) = \frac{2\sigma^2_e}{n_j} \quad (25)$$

and

$$MSE\ (V_{\cdot j2}) = \frac{\sigma^2_U + \sigma^2_e}{n_j} + (U_j)^2 \quad (26)$$

Thus, in general, $V_{\cdot j1}$ is superior if

$$\sigma^2_e < \sigma^2_U + n_j\ (U_j)^2 \quad (27)$$

As we defined U_j in our simple growth model of equation (5), it can be interpreted as the expected residual pre-test score for individuals in group j over and above that which is linearly related to all measured variables. Since the model includes a constant term μ, we would expect U_j to differ from 0 only if group j differs from the other groups in ways unrelated to measured covariates. If the covariates are selected carefully, however, it is unlikely that any U_j will differ substantially from 0. Of course, if subjects are randomly assigned to treatment groups, there should not exist <u>a priori</u> differences between treatment groups in the expected values on any variables. As a result, U_j will equal 0.

4. <u>A Combined Estimate</u>

If $V_{\cdot j1}$ and $V_{\cdot j2}$ were independent, then an appropriate linear combination of these two estimates would have a smaller mean square error than either one individually. Although $V_{\cdot j1}$ and $V_{\cdot j2}$ are clearly not independent, we may still realize a gain in accuracy by creating such a combined estimate. This estimate would take the following general form:

$$V_{\cdot jc} = w\ V_{\cdot j1} + (1 - w)\ V_{\cdot j2} \quad 0 \le w \le 1 \quad (28)$$

It would clearly be desirable to choose w so as to minimize the mean square error of $V_{\cdot jc}$. Note that

$$E(V_{\cdot jc})=wV_j + (1-w)(V_j + U_j)=V_j + (1-w)U_j \quad (29)$$

which indicates that the combined estimate, $V_{\cdot jc}$, has bias $(1 - w)U_j$. Also,

$$Var(V_{\cdot jc})=w^2\ \frac{(2\sigma^2_e)}{n_j} + (1-w)^2\ \frac{(\sigma^2_U+\sigma^2_e)}{n_j} +$$

$$2w(1-w)\ Cov(V_{\cdot j1},V_{\cdot j2}) \quad (30)$$

For any treatment group j, if we assume that the $Cov(V_{ij1},V_{i*j2}) = 0$ for $i \neq i^*$ then

$$Cov(V_{\cdot j1},V_{\cdot j2}) = \sum_{i=1}^{n_j} Cov\left(\frac{V_{ij1}}{n_j},\frac{V_{ij2}}{n_j}\right) \quad (31)$$

It follows that

$$Var(V_{\cdot jc}) = (1+w^2)\ \frac{\sigma^2_e}{n_j} + (1-w)^2\ \frac{\sigma^2_U}{n_j} \quad (32)$$

Thus, we can now write out the mean square error for $V_{\cdot jc}$:

$$MSE(V_{\cdot jc})=(1-w^2)\ \frac{\sigma^2_e}{n_j} + (1-w)^2 \left[\frac{\sigma^2_U}{n_j} + U_j^2\right] \quad (33)$$

If we minimize this with respect to w, we find

$$w_{opt} = \frac{\sigma^2_U + n_j (U_j)^2}{\sigma^2_U + n_j (U_j)^2 + \sigma^2_e} \tag{34}$$

If, as the result of randomization or by chance $U_j = 0$, the expression for w_{opt} simplifies to:

$$w_{opt} = \frac{\sigma^2_U}{\sigma^2_U + \sigma^2_e} \tag{35}$$

Examination of this expression suggests that w_{opt} can be interpreted in this case as a "residual reliability" of the test after variation related to age and other measurable background variables has been removed. Thus, the weight to be placed on method 1, which uses the observed pre-test score, is simply the residual reliability of this score.

Finally, up to this point, we have assumed that the values of σ^2_U, σ^2_e, and U_j are known. In most quasi-experimental settings, however, these values are unknown. Thus, w_{opt} must be estimated from the data. Two alternative procedures for estimating w_{opt} are presented in the Appendix.

5. Growth Model Unknown

To apply the value-added approach in the analysis of an educational quasi-experiment, we need a growth model which accurately describes the process of "natural" maturation. The second major assumption in our theoretical model which we now lift is that the natural growth model is known.

In this situation, we must develop a growth model and estimate the parameters of the model from either the quasi-experimental data or some alternative data set. This is a complex problem because the form of the growth model depends both upon the particular phenomenon under study and the conditions in the setting which bound this investigation. As a result, no single approach to the construction of such models is likely to be uniformly successful. We present here an approach developed by Smith (1973) and Weisberg (1973) which may prove useful in certain circumstances.

The key assumption in this approach to estimating the growth model is that the variation displayed in pre-test scores reflects developmental trends which can be directly related to age and other background variables. The heart of this approach involves a causal linkage of the variation in age and the background variables to the variation in pre-test scores. In particular, if the background variables are held constant, then the variation in pre-test scores may be expected to reflect only differences in the length of time exposed to the natural learning environment, or what we term age. More specifically, suppose we could look at a sub-sample of children with identical values on all measured background variables except age. Suppose then that we observe the mean score for such individuals as a function of age. The resulting curve based on this cross-sectional data is an approximation to the longitudinal growth curve that these children would actually display as they grow older.

This approach of attempting longitudinal inference from cross-sectional data is a general and well-known strategy. Kodlin and Thompson (1958) have considered the limitations of this approach in some detail. Under the conditions of a "stable universe" - i.e., a stable growth process - and a stable population across age levels, the cross-sectional approach can be used in place of the longitudinal for the estimation of mean growth. Thus, in settings where: 1) there are no significant external influences - other than the experimental intervention - to disturb the natural growth process; and 2) the pre-test sample is selected effectively at random with respect to age and the background variables; the cross-sectional approach should provide an excellent approximation to the natural growth curves.

VI. Summary and Conclusions

The basic idea of the value-added analysis is quite intuitive. We develop a growth model to predict normal growth in the absence of an experimental treatment. By combining this expected outcome with the observed outcome, we estimate the program effect - the growth increment over and above natural maturation.

As one approach to estimating the natural growth model, we can assume a causal linkage between the variation in age and background variables, and the variation in pre-test scores. If the research setting is stable, variation as a function of age will be attributable solely to growth, and not to other differences among age cohorts. In such a situation, we can estimate the effect of natural maturation for a treatment group, and the program effect over and above this natural growth.

A particularly perplexing problem for traditional techniques, such as ANCOVA, is the biased and inconsistent estimation of program effects resulting from measurement errors in the covariates (e.g., pre-test scores). Under the value-added model, however, measurement error in the pre-test will not bias estimates of the program effects.

Further, we believe that the concept of the value-added effect is more meaningful than the adjusted treatment mean differences in ANCOVA. The value-added is an absolute measure of program effect. Although control group data is useful as a check on the fitted mathematical model, it is not mandatory. Under the strong assumption that the growth model is correct, the value-added technique generates a statistical control group.

In a pre-school setting such as Head Start, for example, child's age from birth seems suitable to use in the growth model. In certain situations, our model may be applicable with other measures of "age." For example, suppose we were comparing a traditional and an experimental high school foreign language program. A logical choice here for "age at pre-test" might be the length of time studying the language via the traditional approach.

Although we feel that the value-added approach is quite general, two factors cause difficulties in applying it in a school context. First, much testing in schools uses a standardized metric. Standardized tests do not allow measurement of absolute growth, only "relative standing." Our model in its present form would not be suitable for such a metric. Second, most schools have summer vacations, so that growth is not a strictly monotone function of chronological age or length of exposure to schooling. We need to investigate thoroughly the effects of this summer discontinuity. The development of complex growth models incorporating such summer effects would make the value-added approach feasible.

Although we hope that the methods described in this paper will provide analysts with a useful alternative to traditional adjustment strategies, we view this work as only a beginning. There are many possible extensions and refinements. For example, we need to develop realistic models to represent growth in educational settings, and practical ways of estimating model parameters. A better understanding of the sampling theory associated with our methods is needed, so that significance tests and confidence intervals can be obtained. Also, the development of quasi-experimental designs which facilitate this type of analysis should be pursued. Lastly, a critical problem with traditional adjustment strategies is that they implicitly embody a static model which conceptualizes a program effect as a constant increment to a static base. Educational programs, on the other hand, are usually attempts to alter some developmental growth process. An intervention is typically a dynamic change in an on-going process. We view the value-added approach as a tentative effort to operationalize this conceptual approach in the analysis of quasi-experiments.

*The use of a predicted post-test score might also be considered here. This is problematic because treatment effects must be included in the predicted post-test. In reality, there are numerous possible predicted or estimated true scores, such as the Cronbach and Furby estimates, which could be utilized in the value-added setting. Because of the introductory nature of this paper, we consider here only two simple approaches.

REFERENCES:

Campbell, D.T. and Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Compensatory Education: A National Debate. Vol. III. Disadvantaged Child. New York: Brunner-Mazel, 1970.

Campbell, D.T. and Stanley, J.C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963.

Cochran, W.G. Errors of measurement is statistics. Technometrics. 1968, 10, 637-660.

Cohen, D. Social experiments with schools: what has been learned? to be published in Planned Variation Experiments: Should We Give Up or Try Harder?, A Rivlin (Ed.), Brookings Institute.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968-70, 426-443.

Kodlin, D. and Thompson, D.J. An appraisal of the longitudinal approach to studies of growth and development. Monographs of the Society for Research in Child Development, 1958, 23 (No. 67).

Lord, F.M. A paradox in the interpretation of group comparisons. Psychological Bulletin, 1967-68, 304-405.

Lord, F.M. and Novick, M.R. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley, 1968.

Smith, M.S. and Bissell, J.S. Report analysis: the impact of Head Start. Harvard Educational Review, 1970, 40, 51-104.

Weisberg, H.I. Short term cognitive effects of Head Start programs: a report of the third year of planned variation - 1971-1972. Cambridge, Massachusetts: Huron Institute, 1973.

Wiley, D.E. Approximations to ceteris paribus: Data adjustment in educational research. Paper presented at International Research Seminar on Learning and the Educational Process, Munich, Germany, 1971.

APPENDIX

In this Appendix, we present two methods of estimating the optimal weight w in our combined estimator $V_{.jc}$. Both of these methods assume that $U_j = 0$ for all j, so that

$$w_{opt} = \frac{\sigma^2_U}{\sigma^2_U + \sigma^2_e}.$$

It is this quantity we wish to estimate. If U_j is substantially different from 0 for some j's, w_{opt} may differ substantially from the true optimum given by equation (34). If measured covariates are selected judiciously, this is unlikely to occur.

Estimator #1

Suppose the reliability ρ of our outcome measure Y (and Y') is known. Then,

$$\rho = \frac{V(T_{ij})}{V(Y_{ij})} = \frac{V(\alpha + \beta a_{ij} + M_{ij}) + \alpha^2_U}{V(Y_{ij})}$$

Let

$$\frac{V(\alpha + \beta a_{ij} + M_{ij})}{V(Y_{ij})} \equiv \eta^2$$

Then

$$\rho = \eta^2 + \frac{\sigma^2_U}{V(Y_{ij})}$$

$$\frac{\rho - \eta^2}{1 - \eta^2} = \frac{\sigma^2_U}{V(Y_{ij})(1-\eta^2)} = \frac{\sigma^2_U}{\sigma^2_U + \sigma^2_e} = w_{opt}$$

From the regression analyses used to produce our model, we obtain an estimate R^2 of η^2. A natural estimator of w_{opt} is thus given by

$$w_{opt} = \frac{\rho - R^2}{1 - R^2}$$

<u>Estimator #2</u>

Our second estimator does not require independent information about ρ. From equations (29) and (32), we see that for any individual i in group j, assuming $U_j = 0$:

$$E(V_{ijc}) = V_j$$

$$Var(V_{ijc}) = (1+w^2)\,\sigma^2_e + (1-w^2)\,\sigma^2_U$$

Ignoring sampling variation in the regression coefficients of our predictor equation, the V_{ijc}'s (for any value of w) are independent random variables. Thus, if we perform a one-way analysis of variance, using these as outcomes, the mean square error term provides an unbiased estimate, $\hat{Var}(V_{ijc})$ of $Var(V_{ijc})$. Since

$$Var(V_{\cdot ij}) = \frac{Var(V_{ijc})}{n_j}$$

w_{opt} is the value of w which minimizes $Var(V_{ijc})$. It seems reasonable, then, to estimate w_{opt} by the value of w which minimizes $\hat{Var}(V_{ijc})$. Let

$$SS_{11} = \sum_j\sum_i (V_{ij1} - V_{\cdot j1})^2$$

$$SS_{22} = \sum_j\sum_i (V_{ij2} - V_{\cdot j2})^2$$

$$SS_{12} = \sum_j\sum_i (V_{ij1} - V_{\cdot j1})(V_{ij2} - V_{\cdot j2})$$

Then the mean square error for the one-way ANCOVA is easily shown to be

$$MSE = w^2 SS_{11} + (1-w)^2\, SS_{22} + 2w(1-w)\, SS_{12}$$

Minimizing this with respect to w, we find

$$w_{opt} = \frac{SS_{22} - SS_{12}}{SS_{22} + SS_{11} - 2SS_{12}}\ .$$

SAMPLING FROM TWO OVERLAPPING FRAMES WHEN THE DOMAIN SIZES ARE KNOWN

Leon F. Burmeister, University of Iowa

It may be occasionally necessary or economically convenient to utilize more than one frame in drawing a sample from a population. This paper assumes two frames cover the population and the observational units for the two frames are identical. If the sizes of both frames and the population size are known, the number of elements common to the two frames is also known, as well as the number of elements included only on each individual frame.

Estimators will be developed that have smaller variance than those previously suggested. However, since the number of duplicated elements included in the overall sample is utilized in the proposed estimators, the cost is greater than for previous estimators. Consequently, allocation procedures and situations likely to result in appreciably higher cost will be considered.

The adaptation of a previously suggested method to construct unbiased estimators of the population total with empty domain sample sizes is illustrated.

1.1 Examples

The opportunity to utilize two or more frames of known sizes does occur in practice. Comstock, et. al. (3) describe a study designed to evaluate immunization histories obtained from a sample of the population of Washington County, Maryland. The evaluation was completed by the comparison of historical information from interviews with the results of serologic determinations.

The study was done during the summer of 1968. A 1% systematic sample of the county population was drawn from a list of households obtained in a non-official census conducted in 1963, supplemented by a similar sample of dwelling units added since that time. It seems reasonable to assume that the number of households on both frames were known as well as the total number of households in the county. From this knowledge it is easily determined if the two frames cover the population and if they overlap. If there is an overlap, consideration should be given to the determination of the duplicated households. Since it is not obvious that the costs of doing so can be justified, the improvements in the resulting estimators need to be carefully evaluated.

Other examples of studies that have been conducted based on multiple and possibly overlapping frames include those described by Serfling, Cornell and Sherman (8), Bershad (1) and Cochran (2).

1.2 Notation

It is assumed two frames, A and B, containing N_A and N_B elements respectively are available. The notation of Hartly (5) is adopted and N_{ab} denotes the number of elements included on both frame A and Frame B. N_a is the number of elements occurring only on frame A and N_b is the number of elements occurring on Frame B. Thus:

$$N_A = N_a + N_{ab}, \qquad (1.1)$$

$$N_B = N_b + N_{ab} \qquad (1.2)$$

and the total number of elements in the population, N, is given by

$$N = N_a + N_b + N_{ab} = N_a + N_B = N_b + N_A. \qquad (1.3)$$

The elements contained only on frame A are called domain a, the elements only on frame B domain b and those elements on both frames A and B domain ab. It is assumed that a simple random sample of size n_A is selected from Fr. A & and simple random sample of size n_B is selected from FrB. The number of elements sampled from frame A and contained in domain a is denoted by n_a. The number of elements sampled from frame A and contained in domain ab is denoted by n'_{ab}. The number of sampled elements in domains ab and b drawn from frame B are denoted by n''_{ab} and n_b respectively. Thus

$$n_A = n_a + n'_{ab} \qquad (1.4)$$

and

$$n_B = n''_{ab} + n_b \qquad (1.5)$$

This completes the description of the problem. It is only one of several frame problems that the sample designer may face. Kish (6) gives several interesting and informative discussions of these additional problems.

2. ESTIMATORS OF THE POPULATION TOTAL

Assume a sample of size n_A is drawn without replacement from frame A, and a sample of size n_B is drawn without replacement from frame B. Assuming simple random sampling in both frames, the probability of being included in the sample, Π_i, can be calculated for elements in each domain. These probabilities will be utilized in the construction of alternative estimators.

$$\text{Prob}\left[i^{th} \text{ element in domain a is included in the sample} \right] = \frac{n_A}{N_A}, \qquad (2.1)$$

$$\text{Prob}\left[i^{th} \text{ element in domain b is included in the sample} \right] = \frac{n_B}{N_B}, \qquad (2.2)$$

and

$$\text{Prob}\left[i^{th} \text{ element in domain ab is included in the sample at least once} \right] = \frac{n_A N_B + n_B N_A - n_A n_B}{N_A N_B}. \qquad (2.3)$$

Lund (7) proposed the estimator

$$\hat{Y}_L = N_a \bar{y}_a + N_{ab} \bar{y}^*_{ab} + N_b \bar{y}_b \qquad (2.4)$$

for the case of known domain sizes. The sample total for domain ab,

$$y^*_{ab} = \sum_{i=1}^{n'_{ab}} y_i + \sum_{i=1}^{n''_{ab}} y_i, \qquad (2.5)$$

in $\hat{Y}_L$ is based on the $n'_{ab} + n''_{ab}$ elements sampled from frames A and B. If the duplicated elements in domain ab are excluded, the sample mean $\bar{y}^*_{ab}$ becomes

$$\bar{y}_{ab} = \frac{\sum_{i=1}^{n'_{ab}} y_i + \sum_{i=1}^{n''_{ab}} y_i - \sum_{i=1}^{n_d} y_i}{n'_{ab} + n''_{ab} - n_d}, \qquad (2.6)$$

where n_d is the number of duplicated items.

Consideration is thus given to

$$\hat{Y}_d = N_a\bar{y}_a + N_{ab}\bar{y}_{ab} + N_b\bar{y}_b \ . \qquad (2.7)$$

The notation $\hat{Y}_d$ in (2.7) indicates that this estimator of the total is based on the distinct elements included in the sample.

If it can be assumed that n_a, n_{ab} and n_b are each greater than zero, $\hat{Y}_d$ is unbiased. The conditional expectation becomes

$$E(\hat{Y}_d \mid n_a>0,\ n_{ab}>0,\ n_b>0)$$

$$= N_aE(\bar{y}_a|n_a>0) + N_{ab}E(\bar{y}_{ab}|n_{ab}>0) + N_bE(\bar{y}_b|n_b>0)$$

$$= N_a\bar{Y}_a + N_{ab}\bar{Y}_{ab} + N_b\bar{Y}_b = Y. \qquad (2.8)$$

It should be noted that Fuller (4) has devised a method of constructing unbiased post-stratified estimators that does not require each domain size to be positive. This method will be considered below.

A comparison of the variances of $\hat{Y}_L$ and $\hat{Y}_d$ reduces to a comparison of the variances of $\bar{y}_{ab}$ and $\bar{y}^*_{ab}$ since the estimators differ only in the estimated mean of the overlap domain. To facilitate this comparison, let

$$\bar{y}^*_{ab} = \frac{y_{ab} + n_d\bar{y}_d}{n_{ab} + n_d} \qquad (2.9)$$

and

$$\bar{y}_{ab} = \frac{y_{ab}}{n_{ab}} \qquad (2.10)$$

where $n_{ab} = n'_{ab} + n''_{ab} - n_d$, y_{ab} is the total of the n_{ab} distinct elements sampled from domain ab and $\bar{y}_d$ is the mean of the n_d duplicated elements. It is assumed that n_{ab} is greater than zero. Conditional on n_{ab} and n_d,

$$\mathrm{Var}(\bar{y}^*_{ab}) = \frac{\mathrm{Var}(y_{ab})+n_d^2\mathrm{Var}(\bar{y}_d)+2n_d\mathrm{Cov}(y_{ab},\bar{y}_d)}{(n_{ab} + n_d)^2} \qquad (2.11)$$

and

$$\mathrm{Var}(\bar{y}_{ab}) = \frac{\mathrm{Var}(y_{ab})}{n_{ab}^2} \ . \qquad (2.12)$$

To evaluate (2.11) and (2.12), assume the variance-covariance structure of the y_i's to be

$$\begin{bmatrix} 1-1/N_{ab} & -1/N_{ab} & -1/N_{ab} & \cdots & -1/N_{ab} \\ -1/N_{ab} & 1-1/N_{ab} & -1/N_{ab} & & -1/N_{ab} \\ -1/N_{ab} & -1/N_{ab} & 1-1/N_{ab} & & -1/N_{ab} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ -1/N_{ab} & -1/N_{ab} & -1/N_{ab} & \cdots & 1-1/N_{ab} \end{bmatrix} S_{ab}^2 \qquad (2.13)$$

Then the variance of $\bar{y}^*_{ab}$ is greater than or equal to that of $\bar{y}_{ab}$ if

$$n_{ab}^2 \ \mathrm{Var}\ (y_d) + 2n_{ab}^2\ \mathrm{Cov}(y_{ab},\ y_d)$$

$$\geq 2n_dn_{ab}\ \mathrm{Var}(y_{ab}) + n_d^2\mathrm{Var}(y_{ab}). \qquad (2.14)$$

Utilization of the variance-covariance matrix, (2.13), allows us to express each term of (2.14) as follows:

$$n_{ab}^2\mathrm{Var}(y_d) = n_{ab}^2n_d\left[1-\frac{(n_d+1)}{N_{ab}}\right]S_{ab}^2 \qquad (2.15)$$

$$2n_{ab}^2\mathrm{Cov}(y_{ab},y_d) = 2n_{ab}^2n_d\left[1-\frac{n_{ab}}{N_{ab}}\right]S_{ab}^2 \qquad (2.16)$$

$$2n_{ab}n_d\mathrm{Var}(y_{ab}) = 2n_{ab}^2n_d\left[1-\frac{(n_{ab}+1)}{N_{ab}}\right]S_{ab}^2 \qquad (2.17)$$

and

$$n_d^2\mathrm{Var}(y_{ab}) = n_{ab}n_d^2\left[1-\frac{(n_{ab}+1)}{N_{ab}}\right]S_{ab}^2 \ . \qquad (2.18)$$

If n_d equals zero, equality holds in expression (2.14) since the estimators $\bar{y}_{ab}$ and $\bar{y}^*_{ab}$ are identical. If n_{ab} and n_d are both greater than zero, (2.14) may be written as

$$(n_{ab}-n_d)> - \frac{(n_{ab}+n_d)}{N_{ab}}. \qquad (2.19)$$

Thus $\bar{y}_{ab}$ has smaller variance than does $\bar{y}^*_{ab}$ for all positive values of n_{ab} and n_d.

One of the reasons for calculating the probability of selection, Π_i, for each element in the population now is evident. The Π_i's clearly demonstrate that every element in domain ab has the same probability of being included in the sample. This is required if the simple mean of the distinct elements sampled from domain ab is to be used as an unbiased estimator of the domain mean.

3. UNBIASED ESTIMATOR OF Y WITH EMPTY DOMAINS

It was noted above that Fuller (4) has devised a scheme to construct unbiased post-stratified estimators. This scheme does not require the usual assumption of non-empty strata. Fuller's general construction will be reviewed and then his approach will be applied to this problem.

Assume a random sample of size n has been drawn from a population. After the sample has been taken, the sampled elements are classified as members of two strata. Assume that the population is such that the population proportion of elements contained in stratum one, P_1, and the proportion contained in stratum two, $P_2 = 1 - P_1$, are known. Fuller then considers the general estimator

$$A_i\bar{y}_1 + (1-A_i)\bar{y}_2 \ , \qquad (3.1)$$

where

$\bar{y}_1$ = the sample mean of the characteristic y for stratum one,

$\bar{y}_2$ = the sample mean of the characteristic y for stratum two

and

A_i = the weight applied to the mean of stratum one for samples with i (i = 0, 1, . . . , n) sample elements in stratum one.

It is further assumed that $0 \leq A_i \leq 1$ for all i, $A_o = 0$ and $A_n = 1$.

Minimization of the conditional mean square error of (3.1) yields

$$A_i = \frac{i f_{2i} S_2^2 + i(n-i)P_1(\bar{Y}_1-\bar{Y}_2)^2}{(n-i)f_{1i}S_1^2 + i f_{2i}S_2^2 + i(n-i)(\bar{Y}_1-\bar{Y}_2)^2}, \quad (3.2)$$

where $f_{1i} = \frac{N_1 - i}{N_1}$

and $f_{2i} = \frac{N_2 - n + i}{N_2}$.

The estimator (3.1) employing the weight (3.2) is in general biased. However, Fuller shows it is possible to derive weights A_i such that the estimator (3.1) is unbiased.

Fuller also extends the above development for two post-strata to the general case of more than two strata. First the strata must be arranged in a natural order. Then the strata are repeatedly divided into groups of two. Beginning at the finest subdivision, an unbiased estimator is constructed for each pair of strata.

In this problem, the situation is as shown in Table 1.

Table 1 A DESCRIPTION OF THE POST-STRATA RESULTING FROM TWO OVERLAPPING SAMPLING FRAMES

Stratum	Proportion of pop. in stratum	Stratum ID	Sample number	Sample mean
a	N_a/N	1 1	n_a	$\bar{y}_a$
ab	N_{ab}/N	2 1	n_{ab}	$\bar{y}_{ab}$
b	N_b/N	2 2	n_b	$\bar{y}_b$

Table 1 indicates that the first division of the strata is into stratum a and strata ab and b. Thus an unbiased estimator will be constructed first for strata ab and b. It is to be remembered that n_a, n_{ab} or n_b may be zero in the development of this unbiased estimator.

In the development of the scheme, it is convenient to let $n = n_a + n_{ab} + n_b$, $n_1 = n_a$, $n_2 = n_{ab} + n_b$, $P_{21} = N_{ab}/N_B$, $P_{22} = N_b/N_B$ and $P_1 = N_a/N$.

The estimator for this specific problem is

$$N_{\bar{y}} = N\left[A_1\bar{y}_a + A_2(A_{21}\bar{y}_{ab} + A_{22}\bar{y}_b)\right], \quad (3.3)$$

where the general expressions for the weights in (3.3) are

$$A_1 = (1-A_2) = \frac{n_a + n_a(n_{ab}+n_b)\left[\frac{N_a M_1}{N} + \lambda_{M_1}\right]}{n + n_a(n_{ab}+n_b)M_1} \quad (3.4)$$

and

$$A_{21} = (1-A_{22}) = \frac{n_{ab} + n_{ab}n_b\left[\frac{N_{ab}}{N_B}M_2 + \lambda_{M_2}\right]}{(n_b + n_{ab}) + n_{ab}n_b M_2}, \quad (3.5)$$

where

$$\lambda_{M_1} = \frac{(\bar{Y}_a - \bar{Y}_b)^2}{S_w^2},$$

$$\lambda_{M_2} = \frac{(\bar{Y}_{ab} - \bar{Y}_b)^2}{S_w^2},$$

$$S_1^2 = S_2^2 = S_w^2 > 0,$$

$$\lambda_{M_1} = \frac{F_1 - \sum_{i=1}^{n-1} \frac{i + i(n-i)P_1M_1}{n + i(n-i)M_1}}{\sum_{i=1}^{n-1} \frac{P_i\, i(n-i)}{n + i(n-i)M_1}}$$

and

$$\lambda_{M_2} = \frac{F_{21} - \sum_{i=1}^{n_2-1} \frac{i + i(n_2-i)P_{21}M_2}{n_2 + i(n_2-i)M_2}}{\sum_{i=1}^{n_2-1} \frac{P_i\, i(n_2-i)}{n_2 + i(n_2-i)M_2}}. \quad (3.6)$$

F_{21} in (3.6) is P_{21}-Prob $[n_{ab} = (n_{ab}+n_b)$ given strata ab and b contain $n_{ab}+n_b$ sample elements$]$. That is, F_{21} is N_{ab}/N_B minus the probability that all n_B sampled elements fall in domain ab. P_i in expression (3.6) is the probability domain ab contains i units given that n_2 units have been selected from domains ab and b.

4. COST CONSIDERATIONS

Cost considerations are now introduced into the efficiency comparisons. The employment of the proposed estimator, $\hat{Y}_d$, necessitates the identification of duplicated elements in domain ab. Lund's estimator (2.4) is a special case of Hartley's(5) procedure which utilizes a weighted average of $\bar{y}'_{ab}$ and $\bar{y}''_{ab}$ and does not require this identification. Hartley gave expressions for the sampling fractions n_A/N_A and n_B/N_B that minimize the variance of his estimator subject to the cost restraint

$$C = c_A n_A + c_B n_B . \quad (4.1)$$

Lund's estimator results when these optimum sampling fractions are used to solve the bi-quadratic equation given by Hartley for the value of the weight p. Another possibility is the retention of elements from domain ab from one frame only. This procedure was employed by the Bureau of the Census in a 1949 study (1). This procedure is a special case of Hartley's procedure with the weight p = 1.

A third procedure is to merge the two frames before sampling and remove the duplicated elements Once the merging has been completed, any number of sampling schemes could be utilized. For example, one could employ stratified random sampling where the strata are the three domains a, b and ab.

The variance of $\hat{Y}_d$ (2.7) can be minimized subject to the cost constraint (4.1). A system of three equations in three unknowns, f_A, f_B and λ results. Utilization of the ratio of two of these equations reduces the system to the following system of two equations in two unknowns, f_A and f_B:

$$\frac{c_A N_A}{c_B N_B} = \frac{f_B^2\left[N_a\sigma_a^2(f_A+f_B-f_Af_B)^2 + N_{ab}\sigma_{ab}^2 f_A^2(1-f_B)\right]}{f_A^2\left[N_b\sigma_b^2(f_A+f_B-f_Af_B)^2 + N_{ab}\sigma_{ab}^2 f_B^2(1-f_A)\right]} \qquad (4.2)$$

and

$$C = c_A f_A N_A + c_B f_B n_B \qquad (4.3)$$

Solution of these two equations requires the solution of a sixth degree equation.

Rather than solving a sixth degree equation, the following iterative procedure may be used. Let

$$r = \frac{f_A}{f_B} .$$

An initial value for r is

$$r_1 = \sqrt{\frac{c_B}{c_A}} .$$

Note that (4.2) may be expressed as (4.4)

$$\frac{n_A^2}{n_B^2} = \frac{c_B N_A\left[N_a\sigma_a^2\ [f_B+f_A(1-f_B)]^2 + N_{ab}\sigma_{ab}^2 f_A^2(1-f_B)\right]}{c_A N_B\left[N_b\sigma_b^2\ [f_B+f_A(1-f_B)]^2 + N_{ab}\sigma_{ab}^2 f_B^2(1-f_A)\right]}$$

Set $1-f_A$ and $1-f_B$ equal to 1 and divide each term of the right hand side of (4.4) by f_B^2 to obtain

$$r_i^2+1 = \frac{C_B N_B\left[N_a\sigma_a^2(1+r_i)^2 + N_{ab}\sigma_{ab}^2 r_i^2\right]}{C_A N_A\left[N_b\sigma_b^2(1+r_i)^2 + N_{ab}\sigma_{ab}^2\right]} \qquad (4.5)$$

Thus this procedure is repeated until r_i shows an arbitrarily small change from one iteration to the next.

The c_A term in (4.1) includes all the costs involved in taking a sample of size n_A from frame A. Therefore, define c_A as

$$c_A = c_{sA} + c_{cA} + c_{oA} \qquad (4.6)$$

That is, c_A includes the cost of selection, the cost of classification into the proper domain and the cost of observation. The scheme used in constructing Y_d includes the costs of selection and classification from both frames. It may be possible to select elements to be included in the sample and remove the duplicated elements before the actual observations are made. In this situation, the costs of observation are diminished since n_d fewer observations are made; however, the elements drawn from domain ab must now be checked for duplication. The new cost equation then becomes

$$C' = (c_{sA}+c_{cA}+c_{oA})n_A + (c_{sB}+c_{cB}+c_{oB})n_B - n_d c_o' + c_d n_{ab}' n_{ab}'' ,$$

where

$$c_o' = \max(c_{oA}, c_{oB}) \qquad (4.7)$$

and c_d is the cost of determining duplications. In (4.7) it is assumed that each sampled element is classified into its proper domain upon selection. Thus c_d is multiplied by the product of n_{ab}' and n_{ab}''.

We conclude that if (4.1) is the correct cost equation, $\hat{Y}_d$ is preferred to $\hat{Y}_L$. The superiority of $\hat{Y}_d$ over $\hat{Y}_L$ is not so clear if (4.7) is the appropriate cost equation, but it should be noted that the increase in cost when employing $\hat{Y}_d$ may be small. It is quite likely the cost of observation, which would include travel costs and expenses of an enumerator, would be larger than the cost of checking elements for duplication in the office before the fieldwork is started. That is, if it is reasonable to assume that $n_d c_o'$ is of the same magnitude as $n_{ab}' n_{ab}'' c_d$, $\hat{Y}_d$ would be superior to $\hat{Y}_L$ or any of the other special forms of Hartley's original estimator. However, it may be impossible to determine the duplicated elements before the fieldwork is done. In this case, $\hat{Y}_d$ would require more expense than $\hat{Y}_L$.

REFERENCES

1. Bershad, M.A., "A Sample Survey of Retail Stores," Sample Survey Methods and Theory, Vol. I, New York: John Wiley and Sons, Inc., 1953, 516-58.

2. Cochran, R. S., "The Estimation of Domain Sizes when Sampling Frames are Interlocking," Proceedings of the Social Science Section of the American Statistical Association Meeting, Washington, D.C., 1967.

3. Comstock, G. W., et. al., "Validity of Interview Information in Estimating Community Immunization Levels," Health Services Reports 88 (October, 1973), 750-7.

4. Fuller, W. A., "Estimation Employing Post Strata," Journal of the American Statistical Association, 61 (December, 1966), 1172-83.

5. Hartley, H. O., "Multiple Frame Surveys," Proceedings of the Social Science Section of the American Statistical Association meeting, Minneapolis, Minnesota, 1962.

6. Kish, L., Survey Sampling, New York: John Wiley and Sons, Inc., 1965.

7. Lund, R. E., "Estimators in Multiple Frame Surveys," Proceedings of the Social Science Section of the American Statistical Association meeting, Pittsburg, Pennsylvania, 1968.

8. Serfling, R.E., Cornell, R.G. and Sherman, I. L., "The CDC Quota Sampling Technique with Results of 1959 Poliomyelitis Vaccination Surveys," American Journal of Public Health, 50 (December, 1960), 1847-57.

A LATENT DEMAND FOR HOUSING: A PRELIMINARY STUDY

J.W. Byler and Stephen Gale, University of Pennsylvania

1. INTRODUCTION

Population movements are obviously related to changes in the demand for housing: they are, for example, both a cause of housing starts, vacancies, and changes in occupancy, as well as an effect of shifts in housing policies, zoning changes, and the cost of housing units. Aside from studies on interregional migration patterns, however, very little is known about the detailed character of the linkages between individual (and household) movement propensities and related shifts in occupancy conditions.[1] As part of a general research effort on this issue, in this paper we shall offer a preliminary analysis of one portion of the population which may be regarded as having a high latent demand for housing, (i.e., that subpopulation whose potential for demanding future housing is not necessarily reflected by their present status in the housing market).[2]

Current research on migration and the demand for housing has usually assumed that the migrating or housing consuming unit is relatively homogeneous in the sense that the family unit acts as a whole. As such, the determination and interpretation of the causes and effects of migration has usually been based on information concerning the economic, demographic, and social conditions of heads-of-households; the head-of-household is, in effect, treated as if he or she embodies the family unit's stage in the life cycle, income, educational attainment, occupational class, and so on.[3] For certain purposes, this may provide a sufficient characterization, but at the local level where a knowledge of the short-term, sequential variations in housing demand is crucial for many planning purposes, there is also a need to understand the differential characteristics of the several parts of households, their predilections toward independent household formation, and their concomitant effects on the demand for housing.

The process of household formation is clearly very complicated. On the one hand, it is a function of the aging process, marital patterns, and other purely demographic conditions. And, on the other hand, it is related to the economic, social, and kinship status of the members of a household unit: children over 18 years of age may reside at home during their post-secondary education; widowed relatives may join an extended family unit to gain economic security; unrelated individuals may form some sort of cooperative arrangement, and so on. In each case, the composition of the household cannot be treated in the usual homogeneous fashion. The varying needs and propensities for movement and housing consumption of the several parts of the household must all be treated separately in order to assess the overall prospects for future housing consumption. In effect, within each household there exists the components for several kinds of latent conditions for housing demands which must be accounted for in any assessment of present and future occupancy conditions.

One further point should be mentioned here. Within what we have called lodgers with respect to a latent demand for housing, there is also a "true" lodger population: individuals unrelated to other members of a household who, even while comprising family units of their own, have chosen to obtain dwelling accomodations (such as a room) from a family in return for a rental fee. Though the present study includes these individuals, it should be kept in mind that the notion of latent demand is not directly applicable in such cases; "true" lodgers have already expressed their housing consumption needs and, while opting for somewhat similar accomodation to those persons we have termed lodgers, they cannot be presumed to be acting for the same kind of reasons as the remainder of the sub-population.

The purposes of the present study are thus: to describe the characteristics of the lodger population for one metropolitan area in order to identify its component subgroups and to examine the attributes of the households and dwelling units within which these subgroups reside. Of specific interest are the age, sex, and employment distributions of the lodgers. In addition, this investigation will explore sources of variation in the occupancy patterns of the lodger subpopulations which result from differences in the value and tenure status of their dwelling unit and in the race, sex, age, income, and marital status of the heads-of-households of these units.

2. DATA AND METHODS

Until quite recently, the analysis of the links between individual movement propensities and the characteristics of the housing units consumed has depended on rather oblique inferences from the aggregate cross-sectional properties of data from the decennial federal census. However, under funding from national, state, and local agencies, Wichita-Sedgwick County (Kansas) has recently developed a yearly enumeration of population and housing characteristics for the whole of the city-county area which provides an extremely rich micro-level data base for the analysis of population movements, household formation patterns, changes in occupancy structure, and so on.[4] More specifically, the records currently available permit the identification of those individuals who reside as lodgers within a household unit, together with their economic and demographic characteristics, the nature of the household unit and dwelling unit in which they reside, and portions of their past residential mobility experience. In effect, this data source provides the basis for the development of descriptive indices of the joint distribution of lodgers, their characteristics, their kinship patterns, and the kinds of structures they occupy—as well as the capability of identifying the changing characteristics of these individuals.

For the present examination we have obtained the following information for all lodgers for the enumeration year 1973:

(i) The name, age, sex, and employment status of each individual in Wichita-Sedgwick County who is eighteen or over, resides within a household of which he or she is not the head, and is not the spouse of the head-of-household.

(ii) The race, sex, age, income, and marital status of the head-of-household.

(iii) The relationship of the lodger to the head-of-household.

(iv) The number of lodgers in the household.

(v) The value, tenure status, and the number of families in the dwelling unit.

From each of these characteristics, models of data, in the form of multi-way contingency tables, have been tabulated as the principal descriptive indicators.[5] The relevant tables are presented in Section 3. Analysis of these data includes (a) an interpretation of the tables and (b) an examination of classes of homogeneities in the tables using the methods developed by Goodman and others.[6] Again, our intent here is to provide only some preliminary insights into the characteristics of the lodger subpopulation and some characteristic differences among the several parts of this subpopulation.

3. INTERPRETATION AND ANALYSIS

The magnitude of the lodger population is, perhaps, its most important attribute with respect to its potential for effecting the future demand for housing. The 1973 Wichita-Sedgwick County Enumeration recorded 27,391 persons fitting the definition of what we have termed lodgers. This is approximately 13% of the total population of Sedgwick County, and 22% of that portion of the total population of the county over 18 years of age. Additionally, it represents a potential 25% increase in the total number of households in the county—i.e., if all of the lodgers were to form their own single-person households. Though this eventuality is improbable, it is quite likely that a significant fraction of this group of people will seek independent accomodations within the next few years. With respect to their latent demand for housing, then, it is clearly of importance to determine which kinds of current lodgers will move into the housing market and in what numbers. And, while the available cross-sectional data cannot provide estimates of this longitudinal aspect of the latent demand for housing, we are able to identify several subgroups of lodgers which appear to have differential rates of demand for independent housing accomodations.

Even a cursory inspection of the age distribution of the lodger population suggests the nonhomogeneity of this group (See Figure1.). Forty-four percent of the lodgers are under 21 years old; fully two-thirds are less than 26; another seven percent are between 26 and 36; the remaining 30% of the lodger population is fairly equally apportioned in the age range from 36 to 85. When disaggregated by the relationship of the lodger to his head-of-household, the graph of the age distribution provides an indication of the existence of several distinct subgroups within the lodger population. Thus, nearly 38% of the entire group is made up of children, aged 18 to 20, living with at least one of their parents. Another 16% of the group are children aged 21 to 25, living with at least one of their parents. From age 51 on, however, children living at home are increasingly out-numbered by other related lodgers. The age distribution of related individuals other than children has a bimodal form somewhat different from that of unrelated individuals: it drops off from a peak in the lowest age group, rising again after age 35 to another peak between 76 and 81. The distribution of the non-related lodgers, on the other hand, peaks early, between age 21 and 25, drops until age 46 when it again begins to rise, reaching another peak between age 61 and 65. The steep drop in numbers of both related and non-related lodgers after the age of 25 suggests high rates of movement into the housing market for those beyond this age.

Clearly, the trough in the middle years, between age 36 and 46, reflects that period in one's life when one is least likely to be a member of the lodger population; by then most children appear to have left their parents' home to establish their own. Moreover, the large group of non-related boarders 18 to 25 years of age suggests a staging period during which many young people leave their parents' home, yet are either unwilling or unable to establish their own independent households. The second peak in the distribution of non-related lodgers indicates older persons moving into true boarding situations, while the corresponding peak in the distribution of related lodgers other than children indicates elderly persons moving into extended family situations.

The joint distribution of the lodger subpopulation's age and sex with their employment and kinship status (Table 1) provides additional information on the characteristics of the subgroups of the lodgers. Marked differences now appear in the distributions with regard to the sex of the lodgers. For example, female lodgers constitute a smaller persentage of the total female population age 18 to 45 than do male lodgers in the same age group. After age 45, however, the reverse is true: females are more likely than males to be lodgers. Additionally, for lodgers related to the head-of-household a higher percentage of the women than the men are not in the labor force, regardless of the age category; the accompanying absence of income also suggests that female lodgers would be less likely to enter the housing market to seek independent accomodations in the future.

Of the lodgers over 18 who are living with their parents, 40% are males between age 18 and 21 and 30% are females between age 18 and 21. While the number of both male and female lodgers drops off beyond this age, it falls more rapidly for males than for females. (See Table 2.) Also, the distribution of females in this group is bimodal as opposed to the unimodal distribution of the comparable group of males.

Elderly female relatives other than children of the head-of-household constitute the largest group of lodgers other than children age 18 to 35. Women lodgers far out-number male lodgers in the over 65 age class. Note, however, that this difference is not merely a reflection of the differential survival rate for males and females. Women in this age class represent 5% of the total female population over age 65, whereas the men represent only 1.9% of the total males over age 65. Two factors other than the differential survival rate might be presumed to be associated with this discrepancy. First, there may be a cultural bias in the definition of the head-of-household; in those households where an older man is living with his son or daughter, there may be a tendancy to regard the older man as the head of the household and the child as the lodger, while in similar households where the elderly parent is female she may be regarded as the lodger and her child as the head of the household. Second, elderly men may live by themselves to a greater extent than do older women.

For the non-related boarders, what we have termed the "true" lodgers, differences with respect to sex are not as apparent as those exhibited in Table 1. For example, the age distributions are bimodal for both males and females and the distribution peaks in the 18 to 21 year old category for the females, while it crests in the 22 to 35 category for the males. Furthermore, the distribution with regard to employment is virtually the same for all age categories except over age 65 where a larger percentage of men are in the labor force. Otherwise, most members of each age and sex class are active in the labor market; a majority of these are employed full time.

As the foregoing discussion indicates, the lodger population consists of several subgroups which may be characterized by different distributions with respect to age, sex, employment and kinship status. Moreover, it also appears that there are strong dependencies among these variables (i.e., in the sense that any one or even any single combination of two or three attributes is not sufficient to describe the general characteristics of the lodger population as a whole). This point is further supported by the application of Goodman's tests for properties of multi-way contingency tables; with the data in Table 1 we obtain a χ^2 value of 34.33 with 30 degrees of freedom for the model which includes all second order interactions; the best model with fewer terms gives a p-value of only .003, thus indicating that only the model in which each triple of variables are jointly co-dependent would suffice to estimate the table.

Given the non-homogeneity of the lodger population, we now turn our attention to the occupancy patterns exhibited by this group. In terms of the future demand for housing, the current occupancy patterns of lodgers are important for at least two reasons. First, previous experience to some degree conditions one's expectations and desires; the future demand for housing by present members of the lodger population may therefore be in part regarded as a function of current occupancy patterns. Second, the heterogeneity of the lodger population suggests that the housing market is currently meeting a variety of needs for this group; lodgers may therefore be presumed to respond in a variety of ways to differential changes in the supply of the various components of the housing market. In short, shifts in supply may both encourage and discourage shifts in the demand for independent accomodations by both the lodgers themselves and the population as a whole.

Our investigation of the current occupancy patterns of lodgers falls into two parts: a discussion of the kinds of <u>households</u> within which lodgers reside and an examination of the kinds of <u>dwelling units</u> characteristic of these households.

The previous discussion of the subgroups of the lodger population suggests that the propensity to be a lodger is, in part, a function of one's life cycle stage. Using the head-of-household as a surrogate for his life cycle status, it appears from Figure 2 that the propensity to provide housing for lodgers is similarly related to the household's status in the life cycle. By dividing the total lodger population into the kinship classes, it becomes evident that the various age groups of householders contribute differentially to previding housing for each of these groups. Young persons aged 18 to 25, for example, provide the bulk of the housing for non-related individuals. After this initial peak, the distribution of non-related lodgers drops off to a fairly constant level with respect to the age of the householder. The graph of the number of children lodging with parents of a given age, on the other hand, peaks much later, in the age 46 to 50 year old range, falling off rapidly after age 50. The bulk (64%) of the children living with their parents after age 18 are, as one would expect from their own age distribution, 23 to 30 years younger than their parents; the drop in the number of children living at home is then reflected by the corresponding decline in the number of children residing with parents over age 50. The number of related individuals other than children rises steadily, peaking in the 56 to 60 year old range (for the heads-of-households). Assuming a generational period of approximately 25 years as suggested above, this bulge probably reflects elderly parents sharing accomodations with their off-spring.

In addition to the age of the head-of-household, marital status is an important component of life cycle status. We might, for example, expect that the presence or absence of a spouse for the head-of-household would influence the propensity to house lodgers. The distribution of the relationship of the lodger to the householder with respect to the presence or absence of a spouse for the householder (Table 3) indicates no real differences between the three groups when a spouse is not present. When a household's spouse is present, however, the lodgers are almost exclusively (96%) related individuals; by far the largest group of these lodgers are children of the head-of-household. The distribution of the number of lodgers in a household is also highly dependent on the presence of the spouse. Households where no spouse is present are more likely to have three or more lodgers than are those where the householder's spouse is present. (See Table 4.)

The sex of the head-of-household appears to have little effect on the number of lodgers housed, though it does appear to influence the <u>kind</u> of lodgers one takes in. The distribution of the number of lodgers in a household is virtually the same for both male and female headed households (Tables 5 and 6): approximately 82% of the households headed by each sex have no lodgers; 16.7% have one or two lodgers; around 1% of each group house three or more lodgers. Female headed households are much more likely than are male households to consist of non-related lodgers and relatives other than children over age 18.

The differences in the kind of lodgers a householder boards is additionally related to his own race. In households headed by whites, 68% of the lodgers are children over age 18, while this is true for only 60% of black households. Instead, 30% of the lodgers in black homes are relatives other than children, while the comparable figure for whites is only 18.5%. This suggests that blacks are more likely to live in extended family situations than are whites. Additionally, the distribution of the number of lodgers by race of the head-of-household indicates that a higher percentage of black households board lodgers in all kinship classes than do white households.

The distribution of the number of lodgers by the relationship of the lodger to the head-of-household (Table 7) indicates that in a majority of cases the lodger is a child living in a household with just his parents. The next largest group is households with one child over 18 living with his parents and one other lodger. Relatives other than children and non-related lodgers similarly tend to be the only boarders in a household.

The heterogeneity of the lodger population would not lead us to expect the existence of a simple relationship between family income and either the number or kinship relation of the lodgers in a household. Nevertheless, both lower and upper income families house lodgers as a higher rate than do families in the lower-middle income bracket (Table 8.). As income increases, however, the percentage of the lodgers who are children over 18 tends to increase (Table 9.), while the percentages of both other related and non-related lodgers decrease as income increases. We should emphasize that the data reflect total family income and not solely that of the head-of-household; a large percentage of the lodgers who are children over 18 are employed and would be contributed to the family income, while elderly, unemployed other relatives and non-related individuals would usually not be contributing directly to the total family income.

Having examined the kinds of households within which lodgers reside, we now turn to an examination of the kinds of dwelling units characteristic of these households. Nearly 80% of the housing units in Wichita-Sedgwick County are single family homes. Thus, it is not surprising that most of the lodgers live in single family units. It is thus interesting to note that, even in light of this fact, a higher percentage of the single family homes house lodgers than do multiple family units. Children over 18, for example, are more likely to reside in single family dwellings than are other relatives or non-related lodgers (Table 10.). Controlling for the total number of units which are owned and of those which are rented, households with lodgers are about half as likely to live in rented accomodations as in owned quarters (11.5% and 22.6%, respectively). Related lodgers tend to live in accomodations owned by the householder whereas non-related lodgers are more likely to live in rented quarters (Table 11.). The latter trend is also reflected in the distribution of the number of lodgers with respect to the value of owned housing units and with respect to the monthly rent of leased accomodations: for owned units, the percentage of households with three or more lodgers remains more or less constant, while that with one or two lodgers increases as the value of the home increases (Table 12.). For rented units the distribution of households in terms of the number of lodgers is virtually the same for all rental categories (Table 13.).

4. CONCLUSIONS

The foregoing analysis has detailed the non-homogeneities in the characteristics of the lodger population for Wichita-Sedgwick County, Kansas. In addition, it has indicated some of the characteristic occupancy patterns of the lodgers relative to demographic and economic attributes of the households and dwelling units within which they reside. Other available evidence suggests, for the most part, that, in terms of a large number of socioeconomic and demographic characteristics, the Wichita area is typical of many other American cities of similar size.[7] Therefore, while the inferences of this study are specific to Wichita in 1973, it may be surmised that these results are applicable to other cities as well. Note also that, due to the cross-sectional nature of the available data, we have not attempted to estimate the differential latent demand for housing of the various sub-groups identified for the lodger population. Although such an analysis would be difficult with the kinds of data available from the Federal decennial census (i.e., in that it is neither sufficiently detailed nor capable of being structured as longitudinal records), the records of the Wichita Enumeration provide such a data base. As the files for the enumeration become available for years subsequent to 1973, it will thus be possible to calculate rates of movement of the subgroups of lodgers into the housing market.

FOOTNOTES

[1] See W. G. Grigsby, Housing Markets and Public Policy, (Philadelphia: University of Pennsylvania Press, 1963).

[2] For a description of this program see S. Gale and E. G. Moore, "A Research Program for the Description and Examination of Occupancy Shifts and Neighborhood Change," Working Paper No. 1, Research on Metropolitan Change and Conflict Resolution, Peace Science Department, University of Pennsylvania, (Philadelphia: 1973).

[3] M. H. David, Family Composition and Consumption, (Amsterdam: North-Holland Publishing Co., 1962).

[4] See R. A. Gschwind, "The Intergovernmental Enumeration, Wichita-Sedgwick County, Kansas, 1971-1973", Working Paper No. 2, Research on Metropolitan Change and Conflict Resolution, Peace Science Department, University of Pennsylvania, (Philadelphia: 1973), for a detailed account of the properties of these data.

[5] See E. G. Moore and S. Gale, "Comments on Models of Occupancy Shifts and Neighborhood Change," In E. G. Moore (ed.), Models of residential Location and Relocation in the City (Evanston: Northwestern University Studies in Geography, No. 20, 1973), pp. 135-173.

[6] Goodman, L., "Analysis of Cross-Classified Data: Independence, Quasi-independence and Interactions in Contingency Tables with or without Missing Events," Journal of the American Statistical Association, vol. 63 (1968), pp. 1091-1131 and also S. E. Fineberg, "The analysis of incomplete multi-way contingency tables," Biometrics, Vol. 28, (1972), pp. 177-201.

[7] R.M. Berger, "Wichita in a System of Cities," Working Paper No. 6, Research on Metropolitan Change and Conflict Resolution, Peace Science Department, University of Pennsylvania, (Philadelphia: 1974).

Table 1A. Distribution of children over 18 with respect to age, sex, and employment status.

		EMPLOYMENT			
		full-time	part-time	unemp	not in force
MALE	18	627	631	83	688
	19-21	1625	897	122	960
	22-35	1408	349	88	404
	36-45	137	49	13	47
	46-65	78	10	4	40
	66+	2	0	0	5
FEMALE	18	342	569	57	778
	19-21	858	762	180	1041
	22-35	658	142	49	330
	36-45	90	31	2	29
	46-65	107	19	3	73
	66+	7	1	1	51

Table 1B. Distribution of other related lodgers with respect to age, sex, and employment status.

		EMPLOYMENT			
		full-time	part-time	unemp	not in force
MALE	18	52	41	12	53
	19-21	112	24	11	26
	22-35	224	19	12	34
	36-45	47	1	0	13
	46-65	69	4	2	55
	66+	33	5	4	164
	18	16	32	11	112
	19-21	63	36	11	55
	22-35	101	19	11	64
	36-45	31	4	1	22
	46-65	123	15	3	201
	66+	53	19	11	818

Table 1C. Distribution of non-related lodgers with respect to age, sex, and employment status.

		EMPLOYMENT			
		full-time	part-time	unemp	not in force
MALE	18	43	14	10	27
	19-21	292	71	11	63
	22-35	461	68	16	55
	36-45	49	4	1	0
	46-65	41	6	0	23
	66+	58	5	1	39
FEMALE	18	48	20	9	33
	19-21	211	67	17	66
	22-35	256	30	11	32
	36-45	27	1	0	6
	46-65	64	6	1	34
	66+	54	3	1	85

Table 2. Distribution of lodgers with respect to age, sex, and relationship to head-of-household.

	AGE					
	18	19-21	22-35	36-45	46-65	66+
CHILDREN OVER 18						
male	2561	3812	2349	246	132	11
female	1850	2916	1240	152	255	234
OTHER RELATIVES						
male	151	168	282	61	130	206
female	181	159	363	56	343	911
NON-RELATIVES						
male	103	450	626	57	77	147
female	114	371	346	38	117	168

Table 3. Distribution of lodgers with respect to marital status of head-of-household.

	MARITAL STATUS	
	Spouse not present	Spouse present
CHILDREN OVER 18	3304 (37.5)	15090 (81.2)
OTHER RELATIVES	2560 (29.1)	2830 (15.2)
NON-RELATIVES	2943 (33.4)	664 (3.8)
TOTAL	8807	18584

Table 4. Distribution of number of lodgers in household by marital status of head-of-household.

		MARITAL STATUS	
		Spouse not present	Spouse present
NUMBER OF LODGERS IN HOUSEHOLD	1	4806 (74.4)	10360 (74.1)
	2	1197 (18.5)	2864 (20.5)
	3	302 (4.7)	581 (4.2)
	4+	151 (2.3)	165 (1.2)
	total	6456	13970

Table 5. Distribution of types of lodgers with respect to sex of head-of-household.

	SEX OF HEAD-OF-HOUSEHOLD	
	male	female
CHILDREN OVER 18	15409 (73.0)	2829 (46.9)
OTHER RELATIVES	3513 (16.7)	1821 (30.2)
NON-RELATIVES	2172 (10.3)	1386 (23.0)
TOTAL	21094	6036

Table 6. Distribution of number of lodgers with respect to sex of head-of-household.

		SEX OF HEAD-OF-HOUSEHOLD	
		male	female
NUMBER OF LODGERS IN HOUSEHOLD	1	11728 (74.1)	3286 (74.6)
	2	3229 (20.4)	795 (18.0)
	3	655 (4.1)	218 (4.9)
	4+	207 (1.3)	108 (2.5)
	total	15819	4407

Table 7. Distribution of number of lodgers in household by the relationship of the lodger to the head-of-household.

	NUMBER OF LODGERS IN HOUSEHOLD				
	1	2	3	4+	total
CHILDREN OVER 18	10643	2822	533	111	14109
OTHER RELATIVES	2626	720	223	141	3710
NON-RELATIVES	1897	519	127	60	2603

Table 8. Distribution of number of lodgers in household with respect to family income.

		FAMILY INCOME		
		under $4000	$4000–$9,999	$10,000+
NUMBER OF LODGERS IN HOUSEHOLD	0	85.8%	92.2%	86.6%
	1–2	13.2%	7.4%	12.7%
	3+	1.0%	0.4%	0.7%

total sample size: 101,681

Table 9. Distribution of types of lodgers with respect to family income.

	FAMILY INCOME		
	under $4000	$4000–$9,999	$10,000+
CHILDREN OVER 18	733	2756	6878
OTHER RELATIVES	526	1146	1026
NON-RELATIVES	414	738	722
TOTAL	1673	4640	8626

Table 10. Distribution of types of lodgers with respect to the number of units in structure.

	NUMBER OF UNITS	
	SINGLE FAMILY	MULTI-FAMILY
CHILDREN OVER 18	13038	815
OTHER RELATIVES	3709	676
NON-RELATIVES	1880	1293
TOTAL	18627	2784

Table 11. Distribution of types of lodgers with respect to tenure status of dwelling unit.

	TENURE	
	OWNED	RENTED
CHILDREN OVER 18	14627 (75.8)	2414 (39.3)
OTHER RELATIVES	3566 (18.5)	1495 (24.3)
NON-RELATIVES	1116 (5.8)	2231 (36.3)
TOTAL	19309	6140

Table 12. Distribution of number of lodgers in household with respect to the value of owned units.

	NUMBER OF LODGERS		
	0	1–2	3+
VALUE $0–$4,999	88.4%	10.7%	0.9%
$5,000–$9,999	85.3%	13.7%	1.0%
$10,000–$14,999	83.3%	15.8%	0.9%
$15,000–$19,999	82.6%	16.5%	0.9%
$20,000–$24,999	80.9%	18.1%	1.0%
$25,000+	80.1%	18.8%	1.1%

total sample size: 74,476

Table 13. Distribution of number of lodgers in household by monthly rent.

	NUMBER OF LODGERS		
	0	1–2	3+
RENT $0–$49	88.3%	11.1%	0.6%
$50–$99	90.2%	9.2%	0.6%
$100–$149	88.6%	10.6%	0.8%
$150–$199	86.7%	12.7%	0.6%
$200+	88.1%	10.8%	1.1%

total sample size: 39,625

ERRORS IN VARIABLES FOR THE SIMPLE LINEAR MODEL: THE EFFECTS OF CORRELATED ERRORS OF MEASUREMENT ON INTERVAL ESTIMATION AND HYPOTHESIS TESTING

John J. Chai and George Frankfurter, Syracuse University

1. INTRODUCTION

1.1 Purpose and scope: It is well known that the Ordinary Least Squares Estimator (OLSE) of the regression coefficient for the simple linear regression model

$$Y = \alpha + \beta X + \varepsilon$$

is biased and not consistent when both variables are subject to errors of measurement [9]. A number of studies in this area largely deal with simple models [e.g., 1, 3, 6, 8, 10]. In these studies each of the "true" values and the corresponding error terms are assumed to be uncorrelated. Also, the two error terms are assumed to be uncorrelated. There were some exceptions however. Chai [1] and Cochran [3] disucss some effects of correlation between the two error terms and between the "true" independent variable (X) and its errors of measurement. Although no empirical results have been published, some have mentioned possible cases where correlation between the "true" independent variable and its measurement errors may exist.

The purpose of this paper is to examine, for a range of different values of ρ_{XY} and other relevant parameters, the bias of the Ordinary Least Squares Estimator of β due to

(1) Correlation between the errors of measurement of the "true" values (i.e., ρ_{de}),

(2) Correlation between each of the errors of measurement and the respective "true" values (i.e., ρ_{Xd}, ρ_{Xe}, ρ_{Yd}, ρ_{Ye}).

Further, the paper shows the effect of the bias of the OLSE of β on the confidence interval estimation for β and the hypothesis testing for $\beta=0$ v.s. $\beta\neq0$.

1.2 The rationale for this study: In econometric research it quite often happens that there would be certain relationships between errors of measurement and the "true" variables and between measurement errors themselves. For example, in measuring the values of housing, the error of measuring expensive units may be positively correlated with the "true" values of the units, because there exists a tendency to undervalue housing in ghetto areas and overestimate property value in suburban areas. Another case may be measurement of family income. The error of measurement might be negatively correlated with the "true" family income, because of the tendency of under-reporting of higher income and of over-reporting lower income.

Still another group of examples could be found in time series data. If the independent variable is a function of time, and if the errors of measurement for the "true" value for the earlier period are greater (because of lack of memory, for example) than the ones for the recent measurement, then one would expect a correlation not only between the "true" value and the error of measurement, but also a correlation between errors over successive time periods.

In the case of a stock market index it may happen that the majority of the securities in the index over-react to the general trend of the market. They go up more on an "up-market" and go down more on a "down-market" than the average movement of the universe of all securities.

All these socio-economic interrelationships between the "true" values of the two variables and between their respective errors of measurement warrant a careful examination of the effects of various errors of measurement on the OLSE of β. In section 2 the mathematical model, the bias, and the bias relative to the true β are presented. A discussion on the effect of the bias on the OLSE is presented in Section 3.

2. THE MODEL

2.1 Bias: Let Y and X respectively be the "true" values for a population element and let the relationship between Y and X be given by a linear regression model

$$Y = \alpha + \beta X + \varepsilon \tag{1}$$

where ε is a random residual of the regression. We assume that both X and Y are subject to errors of measurement and let the values actually obser observed be:

$$x = X + d \tag{2}$$

$$y = Y + e \tag{3}$$

where d and e represent measurement errors. Substituting (2) and (3) into (1) results in

$$y = \alpha + \beta x + u \tag{4}$$

$$\text{where } u = -\beta d + e + \varepsilon \tag{5}$$

we assume that X, Y, d, and e jointly follow a multi-variate normal process with mean vector $[\mu_X, \mu_Y, \mu_d, \mu_e]$ and variance-covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{Xd} & \sigma_{Xe} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YD} & \sigma_{Ye} \\ \sigma_{Xd} & \sigma_{Yd} & \sigma_d^2 & \sigma_{de} \\ \sigma_{Xe} & \sigma_{Ye} & \sigma_{de} & \sigma_e^2 \end{pmatrix}$$

Hence, the regression given by (4) above is linear according to Lindley [11].

Without loss of generality we assume that all means are put equal to zero ($\alpha=0$) for the purpose of deriving the bias of the OLSE. Let the OLSE of β be denoted by b. Then $b = \Sigma xy/\Sigma x^2$ and for the regression of y on x, $E(b) = E\{xE(y|x)/\Sigma x^2\}$. Furthermore, the regression of y on x is linear so if we write $E(y|x) = \lambda x$ we have $E(b) = \lambda$. λ is found, therefore, by evaluating $E(y|x)/x$. But first, evaluate $E(y|x)$.

$E(y|x) = \beta x + E(u|x) = \beta x + E(e|x) - \beta E(d|x) + E(\varepsilon|x)$. Since $\sigma_{X\varepsilon} = 0$ and $\sigma_{d\varepsilon} = 0$, we have

$$E(y|x) = \beta x + E(e|x) - \beta E(d|x). \tag{6}$$

We now evaluate the last two terms of the right-hand side of equation (6) each divided by x as follows:

$$\frac{E(d|x)}{x} = \frac{E(dx)}{E(x^2)} = \frac{E\{d(X+d)\}}{E(X+d)^2} = \frac{\sigma_{Xd} + \sigma_d^2}{\sigma_X + 2\sigma_{Xd} + \sigma_d^2}$$

$$\frac{E(e|x)}{x} = \frac{E(ex)}{E(x^2)} = \frac{E\{e(X+d)\}}{E(X+d)^2} = \frac{\sigma_{Xe} + \sigma_{de}}{\sigma_X^2 + 2\sigma_{Xd} + \sigma_d^2}$$

Thus, by substitution

$$\lambda = \beta + \frac{\sigma_{Xe} + \sigma_{de} - \beta(\sigma_{Xd} + \sigma_d^2)}{\sigma_X^2 + 2\sigma_{Xd} + \sigma_d^2} \tag{7}$$

The bias is therefore the second term on the right-hand side of (7). Let the bias be denoted by B_b. Rewriting it in terms of correlation coefficients ρ_{Xd}, ρ_{Xe}, and ρ_{de}, we obtain

$$B_b = \frac{\rho_{Xe}\sigma_X\sigma_e + \rho_{de}\sigma_d\sigma_e - \beta(\rho_{Xd}\sigma_X\sigma_d + \sigma_d^2)}{\sigma_X^2 + 2\rho_{Xd}\sigma_X\sigma_d + \sigma_d^2} \tag{8}$$

We define the relative biases as follows:

(1) Bias relative to β: B_b/β (9)

(2) Bias relative to σ_b: B_b/σ_b (10)

where σ_b is the standard error of b, i.e.,

$$\sigma_b = \sigma_u / \sqrt{\Sigma(x-\bar{x})^2}$$

and from (4) and (5)

$$\sigma_u^2 = \Sigma(y - \alpha + \beta x)^2 = \Sigma(\varepsilon + e - \beta d)^2$$
$$= \sigma_\varepsilon^2 + \sigma_e^2 + \beta^2\sigma_d^2 - \beta\sigma_{de} \tag{11}$$

2.2 Effect of bias on confidence interval for β: Assuming that we have a multivariate normal distribution of X, Y, d, and e, the OLSE of β, b is normal. Let b_L and b_U respectively be defined as follows:

$$b_L = \beta - |z|\ \sigma_b$$

$$b_U = \beta + |z|\sigma_b$$

where

$$z = \{b - E(b)\}/\sigma_b$$

Now let Z_L and Z_U respectively be the standardized values of b_L and b_U. Then

$$Z_L = -|z| - B_b/\sigma_b$$

$$Z_U = |z| - B_b/\sigma_b$$

The z indicates the <u>desired</u> level of confidence and Z_L and Z_U indicate the actual level of confidence realized.

2.3 Effect of bias on testing of hypotheses $\beta=0$ v.s. $\beta\neq0$: In testing of the null hypothesis $H_0: \beta=0$ against the alternative $H_1: \beta\neq0$ with Type I error controlled, the critical values (action limits) in terms of standardized values, A_1 and A_2 actually realized are

$$A_1 = 0 - |z| - B_b/\sigma_b$$

$$A_2 = 0 + |z| - B_n/\sigma_b$$

3. DISCUSSION

First we consider bias relative to β. The relative biases for different parameters are shown by three figures attached. Each figure contains three graphs -- the one on the left is for $\rho_{XY} = .1$, the center one is for $\rho_{XY} = .5$, and the one on the right is for $\rho_{XY} = .9$. Each graph shows the relative biases for $\rho_{de} = \pm .1, .5$, and $.9$ and for various values of the error variances σ_d^2/σ_X^2 and σ_e^2/σ_Y^2. Figure 1 is for $\rho_{Xd} = 0$ and $\rho_{Xe} = 0$, Figure 2 is for $\rho_{Xd} = 0.9$ and $\rho_{Xe} = 0.9$, and Figure 3 is for $\rho_{Xd} = -0.5$ and $\rho_{Xe} = -0.5$.

The main points of these graphs are: (1) for given ρ_{Xd}, ρ_{Xe}, and ρ_{de}, the greater the ρ_{XY}, the smaller the relative bias for the ranges of the error variances considered. But the magnitude of the relative bias is substantial even when $\rho_{XY}=0.9$ $\rho_{Xd} = 0$, and $\rho_{Xe} = 0$ (see Figure 1). As ρ_{Xd} and ρ_{Xe} increase, the relative biases increase -- with much more variation over different ρ_{de}'s for low ρ_{XY} than for high ρ_{XY}. For $\rho_{Xd}<0$ and $\rho_{Xe}<0$, the relative bias varies more over different ρ_{de}'s for given ρ_{XY} than the relative bias does for $\rho_{Xd}>0$ and $\rho_{Xe}>0$. (2) For given ρ_{Xd} and ρ_{Xe} the relative bias varies more for low values of σ_d^2/σ_X^2 and ρ_{XY} than for higher values of the same parameters; conversely, the relative bias varies more for high values of ρ_{de} and σ_e^2/σ_Y^2 than for lower values of ρ_{de} and σ_e^2/σ_Y^2. When $\rho_{Xd}<0$ and $\sigma_{Xe}<0$ the relative bias is highly sensitive to change in σ_d^2/σ_X^2, σ_e^2/σ_Y^2 and ρ_{de} (see Figure 3).

Next we consider the effect of the bias on interval estimation of β and on hypothesis testing. It is well known that if an estimator is biased the probability of including the parameter in the confidence interval is reduced and that the amount of loss in probability in general is quite serious as the bias relative to the standard error of the estimator is greater than 0.2. Table 1 presents the actual probability realized for a 95 per cent confidence interval for β. The data presented in this table are for $\rho_{XY} = .55$ only. According to the table, the probability decreases as $|\rho_{Xd}|$ increases for a given sample size (n) and for a given ρ_{de}, except for $\rho_{de} = 0$ and $-.56 \le \rho_{Xd} \le -.86$. The probability also decreases rather rapidly as n increases (except for $\rho_{de} = 0$ and $\rho_{Xd} = -.86$). This is expected since σ_b^2 gets smaller as $\Sigma(x-\bar{x})^2$ decreases when n increases, whereas the bias B_b is independent of n.

In the case of hypothesis testing, the same consequence is realized as in the case of interval estimation. This time, however, the probability of accepting the null hypothesis when it is indeed true.

In short, what we have shown and reemphasized (on the basis of the model which is more general and perhaps more realistic) is the danger of using the OLSE of β when both variables are subject to errors of measurement. As many statisticians are aware, there has been some progress, however limited, in finding ways of improving estimation methods and a way of actually assessing the various error parameters, but there is much need for more research in this area.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the useful comments given by the referees and Professor Herbert Phillips of Temple University and the computer assistance by Mr. Michael Legault, Ph.D. student, School of Management, Syracuse University.

REFERENCES

[1] Chai, John J. (1971). "Correlated Measurement Errors and Least Squares Estimator of the Regression Coefficient," Journal of the American Statistical Association, 66, 478-483.

[2] ____ and Frankfurter, G. M. (1973). "A Simulation Study of Errors of Measurement," Unpublished manuscript, Working Paper Series, School of Management, Syracuse University.

[3] Cochran, W. G. (1970). "Some Effects of Errors of Measurement on Multiple Correlation," Journal of American Statistical Association, 65, 22-34.

[4] ____ (1968). "Errors of Measurement in Statistics," Technometrics, 10, 637-666.

[5] ____ (1963). Sampling Techniques, 2nd ed., New York: John Wiley & Sons, Inc., 1963.

[6] DeGracie, J. S. and Fuller, W. A. (1972). "Estimation of the Slope and Analysis of Covariance when the Concommitant Variable is Measured with Error," Journal of the American Statistical Association, 67, 930-937.

[7] Hansen, M. H., Hurwitz, W. N. and Pritzker,L. (1964) "The Estimation and Interpretation of Gross Differences and the Simple Response Variance," Contributions to Statistics Presented to Professor P. S. Mahalanobis on the Occasion of His 70th Birthday, Calcutta: Pergamon Press.

[8] Horwitz, D. G., and Koch, Gary G. (1969). "The Effect of Response Errors on Measures of Association," in N. L. Johnson and H. Smith, Jr., eds, New Development Sampling, New York: John Wiley & Sons, Inc., 247-282.

[9] Johnston, J. (1960). Econometric Methods, first ed., 148-150, New York: McGraw-Hill Book Company.

[10] Koch, Gary G. (1969). "The Effects of Non-Sampling Errors on Measures of Association in 2 x 2 Contingency Tables," Journal of the American Statistical Association, 64, 851-864.

[11] Lindley, D. V. (1947). "Regression Lines and the Linear Functional Relationship," Journal of the Royal Statistical Society, Supplement, 218-244.

TABLE 1 ACTUAL CONFIDENCE LEVEL

REALIZED FOR 95 PER CENT CONFIDENCE

(For $\rho_{XY} = .55$)

		Sample Size			
ρ_{de}	ρ_{Xd}	10	25	50	100
.43	.86	.7985	.4840	.2709	.0505
	.56	.8555	.6406	.3707	.1075
	0	.9305	.8835	.7612	.6141
	-.56	.9333	.8817	.8300	.6985
	-.86	.4920	.0250	0	0
0	.86	.7157	.2946	.0401	0
	.56	.7580	.3783	.0401	.0000
	0	.8395	.5987	.2062	.0708
	-.56	.9219	.8300	.6772	.4840
	-.86	.9498	.9477	.9449	.9441
-.43	.86	.5714	.1492	.0150	0
	.56	.6217	.1922	.0078	0
	0	.6844	.3372	.0559	.0018
	-.56	.7454	.2451	.0582	.0021
	-.86	.7190	.1635	.0154	0

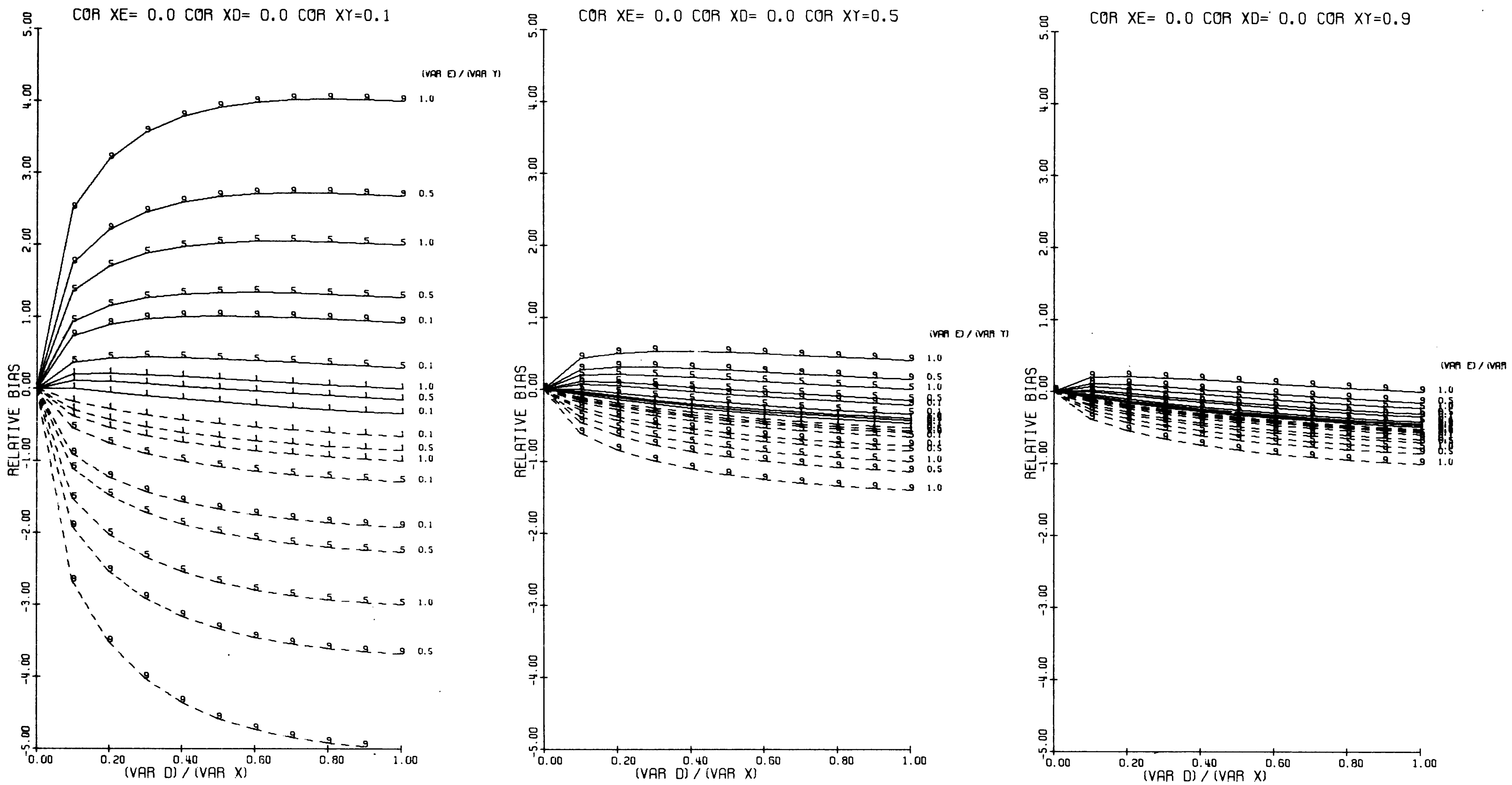


Figure 1 The relative bias of the OLSE for $\rho_{Xe} = 0$, $\rho_{Xd} = 0$, and for selected values of ρ_{de}.

*Remark: Solid lines in Figures represent the relative bias for positive ρ_{de} and the broken lines show the relative bias for negative ρ_{de}. The numbers on the lines indicate the magnitudes of ρ_{de}. For example, 9 means $\rho_{de} = .9$ if it is on a solid line, it means $\rho_{de} = -.9$ if on a broken line. The numbers at the end of curves indicate σ_e^2/σ_Y^2.

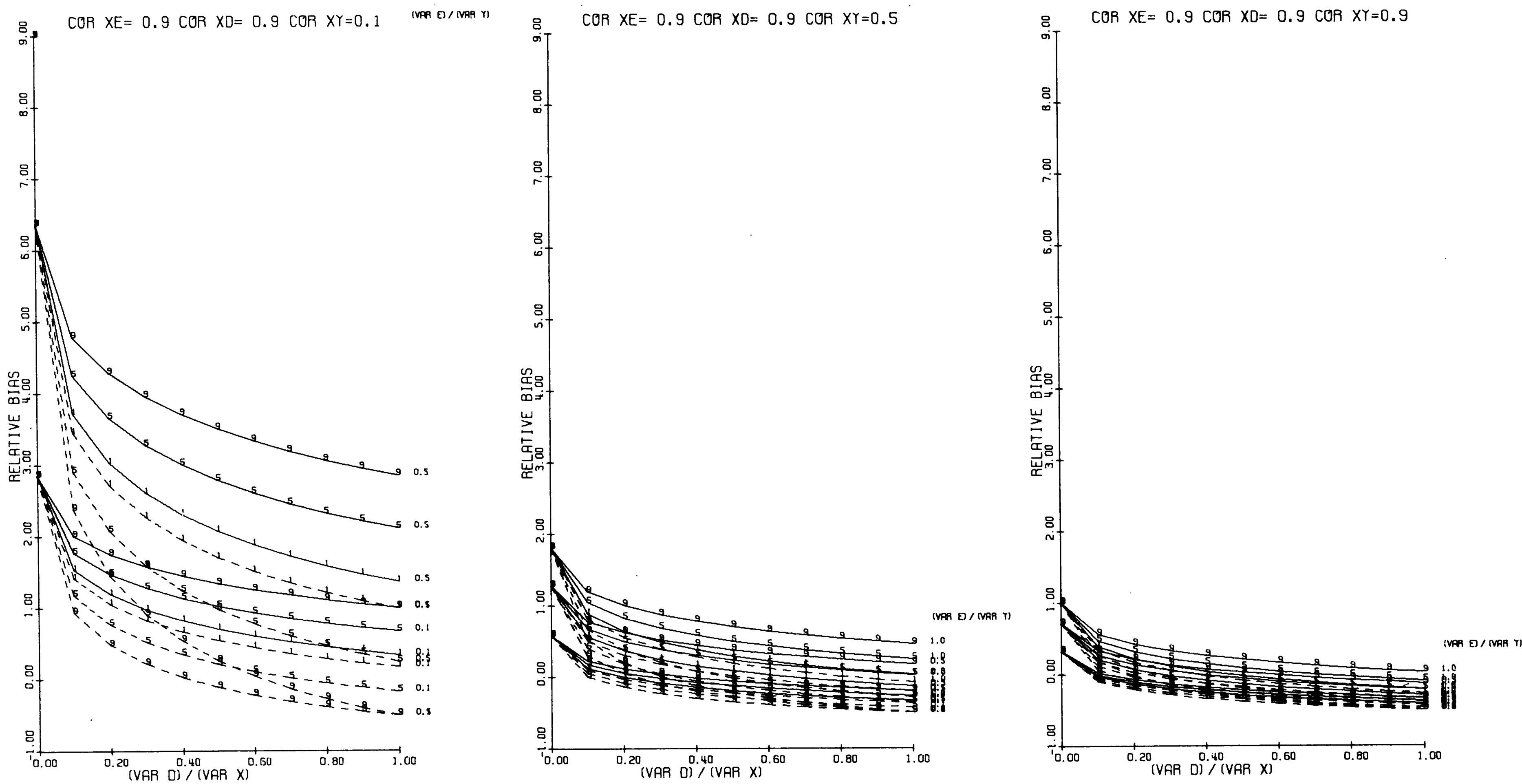


Figure 2 The relative bias of the OLSE for $\rho_{Xe} = .9$, $\rho_{Xd} = .9$, and for selected values of ρ_{de}.

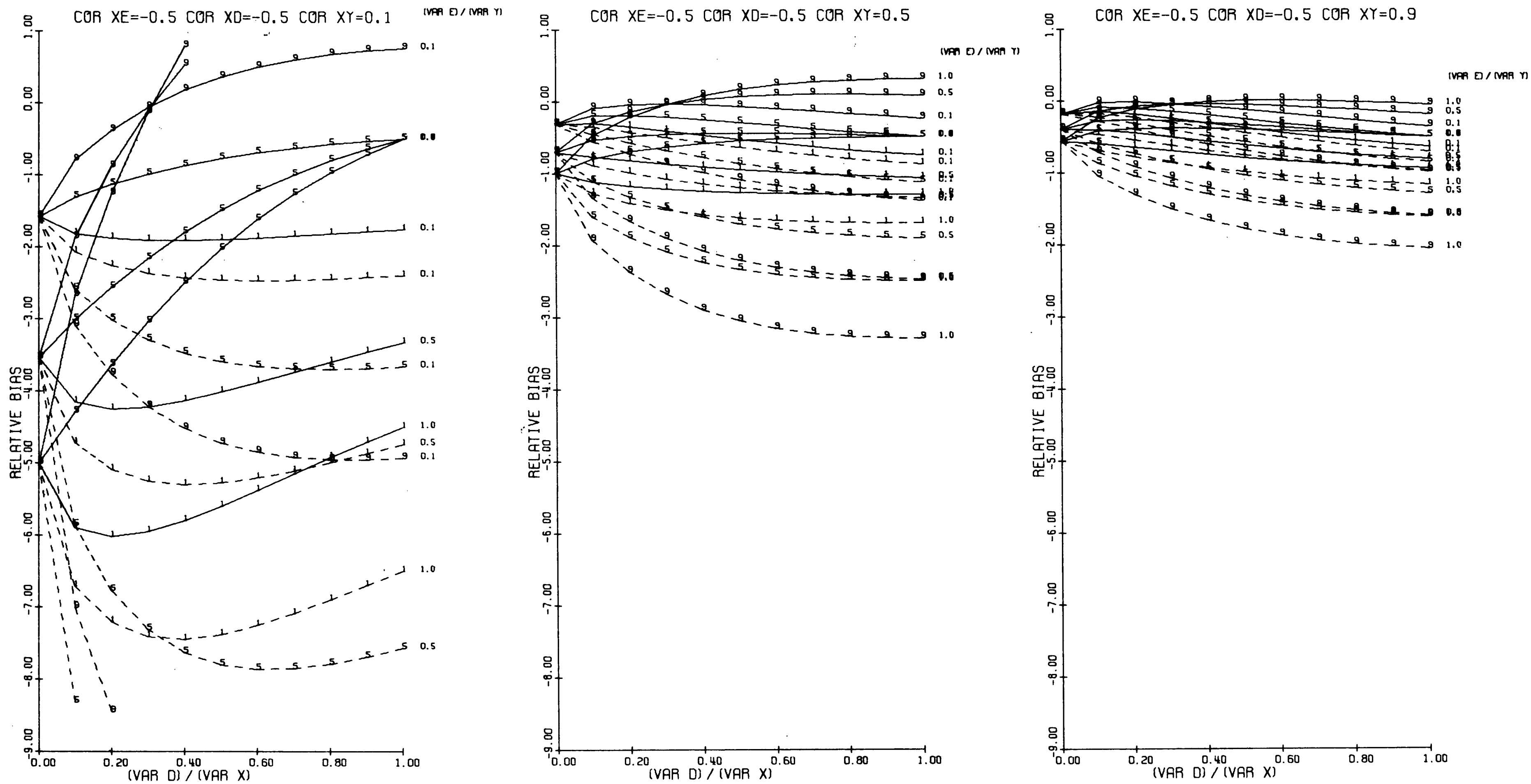


Figure 3 The relative bias of the OLSE for $\rho_{Xe} = -.5$, $\rho_{Xd} = -.5$, and for selected values of ρ_{de}.

MULTI-STAGE SAMPLING WITH PARTIAL REPLACEMENT OF THE SAMPLE ON SUCCESSIVE OCCASIONS

R. P. Chakrabarty and D. S. Rana, University of Georgia

Summary

In this paper, we develop the theory of sampling on successive occasions when the sampling design is multi-stage. We consider four different sampling plans, with different combinations of replacement of primary and secondary sampling units for the case of the two-stage sampling on two occasions. The best linear unbiased estimates of the population mean on the current occasion, the change from one occasion to another and the overall mean over two occasions are given. The relative efficiencies of the four sampling plans are also given. It is shown that the partial replacement of the primary and secondary sampling units is generally efficient for the estimation of mean on the current occasion.

1. INTRODUCTION

In the study of any dynamic population that changes with time, it is necessary to sample the population at different time points to obtain reliable estimates of the population parameters. Generally speaking, we may be interested in estimating three different parameters: (1) The population mean on the most recent occasion, (2) the change in the mean from one occasion to the next, (3) The mean over all occasions in a given period of time. For the special case of the simple random sampling the theory of successive sampling has been developed by Jessen (1942), Yates (1960), Patterson (1950), Eckler (1955), and Rao and Graham (1964). Recently, Singh (1968) and Singh and Kathuria (1969) have obtained some results on the successive sampling for a two-stage sampling design. In this paper, we develop the theory of successive sampling in the case of two-stage sampling. The extension of the theory of successive sampling from the single stage to the two-stage sampling is of considerable practical importance since usually a survey design is two-stage or multi-stage. We consider four different sampling procedures for the case of two-stage sampling on two occasions only. In all sampling procedures, we restrict ourselves to the following two-stage design for simplicity:

N=the number of primary sampling units (psu's) in the population

M=the number of the secondary sampling units (ssu's) in each psu in the population

n=the number of psu's in the sample

m=the number of ssu's in each psu in the sample.

The psu's and ssu's are selected by simple random sampling without replacement on each occasion and n<<N&m<<M so that the finite population correction factors in the variance formulas can be ignored. In deriving the variances of the estimates we shall also ignore the covariance terms which are of order 1/N. The sample size on each occasion is nm. Some notations are introduced here that will be used in the subsequent sections.

$y_{ik\ell}$=the value of y for the ℓ-th secondary in the k-th primary on the i-th occasion (i=1,2).

$$\bar{Y}_{i..}=\sum_{k=1}^{N}\sum_{\ell=1}^{M} y_{ik\ell}/NM$$

=the population mean of y on the ith occasion (i=1,2).

$$S_{bi}^2=\sum_{k=1}^{N}(\bar{Y}_{ik.}-\bar{Y}_{i..})^2/(N-1)$$

=mean square error between psu means in the population on the i-th occasion (1=1,2)

$$S_{wi}^2=\frac{\sum_{k=1}^{N}\sum_{\ell=1}^{M}(y_{ik\ell}-\bar{Y}_{ik.})^2}{N(m-1)}$$

=mean square error between ssu's within psu's in the population on the i-th occasion (i=1,2)

$$\rho_{bij}S_{bi}S_{bj}=\frac{1}{(N-1)}\sum_{k=1}^{N}(\bar{Y}_{ik.}-\bar{Y}_{i..})(\bar{Y}_{jk.}-\bar{Y}_{j..})$$

=true covariance between the psu means on the i-th and j-th occasions (i≠j=1,2).

$$\rho_{wij}S_{wi}S_{wj}=\frac{1}{N(M-1)}\sum_{k=1}^{N}\sum_{\ell=1}^{M}(y_{ik\ell}-\bar{Y}_{ik.})(y_{jk\ell}-\bar{Y}_{jk.})$$

=true covariance between ssu's within psu's on i-th and j-th occasions (i≠j=1,2).

The four sampling procedures that we consider for the two-stage sampling design are as follows:

Procedure I: Retain all primary sampling units (psu's) and from each of these psu's make a fresh selection of secondary sampling units (ssu's) on the current occasion.

The pattern of sampling may be illustrated as follows:

Sample fraction of Secondaries	nm	nm	
First Occasion	xxxx $\bar{y}^*_{1..I}$		psu's same but ssu's different
Second Occasion		xxxx $\bar{y}^*_{2..I}$	

where $\bar{y}^*_{i..I}$=mean per secondary on the i-th occasion based on nm ssu's that are present only on the i-th occasion (i=1,2)

Procedure II: Retain only a fraction p of the psu's with their samples of ssu's from the previous occasion and make a new selection of fraction q of the psu's on the current occasion, where p+q=1.

Sampling Pattern

Sample fraction of psu's	np	nq	nq
First Occasion	xxx $\bar{y}_{1..II}'$	xxxx $\bar{y}_{1..II}''$	

Second Occasion	xxx $\bar{y}'_{1..II}$	xxxx $\bar{y}''_{2..II}$
	psu's & ssu's same	psu's (& ssu's) different

where

$\bar{y}'_{i..II}$ = mean per secondary on the i-th occasion (i=1,2) based on npm units common to both occasions

$\bar{y}''_{i..II}$ = mean per secondary on the i-th occasion based on nqm units that are present on the i-th occasion only (i=1,2).

Procedure III: Retain all the psu's from the preceeding occasion, but retain only a fraction p of ssu's within each of the psu's retained and make a fresh selection of the remaining fraction q of ssu's on the current occasion, (p+q=1).

Sample fraction of ssu's	npm	nqm	nqm
First Occasion	xxx $\bar{y}'_{1..III}$	xxxx $\bar{y}^*_{1..III}$	
Second Occasion	xxx $\bar{y}'_{2..III}$		xxxx $\bar{y}^*_{2..III}$
	psu's & ssu's same		psu's same but ssu's different

where

$\bar{y}'_{i..III}$ = mean per secondary on the i-th occasion (i=1,2) based on npm units present on both occasions.

$\bar{y}^*_{i..III}$ = mean per secondary on the i-th occasion for nqm units that are present only on the i-th occasion (i=1,2).

Procedure IV: Retain a fraction p of psu's from the previous occasion and in each of these psu's retain only a fraction r of the ssu's and make a fresh selection of fraction s of ssu's (r+s=1). Also select the remaining fraction q of the psu's on the current occasion.

Sampling Pattern

	Primaries np		nq
Secondaries	nprm	npsm	nqm
First occasion	xxx $\bar{y}'_{1..IV}$	xxxx $\bar{y}^*_{1..IV}$	xxxxx $\bar{y}''_{1..IV}$
Second Occasion	xxx $\bar{y}'_{2..IV}$	xxxx $\bar{y}^*_{2..IV}$	xxxxx $\bar{y}''_{2..IV}$
	psu's,ssu's same	psu's same ssu's diff.	both psu's and ssu's different

where

$\bar{y}'_{i..IV}$ = mean per secondary on the i-th occasion (i=1,2) based on nprm units that are present on both occasions.

$\bar{y}^*_{i..IV}$ = mean per secondary on the i-th occasion (i=1,2) based on npsm units that are present only on the i-th occasion (i=1,2).

$\bar{y}''_{i..IV}$ = mean per secondary on the i-th occasion based on the nqm units that are present only on the i-th occasion (i=1,2).

We obtain the best linear unbiased estimates of the change and over all mean under procedures I, II & III. The best linear unbiased estimate of the population mean on the current occasion is given under procedure IV also. The relative efficiencies of suggested sampling procedures are given.

2. ESTIMATES OF THE CHANGE

2.1 Procedure I: A linear unbiased estimate of the change in the population mean $\bar{Y}_{2..}-\bar{Y}_{1..}$ is given by

$$\Delta_I = \bar{y}^*_{2..I} - \bar{y}^*_{1..I} \qquad (2.1)$$

and its variance is

$$Var(\Delta_I) = \frac{S_{b1}^2}{n} + \frac{S_{w1}^2}{nm} + \frac{S_{b2}^2}{n} + \frac{S_{w2}^2}{nm} - 2\rho_{b12}\frac{S_{b1}S_{b2}}{n}$$

If the variance components on two occasions remain same, that is $S_{b1}^2 = S_{b2}^2 = S_b^2$; $S_{w1}^2 = S_{w2}^2 = S_w^2$ and $\rho_{b12} = \rho_b$ then the variance of Δ_I is further simplified to

$$Var(\Delta_I) = \frac{2}{n}[(1-\rho_b)S_b^2 + \frac{S_w^2}{m}] \qquad (2.2)$$

If independent samples are drawn on each occasion, the variance of Δ_I is

$$Var(\Delta_1) = \frac{2}{n}(S_b^2 + \frac{S_w^2}{m}) \qquad (2.3)$$

If the same sample is repeated on each occasion, the variance of Δ_I is

$$Var(\Delta_I) = \frac{2}{n}[(1-\rho_b) S_b^2 + (1-\rho_w)\frac{S_w^2}{m}] \qquad (2.4)$$

A comparison of (2.2),(2.3)&(2.4) shows that for $\rho_b > 0, \rho_w > 0$, the most precise estimate of change is obtained by observing the sample on each occasion. The partial replacement policy is, however, superior to the policy of complete replacement.

2.2. Procedure II: A general linear estimate of the change $\bar{Y}_{2..} - \bar{Y}_{1..}$ may be written in the following form:

$$\Delta_{II} = a\bar{y}''_{1..II} + b\bar{y}'_{1..II} + c\bar{y}'_{2..II} + d\bar{y}''_{2..II} \qquad (2.5)$$

Now $E(\Delta_{II}) = (a+b)\bar{Y}_{1..} + (c+d)\bar{Y}_{2..}$

If Δ_{II} is to be an unbiased estimate of the change $\bar{Y}_{2..} - \bar{Y}_{1..}$, then a+b = -1 and c+d = 1, and

$$\Delta_{II} = a\bar{y}''_{1..II} - (1+a)\bar{y}'_{1..II} + c\bar{y}'_{2..II} + (1-c)\bar{y}''_{2..II} \qquad (2.6)$$

Consequently, the variance of Δ_{II} is

$$Var(\Delta_{II}) = a^2\frac{\alpha_1}{nq} + (1+a)^2\frac{\alpha_1}{np} + c^2\frac{\alpha_2}{np} + (1-c)^2\frac{\alpha_2}{nq} - 2c(1+a)\frac{\beta_{12}}{np} \qquad (2.7)$$

where

$$S_{bi}^2 + \frac{S_{wi}^2}{m} = \alpha_i, \quad i=1,2$$

$$\rho_{bij}\frac{S_{bi}S_{bj}}{} + \rho_{wij}\frac{S_{wi}S_{wj}}{m} = \beta_{ij},\ i\neq j = 1,2$$

The values of a and c that minimize the variance of Δ_{II} are obtained by solving the following equations

$$\frac{\partial}{\partial a}\text{Var}(\Delta_{II}) = 0 \text{ and } \frac{\partial}{\partial c}\text{Var}(\Delta_{II}) = 0$$

as

$$a_0 = \frac{\beta_{12}pq\,\alpha_2+\beta_{12}^2q^2-\alpha_1\alpha_2q}{\alpha_1\alpha_2-\beta_{12}^2q^2};\quad c_0=\frac{\alpha_1p(\alpha_2+\beta_{12}q)}{\alpha_1\alpha_2-\beta_{12}^2q^2} \tag{2.8}$$

If we assume that

$$S_{b1}^2= S_{b2}^2= S_b^2 \text{ and } S_{w1}^2= S_{w2}^2= S_w^2 \quad \text{(say)}$$

then $\alpha_1 = \alpha_2 = \alpha$ and $\beta_{12} = \beta$ (2.9)

Now, using (2.6), through (2.9) we get

$$\Delta_{II} = [\frac{(\alpha-\beta)q}{\alpha-\beta q}]\ (\bar{y}''_{2..II}-\bar{y}''_{1..II}) + \frac{\alpha p}{\alpha-\beta q}\ (\bar{y}'_{2..II}-\bar{y}'_{1..II}) \tag{2.10}$$

and

$$\text{Var}(\Delta_{II})=2\frac{\alpha}{n}\ (\frac{\alpha-\beta}{\alpha-\beta q});\ 0<q<1 \text{ and } \alpha>\beta \tag{2.11}$$

Thus, $\text{Var}(\Delta_{II})$ is minimum when q=0 and is maximum when q=1, for $\beta>0$. Hence to estimate change it is best to repeat the same sample on the second occasion if $\beta>0$.

Relative Efficiency: Let REC21 denote the relative efficiency of procedure II w.r.t. procedure I for the estimation of the change. From the equations (2.2) and (2.11) it can be shown that

$$\text{REC21} = \frac{(1-\rho_b+\frac{\phi}{m})[(1-\rho_bq)+(1-\rho_wq)\frac{\phi}{m}]}{(1+\frac{\phi}{m})[(1-\rho_b)+(1-\rho_w)\frac{\phi}{m}]} \tag{2.12}$$

where $\phi = S_w^2/S_b^2$.

It is interesting to note that for q=0 (complete Matching), $\rho_b>0$ and $\rho_w>0$, REC21<1, and REC21<1 if q = 0, $\rho_b<0$ and $\rho_w<0$. But for 0<q<1, it is difficult to make analytic comparisons. The numerical evaluation of REC21 given by us in a technical report (1974) shows that for small values of m and q, and large values of ϕ, ρ_b and ρ_w, procedure II is generally superior to procedure I for estimating the change in the population mean.

2.3 Procedure III: Following the methods of the previous section and under the assumption (2.9) it can be shown that the best estimate of the change is given by

$$\Delta_{III} = \frac{(1-q)}{1-\rho_wq}\ (\bar{y}'_{2..III}-\bar{y}'_{1..III}) + \frac{q(1-\rho_w)}{1-\rho_wq}\ (\bar{y}^*_{2..III}-\bar{y}^*_{1..III}) \tag{2.13}$$

and its variance after some algebraic manipulation can be obtained as

$$\text{Var}(\Delta_{III})= \frac{2}{n}[(1-\rho_b)S_b^2+(\frac{1-\rho_w}{1-\rho_wq})\frac{S_w^2}{m}] \tag{2.14}$$

Special Cases

(i) For q=0, $\text{Var}(\Delta_{III})= \frac{2}{n}[(1-\rho_b)S_b^2+(1-\rho_w)\frac{S_w^2}{m}]$ (2.15)

(ii) For q=1, $\text{Var}(\Delta_{III})=\frac{2}{n}[(1-\rho_b)S_b^2+\frac{S_w^2}{m}]$ (2.16)

It may be noted that for q=1, the procedure III becomes identical to procedure I, and the variance in (2.16) checks with that in (2.2).

Now, from (2.14), (2.15) and (2.16) it can be seen that for q=0, the variance of Δ_{III} is minimized. Therefore, the most precise estimate of change is obtained by repeating the same sample on each occasion. It follows from (2.14) and (2.16) that a policy of partial retention is superior to the policy of selecting independent samples on each occasion if $\rho_w>0$.

Relative Efficiency: (i) Denoting the relative efficiency of procedure III with respect to procedure I by REC31, it can be seen from (2.2) and (2.15) that

$$\text{REC31}= \frac{1-\rho_b+\frac{\phi}{m}}{1-\rho_b+(\frac{1-\rho_w}{1-\rho_wq})\frac{\phi}{m}} >1, \text{ for } \rho_b,\rho_w>0. \tag{2.17}$$

Thus, procedure III is superior to procedure I for estimating change when $\rho_b,\rho_w>0$.

(ii) let REC32 denote the relative efficiency of procedure III w.r.t. II for estimating the change. It follows from (2.11) and (2.15) that

$$\text{REC32}= \frac{(1-\rho_wq)(1+\frac{\phi}{m})[1-\rho_b+(1-\rho_w)\frac{\phi}{m}]}{[(1-\rho_b)(1-\rho_wq)+(1-\rho_w)\frac{\phi}{m}][1-\rho_bq+(1-\rho_wq)\frac{\phi}{m}]} \tag{2.18}$$

It may be noted that REC32=1 for q=0 as it is expected and for q=1 REC32>1 if $\rho_b>0$. For 0<q<1, it is difficult to make analytical comparisons of the efficiencies. From the numerical evaluation of REC32, we find that the procedure III generally gives a better estimate of change than procedure II, and the gains are appreciable when q, ρ_b, ρ_w and m are large and ϕ is small.

3. ESTIMATES OF THE OVER-ALL MEAN FOR TWO OCCASIONS

3.1 Procedure I: A general linear estimate of the over-all mean on two occasions, $\bar{Y}=\frac{1}{2}(\bar{Y}_{1..}+\bar{Y}_{2..})$ is given by

$$\bar{y}_I=\theta_1\bar{y}^*_{1..I}+\theta_2\bar{y}^*_{2..I} \tag{3.1}$$

where θ_1 and θ_2 are suitable weights depending on the relative importance of the two occasions and $\theta_1+\theta_2=1$. Suppose in the study of a disease it is intended to estimate the average number of persons affected by it, then θ_1 and θ_2 may be taken in proportion to the number of affected persons on each occasion. Under the assumption (2.9) the variance of $\bar{y}_I$ is obtained as

$$\mathrm{Var}(\bar{y}_I)=[\theta_1^2+\theta_2^2](\frac{S_b^2}{n}+\frac{S_w^2}{nm})+2\theta_1\theta_2\rho_b\frac{S_b^2}{n} \qquad (3.2)$$

If $\theta_1=\theta_2=\frac{1}{2}$ then

$$\mathrm{Var}(\bar{y}_I)=\frac{1}{2n}[(1+\rho_b)S_b^2+\frac{S_w^2}{m}] \qquad (3.3)$$

If new samples are drawn on each occasion, then

$$\mathrm{Var}(\bar{y}_I)=\frac{1}{2n}[S_b^2+\frac{S_w^2}{m}]$$

Thus, for $\rho_b>0$ procedure I does not provide an efficient estimate of the over-all mean. However, if $\rho_b<0$ procedure I is perferable to drawing different samples on each occasion.

3.2 Procedure II: One possible linear form of the estimate of the over-all population mean may be

$$\bar{y}_{II}=\theta_1[a\bar{y}'_{1..II}+(1-a)\bar{y}''_{1..II}]+\theta_2[b\bar{y}'_{2..II}+(1-b)\bar{y}''_{2..II}] \qquad (3.4)$$

Where θ_1 and θ_2 are suitable weights that may be chosen according to the relative importance of the two occasions so that $\theta_1+\theta_2=1$; a and b are constants to be chosen in such a way that Var $(\bar{y}_{II})$ is minimum. From Singh (1968) who developed this procedure we have the variance of $\bar{y}_{II}$ under assumption (2.9) as

$$\mathrm{Var}(\bar{y}_{II})=\frac{1}{2n}\frac{(S_b^2+\frac{S_w^2}{m})[(1+\rho_b)S_b^2+(1+\rho_w)\frac{S_w^2}{m}]}{(1+\rho_b q)S_b^2+(1+\rho_w q)\frac{S_w^2}{m}} \qquad (3.5)$$

$\mathrm{Var}(\bar{y}_{II})$ is minimum when q=1 and maximum when q=0 for $\rho_b>0$ and $\rho_w>0$. Hence, it follows, that to estimate the over-all mean for two occasions it is best to draw independent samples. However, if $\rho_b<0$, $\rho_w<0$, a partial replacement procedure will be more efficient.

Relative Efficiency: The relative efficiency of procedure II w.r.t. procedure I for the estimation of the over-all mean from (3.3) and (3.5) is given by

$$\text{REOM21}=\frac{(1+\rho_b+\frac{\phi}{m})[(1+\rho_b q)+(1+\rho_w q)\frac{\phi}{m}]}{(1+\frac{\phi}{m})[(1+\rho_b)+(1+\rho_w)\frac{\phi}{m}]} \qquad (3.6)$$

Thus, REOM21>1 for q=1, $\rho_b>0$ and REOM21<1 for q=0 & ρ_b, $\rho_w>0$. Further, the numerical evaluation of REOM21 shows that procedure II is superior to procedure I for ρ_b, ρ_w >.5, $1\leq\phi\leq 5$ and q> .5.

3.3 Procedure III: A general linear estimate of the over-all mean, under procedure III, may be written as

$$\bar{y}_{III}=\theta_1[a\bar{y}'_{1..III}+(1-a)\bar{y}^*_{1..III}]+\theta_2[c\bar{y}'_{2..III}+(1-c)\bar{y}^*_{2..III}] \qquad (3.7)$$

where θ_1 and θ_2 are suitable weights $(\theta_1+\theta_2=1)$. It can be shown that under the assumption (2.9) the values of a and c that minimize $\mathrm{Var}(\bar{y}_{III})$ are given by

$$a_0=\frac{p(\theta_1-\theta_2\rho_w q)}{\theta_1(1-\rho_w^2q^2)}\ ;\ c_0=\frac{p(\theta_2-\theta_1\rho_w q)}{\theta_2(1-\rho_w^2q^2)} \qquad (3.8)$$

Therefore, with optimum values of a and c, $\bar{y}_{III}$ becomes

$$\bar{y}_{III}=\theta_1[\frac{p(\theta_1-\theta_2\rho_w q)}{\theta_1(1-\rho_w^2q^2)}(\bar{y}'_{1..III}-\bar{y}^*_{1..III})+\bar{y}^*_{1..III}]$$
$$+\theta_2[\frac{p(\theta_2-\theta_1\rho_w q)}{\theta_2(1-\rho_w^2q^2)}(\bar{y}'_{2..III}-\bar{y}^*_{2..III})+\bar{y}^*_{2..III}] \qquad (3.9)$$

and its variance is,

$$\mathrm{Var}(\bar{y}_{III})=[(2\theta_1\theta_2\rho_w q-\theta_1^2-\theta_2^2)p+(\theta_1^2+\theta_2^2)(1-\rho_w^2q^2)]\frac{S_w^2}{nqm(1-\rho_w^2q^2)}+(\theta_1^2+\theta_2^2+2\theta_1\theta_2\rho_b)\frac{S_b^2}{n} \qquad (3.10)$$

If $\theta_1=\theta_2=\frac{1}{2}$, then

$$\mathrm{Var}(\bar{y}_{III})=\frac{1}{2n}[(1+\rho_b)S_b^2+(\frac{1+\rho_w}{1+\rho_w q})\frac{S_w^2}{m}] \qquad (3.11)$$

Relative Efficiency: Note that for q=1, $\mathrm{Var}(\bar{y}_{III})=\mathrm{Var}(\bar{y}_I)$ given in (3.3); but for $0\leq q<1$ and $\rho_b>0$, $\rho_w>0$ procedure I is more efficient than procedure III. The relative efficiency of procedure II w.r.t III for the estimation of the over-all mean is given by

$$\text{REOM23}=\frac{[1+\rho_b+(\frac{1+\rho_w}{1+\rho_w q})\frac{\phi}{m}][(1+\rho_b q)+(1+\rho_w q)\frac{\phi}{m}]}{(1+\frac{\phi}{m})[1+\rho_b+(1+\rho_w)\frac{\phi}{m}]} \qquad (3.12)$$

We note that REOM23=1 for q=0 and for q=1, REOM23 >1 if $\rho_b>o$ & REOM23<1 if $\rho_b<0$. For 0<q<1 the numerical evaluation of REOM23 shows that procedure II is more efficient than procedure III and the gains in efficiency are appreciable for ρ_b,ρ_w >.5 and q>.5.

We find from the above results that for the estimation of over-all mean it will be statistically more efficient to draw independent samples on each occasion when the correlation between the values of the sampling units on successive occasions is positive. However, practical considerations such as construction of frame, operational problems and costs may often weigh in favor of retaining a part of the sample from one occasion to the next. This policy of partial replacement is, of course, better if the correlation is negative.

4. ESTIMATES OF THE MEAN ON THE CURRENT OCCASION

4.1 Procedure I: A general linear estimate of $\bar{Y}_{2..}$, the population mean on the second occasion may be written as

$$\bar{y}_{2..I}=a\bar{y}^*_{1..I}+b\bar{y}^*_{2..I}$$

If $\bar{y}_{2..I}$ is to be an unbiased estimate of $\bar{Y}_{2..}$, we must have a=0 and b=1. Thus procedure I becomes a trivial case.

4.2 Procedure II: A linear unbiased estimate of $\bar{Y}_{2..}$ may be expressed as

$$\bar{y}_{2..II}=a(\bar{y}'_{1..II}-\bar{y}''_{1..II})+c\bar{y}'_{2..II}+(1-c)\bar{y}''_{2..II} \quad (4.1)$$

From Singh(1968) we have the optimum values of a and c under the assumption (2.9) as

$$a_0=pq\alpha\beta/(\alpha^2-q'^2\beta^2) \qquad c_0=p\alpha^2/(\alpha^2-q'^2\beta^2) \quad (4.2)$$

and $$\mathrm{Var}(\bar{y}_{2..II})=\alpha(\alpha^2-q'\beta^2)/n(\alpha^2-q'^2\beta^2) \quad (4.3)$$

Where q' denotes the replacement fraction for psu's. Now $\mathrm{Var}(\bar{y}_{..II})=\alpha/n$ for q'=0 or q'=1 and for all other values of q'(0<q'<1) the partial replacement will be more efficient if $\beta\neq 0$. Singh also shows that q' should always exceed $\frac{1}{2}$.

4.3 Procedure III: The estimation of the mean on the current occasion under this procedure was investigated by Singh and Kathuria (1969). Under the assumption (2.9) they obtained the estimate and its variance as

$$\bar{y}_{2..III}=\frac{p}{1-\rho_w q'^2}[\bar{y}'_{2..III}+\rho_w q'(\bar{y}^*_{1..III}-\bar{y}'_{1..III})]+\frac{q'(1-\rho_w^2 q')}{(1-\rho_w^2 q'^2)}\bar{y}^*_{2..III} \quad (4.4)$$

and

$$\mathrm{Var}(\bar{y}_{2..III})=\frac{S_b^2}{n}+[\frac{(1-\rho_w^2 q')}{1-\rho_w^2 q'^2}]\frac{S_w^2}{nm},\ 0<q'<1$$

$$=\frac{S_b^2}{n}+\frac{S_w^2}{nm}\quad \text{for } q'=0,1 \quad (4.5)$$

where q' denotes replacement fraction for ssu's. From the results in (4.5) it is easily seen that it is advantageous to retain a part of the sample for estimation of the population mean on the second occasion.

4.4 Procedure IV: A general linear estimate of $\bar{Y}_{2..}$, the population mean on the second occasion may be written as

$$\bar{y}_{2..IV}=a\bar{y}'_{1..IV}+b\bar{y}^*_{1..IV}+e\bar{y}''_{1..IV}+c\bar{y}'_{2..IV}+d\bar{y}^*_{2..IV}+f\bar{y}''_{2..IV} \quad (4.6)$$

Now $E(\bar{y}_{2..IV})=(a+b+e)\bar{Y}_{1..}+(e+d+f)\bar{Y}_{2..}$

If $\bar{Y}_{1..}\neq\bar{Y}_{2..}$, and $\bar{y}_{2..IV}$ is to be an unbiased estimate of $\bar{Y}_{2..}$, we must have a+b+e=0 and c+d+f=1 . Therefore,

$$\bar{y}_{2..IV}=a\bar{y}'_{1..IV}+b\bar{y}^*_{1..IV}-(a+b)\bar{y}''_{1..IV}+c\bar{y}'_{2..IV}+d\bar{y}^*_{2..IV}+(1-c-d)\bar{y}''_{2..IV} \quad (4.7)$$

The variance of $\bar{y}_{2..IV}$, ignoring the finite population correction terms and covariance terms which are of order 1/N, is given by

$$V=\mathrm{Var}(\bar{y}_{2..IV})=a^2(\frac{S_{b1}^2}{np}+\frac{S_{w1}^2}{nprm})+b^2(\frac{S_{w1}^2}{np}+\frac{S_{w1}^2}{npsm})+(a+b)^2(\frac{S_{b1}^2}{nq}+\frac{S_{w1}^2}{nqm})+c^2(\frac{S_{b2}^2}{np}+\frac{S_{w2}^2}{nprm})+d^2(\frac{S_{b2}^2}{np}+\frac{S_{w2}^2}{npsm})+(1-c-d)^2(\frac{S_{b2}^2}{nq}+\frac{S_{w2}^2}{nqm})+2ab\frac{S_{b1}^2}{np}+2ac(\rho_{b12}\frac{S_{b1}S_{b2}}{np}+\rho_{w1}\frac{S_{w1}S_{w2}}{nprm})+2ad\rho_{b12}\frac{S_{b1}S_{b2}}{np}+2bc\rho_{b12}\frac{S_{b1}S_{b2}}{np}+2bd\rho_{b12}\frac{S_{b1}S_{b2}}{np}+2cd\frac{S_{b2}^2}{np} \quad (4.8)$$

The optimum values of the weights a,b,c and d that will minimize $\mathrm{Var}(\bar{y}_{2..IV})$ are obtained by solving the following system of linear equations:

$\frac{\partial V}{\partial a}=0$, $\frac{\partial V}{\partial b}=0$, $\frac{\partial V}{\partial c}=0$ and $\frac{\partial V}{\partial d}=0$.

Under the assumption (2.9) these lead to the following equations.

$$a[m(1-s)+(q+(1-q)(1-s))\phi]+b[m(1-s)+(1-q)(1-s)\phi]+c[\rho_b mq(1-s)+\rho_w q\phi]+d[mq(1-s)\rho_b]=0 \quad (4.9)$$

$$a[ms+(1-q)s\phi]+b[ms+(q+(1-q)s)\phi]+cmqs\rho_b+dmqs\rho_b=0 \quad (4.10)$$

$$a[mq(1-s)\rho_b+q\rho_w\phi]+b[mq(1-s)\rho_b]+c[m(1-s)+(q+(1-q)(1-s))\phi]+d[m(1-s)+(1-q)(1-s)\phi]=m(1-q)(1-s)+(1-q)(1-s]\phi \quad (4.11)$$

$$amqs\rho_b+bmqs\rho_b+c[ms+(1-q)s\phi]+d[ms+(q+(1-q)s)\phi]=m(1-q)s+(1-q)s\phi \quad (4.12)$$

Solving the equations (4.9) through (4.12) the optimum values of a,b,c, and d can be expressed in terms of m,q and s, and parameters ρ_b,ρ_w and $\phi=S_w^2/S_b^2$ but they are too cumbersome to present here. The numerical values of the weights for selected values of the parameters are presented in tables 1 and 2. Substitution of the optimum values of the weights a,b,c and d in (4.8) provide $\text{Var}(\bar{y}_{2..IV})$ which is too complicated for analytical investigation of the replacement policy. To compare procedure IV with procedures II and III the over-all replacement fraction should be same in each case. By equating the number of unmatched units we obtain

$$nq'm = npsm + nqm$$

$$\text{or } q' = q+s-qs \qquad (4.13)$$

where q' is the replacement fraction of psu's in procedure II and that of ssu's in procedure III. The relative efficiencies, REM42 and REM43 of procedure IV w.r.t procedures II and III respectively have been evaluated numerically for selected values of m,q,s and parameters ρ_b, ρ_w and ϕ. The results are given in tables 3 and 4. It can be seen from tables 3 and 4 that (1) procedures II and IV are superior to procedure III, and (2) for $q \geq .5$ and $s > .5$, REM42>1, but the gains achieved are not high. Thus the choice between procedure II and IV for estimating mean on the current occasion may depend on other practical considerations such as cost and operational problems in a given survey situation.

We may mention that the efficiencies of the suggested sampling procedures have been illustrated using 'live data' from the Georgia Agricultural Facts (1964) for the estimation of average corn-yield per acre on the current occasion. The procedure IV was generally superior to procedures II & III but the gains achieved were modest. Further, the assumption (2.9)was found to be nearly valid in this case. These results are given in a technical report (1974).

REFERENCES

[1] Eckler, A. R. (1955). "Rotation Sampling". Ann. Math. Stat., 26, pp. 664-685.

[2] Georgia Agricultural Facts (August, 1964) Crop Reporting Service, Athens, Georgia.

[3] Jessen, R. J. (1942. "Statistical Investigation of a sample survey for obtaining farm facts." Iowa Agri. Exp. Stn. Res. Bul. 304.

[4] Patterson, H.D.(1950). "Sampling on successive occasions with partial replacement of units." Jour. Roy. Stat. Soc., Series B, pp. 241-255.

[5] Rana, D. S. and Chakrabarty, R. P. (1974). Two-stage sampling on successive occasions with partial replacement of units. Technical Report No. 104. Statistics & Computer Science Dept. University of Georgia.

[6] Rao, J.N.K. and Graham, J.E.(1964). "Rotation designs for sampling on repeated occasions." Jour. Amer. Stat. Asso., 59, PP. 492-509.

[7] Singh. D. (1968). "Estimates in successive sampling using a multistage design." Jour. Amer. Stat. Assoc., 63, pp. 99-112.

[8] Singh, D. and Kathuria, O.P.(1969). "On two-stage successive sampling." The Australian Journal of Statistics, 11, No. 2, pp. 59-66.

[9] Yates, F. (1960). Sampling methods for censuses and surveys. Third edition, Charles Griffin and Co., Ltd., London.

Table 1. Values of the weights a, b, c and d for ρ_b = .7, ϕ = .5, m = 4 and a set of values of ρ_w q and s.

	q = .3				q = .5				q = .7				
s	a	b	c	d	a	b	c	d	a	b	c	d	
.1	-.1613	.0190	.6639	.0648	-.1901	.0074	.5141	.0466	-.1738	.0002	.3509	.0293	
.3	-.1774	.0389	.5366	.1920	-.1872	.0097	.4194	.1396	-.1608	-.0072	.2881	.0890	
.5	-.1687	.0343	.4060	.3216	-.1658	-.0060	.3196	.2367	-.1359	-.0257	.2204	.1525	ρ_b=.7
.7	-.1312	.0014	.2637	.4621	-.1225	-.0426	.2086	.3438	-.0966	-.0574	.1439	.2232	ρ_b=.5
.9	-.0564	-.0680	.0976	.6250	-.0505	-.1064	.0774	.4691	-.0384	-.1062	.0532	.3062	
.1	-.1834	.0319	.6727	.0605	-.2139	.0173	.5280	.0402	-.1973	.0067	.3684	.0256	
.3	-.2204	.0736	.5600	.1739	-.2271	.0365	.4462	.1231	-.1951	.0109	.3149	.0764	
.5	-.2271	.0852	.4462	.2872	-.2178	.0339	.3596	.2071	-.1777	.0011	.2555	.1311	ρ_b=.7
.7	-.1951	.0591	.3149	.4161	-.1777	.0026	.2555	.3059	-.1389	-.0274	.1817	.1971	ρ_b=.7
.9	-.0969	-.0306	.1336	.5919	-.0844	-.0775	.1083	.4429	-.0631	-.0875	.0763	.2889	
.1	-.2063	.0454	.6829	.0553	-.2394	.0278	.5443	.0365	-.2241	.0141	.3897	.0209	
.3	-.2691	.1127	.5912	.1496	-.2747	.0682	.4819	.1006	-.2387	.0334	.3518	.0587	
.5	-.3047	.1526	.5081	.2339	-.2904	.0893	.4219	.1605	-.2403	.0408	.3124	.0961	ρ_b=.7
.7	-.3053	.1584	.4147	.3265	-.2778	.0842	.3488	.2304	-.2209	.0304	.2599	.1426	ρ_b=.9
.9	-.2127	.0761	.2485	.4887	-.1849	.0078	.2064	.3599	-.1398	.0299	.1517	.2325	

Table 2. Values of the weights a, b, c and d for $\rho_b = .9$, $\phi = .5$, m = 4 and a set of values of ρ_w, q and s.

	q = .3				q = .5				q = .7				
s	a	b	c	d	a	b	c	d	a	b	c	d	
.1	-.1972	.0159	.6804	.0646	-.2469	.0033	.5524	.0477	-.2485	-.0047	.4129	.0321	
.3	-.2070	.0296	.5522	.1923	-.2344	-.0033	.4533	.1440	-.2228	-.0224	.3408	.0983	
.5	-.1914	.0183	.4194	.3237	-.2023	-.0287	.3472	.2461	-.1834	-.0525	.2617	.1700	ρ_b=.9
.7	-.1461	-.0221	.2734	.4674	-.1465	-.0764	.2276	.3600	-.1274	-.0974	.1712	.2510	ρ_b=.5
.9	-.0619	-.1003	.1016	.6353	-.0594	-.1535	.0847	.4949	-.0496	-.1614	.0633	.3465	
.1	-.2204	.0293	.6909	.0596	-.2746	.0139	.5710	.0421	-.2823	.0030	.4414	.0271	
.3	-.2520	.0655	.5785	.1723	-.2800	.0260	.4865	.1245	-.2692	-.0013	.3804	.0823	
.5	-.2529	.0715	.4634	.2864	-.2613	.0152	.3949	.2120	-.2383	-.0210	.3104	.1436	ρ_b=.9
.7	-.2136	.0385	.3287	.4181	-.2089	-.0266	.2822	.3172	-.1815	-.0623	.2211	.2195	ρ_b=.7
.9	-.1048	-.0610	.1400	.6002	-.0975	-.1217	.1200	.4657	-.0803	-.1399	.0924	.3264	
.1	-.2445	.0431	.7030	.0537	-.3048	.0254	.5927	.0354	-.3224	.0118	.4765	.0207	
.3	-.3035	.1065	.6132	.1457	-.3352	.0610	.5307	.0982	-.3308	.0264	.4359	.0592	
.5	-.3352	.1424	.5307	.2290	-.3452	.0770	.4691	.1585	-.3241	.0278	.3909	.0993	ρ_b=.9
.7	-.3308	.1435	.4359	.3224	-.3241	.0648	.3909	.2316	-.2907	.0081	.3269	.1522	ρ_b=.9
.9	-.2279	.0518	.2597	.4907	-.2119	-.0268	.2321	.3729	-.1782	-.0716	.1892	.2591	

Table 3. REM42 and REM43 for $\rho_b = .7$, $\phi = .5$, m = 4 and a set of values of ρ_w, q and s.

	q = .3			q = .5			q = .7			
s	q'	REM42	REM43	q'	REM42	REM43	q'	REM42	REM43	
.1	.37	99	110	.55	100	114	.73	100	113	
.3	.51	97	110	.65	100	114	.79	102	113	
.5	.65	97	110	.75	101	113	.85	103	113	ρ_b=.7
.7	.79	99	110	.85	104	113	.91	106	112	ρ_b=.5
.9	.93	104	109	.95	108	112	.97	109	112	
.1	.37	99	111	.55	100	114	.73	100	114	
.3	.51	97	111	.65	100	114	.79	102	114	
.5	.65	97	111	.75	101	114	.85	104	114	ρ_b=.7
.7	.79	99	110	.85	104	114	.91	106	113	ρ_b=.7
.9	.93	104	110	.95	109	113	.97	109	112	
.1	.37	99	111	.55	100	114	.73	101	114	
.3	.51	97	111	.65	100	114	.79	102	114	
.5	.65	97	111	.75	102	114	.85	105	114	ρ_b=.7
.7	.79	99	111	.85	105	114	.91	108	114	ρ_b=.9
.9	.93	106	110	.95	110	114	.97	111	113	

Table 4. REM42 and REM43 for $\rho_b = .9$, $\phi = .5$, m = 4 and a set of values of ρ_w, q and s.

	q = .3			q = .5			q = .7			
s	q'	REM42	REM43	q'	REM42	REM43	q'	REM42	REM43	
.1	.37	97	119	.55	99	128	.73	100	130	
.3	.51	93	119	.65	97	127	.79	101	129	
.5	.65	90	118	.75	98	126	.85	103	128	ρ_b=.9
.7	.79	92	118	.85	101	126	.91	108	127	ρ_b=.5
.9	.93	102	117	.95	112	124	.97	117	125	
.1	.37	97	120	.55	99	129	.73	100	133	
.3	.51	92	119	.65	97	129	.79	101	132	
.5	.65	90	119	.75	97	128	.85	103	131	ρ_b=.9
.7	.79	91	119	.85	100	127	.91	108	129	ρ_b=.7
.9	.93	100	118	.95	111	125	.97	117	127	
.1	.37	97	120	.55	99	130	.73	100	135	
.3	.51	92	120	.65	97	130	.79	101	135	
.5	.65	89	120	.75	97	130	.85	104	134	ρ_b=.9
.7	.79	90	120	.85	100	129	.91	109	133	ρ_b=.9
.9	.93	100	119	.95	111	127	.97	119	130	

PAIRWISE PROBABILITIES IN PROBABILITY NONREPLACEMENT SAMPLING

James R. Chromy, Research Triangle Institute

1. INTRODUCTION

The problem of controlling the pairwise probabilities, P(ij), in probability nonreplacement sampling has existed since the development of a general theory of probability nonreplacement sampling in conjunction with the use of the Horvitz-Thompson estimator (ref. 1). A number of researchers have treated the subject directly or indirectly including Raj (ref. 2), Hanurav (ref. 3), Brewer (ref. 4), Durbin (ref. 5), Sampford (ref. 6), Jessen (ref. 7), Rao and Bayless (ref. 8), and Dodds and Fryer (ref. 9). Some of these approached the problem through a superpopulation model and examined such criteria as the superpopulation expectation of the variance of the Horvitz-Thompson estimator or the stability of the variance estimator. Jessen posed an intuitive criterion which relates most closely to the approach followed in this research, but he did not solve for optimum designs directly.

A general superpopulation model with size measures defined to potentially achieve the principal advantages of probability nonreplacement sampling is described. Under this model the superpopulation expectation of the variance is not a function of the pairwise probabilities, P(ij). The criterion employed in seeking optimum designs is the superpopulation variance of the finite population variance. A general solution for the P(ij) in terms of the variances of the error terms in the superpopulation model is developed. Although the solution can be generalized to any sample size, n (within certain restrictions), the problem of choosing three-wise to n-wise probabilities may still be difficult. A method for obtaining three-wise probabilities, P(ijk), consistent with the pairwise probabilities, P(ij), has been developed.

2. THE HORVITZ-THOMPSON ESTIMATOR

A theory for PPS nonreplacement sampling for within stratum samples of two or more elements was presented by Horvitz and Thompson (ref. 1). The unbiased estimator within Horvitz and Thompson's class two of linear estimators for probability nonreplacement (pnr) designs has come to be known as the Horvitz-Thompson estimator and is of the form

$$t(s|pnr) = \sum_{u=1}^{n} Y[s(u)]/P[s(u)]$$

where

s = [s(1),s(2),...,s(n)], a sample of n distinct elements from a sampling frame of N elements,

Y(i) = the variate of interest in the sampling investigation for the i-th element in the universe, and

P(i) = probability that the i-th element is included in the sample.

In practice, the P(i) are usually determined so that

$$P(i) = nX(i)/X(+),$$

where

X(i) = a known positive-valued variate associated with the i-th element in the universe, and

$$X(+) = \sum_{i=1}^{N} X(i).$$

Horvitz and Thompson presented a formula for the variance of t(s|pnr) under PPS sampling without replacement given by

$$V[t(s|pnr)] = \sum_{i=1}^{N} Y^2(i)[1-P(i)]/P(i)$$

$$+ \sum_{i \neq j}^{N}\sum^{N} Y(i)Y(j)[P(ij) - P(i)P(j)]/[P(i)P(j)],$$

where

P(ij) = the probability that the i-th and j-th units are both included in the sample (or the pairwise probability for the i-th and j-th units).

Yates and Grundy (ref. 10) presented an alternate expression for the variance of the Horvitz-Thompson estimator, namely

$$V[t(s|pnr)] = \sum_{i<j}^{N}\sum^{N} W(ij)D^2(ij)$$

where

$$W(ij) = P(i)P(j)-P(ij),$$

and

$$D(ij) = Y(i)/P(i)-Y(j)/P(j).$$

They also presented an alternate variance estimator which may be expressed as

$$v[t(s|pnr)] = \sum_{u<v}^{n}\sum^{n} w[s(u),s(v)]D^2[s(u),s(v)]$$

where

$$w(ij) = W(ij)/P(ij).$$

3. THE BASIC SUPERPOPULATION MODEL

The model considered most relevant for study of the Horvitz-Thompson estimator and probability nonreplacement sampling is

$$Y(i) = BX(i) + e(i)$$

where

B = an unknown constant, and

e(i) = the error term which measures the deviation from the model.

As pointed out by Horvitz and Thompson (ref. 1), t(s|pnr) will have zero variance if P(i) is exactly proportional to Y(i) for all i. The proposed model is designed to take advantage of this possibility. By making the P(i) proportional to X(i), a nonzero variance of t(s|pnr) occurs only if some of the e(i) are nonzero.

Suppose one believed that a linear model with a nonzero intercept was more realistic for the population to be sampled. Such a model may be written as

$$Y(i) = A + BZ(i) + e(i)$$

where A and B are both unknown but nonzero constants and Z(i) is a measure of size. Use of the Horvitz-Thompson estimator and probability nonreplacement sampling with P(i) proportional to Z(i) would not provide the opportunity for a zero variance of $t(s|pnr)$ even if the e(i) were all zero. A zero variance would occur with the e(i) equal to zero only if the P(i) were exactly proportional to [A/B + Z(i)]. Unless one wishes to rely upon an extremely fortuitous set of error terms, e(i), to rescue the design and estimation plans from producing a large variance of $t(s|pnr)$, it appears more appropriate to use the no-intercept model

$$Y(i) = BX(i) + e(i)$$

where

$$X(i) = A/B + Z(i).$$

To define X(i), it is not necessary to know A and B individually, but only their ratio. Even a poor guess should lead to a better design than one that sets the P(i) in such a way that the principal advantage of PPS sampling cannot be achieved.

In summary, the premise put forth here is that if one uses the Horvitz-Thompson estimator in conjunction with probability nonreplacement sampling, the proper course to follow is to set P(i) exactly proportional to some known X(i) which is believed to be approximately proportional to Y(i). If the no-intercept model does not present a good approximation of the Y(i) to X(i) relationship, the proper course to follow does not involve a magical choice of P(ij) to compensate for this shortcoming. It does involve the selection of some other variate as a measure of size. It should be noted that X(i) could be defined as a function of several other known variables.

Consideration of an intercept model can be defended in the instance of multi-purpose sample surveys for modeling the variance of estimates associated with some less major purposes of the survey. The choice of the P(i) and P(ij) should probably still be based on some known variate, X(i), which fits the no-intercept model for the particular Y(i) that is associated with the major purpose of the survey.

Suppose that the sampling statistician assumes the no-intercept model and utilizes as much information as is available to him to construct X(i) approximately proportional to the unknown Y(i). The following assumptions about the error terms appear reasonable in the predata collection stage of the sample design process:

(1) $Ee(i) = 0$ for $i = 1,2,\ldots,N$,

(2) $E[e(i)/X(i)]^2 = s^2(i)$ (finite), for $i = 1,2,\ldots,N$, and

(3) the error terms are independent.

The assumption of independence is not unreasonable if the X(i) are properly chosen and if stratification is used to its fullest extent consistent with requirements for the estimability of the variance. For most applications where PPS sampling is applied, deep stratification is also utilized.

If one first assumes that the value of B is fixed and that the error terms sum to zero over the N population elements, then the independence assumption is errorneous. Prior to sampling, however, B is not known and cannot be estimated. In this *a priori* sense, independence of error terms may be assumed as an indication of complete lack of further knowledge about the Y(i) given the X(i).

4. THE SUPERPOPULATION VARIANCE OF THE FINITE POPULATION VARIANCE

A particular outcome of the superpopulation model may be represented by an N-dimensional vector of error terms and designated as $\underline{e}$. The transpose of $\underline{e}$ may be written as

$$[e(1),e(2),e(3),\ldots,e(N)].$$

For brevity, the Horvitz-Thompson estimator will be denoted by t(s) and its variance for a particular finite population (i.e., for a particular $\underline{e}$) will be denoted by $V[t(s)|\underline{e}]$. The $\underline{e}$ indicates a particular sample of N error terms selected from the hypothetical superpopulation. The s indicates a particular sample of n elements selected from the finite population of N elements.

In the finite population sense, the variance is computed over all possible samples s. The finite population variance is still a function of the error vector $\underline{e}$. It is reasonable under the model to consider taking the expectation of $V[t(s)|\underline{e}]$ over all possible outcomes $\underline{e}$. The finite population variance for the Horvitz-Thompson estimator may be expressed in terms of the superpopulation model by selectively substituting nX(i)/X(+) for P(i) and BX(i) + e(i) for Y(i) in the expression given in section 1. After some simplification this yields

$$n^2X^{-2}(+)V[t(s)|\underline{e}] = \sum_{i=1}^{N} [B+e(i)/X(i)]^2P(i)[1-P(i)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N} [B+e(i)/X(i)][B+e(j)/X(j)][P(ij)-P(i)P(j)].$$

After collecting terms in B^2 and B, this expression becomes

$$B^2\{\sum_{i=1}^{N} P(i)[1-P(i)] + \sum_{i\neq j}^{N}\sum^{N} [P(ij)-P(i)P(j)]\}$$

$$+ B\{2\sum_{i=1}^{N} [e(i)/X(i)]P(i)[1-P(i)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N} [e(i)/X(i)][P(ij)-P(i)P(j)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N} [e(j)/X(j)][P(ij)-P(i)P(j)]\}$$

$$+ \sum_{i=1}^{N} [e(i)/X(i)]^2P(i)[1-P(i)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N} [e(i)/X(i)][e(j)/X(j)][P(ij)-P(i)P(j)].$$

After noting that

$$\sum_{j\neq i}^{N} [P(ij)-P(i)P(j)] = -P(i)[1-P(i)],$$

the expression simplifies further to

$$n^2X^{-2}(+)V[t(s)|\underline{e}] = \sum_{i=1}^{N} [e(i)/X(i)]^2P(i)[1-P(i)]$$

$$+ \sum_{i\neq j}^{N}\sum^{N}[e(i)/X(i)][e(j)/X(j)][P(ij)-P(i)P(j)].$$

Given the assumptions about the error terms, e(i), as stated above, the superpopulation expectation of the variance is simply

$$EV[t(s)|\underline{e}] = n^{-2}X^2(+) \sum_{i=1}^{N} P(i)[1-P(i)]s^2(i).$$

Under the assumed model, the expectation of the variance is not a function of the pairwise probabilities and does not, therefore, provide any basis for selection of the pairwise probabilities.

Since the particular choice of pairwise probabilities cannot be used to reduce the expected variance, another criterion must be developed. Intuitively, the sampling statistician should want to protect against an extremely large variance. To protect against a large variance occurring most of the time, the sample design should produce variances that have a tight superpopulation distribution about their expected value. This intuitive criterion suggests minimizing the superpopulation variance of the finite population variance; this variance may be written as

$$V\{V[t(s)|\underline{e}]\} = E\{V^2[t(s)|\underline{e}]\}-E^2\{V[t(s)|\underline{e}]\}.$$

Using the results obtained above, the square of the variance, $V^2[t(s)|\underline{e}]$, may be represented as

$$n^4X^{-4}(+)V^2[t(s)|\underline{e}] =\{ \sum_{i=1}^{N} [e(i)/X(i)]^2P(i)[1-P(i)]\}^2$$

$$+ \{\sum_{i\neq j}^{N}\sum^{N} [e(i)/X(i)][e(j)/X(j)][P(ij)-P(i)P(j)]\}^2$$

$$+ 2\{ \sum_{i=1}^{N} [e(i)/X(i)]^2P(i)[1-P(i)]\} \times$$

$$\{\sum_{i\neq j}^{N}\sum^{N} [e(i)/X(i)][e(j)/X(j)][P(ij)-P(i)P(j)]\}.$$

The third term, or the cross product term can be seen to have a zero expectation under the assumed superpopulation model. After expanding the first two terms and noting which terms in the products have nonzero expectation, the superpopulation expectation of the squared variance, $EV^2[t(s)|\underline{e}]$, may be represented as

$$EV^2[t(s)|\underline{e}] =$$

$$n^{-4}X^4(+)\{ \sum_{i=1}^{N} E[e(i)/X(i)]^4P^2(i)[1-P(i)]^2$$

$$+ \sum_{i\neq j}^{N}\sum^{N} s^2(i)s^2(j)P(i)P(j)[1-P(i)][1-P(j)]$$

$$+ 2 \sum_{i\neq j}^{N}\sum^{N} s^2(i)s^2(j)[P(ij)-P(i)P(j)]^2\}.$$

It must be further assumed that $E[e(i)/X(i)]^4$ exists and is finite.

The square of the superpopulation expectation of the variance, $E^2V[t(s)|\underline{e}]$, may be written as

$$E^2V[t(s)|\underline{e}] = n^{-4}X^4(+)\{ \sum_{i=1}^{N} s^4(i)P^2(i)[1-P(i)]^2$$

$$+ \sum_{i\neq j}^{N}\sum^{N} s^2(i)s^2(j)P(i)P(j)[1-P(i)][1-P(j)]\}.$$

Combining the above results for $EV^2[t(s)|\underline{e}]$ and $E^2V[t(s)|\underline{e}]$ to obtain the superpopulation variance of the finite population variance produces

$$n^4X^{-4}(+)VV[t(s)|\underline{e}] =$$

$$\sum_{i=1}^{N} \{E[e(i)/X(i)]^4-s^4(i)\}P^2(i)[1-P(i)]^2$$

$$+ 2 \sum_{i\neq j}^{N}\sum^{N} s^2(i)s^2(j)[P(ij)-P(i)P(j)]^2.$$

To minimize this variance by choice of P(ij), it is only necessary to consider those terms in the expanded form that are a function of the P(ij). This approach reduces to minimizing

$$\sum_{i\neq j}^{N}\sum^{N}[P(i)P(j) - P(ij)]^2s^2(i)s^2(j),$$

or, in more compact notation

$$\sum_{i\neq j}^{N}\sum^{N} W^2(ij)s^2(i)s^2(j).$$

If the $s^2(i)$ are constant over all elements, i, this criterion reduces to the one suggested by Jessen (ref. 7).

5. THE ALGEBRAIC SOLUTION FOR OPTIMAL VALUES OF P(ij)

The solution for the P(ij) must satisfy the set of constraints

$$\sum_{j\neq i}^{N} P(ij) = (n-1)P(i).$$

In terms of the W(ij), these constraints are equivalent to requiring that

$$\sum_{j\neq i}^{N} W(ij) = P(i)[1-P(i)].$$

The method of Lagrange multipliers may be used to obtain the constrained solution for the W(ij) and subsequently for the P(ij). The function to be minimized may be written as

$$F\{W(ij),K(i):i\neq j\} = \sum_{i\neq j}^{N}\sum^{N} W^2(ij)s^2(i)s^2(j)$$

$$- 4 \sum_{i=1}^{N} K(i)\{ \sum_{j\neq i}^{N} W(ij)-P(i)[1-P(i)]\}.$$

In the above, the K(i) are the Lagrange multipliers. Taking derivatives with respect to the W(ij) and setting them equal to zero as a necessary condition for obtaining a minimum produces N(N-1)/2 independent equations of the form

$$W(ij) = s^{-2}(i)s^{-2}(j)\{K(i)+K(j)\}.$$

Taking derivatives with respect to the K(i) and setting them equal to zero yields N additional equations of the form

$$\mathop{S}_{j\neq i}^{N} W(ij) = P(i)[1-P(i)].$$

The entire process yields a set of N(N+1)/2 simultaneous equations. The number of equations to be solved initially can be reduced by applying the second set of equations to the first set to obtain N equations in K(i) only:

$$P(i)[1-P(i)]s^{2}(i) = K(i) \mathop{S}_{j\neq i}^{N} s^{-2}(j) + \mathop{S}_{j\neq i}^{N} K(j)s^{-2}(j)$$

$$= K(i)[\mathop{S}_{j=1}^{N} s^{-2}(j) - 2s^{-2}(i)] + \mathop{S}_{j=1}^{N} K(j)s^{-2}(j).$$

This set of equations may be represented in matrix notation as

$$\underline{c} = M\,\underline{k}.$$

The N x 1 vector, $\underline{c}$, has elements c(i) defined as

$$c(i) = P(i)[1-P(i)]s^{2}(i).$$

The N x 1 vector $\underline{k}$ is the vector of unknown Lagrange multipliers, K(i). The N x N matrix, M, may be represented as

$$M = D + \underline{1}\,\underline{b}'$$

where D is a diagonal matrix. The diagonal elements of the matrix D may be expressed as

$$d(ii) = S^{(1)} - 2s^{-2}(i).$$

The term $S^{(1)}$ is defined as the following sum:

$$S^{(1)} = \mathop{S}_{k=1}^{N} s^{-2}(k).$$

The vector $\underline{1}$ consists of N ones. The vector $\underline{b}$ has elements

$$b(i) = s^{-2}(i).$$

If the two sums $S^{(2)}$ and $S^{(3)}$ are defined as

$$S^{(2)} = \mathop{S}_{k=1}^{N} b(k)d^{-1}(kk) \text{ and}$$

$$S^{(3)} = \mathop{S}_{k=1}^{N} b(k)c(k)d^{-1}(kk),$$

and if $S^{(2)} \neq -1$, then the simultaneous equations in k(i) have solutions given by

$$k(i) = d^{-1}(ii)[c(i)-(1+S^{(2)})^{-1}S^{(3)}].$$

Having obtained algebraic solutions for the K(i), the algebraic solutions for the W(ij) may be obtained by substituting the solved K(i) values in the first N(N-1)/2 equations, namely as

$$W(ij) = s^{-2}(i)s^{-2}(j)\{K(i) + K(j)\}.$$

The P(ij) may be obtained as

$$P(ij) = P(i)P(j)-W(ij).$$

To be valid solutions in the sense of yielding a probability sampling design, all of the P(ij) must satisfy the conditions

$$0 \le P(ij) \le \text{minimum}\{P(i),P(j)\}.$$

These may be called the "weak constraints".

An even stronger set of constraints on the P(ij) may be considered, namely,

$$CP(i)P(j) \le P(ij) \le P(i)P(j)$$

for some small C > 0. These may be called the "strong constraints".

6. ADJUSTMENT PROCEDURES

Three alternative methods for adjusting the P(ij) when the algebraic solutions fall outside of prescribed boundary values may be applied:

(1) constrained minimization,
(2) compensating additive adjustments, and
(3) sampling unit redefinition.

All three adjustment methods may be handled as methods for adjusting the W(ij) rather than the P(ij). Only method 2, compensating additive adjustments, is discussed in this paper.

Method 2 may be applied without accuracy limitations of computer matrix inversion procedures. The method treats one W(ij) at a time as follows. Suppose W(ij) exceeds the prescribed upper bound, i.e.,

$$W(ij) - U(ij) = a$$

and a > 0. Corrective adjustments may be made in the W(ij) without violating the N constraints

$$\mathop{S}_{j\neq i}^{N} W(ij) = P(i)[1-P(i)].$$

The recommended procedure is:

(1) adjust W(ij) by -a,
(2) adjust W(ik) by +a/(N-2) for $k\neq i$ or j,
(3) adjust W(jk) by +a/(N-2) for $k\neq i$ or j, and
(4) adjust W(km) by -2a/[(N-2)(N-3)] for $k\neq i$ or j and $m\neq i$ or j.

If several W(ij) must be adjusted, they can be handled individually in an iterative fashion. Since the adjustments for different W(ij) may tend to work against each other in the iterative process, it is recommended that the iteration continue until sufficient numerical accuracy is obtained.

7. SOME COMMENTS ON THE CHOICE OF CONSTRAINTS

Requiring that

$$P(ij) \le P(i)P(j)$$

guarantees that all terms in the Yates-Grundy form of the variance will be non-negative and the Yates-Grundy variance estimator also will be non-negative for all possible samples.

The initial form of the strong constraints was discussed in section 5 and was stated as

$$CP(i)P(j) \le P(ij) \le P(i)P(j)$$

for all P(ij) and some small C > 0. If one considers the left-hand part of the inequalities and sums both sides over j not equal i, the result is

$$CP(i)[n-P(i)] \le (n-1)P(i)$$

for all i, or after factoring out P(i) and dividing by [n-P(i)],

$$C \le (n-1)/[n-P(i)]$$

for all i. If a single value of C is to be used, it must be less than the minimum value over all i of

$$(n-1)/[n-P(i)].$$

Setting C less than or equal to (n-1)/n satisfies this requirement for an upper bound of C.

If the strong constraints are applied with C set at (n-1)/n, some immediate effects on the finite population variance and variance estimates may be noted.

Stating the constraints in terms of the W(ij) gives

$$C \le W(ij) \le P(i)P(j)/n$$

The right hand term may also be expressed as $nX(i)X(j)/X^2(+)$. The variance of the Horvitz-Thompson estimator for designs employing this constraint will then be bounded above by

$$[n/X(+)]^2 \mathop{S}^{N}\mathop{S}^{N}_{i \ne j} X(i)X(j)D^2(ij)$$

where

$$D(ij) = Y(i)/P(i) - Y(j)/P(j)$$

or equivalently

$$D(ij) = [X(+)/n][Y(i)/X(i) - Y(j)/X(j)].$$

This upper bound is exactly equivalent to the variance of the total estimates for the standard Hansen-Hurwitz probability replacement design. This particular form of the variance for probability replacement sampling is developed as an analogue to the nonreplacement form by Kendall and Stuart (ref. 11).

The same constraint on the P(ij) may be expressed in terms of coefficients, w(ij), of the Yates-Grundy variance estimator by first noting that

$$w(ij) = P(i)P(j)/P(ij)-1.$$

Then the effects of the constraints on the w(ij) can be noted in the following steps:

$$(n-1)P(i)P(j)/n \le P(ij) \le P(i)P(j),$$

$$1/[P(i)P(j)] \le 1/P(ij) \le n/[(n-1)P(i)P(j)],$$

$$1 \le P(i)P(j)/P(ij) \le n/(n-1),$$

and

$$0 \le w(ij) \le 1/(n-1).$$

The variance estimators for designs employing this constraint will be bounded above for each sample by

$$(n-1)^{-1} \mathop{S}^{n}\mathop{S}^{n}_{i<j} D^2(ij).$$

This expression is exactly equal to the variance estimate for the Hansen-Hurwitz probability replacement design when applied to samples of distinct elements (ref. 11).

It may then be concluded that application of the strong constraints with C equal to (n-1)/n will produce designs with variances and variance estimates that are never greater than those that would be obtained under probability replacement designs.

8. SOME EMPIRICAL COMPARISONS

Two general questions cannot be answered directly by examination of the theoretical development in the preceding sections. The first question is concerned with the sensitivity of the superpopulation variance criterion to poor or wrong initial assumptions about the error structure of the population to be sampled. The second question is concerned with the effects of the choice of constraints on the superpopulation variance criterion and on other statistical properties such as the stability of the variance estimator as measured by the superpopulation expectation of the finite population variance of the Yates-Grundy variance estimator. These two questions were studied empirically in a small simulation experiment.

Two simple population simulation models based on independent Bernoulli trials were utilized. Since moments up to the fourth order could be computed for these models, theoretical superpopulation results could be obtained as well as results of repeated simulation trials. Only the theoretical results are presented in this paper.

Populations of size 10 were considered with size measures, X(i), being the numbers one through 10. The observed values, Y(i), were generated by the simulation models with the expected squared error, $Ee^2(i)$, proportional to X(i) for the first population and proportional to $X^2(i)$ for the second population.

Twelve designs were studied based on four different assumptions about the expected squared error term in the superpopulation model and three successively more stringent sets of constraints on the algebraic solutions.

The results comparing the superpopulation variances of the finite population variance are presented in table 1. Little variation in the criterion function occurs due to erroneous assumptions about the true error model. The imposition of strong constraints increases the obtainable minimum slightly and decreases the variation due to error model assumed.

Table 2 shows the behavior of the superpopulation expectations of the finite population variance of the Yates-Grundy variance estimator. Among the four model assumptions studied the assumption of squared error proportional to X(i) yields the minimum value of the function for either model and for any of the three constraints. Imposition of the strong constraints can be seen to be very effective in controlling the stability of the variance estimator.

9. EXTENSIONS

The solutions for P(ij) may be obtained for any sample size, n. However, higher order probabilities must be obtained to completely specify the sample design if the sample size is greater than two. An extension to sample size three has been obtained (ref. 12). The problem remains

TABLE 1 - THEORETICAL SUPERPOPULATION VARIANCES OF THE FINITE POPULATION VARIANCE

Assumed model for design purposes	Constraints applied to the P(ij)		
	.01P(i)(j) ≤ P(ij) ≤ Min{P(i),P(j)}	.01P(i)P(j) ≤ P(ij) ≤ P(i)P(j)	.50P(i)P(j) ≤ P(ij) ≤ P(i)P(j)
		True model: $Ee^2(i) = .16X(i)$	
$Ee^2(i) = k$	288.68	288.68	288.26
$Ee^2(i) = kX(i)$	287.95	287.95	287.96
$Ee^2(i) = kX^2(i)$	292.81	289.17	288.14
$Ee^2(i) = kX^3(i)$	305.58	291.20	288.28
		True model: $Ee^2(i) = .16X^2(i)$	
$Ee^2(i) = k$	9,686.55	9,686.55	9,687.80
$Ee^2(i) = kX(i)$	9,692.67	9,692.67	9,690.86
$Ee^2(i) = kX^2(i)$	9,664.04	9,671.54	9,683.12
$Ee^2(i) = kX^3(i)$	9,728.40	9,689.40	9,685.03

to extend the procedure to samples of size four or greater.

8. LIST OF REFERENCES

[1] D. G. Horvitz and D. J. Thompson, "A Generalization of Sampling from a Finite Universe," Journal of the American Statistical Association, 47:663-685, 1952.

[2] Des Raj, "A Note on the Determination of Optimum Probabilities in Sampling Without Replacement," Journal of the American Statistical Association, 51:269-284, 1956.

[3] T. V. Hanurav, "Optimum Utilization of Auxiliary Information: πPS Sampling of Two Units from a Stratum," J. R. Statistical Society, B29:371-391, 1967.

[4] R. K. W. Brewer, "A Model of Systematic Sampling with Unequal Probabilities," Australian Journal of Statistics, 5:5-13, 1963.

[5] J. Durbin, "Design of Multi-stage Surveys for the Estimation of Sampling Errors," Applied Statistics 16:152-164, 1967.

[6] M. R. Sampford, "On Sampling Without Replacement with Unequal Probabilities of Selection," Biometrika 54:499-513, 1967.

[7] R. J. Jessen, "Some Methods of Probability Nonreplacement Sampling," Journal of the American Statistical Association 64:175-193, 1969.

[8] J. N. K. Rao and D. L. Bayless, "An Empirical Study of the Stabilities of Estimators and Variance Estimators in Unequal Probability Sampling of Two Units Per Stratum," Journal of the American Statistical Association 64:540-559, 1969.

[9] D. J. Dodds and J. G. Fryer, "Some Families of Selection Probabilities for Sampling With Probability Proportional to Size," J. R. Statistical Society 33(2):263-274, 1971.

[10] F. Yates and P. M. Grundy, "Selection Without Replacement from Within Strata and with Probability Proportional to Size," J. R. Statistical Society B15:253-261, 1953.

[11] Maurice G. Kendall and Alan Stuart, The Advanced Theory of Statistics, Volume 3, Design and Analysis, and Time Series. Charles Griffin and Company Limited, London, 1966.

[12] James R. Chromy, Pairwise Probabilities in Probability Nonreplacement Sampling, Ph.d. Thesis, North Carolina State University, Raleigh, North Carolina, 1974.

TABLE 2 - THEORETICAL SUPERPOPULATION EXPECTATIONS OF THE FINITE POPULATION VARIANCE OF THE YATES-GRUNDY VARIANCE ESTIMATOR

Assumed model for design purposes	Constraints applied to the P(ij)		
	.01P(i)P(j) ≤ P(ij) ≤ Min{P(i),P(j)}	.01P(i)P(j) ≤ P(ij) ≤ P(i)P(j)	.50P(i)P(j) ≤ P(ij) ≤ P(i)P(j)
		True model: $Ee^2(i) = .16X(i)$	
$Ee^2(i) = k$	4,647	4,647	3,979
$Ee^2(i) = kX(i)$	4,014	4,014	3,944
$Ee^2(i) = kX^2(i)$	35,252	7,735	4,410
$Ee^2(i) = kX^3(i)$	393,893	27,701	4,468
		True model: $Ee^2(i) = .16X^2(i)$	
$Ee^2(i) = k$	72,003	72,003	66,085
$Ee^2(i) = kX(i)$	62,815	62,815	62,826
$Ee^2(i) = kX^2(i)$	170,775	81,915	65,037
$Ee^2(i) = kX^3(i)$	1,549,399	166,684	66,517

RESULTS OF A SURVEY ON HOUSEHOLD PARTICIPATION IN THE FOOD STAMP PROGRAM: DATA FROM THE JUNE 1973 CURRENT POPULATION SURVEY

by
John F. Coder
Bureau of the Census

Introduction

Since June 1961, the Federal Food Stamp Program has been in operation assisting low income families and individuals in the purchase of food. Households which qualify under the eligibility requirements established by the United States Department of Agriculture and apply for certification become eligible to purchase food stamps. Participants purchase food stamps at prices below the face value of the stamps then redeem the stamps for food at supermarkets or other stores which have been certified to handle transactions involving food stamps. The difference between the purchase price and face value of the stamps is termed bonus value. The purchase price of the food stamps is based on the household's net monthly income while the monthly coupon allotment (face value) is determined by the size of the household.

The food stamp program has grown from approximately 50,000 participants receiving $381,000 in food stamp bonus value in June 1961 to nearly 13,600,000 participants receiving approximately $271 million in bonus value in March 1974. Throughout this twelve year period of growth little information has been available concerning the characteristics of households participating in the food stamp program. To provide such data, the Bureau of the Census, under sponsorship of the Office of Economic Opportunity, conducted a household survey in June 1973. The questions concerning food stamps were designed mainly to collect information concerning household participation in the food stamp program during the month of May 1973 with additional questions on participation during the previous twelve month period.

This presentation which is largely based on information collected by this survey has two objectives. The first objective is to provide a brief analysis covering the quality of the food stamp information collected using the June 1973 Current Population Survey (CPS). The second objective is to develop a profile of food stamp households with respect to their economic and demographic characteristics including their annual income in relation to official Federal poverty levels.

Quality of Food Stamp Survey Data

The survey data presented here pertaining to food stamp recipients were collected by supplementing the June 1973 Current Population Survey questionnaire with a group of questions covering household purchases of food stamps during the month of May 1973 and during the previous twelve month period, June 1, 1972 to May 31, 1973. The Current Population Survey is a monthly household survey of approximately 50,000 households designed primarily to provide national estimates of employment, unemployment, and other characteristics of the labor force. It has also been widely used to provide data covering income, migration, educational attainment, and many other demographic, economic, and social characteristics of the population. To evaluate the quality of the food stamp data collected using the June 1973 CPS, a comparison was made between the survey data and data published by the United States Department of Agriculture.

Comparison of published information from the Department of Agriculture and data from the June 1973 CPS indicate that, although the survey data apparently underestimate the number of persons in households purchasing food stamps in May 1973, the survey universe of food stamp households is representative of the USDA universe of food stamp households with respect to household size, bonus values, and participation in public assistance programs.

The number of persons in households reporting the purchase of food stamps in May 1973 on the CPS was 9,881,100. This figure compares to a USDA published figure of about 12,358,200. The CPS figure which is approximately 80 percent of the USDA estimate differs from the USDA figures for several reasons. As in all household surveys, the data are subject to sampling and nonsampling errors. Of the factors causing the discrepancy between the two estimates, nonsampling errors such as misreporting and nonreporting are probably the most important.

Two other factors which may also contribute to the 20 percent difference between these estimates are possible errors in the USDA administrative data and the occurrence of multiple USDA food stamp households within the CPS household structure. Because the USDA estimates are derived from information provided by local (in most cases counties) food stamp agencies, some of which undoubtedly do a better record keeping job than others, the USDA estimates are also subject to error and should not be used as a tool for evaluation without this fact in mind. Whether errors in the administrative data tend to lessen or widen the gap between these estimates has not been determined. Evidence from a record check study indicate that some of the CPS households contain two or more USDA food stamp units. This situation may occur, for example, in the case of a CPS household containing both a primary family and a subfamily consisting of the primary family head's son and the son's wife and child. Although this group represents one CPS household, it is conceivable that both the primary family and the subfamily could receive food stamps as separate units. The design of the questionnaire, however, allowed only one food stamp household for each CPS defined household. This fact caused a portion of the difference in the number of participating households (see table 1) but should have had little effect on the count of the number of persons in food stamp households, a figure which could be recorded correctly regardless of the number of food stamp households within the CPS household. Although these facts would suggest that the underestimate of the number of food stamp households would be greater than that for persons in the households, the underestimate for both figures was about the same, 20 percent.

An evaluation of the CPS estimate of the total number of households purchasing food stamps in the twelve month period, June 1, 1972 to May 31, 1973 was also made. The June CPS data show a total of about 3,938,000 different households participating during this period. Of this total, 829,000 households reported purchasing food stamps during the specified twelve month period but not in May 1973. A method offered by USDA for determining the number of different households participating in the food stamp program in a given twelve month period was 1.5 times the number participating in any single month. If this figure is used and May 1973 is taken as the base month then an estimate of the total number of households derived from administrative data is 1.5 x 3,941,000 or 5,912,000. The survey underestimate of the number of households participating between June 1, 1972 and May 31, 1973 using this method is about 33 percent.

Estimates derived from the CPS for both bonus value per participant and participation rates in public assistance programs by food stamp households are also very similar to USDA published figures. The aggregate bonus value for May 1973, defined as the difference between the purchase price and coupon allotment, was estimated from the CPS to be $141.7 million.[1/] This figure is about 80 percent of the USDA estimate of $178.5 million. The average bonus value per participant calculated from CPS data, however, was $14.27, about 99 percent of the corresponding figure of $14.45 published by the USDA for May 1973. USDA estimates also show that in May 1973, 61 percent of all persons receiving food stamps were in households which received food stamps as a direct result of their participating in the public assistance program in May. Data compiled from the CPS indicate that in May 1973 about 58 percent of all persons in food stamp households were in households headed by a person reporting the receipt of public assistance in May.

Evaluation of the CPS food stamp data with respect to many important demographic characteristics of the household such as income, age, race, and sex of head, and labor force status of head, etc., could not be made because little or no information is available from administrative sources. An evaluation of the survey data with respect to reliable demographic information from administrative sources could show that the difference is not proportionally distributed between demographic subgroups and that the survey households are not representative of the USDA food stamp households with respect to certain demographic characteristics. If differential underreporting could be determined using reliable administrative figures as a control, adjustments could be made to improve the usefulness of the data by simultaneously correcting for the underreporting of food stamp purchases and the differences in demographic composition.

Characteristics of Food Stamp Households

Shown in tables 2 and 3 are data for families reporting the purchase of food stamps in May 1973 by selected characteristics. Data in table 2 show the proportion of the selected universe reporting the purchase of food stamps in May while the data in table 3 show the distribution of food stamp households for each selected universe. These data have not been adjusted for the 20 percent difference between USDA figure and the survey figure of the number of households purchasing food stamps in May 1973.

Since the data which will be presented in the following sections are based on a sample, they are subject to sampling variability as well as errors of response and nonreporting. None of the statements in this report have been tested for statistical significance, therefore, an attempt has been made to limit comparisons to situations where obvious significant differences exist.

Overall approximately 4.2 percent of all families and 5.7 percent of all households headed by a primary individual reported purchases of food stamps in May 1973. Almost one of every five families headed by a woman purchased food stamps in May while only 2.2 percent of all families headed by men purchased food stamps.

Participation rates for Negro[2/] families were higher than rates for white families for each characteristic shown; the high participation rate for Negro families reflecting the lower income and greater proportion of families headed by women for this group. Participation rates for families were higher in the South than in any other region, a result apparently of the lower incomes in this region.

Participation rates by income class decline, as would be expected, as income increases. The participation rates for Negro families appear to be higher than for white families at each income level. This higher participation rate for Negro families may result from the fact that a larger proportion of Negro families with low incomes receive public assistance than white families and therefore a larger proportion of Negro families are categorically eligible to participate in the food stamp program.

Negro families which comprise about 11 percent of all families and 32 percent of all families with incomes below poverty level in 1972 made up about 42 percent of the families purchasing food stamps in May 1973. Approximately 70 percent of all Negro families purchasing food stamps were headed by a woman. This contrasts with the corresponding figure of 43 percent for white families.

The South which has the largest population of the four regions in the U.S. also accounts for the largest group of food stamp families, 38.4 percent, a rate which slightly exceeds its proportion of the total number of families in March 1973 (31.8 percent).

The distribution of food stamp families by residence areas shown in table 3 varies somewhat with respect to the distribution of all families within these areas. While both 67 percent of all families and families purchasing food stamps reside in metropolitan areas, 30 percent of all families and 48 percent of all food stamp families live in the central city. Farm families which comprise about five percent of all families constitute only 2.5 percent of the total number of food stamp families.

Participation rates for primary individuals which make up approximately 30 percent of all food stamp households were highest for Negro females age 65 or older (25.5 percent) and lowest for white males less than 25 years old (1.4 percent). As was the case with families, Negro primary individuals had higher participation rates than white primary individuals. Of all primary individuals purchasing food stamps in May 1973, 74 percent were women; 41 percent were women age 65 and over.

Available information which have not been included in any of the tables indicate that of the households reporting the purchase of food stamps in May 1973, more than 60 percent reported purchasing food stamps in each of the previous twelve months. Over 80 percent reported purchasing food stamps for six or more months during this period. The mean number of months purchased for this group was approximately 9.3.

For the household which reported purchasing food stamps during the twelve month period June 1, 1972 to May 31, 1973 but not in May 1973 (a total of 828,000 households), the average number of purchase months was 3.8. Sixty percent of these households purchased food stamps for three months or less months with almost 30 percent reporting a purchase in only one month in the past twelve.

Food Stamps and the Low Income Population

One obvious topic of interest concerning the food stamp population is the status of this population with respect to official Federal poverty standards and, conversely, the status of the poverty population with respect to participation in the food stamp program. To provide information on these subjects, a match of two Census Bureau data files, the March 1973 CPS file containing detailed income data and poverty status for calendar year 1972 and the June 1973 CPS file containing food stamp information was made. As a result of the CPS sampling design, households are interviewed for four consecutive months, released from the sample for eight months, and returned to the sample for a final four month period. One fourth of the households in the March 1973 CPS sample, therefore, were also in the June 1973 CPS sample and a match of these data files thus provided information from both data files for the households common to these monthly surveys.[3]

The total number of households available for matching (based on the March 1973 CPS data file) was 11,186 (actual number of sample households). Of these, 8,596, or 77 percent, were termed fully-matched.[4] Only these fully-matched households were used for analytical purposes since households of other matched status had undergone some compositional change between March and June, a fact which would complicate the analysis. Households which moved between March and June were by definition excluded from the analysis since no data from the June CPS was available.

After matching of the data files further screening of the matched records was required before the analysis could begin. Since the food stamp information on the June 1973 CPS was collected for households and poverty status is developed on a basis of families and unrelated individuals, all secondary families and secondary unrelated individuals were eliminated from the matched data file. An assumption was made, therefore, that the unit receiving food stamps within the CPS household was always the primary family or the primary unrelated individual.[5]

Tabulations made from the matched data file indicate that approximately half (52.8 percent) of the families and two-thirds (68.7 percent) of the unrelated individuals purchasing food stamps[6] had annual incomes below the poverty level in 1972. These figures vary somewhat by race and sex of head, with food stamp families headed by a white male having a 42.4 percent poverty rate and food stamp families headed by a Negro female having a poverty rate of 60.2 percent.

The fact that only about 50 percent of all families purchasing food stamps during the specified twelve month period had incomes below the poverty level in 1972 is not surprising for several reasons. First, eligibility for food stamps is based on a net monthly income figure while the poverty status is based on an annual income concept. As a result, a household may be eligible for food stamps for several months because of some short term decrease in income but on an annual basis have income above the poverty level. Secondly, USDA maximum monthly income amounts, which help determine a household's eligibility to participate in the food stamp program are higher than one twelth of the poverty level which is an annual gross income amount. Thirdly, since monthly income used to determine eligibility is a net figure derived after deductions are made for payroll taxes, shelter costs, medical expenses, etc., the annual gross income of a participating household may be considerably higher than the poverty level.

In addition to these two factors which demonstrate that a considerable number of families with annual incomes above the poverty level are eligible for food stamps, guidelines set up for the food stamp program provide that all persons receiving public assistance are also eligible to participate in the food stamp program. About one half of the families and one-third of the unrelated individuals receiving public assistance in 1972 had annual incomes above the poverty level, yet were categorically eligible to participate in the food stamp program while receiving public assistance income.

Estimates of participation in the food stamp program by families and unrelated individuals with incomes below the poverty level in 1972 are shown in table 4. Several adjustments were made to the survey data before the estimates of participation in food stamp program by the poverty population could be derived. First, adjustments were made to the survey data for the underestimate of the number of food stamp households by assuming the underestimate was distributed proportionally between demographic subgroups and between the poverty and nonpoverty population. This adjustment also assumed that the survey underestimated the number of households purchasing food stamps during the 12 month period June 1, 1972 to May 31, 1973 by 33 percent. Because the food stamp program was in operation in only about two-thirds of the counties in the U.S. in May 1973, a second adjustment was made which excluded from the analysis sample households located in sampling areas (primary sampling units) containing one or more counties which did not have the food stamp program in May 1973. This adjustment allowed the derivation of participation rates based only on households residing in areas administering the food stamp program.

The data shown in table 4 indicate that overall, 53.3 percent of all families and 30.8 percent of all unrelated individuals with incomes below the poverty level in 1972 purchased food stamps between June 1, 1972 and May 31, 1973. The participation rate in the food stamp program for female headed families with incomes below the poverty level in 1972 was 71.1 percent while the rate for families headed by an elderly male was 34.2 percent. About 75 percent of the families and 80 percent of the primary individuals below the poverty level who purchased food stamps, purchased food stamps in six or more months during this period.

Participation rates were significantly higher for both Negro families and primary individuals below the poverty level than for white families and individuals. This, again, may result, in part, from the fact that a larger proportion of low income Negro families receive public assistance than white families.

<u>Summary</u>

The June 1973 CPS represents the first attempt by the Census Bureau at collecting detail information concerning household participation in the food stamp program on a national scale. The role of food stamps as well as other sources of what have been termed "nonmoney" or "noncash" income as an important source of income to both families and individuals has been increasing at a rapid rate. More efforts will be needed by the Bureau of the Census and others to expand our knowledge as to the effects of "noncash" income on the total welfare of the population.

Collection of food stamp data using the June 1973 CPS was an experience which resembles our previous survey experiences in collecting public transfer money income data such as public assistance. This experience has been that the number of recipients and the aggregate amount of benefits, in terms of dollars, have been substantially underestimated by the surveys but that the reporting recipients appear to closely represent the total universe in many important respects.

The data presented here pertaining to the relationship between the food stamp and official Federal poverty universes must be used only keeping in mind the assumptions which have been made. Since the data used to develop these estimates were not collected for the specific purpose of estimating the relationship between poverty and food stamps, some useful information pertinent to the analysis such as changes in household composition during the year, financial assets, and purchases of food stamps covering all of calendar year 1972 were not collected. The fact that the analysis required a match of the March 1973 CPS and June 1973 CPS files resulted in some problems. Households which moved during the period between March and June were necessarily excluded from the final sample used in the analysis. The sample size was reduced considerably and, therefore, sampling errors increased to levels which are much higher than those associated with the entire CPS sample. Even in the light of these problems, this study has produced some interesting findings which should serve to increase our knowledge about the food stamp program and to promote more detail and specialized studies in this area.

1/ The estimate of $141.7 million bonus value from the CPS was derived assuming households reporting the purchase of food stamps in May but failing to report the necessary information to calculate bonus value received, on the average, the same bonus values as households reporting both purchase price and coupon allotment. Approximately 10 percent of the food stamp households did not provide all information required to calculate the bonus value.

2/ Negro actually refers to Negro and other races throughout this paper.

3/ As a result of residents moving from the sample address, noninterviews, and errors in recording identifying information, some households could not be matched.

4/ A fully-matched code was assigned to a household only if records for each person 14 years and older within the household were present on both data files. Therefore, these households have undergone no compositional changes among their members 14 years old and over.

5/ Since the number of secondary families is very small (about 100,000 of a total of 54.3 million total families in March 1973) and 75 percent of all unrelated individuals are primary individuals living alone, this assumption will be a correct one in almost all cases.

6/ Families and unrelated individuals were classified as food stamp recipients if any person in the family or anyone in the household headed by the primary individual reported purchasing food stamps in any month between June 1, 1972 and May 31, 1973.

TABLE 1. A COMPARISON OF JUNE 1973 CPS SURVEY DATA AND U.S. DEPARTMENT OF AGRICULTURE DATA ON THE NUMBER OF HOUSEHOLDS PURCHASING FOOD STAMPS IN MAY 1973

Size of Household	Food Stamp Households by Size				Ratio of CPS Households to USDA households (1÷3)
	Number (thous.) (1)	Percent[1/] (2)	Number (thous.) (3)	Percent (4)	
Total	3,110	100.0	3,941	100.0	.79
1	874	29.2	1,186	30.1	.74
2	655	21.9	822	20.9	.80
3	451	15.1	581	14.7	.78
4	305	10.2	445	11.3	.69
5	228	7.6	304	7.7	.75
6	170	5.7	210	5.3	.81
7	131	4.4	143	3.6	.92
8 or more	175	5.6	250	6.3	.70
Size not available	121	3.9[2/]	(X)	(X)	(X)
Mean size	3.19	(X)	3.13	(X)	1.02

1/ Percents are calculated based on the total number of households reporting on the number of persons covered under the food stamp program.

2/ This figure represents the percent of all households reporting the purchase of food stamps in May which did not report the number of persons covered under the food stamp program.

X - Not applicable.

TABLE 2. PERCENT OF FAMILIES PURCHASING FOOD STAMPS DURING MAY 1973 BY SELECTED CHARACTERISTICS

(Data from the June 1973 CPS. Numbers as of June 1973)

Selected Characteristics	All families		White Families		Negro and Other Races Families	
	Number (Thousands)	Percent Purchasing Food Stamps	Number (Thousands)	Percent Purchasing Food Stamps	Number (Thousands)	Percent Purchasing Food Stamps
All families..................	54,309	4.2	48,154	2.8	6,155	15.5
Sex of Head						
Male.........................	47,794	2.2	43,531	1.8	4,262	6.6
Female.......................	6,515	19.0	4,623	12.2	1,892	35.5
Age of Head						
Less than 25.................	4,113	6.7	3,548	4.4	564	21.2
25 to 54.....................	33,779	4.1	29,660	2.6	4,119	15.1
55 to 64.....................	8,699	3.2	7,947	2.1	752	14.6
65 years or older............	7,718	4.2	6,999	3.2	719	14.1
Residence						
Farm.........................	2,538	2.2	2,399	1.6	139	14.1
Nonfarm, total...............	51,771	4.3	45,755	2.8	6,016	15.5
In metropolitan areas......	36,531	4.2	31,733	2.6	4,798	15.0
In central city..........	16,427	6.6	12,751	3.6	3,676	17.0
Outside central city.....	20,104	2.2	18,982	1.8	1,122	8.3
Outside metropolitan areas.	15,240	4.6	14,022	3.4	1,218	17.8
Region						
Northeast....................	12,831	3.8	11,713	2.5	1,117	17.8
North Central................	15,098	3.5	13,871	2.3	1,227	17.3
South........................	17,001	5.2	14,053	2.9	2,948	15.7
West.........................	9,380	4.1	8,517	3.6	863	9.1
Income[1/]						
Under $2,000.................	1,102	31.6	770	24.8	332	45.6
$2,000 - $2,999..............	2,062	21.5	1,637	15.8	425	40.0
$3,000 - $3,999..............	2,285	15.7	1,830	12.5	455	27.9
$4,000 - $4,999..............	2,348	10.6	2,071	6.9	277	33.4
$5,000 - $5,999..............	2,905	5.0	2,456	3.3	449	13.9
$6,000 or more...............	39,003	0.9	35,297	0.6	3,712	3.8
NA...........................	4,537	2.3	4,045	1.8	492	8.0

1/ The money income levels shown here may be somewhat understated. These data which are from the June 1973 CPS control card are based on the respondent's estimate of total family money income for the preceding 12 month period coded in broad, fixed income intervals.

TABLE 3. FAMILIES PURCHASING FOOD STAMPS DURING MAY 1973 BY SELECTED CHARACTERISTICS

(Data from the June 1973 CPS. Numbers as of June 1973)

Selected Characteristics	All Families	White Families	Negro and Other Races Families
Total purchasing food stamps..........	2,281	1,327	954
Sex of Head			
Total..............................	100.0	100.0	100.0
Male...............................	45.8	57.5	29.6
Female.............................	54.2	42.5	70.4
Age of Head			
Total..............................	100.0	100.0	100.0
Less than 25.......................	12.2	11.9	12.6
25 to 54...........................	61.4	58.5	65.3
55 to 64...........................	12.2	12.7	11.5
65 years or older..................	14.2	16.9	10.6
Residence			
Total..............................	100.0	100.0	100.0
Farm...............................	2.5	2.8	2.0
Nonfarm, total.....................	97.5	97.2	98.0
In metropolitan areas..............	67.0	61.2	75.2
In central city..................	47.6	34.7	65.4
Outside central city.............	19.4	26.5	9.8
Outside metropolitan areas.........	33.0	38.9	22.8
Region			
Total..............................	100.0	100.0	100.0
Northeast..........................	21.6	22.1	20.9
North Central......................	23.1	23.8	22.3
South..............................	38.4	31.2	48.6
West...............................	16.8	23.0	8.3
Income[1/]			
Total..............................	100.0	100.0	100.0
Under $2,000.......................	21.7	19.7	24.6
$2,000 - $2,999....................	23.7	22.8	24.9
$3,000 - $3,999....................	17.6	19.1	15.5
$4,000 - $4,999....................	11.8	11.4	12.3
$5,000 - $5,999....................	6.3	6.2	6.4
$6,000 or more.....................	14.3	15.4	12.9
NA.................................	4.6	5.4	3.5

1/ The money income levels shown here may be somewhat understated. These data which are from the June 1973 CPS control card are based on the respondent's estimate of total family money income for the preceding 12 month period coded in broad, fixed income intervals.

TABLE 4. PERCENT OF FAMILIES AND PRIMARY INDIVIDUALS WITH INCOMES BELOW THE POVERTY LEVEL IN 1972 AND PURCHASING FOOD STAMPS DURING THE PERIOD JUNE 1, 1972 TO MAY 31, 1973

Age and Sex of Head	Percent Below the Poverty Level in 1972 Purchasing Food Stamps		
	All Races	White	Negro and Other Races
FAMILIES			
Total..	53.3	49.2	61.1
Male head, total.............................	40.3	38.8	45.6
Under 65 years.............................	42.2	41.2	45.2
65 years or older..........................	34.2	31.6	(B)
Female head, total...........................	71.1	71.8	70.5
Under 65 years.............................	74.4	78.0	69.6
65 years or older..........................	(B)	(B)	(B)
PRIMARY INDIVIDUALS			
Total..	30.8	26.2	59.2
Male, total..................................	38.5	36.5	(B)
Under 65 years.............................	38.6	(B)	(B)
65 years or older..........................	(B)	(B)	(B)
Female, total................................	28.7	24.0	(B)
Under 65 years.............................	38.0	32.1	(B)
65 years or older..........................	23.6	20.1	((B)

B Base less than 50,000

REGRESSION MODELS FOR THE COMBINATION OF TWO "INCOMPLETE" DATA SETS IN SECONDARY ANALYSIS

David Elesh, Temple University
James R. Taylor, University of Wisconsin

1.0. Introduction.

A problem frequently encountered in secondary analysis is that there is no one data file completely adequate to the researcher's needs. To be specific, data files which have detailed and comprehensive information on one topic required for an analysis may have only limited coverage of another. Suppose, for example, that one wanted to investigate the effects of health on labor supply and earnings using existing data resources. Perhaps the best available data on health is the National Health Survey, but it has only limited information on labor supply and earnings. On the other hand, the Michigan Survey of Income Dynamics has excellent coverage of labor supply and earnings and poor health data. How, then, are the files to be combined in an analysis? The traditional answer to this question is simply to analyze the data sets separately and to bridge their inadequacies with a variety of extrapolations, inferences, and "informed judgments" which the data may, to varying degrees, support. The problem with this approach is, of course, the difficulty of assessing the inferences made from it. As often as not, confidence in an author's conclusions comes more from the persuasiveness of his theoretical argument than from the weight of the empirical evidence behind it.

Consequently, we propose a different approach to the problem -- one which attempts to combine the best elements of two or more data sets into a single, analyzable file. For present purposes, we shall assume that such data sets are either samples of the same kind (identical probability, simple random, etc.) from the same population or censuses of the same population. To accomplish this combination of files, we shall also make certain distributional assumptions in the context of which a specific model will be estimated. We want to stress the importance of these assumptions at the outset since the validity of our approach depends directly upon their validity. The use of our approach requires prior investigation of the validity of the assumptions unless there is supporting a priori knowledge.

To clarify the exposition, we shall assume in our description of the approach that we wish to combine only two files, "adding" a variable from one file to the other so that regression or other statistical procedures may be performed which would include that variable. A generalization of the approach will be taken up at a later time.

We shall begin with an examination of a "complete data" model and a "restricted complete data" model, both containing p variables, all of which are jointly observed in the same sample. The information gained from this exercise will then be used to place constraints on an "incomplete data" model in which p-1 variables are jointly observed in one sample and the pth variable is "added" from a second and independent sample from the same population. Given these constraints, we shall then show that parameter estimates for the "incomplete" model (1) heuristically parallel that of the "complete data" model and (2) are maximum likelihood estimates.

1.1. The "Complete Data" Model.

Let us suppose that $\underline{Y}$ is a dependent variable of ultimate interest, $\underline{H}$ is the variable to be "added," $\underline{X}_{11}$ and $\underline{X}_{12}$ are sets of other variables, and $\underline{u}$ is a disturbance vector. The regression of $\underline{Y}$, given $\underline{H}$, on the remaining variable then can be written as

$$(1)\qquad (\underline{Y}|\underline{H}) = (\underline{H}\,|\,\underline{X}_{11}\,|\,\underline{X}_{12})\begin{pmatrix}\theta_{10}\\ \theta_{11}\\ \theta_{12}\end{pmatrix} + \underline{u}$$

Let us also suppose that $\underline{Y}$ and $\underline{H}$ are jointly observed random variables (the $\underline{X}$'s are fixed) with the following likelihood function obtained from a random sampling of a bivariate normal distribution:

$$(2)\qquad \frac{\underline{Y}}{\underline{H}} \sim \text{BVN}\left[\begin{pmatrix}\underline{X}_{11} & 0 & \underline{X}_{12}\\ \hline 0 & \underline{X}_{11} & 0\end{pmatrix}\begin{pmatrix}\underline{\theta}_{11} & \theta_{10}\underline{\beta}\\ \hline & \underline{\beta} \\ \hline & \theta_{12}\end{pmatrix};\ \begin{pmatrix}\left(\sigma_1^2 + \theta_{10}^2\sigma_2^2\right)I_n & \theta_{10}\sigma_2^2 I_n\\ \theta_{10}\sigma_2^2 I_n & \sigma_2^2 I_n\end{pmatrix}\right]$$

From this construction of the joint distribution, we can see that the marginal distributions of the two variables are

$$(3)\qquad \underline{Y} \sim N\left[\left(\underline{X}_{11}\,|\,\underline{X}_{12}\right)\left(\theta_{11} + \theta_{10}\underline{\beta}\right);\ \left(\sigma_1^2 + \theta_{10}^2\sigma_2^2\right) I\right]$$

$$(4)\qquad \underline{H} \sim N[\underline{X}_{11}\ \underline{\beta};\ \sigma_2^2 I]$$

while the conditional distribution for $\underline{Y}$, given $\underline{H}$, is

$$(5)\qquad (\underline{Y}|\underline{H}) \sim N\left[(\underline{H}\,|\,\underline{X}_{11}\,|\,\underline{X}_{12})\begin{pmatrix}\theta_{10}\\ \theta_{11}\\ \theta_{12}\end{pmatrix};\ \sigma_1^2 I\right]$$

being definable from the ratio of the joint distribution to the marginal distribution for $\underline{H}$ (Graybill, 1961:63).

Now suppose only independent samples of the marginal distributions are observed. This we shall call the "restricted complete data" model. Equation (1) cannot be estimated because knowledge of the conditional distribution is necessary, and the marginal distributions do not contain sufficient information to identify the parameters of the conditional distribution. To show this, we need only note that model (3) gives the estimate $\widehat{\underline{\theta}_{11} + \theta_{10}\underline{\beta}}$, $\hat{\underline{\theta}}_{12}$, and $\widehat{\sigma_1^2 + \theta_{10}^2\sigma_2^2}$ while model (4) gives estimators $\hat{\underline{\beta}}$ and $\hat{\sigma}_2^2$ -- all of which are clearly inappropriate in the sense that not all the required parameters are estimable since model (3) is over-parameterized.[1]

Our situation is thus analogous to the "incomplete data" model in that only the marginal distributions of our variables are known. Consequently, if a procedure can be found which permits estimation of model (1) where models (2) and (5) are unobservable it may also be applicable to the analysis of the "incomplete data" model.

What is needed are constraints which specify a relationship between the marginal and conditional (or jointly) distributions so that the latter can be identified from the available data in the former. Consider, for example, the following constraints:

$$(6) \qquad \underline{C}'\underline{\theta}_{11} = 0; \; \underline{C}'\underline{\beta} \neq 0$$

where $\underline{C}$ is a vector of known constants. Reference to models (3) and (4) provides an interpretation for (6): it says that the vectors of variables, $\underline{X}_{11}$, do not have the same effect on $\underline{Y}$ as they have on $\underline{H}$. Since the number of cases in which a set of independent variables has different effects on two different dependent variables probably is larger than the number of cases in which the effects are the same, the constraints are not particularly restrictive.

Now in models (3) and (4) let

$$\underline{\ell}_{11} = \underline{\theta}_{11} + \theta_{10}\underline{\beta}$$

$$\underline{\ell}_{12} = \underline{\theta}_{12}$$

and

$$\eta^2 = \sigma_1^2 + \theta_{10}^2\sigma_2^2 \, .$$

The marginal equations for $\underline{Y}$ and $\underline{H}$ for two independent samples would then be

$$\underline{Y} = \left(\underline{X}_{11} \mid \underline{X}_{12}\right)\begin{pmatrix}\ell_{11} \\ \ell_{12}\end{pmatrix} + \underline{v}$$

$$\underline{H} = \underline{X}_{21}\underline{\beta} + \underline{w}$$

where we emphasize the independence of the marginal distribution by changing the first subscript of the X matrix in the equation for $\underline{H}$. The application of ordinary least squares procedures to these equations thus would give the maximum likelihood estimates $\hat{\underline{\ell}}_{11}$, $\hat{\underline{\ell}}_{12}$, $\hat{\eta}^2$, $\hat{\underline{\beta}}$, and $\hat{\sigma}_2^2$.

However, since by (6) $\underline{C}'\underline{\theta}_{11} = 0$, we can obtain estimates for each of the parameters of models (1) and (5) from them (Scheffe, 1959: 16-19):

$$\hat{\theta}_{10} = C'\hat{\underline{\ell}}_{11} \, / \, C'\hat{\underline{\beta}}$$

$$\hat{\underline{\theta}}_{11} = \hat{\underline{\ell}}_{11} - \hat{\theta}_{10}\hat{\underline{\beta}}$$

$$\hat{\theta}_{12} = \hat{\ell}_{12}$$

$$\hat{\sigma}_1^2 = \hat{\eta}^2 - \hat{\theta}_{10}^2\hat{\sigma}_2^2$$

by the invariance property of maximum likelihood estimators (Graybill, 1961: 36-37). Thus our constraints have permitted us to estimate all the parameters of the conditional model of $\underline{Y}$, given $\underline{H}$, even though only independent samples from the marginal distributions of the two variables were observed.

Moreover, since our estimators in (7) are maximum likelihood estimators, they will have the properties of consistency and asymptotic normality. They may, however, be biased. In particular, $\hat{\theta}_{10}$ is a "ratio estimator," and "ratio estimators" are rarely unbiased (Donahue, 1964). But since $\hat{\theta}_{10} = \underline{C}'\hat{\underline{\theta}}_{11} \, / \, \underline{C}'\hat{\underline{\beta}}$ for some a priori constraint vector $\underline{C}$, there may exist an optimal choice for $\underline{C}$. For example, one might choose $\underline{C}$ so that the mean square error for $\hat{\theta}_{10}$ is at a minimum. This would minimize the variance plus the square of the bias of $\hat{\theta}_{10}$, thus having the desirable effect of maximizing the predictive power of $\underline{H}$. However, further research on the optimal choice of $\underline{C}$ is necessary.

I.2. The "Incomplete Data" Model.

As noted earlier, the "incomplete data" model is analogous to the "restricted complete data" model in that only independent samples from the marginal distributions of $\underline{Y}$ and $\underline{H}$ are known. However, the former differs from the latter in that the sampling frames, sampling procedures, and administrative procedures may differ for the two data files to be combined whereas, in the latter case, where there is only a single parent population, these differences do not exist. Consequently, comparison of the two data files to be combined in these terms is a necessary part of validating the "incomplete data" model.

At the same time, it should be said that, despite these differences, the procedures of the preceding section could be used to solve models based on two data files ("incomplete data" models). However, it is instructive to examine an alternative procedure for solving "incomplete data" models. We shall show that, given our assumptions and constraints, this alternative procedure produces results identical to those in the preceding section.

We begin with a regression model for $\underline{Y}$, given $\underline{H}$, which has a somewhat different form from model (5) due to the fact that $\underline{Y}$ is observed in one sample and $\underline{H}$ in another:

(9) $$\left(\underline{Y}^{n_1 x1} \mid \underline{H}_2^{n_2 x1}\right) =$$

$$\left(\hat{\underline{H}}_1^{n_1 x1} \,\vdots\, \underline{X}_{11}^{n_1 x p_1} \,\vdots\, \underline{X}_{12}^{n_1 x p_2}\right) \begin{bmatrix} \theta_{10}^{1x1} \\ \underline{\theta}_{11}^{p_1 x1} \\ \underline{\theta}_{12}^{p_2 x1} \end{bmatrix} + \underline{u}_1^{n_1 x1}$$

where $\underline{Y}$ is again the dependent variable, a column vector of n_1 observations; $\underline{H}_2$ is the variable to be "added" observed in sample 2, a column vector of n_2 observations. $\hat{\underline{H}}_1$ is the "added" measure, a column vector of n_1 observations; $\underline{X}_{11}$ is a submatrix of independent variables $n_1 x p_1$; $\underline{X}_{12}$ is another submatrix of independent variables $n_1 x p_2$; θ_{10} is the coefficient for $\hat{\underline{H}}_1$; θ_{11} is the column vector $(p_1 x1)$ of coefficients for $\underline{X}_{11}$; θ_{12} is the column vector $(p_2 x1)$ of coefficients for $\underline{X}_{12}$; and $\underline{u}_1$ is the column vector $(n_1 x1)$ of disturbances. The initial subscript indexes the sample in question; the second distinguishes subsets of independent variables. Since, except for $\hat{\underline{H}}_1$, equation (9) represents a conventional regression model estimable by ordinary least squares, the next step is to determine $\hat{\underline{H}}_1$ and its implications for the analysis.

To do this, let us suppose that $\underline{H}_2$ is a measure available in the second sample but not in the first. Now suppose that there are a number of other variables which are common to both surveys. From this common list, we want to find the subset that will predict $\underline{H}_2$ as well as possible. We may write this prediction equation as

(10) $$\underline{H}_2^{n_1 x1} = \underline{X}_{21}^{n_2 x p_1} \, \underline{\beta}^{p_1 x1} + \underline{u}_2^{n_2 x1}$$

where $\underline{H}_2$ is the "added" variable in the second sample, a column vector $n_2 x1$; $\underline{X}_{21}$ is the subset of independent variables described in $\underline{X}_{11}$ but measured in the second sample, a submatric $n_2 x p_1$; $\underline{\beta}$ is the column vector of coefficients, and $\underline{u}_2$ is the disturbance vector $(n_2 x1)$; and where the disturbances, $\underline{u}_1$ and $\underline{u}_2$, have a bivariate normal distribution with zero means, finite variances, and zero covariances.

$$\begin{pmatrix} \underline{u}_1 \\ \underline{u}_2 \end{pmatrix} \sim \text{BVN}\left[\begin{pmatrix} \underline{0}^{n_1 x1} \\ \underline{0}^{n_2 x1} \end{pmatrix} \begin{pmatrix} \left(\sigma_1^2 + \theta_{10}^2 \sigma_2^2\right) I_{n_1} & \underline{0} \\ \underline{0} & \sigma_2^2 I_{n_2} \end{pmatrix}\right]$$

Thus $\hat{\underline{H}}_1$ can be computed as

(11) $$\hat{\underline{H}}_1 = \underline{X}_{11} \, (\underline{X}_{21}'\underline{X}_{21})^{-1} \underline{X}_{21}' \underline{H}_2 \equiv \underline{X}_{11}\hat{\underline{\beta}} \, .$$

Substituting equation (11) into equation (9) we get

(12) $$(\underline{Y} \mid \underline{H}_2) = [\underline{X}_{11}\hat{\underline{\beta}} \,\vdots\, \underline{X}_{11} \,\vdots\, \underline{X}_{12}] \begin{bmatrix} \theta_{10} \\ \underline{\theta}_{11} \\ \underline{\theta}_{12} \end{bmatrix} + \underline{u}_1$$

$$= [\underline{X}_{11} \,\vdots\, \underline{X}_{12}] \begin{bmatrix} \underline{\theta}_{11} + \theta_{10}\hat{\underline{\beta}} \\ \underline{\theta}_{12} \end{bmatrix} + \underline{u}_1$$

In words, what we have done has been to predict a variable, $\underline{H}$, found in the second sample but not in the first from a list of variables common to both surveys; then using the coefficients from the equation run with the second sample data and the scores of the appropriate variables (the $\underline{X}_{11}$) in the first survey, we produced a set of predicted values for the $\underline{H}$ variable for the first sample. The result is equation (9).

However, equation (9) cannot be estimated as it stands because its coefficient matrix is singular as shown below:

$$[\underline{X}_{11} \, (\underline{X}_{21}'\underline{X}_{21})^{-1} \underline{X}_{21}'\underline{H}_2 \,\vdots\, \underline{X}_{11} \,\vdots\, \underline{X}_{12}] \cdot \begin{bmatrix} 1^{1x1} & \underline{0}^{1xp_1} & \underline{0}^{1xp_2} \\ -(\underline{X}_{21}'X_{21})\underline{X}_{21}^{-1}\underline{H}_2 & I_{p_1} & \underline{0}^{p_1 x p_2} \\ \underline{0}^{p_2 x1} & \underline{0}^{p_2 x p_1} & I_{p_2} \end{bmatrix}$$

$$= [\underline{X}_{11} \, (\underline{X}_{21}'\underline{X}_{21})^{-1} \underline{X}_{21}'\underline{H}_2 - \underline{X}_{11} \, (\underline{X}_{21}'\underline{X}_{21})^{-1} \cdot \underline{X}_{21}'\underline{H}_2 \,\vdots\, \underline{X}_{11} \,\vdots\, \underline{X}_{12}] = \left[\underline{0}^{n_1 x1} \,\vdots\, \underline{X}_{11}^{n_1 x p_1} \,\vdots\, \underline{X}_{12}^{n_1 x p_2}\right].$$

Consequently, as was done with the "complete data" model, it will be necessary to place constraints on the coefficients.

A variety of constraints are, of course, possible. For example, we could set $\theta_{10} = 0$, but this has the rather ridiculous effect of asserting that $\underline{Y}$ and $\underline{H}$ are unrelated. Alternatively, we may employ the general constraint:[2]

$$a_1\theta_{10} + \sum_{j=1}^{p_1} c_j\theta_{11j} + \sum_{j=1}^{p_2} d_j\theta_{12j} = 0$$

where the a_1, c_j, and d_j are all known real numbers such that not all are zero. The specification of the above constraint will depend on the particular model being estimated and the reasonableness of any specific constraint in that context. For example, one might choose

$$\sum_{j=1}^{p_1} \theta_{11j} = 0$$

$$\sum_{j=1}^{p_2} \theta 12_j = 0 \ .$$

In our particular case, we shall use neither or these but instead assume that $\underline{C}'\underline{\theta}_{11} = 0$. If we define the first column of the $\underline{X}_{11}$ matrix as a vector containing only unity for elements, then letting

$$\theta_{111} = 0$$

implies that

$$\underline{C}' = (1,0,0,\ldots,0)$$

and that $\underline{C}'\underline{\theta}_{11} = 0$ and $\underline{C}'\underline{\beta} \neq 0$. In other words, the intercept for the $\underline{X}_{11}$ matrix is set to zero. Model (9) can now be solved. The coefficient for $\hat{\underline{H}}_1$ is in fact $\hat{\theta}_{10}$; the last p-1 columns of $\underline{X}_{11}$ give the estimates, $\hat{\underline{\sigma}}_1^2$; and the coefficients of $\underline{X}_{12}$ are the $\hat{\underline{\theta}}_{12}$. The estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ can be computed from the residual sum of squares for equations (9) and (10).

Comparison of equations (9), (10), (12), with (3) and (4) show that our estimates for θ_{10}, $\underline{\theta}_{11}$, and $\underline{\theta}_{12}$ are the same for both the restricted "complete data" and "incomplete data" models. Furthermore, comparing the "incomplete data" model (9) with the "complete data" model (5), it can be seen that (9) is an approximation of (5) in the sense that $\hat{\underline{H}}_1$ has been used as a substitute for $\underline{H}_1$. The constraint $\underline{C}'\underline{\theta}_{11} = \theta_{111} = 0$ was here chosen for illustrative purposes only, and may not be optimal in terms of the mean square error of $\hat{\theta}_{10}$ criterion previously mentioned. However, reference to models (1), (3) and (4) provides an easy interpretation of the constraint $\theta_{111} = 0$. That is to say $\theta_{111} = 0$ implies that the intercept term of model (3) is equal to the product of θ_{10} and the intercept term of model (4), and that the intercept term of model (1) is zero. Because of our distributional assumptions, the estimates are also maximum likelihood estimates. We should note, however, that we do not know how sensitive these estimates are to violations of the assumptions of distributional normality made here. Testing these assumptions will be one of our first research tasks.

FOOTNOTES

[1] Specifically, model (3) contains one too many parameters for unique estimators to exist unless an additional constraint is imposed. Since model (3) is induced from the bivariate model (2), and models (2) and (4) lead to model (5); it is necessary that any such constraint on model (3) be compatible with models (2) and (5). An appropriate constraint is given below.

[2] This alternative constraint derives from a comparison of models (3) and (12). The original constraint was applied to model (3) is identical to the final form of the parameter matrix of (12). The alternative constraint is applied to the initial parameter matrix of model (12) and is thereby equivalent to original constraint.

REFERENCES

Donahue, James D., Products and Quotients of Random Variables and their Applications, Washington, D.C.: Office of Technical Services, U.S. Department of Commerce, Aerospace Research Laboratories Technical Report, ARL 64-115.(1964).

Graybill, Franklin A., An Introduction to Linear Statistical Models, New York: McGraw-Hill, vol. 1. (1961).

Scheffe, Henry, The Analysis of Variance, New York: Wiley. (1959).

DISTRIBUTION OF THE RATIO OF CORRELATED SUMS OF GAMMA VARIABLES

John A. Flueck, Burt S. Holland, and Ru-Ying Lee,* Temple University

1. Introduction

In this paper we present the exact probability distribution of the ratio of two correlated gamma random variables for a given bivariate gamma structure. Interest in the theory and application of ratios of random variables has been present in the literature for many years. The first two authors [5] have reviewed some of these results.

The probability distribution of the ratio of two independent r.v.'s is generally easy to obtain, cf. Kendall and Stuart [11, p. 265]. It is well known that the distribution of the ratio of two independent normals is Cauchy and of two independent gammas is the Inverted Beta Type II, Kullback [12]. However, the literature on the distribution of the ratio of two correlated r.v.'s is less developed.

Geary [7] was apparently the first to present results for the ratio of two correlated normals, under the restriction of an always positive denominator. Fieller [4] obtained more general results for the same problem. Rietz [15] presented results for the ratio of correlated uniform r.v.'s. In 1937, Cramér [2] showed that the p.d.f. of the ratio $r=U/V$ of two continuous r.v.'s with $P(V>0)=1$, $E(V)<\infty$, is given by an inversion formula which is in terms of the joint characteristic function of U and V. Application of this approach requires, of course, knowledge of the joint characteristic function and the ability to perform the indicated integration. Gurland [9] has generalized this result by presenting an inversion formula for the probability distribution of a ratio of linear combinations of the same random variables. This formula is based on an n-variate characteristic function. In 1952, C. R. Rao [14, p. 207] showed that under very general conditions the standardized ratio of two means is asymptotically normally distributed. More recently, Marsaglia [13] has studied the probability distribution of the ratio of correlated normal r.v.'s, providing 63 computer-drawn plots of p.d.f.'s to illustrate that the distribution may be symmetric or skewed and unimodal or bimodal.

2. A Bivariate Gamma Distribution

Let X, Y, Z be independently distributed gamma r.v.'s with skew parameters a,b,c and a common scale parameter λ', e.g.,

$$f_X(x)=(\lambda'x)^{a-1}\lambda' e^{-\lambda' x}/\Gamma(a),\ x>0,\ a>0,\ \lambda'>0. \quad (1)$$

Following Weldon's approach, David and Fix [3] defined that for $\lambda'=1$, $U=X+Y$ and $V=X+Z$ are bivariate gamma distributed with density

$$g_{U,V}(u,v) = \frac{e^{-u-v}}{\Gamma(a)\Gamma(b)\Gamma(c)}\int_0^{\min\{u,v\}} t^{a-1}(u-t)^{b-1}(v-t)^{c-1}e^t dt \quad (2)$$

and correlation coefficient $\rho=a/[(a+b)(a+c)]^{\frac{1}{2}}$.

In this paper we consider the distribution of the ratio of gamma variables

$$r=U/V=(X+Y)/(X+Z)$$

It is clear that the scale parameter λ' does not appear in the p.d.f. of r, hence we may take $\lambda'=1$ in further discussion of the distribution. Flueck and Holland [5, 6] have previously presented results for the moments of r and a first attempt at its distribution.

A disadvantage of the above bivariate gamma formulation (2) is that it does not admit negative values of the correlation coefficient ρ. The authors are also examining other bivariate gamma distributions (see Johnson and Kotz [10]).

The ratio r easily generalizes to the ratio of sums of gamma variates,

$$r^*=\sum_{i=1}^{m}(X_i+Y_i)/\sum_{j=1}^{n}(X_j+Z_j),$$

where the $\{X_i\}$, $\{Y_i\}$ and $\{Z_j\}$ are mutually independent r.v.'s identically distributed within each set. If we let $r(a,b,c)$ denote the probability distribution of r with numerator skew parameters a and b, and denominator skew parameters a and c, it can be shown, using the regenerative property of the gamma distribution, that

$$r^* \sim \begin{cases} r(na,\ (m-n)a+mb,\ nc),\ m>n \\ r(ma,\ mb,\ (n-m)a+nc),\ m\le n, \end{cases} \quad (3)$$

i.e., r^* is a member of the same family of probability distributions as r.

An extension of r is

$$r' = \lambda r = \lambda(X+Y)/(X+Z)$$

so that the numerator is gamma distributed with scale parameter λ'/λ rather than λ'.

3. Derivation of the Probability Density of r.

Let $T=X/(X+Z)$. Then the joint density of r, T, and V is

$$f(r,t,v;a,b,c)=\frac{(tv)^{a-1}e^{-tv}}{\Gamma(a)}\cdot\frac{[v(r-t)]^{b-1}e^{-v(r-t)}}{\Gamma(b)}\cdot\frac{[v(1-t)]^{c-1}e^{-v(1-t)}}{\Gamma(c)}\cdot v^2,\ 0<r,\ 0<v,\ 0<t<\min\{1,r\},$$

and the joint density of r and t is

$$f(r,t;a,b,c)=K(a,b,c)\frac{t^{a-1}(r-t)^{b-1}(1-t)^{c-1}}{(1-t+r)^{a+b+c}},\ 0<r,\ 0<t<\min\{1,r\},$$

where

$K(a,b,c)=\Gamma(a+b+c)/\Gamma(a)\Gamma(b)\Gamma(c)$.

Hence the p.d.f. of r is

$$f(r;a,b,c) = \int_0^{\min\{1,r\}} f(r,t;a,b,c)dt, \quad r>0. \tag{4}$$

Evaluation of (4) by elementary methods appears possible only when a,b,c are all positive integers. For example, when a=b=c=1, (i.e., X, Y, Z each exponential r.v.'s),

$$f(r;1,1,1)=[\max\{1,r\}]^{-2} - (1+r)^{-2}, \quad r>0.$$

We attempted to obtain (4) in closed form using the inversion formula in Gurland [9]; however, the integral in the inversion formula has proved difficult to evaluate.

Next we considered calculation of (4) by numerical methods [6], including Simpson and Romberg integration. However, the p.d.f. (4) often has vertical asymptotes at r=0 and r=1; these created problems in obtaining adequate numerical accuracy within reasonable economic limits.

The approach finally adopted and discussed below involved rearrangement of the integral in (4) so that it may be expressed as a weighted difference of hypergeometric functions. The hypergeometric functions are defined and convergent except possibly at r=1.

In (4), for r≤1, replace t with rt to yield

$$f(r;a,b,c) = \begin{cases} K(a,b,c) \cdot \displaystyle\int_0^1 \frac{r^{a+b-1}t^{a-1}(1-t)^{b-1}(1-rt)^{c-1}}{(1-rt+r)^{a+b+c}} dt, & r\le 1, \\ K(a,b,c) \cdot \displaystyle\int_0^1 \frac{t^{a-1}(r-t)^{b-1}(1-t)^{c-1}}{(1-t+r)^{a+b+c}} dt, & r>1. \end{cases} \tag{5}$$

For r≤1 it follows from the identity

$$r = (1-rt+r) - (1-rt)$$

and 3.211, 9.1821 in [8, pp. 287, 1054] that

$$f(r;a,b,c) = K(a,b,c)\, r^{a+b-2} \int_0^1 \frac{t^{a-1}(1-t)^{b-1}(1-rt)^{c-1}}{(1-rt+r)^{a+b+c-1}} dt - \int_0^1 \frac{t^{a-1}(1-t)^{b-1}(1-rt)^{c}}{(1-rt+r)^{a+b+c}} dt$$

$$= K(a,b,c)B(a,b)r^{a+b-2}(1+r)^{-b-c}[(1+r)F(a,1-c,a+b;r^2) -F(a,-c,a+b;r^2)], \tag{6}$$

where $F(k_1,k_2,k_3;x)$ is the hypergeometric function (see [1]) and

$$B(k_4,k_5)=\Gamma(k_4)\Gamma(k_5)/\Gamma(k_4+k_5), \quad k_4,\ k_5 > 0.$$

Note that F may not converge when r=1. Similarly, when r>1 we use the identity

$$1=(1-t+r)-(r-t)$$

to arrive at

$$f(r;a,b,c)=K(a,b,c)B(a,c)r^{b-a-1}(1+r)^{-b-c}[(1+r) F(a,1-b,a+c;r^{-2})-rF(a,-b,a+c;r^{-2})]. \tag{7}$$

It is interesting to note that if c in (6) or b in (7) is an integer then the series expression of the corresponding hypergeometric function contains only a finite number of terms.

Unless r=0 or r=1, we have found that (6) and (7) are satisfactory for numerically computing f(r;a,b,c). In most cases less than 100 terms of the hypergeometric series are needed to obtain accuracy to five decimal places, the rate of convergence depending on r, a, b, and c. For the two points r=0 and r=1, analytic formulas have been developed for the p.d.f. and its derivatives.

The presentation here is restricted to a presentation of results; derivations of the analytic formulas rely on the Lebesgue Dominated Convergence Theorem and Fatou's Lemma.

The behavior of f(r;a,b,c) at r=0 is principally governed by the value of a+b. We find:

$$\lim_{r\to 0} f(r;a,b,c) = \begin{cases} 0 & \text{if } a+b>1 \\ c & \text{if } a+b=1 \\ \infty & \text{if } a+b<1 \end{cases} \tag{8}$$

and

$$\lim_{r\to 0} f'(r;a,b,c) = \begin{cases} 0 & \text{if } a+b>2 \\ c(c+1) & \text{if } a+b=2 \\ \infty & \text{if } 1<a+b<2 \\ c(2a-1-c) & \text{if } a+b=1 \\ -\infty & \text{if } a+b<1 \end{cases} \tag{9}$$

For k>1, the expression for

$$\lim_{r\to 0} f^{(k)}(r;a,b,c)$$

is quite complicated and we note that the limit may conceivably be positive or negative.

On the other hand, the shape of f(r;a,b,c) in the neighborhood of r=1 is primarily governed by b and c. If b+c>1, then f(r;a,b,c) is continuous at r=1 and

$$f(1;a,b,c)=\frac{\Gamma(b+c-1)}{\Gamma(b)\Gamma(c)} \cdot \frac{2a+b+c-1}{2^{b+c}}, \tag{10}$$

while if $b+c\le 1$ and $a>0$,

$$f(1^-;a,b,c) = f(1^+;a,b,c) = \infty \qquad (11)$$

If $b+c>2$, $f'(r;a,b,c)$ is continuous at $r=1$ and

$$f'(1;a,b,c)=K(a,b,c)B(a,b+c-2)2^{-b-c-1}[(-a+b-c-2)$$

$$+\frac{a(-a+3b-c-4)}{a+b+c-2} + \frac{a(a+1)(2a+3b+c)}{(a+b+c-2)(a+b+c-1)}]. \qquad (12)$$

Note that (12) may be positive or negative. Next, if $0<b+c\le 1$ and $a>0$ we find

$$f'(1^-;a,b,c)=\infty \qquad f'(1^+;a,b,c)=-\infty \qquad (13)$$

The intermediate case $1<b+c\le 2$, $a>0$ breaks down into six subcases as follows:

If $c>1$ and $b<1$,

$$f'(1^-;a,b,c)= -\infty,\ f'(1^+;a,b,c)= -\infty; \qquad (14a)$$

if $c=1$ and $b=1$,

$$f'(1^-;a,b,c)=\tfrac{1}{4}(2a^2-1),\ f'(1^+;a,b,c)=-\tfrac{1}{4}(2a^2+4a+1); \qquad (14b)$$

if $c=1$ and $b<1$,

$$f'(1^-;a,b,c)=\frac{1}{2^{b+2}}(4a^2+4ab+b^2-4a-3b),$$

$$f'(1^+;a,b,c)=-\infty; \qquad (14c)$$

if $c<1$ and $b>1$,

$$f'(1^-;a,b,c)=\infty,\ f'(1^+;a,b,c)=\infty; \qquad (14d)$$

if $c<1$ and $b=1$,

$$f'(1^-;a,b,c)=\infty,\ f'(1^+;a,b,c)=-\frac{1}{2^{c+2}}(3a^2+3ac+c^2+3a+c); \qquad (14e)$$

if $c<1$ and $b<1$,

$$f'(1^-;a,b,c)=\infty,\ f'(1^+;a,b,c)=-\infty. \qquad (14f)$$

For $a=0$, the independent case, the results for $f(r;a,b,c)$ and $f'(r;a,b,c)$ follow from [12], with

$$f(r;0,b,c) = \frac{1}{B(b,c)} \frac{r^{b-1}}{(1+r)^{b+c}}.$$

4. Discussion of the Probability Density of r.

It is seen from (8)-(14) that the shape of $f(r;a,b,c)$ is determined by a, b, and c. To catalogue the situation, we present a 36-cell partitioning of the parameter space $\{(a,b,c): a\ge 0,\ b>0,\ c>0\}$. Table 1 summarizes the implications of (8)-(14) for each cell, while Figures 1-4 contain Calcomp computer plots of $f(r;a,b,c)$ for selected values of a, b, c in order to display the various members of the family. Note that in the Table and in the following discussion we abbreviate $f(r;a,b,c)$ to $f(r)$. Also we use the symbol k to represent a non-negative finite generic constant and k^* to represent a positive finite generic constant.

It is clear from the Figures that the graph of $f(r;a,b,c)$ can assume many unusual shapes and that its appearance is sensitive to small changes in the parameters a, b, c. The following are among the special features of $f(r;a,b,c)$:

1. There is a vertical asymptote at $r=0$ ($f(0;a,b,c)=\infty$ and $f'(0;a,b,c)=-\infty$) iff $a+b<1$.

2. If $a+b\le 1$, the distribution may be bimodal with modes at $r=0$ and $r=1$.

3. There is a vertical asymptote at $r=1$ ($f(1^-;a,b,c)=f(1^+;a,b,c)=f'(1^-;a,b,c)=\infty$ and $f'(1^+;a,b,c)=-\infty$) iff $b+c\le 1$.

4. If $b+c\le 2$, $f'(r;a,b,c)$ is discontinous at $r=1$.

5. The slope of the p.d.f. at $r=0$, $f'(0;a,b,c)$, may be extremely senstitive to small changes in $a+b$. For example, it follows from (9) that

$$f'(0;a,b,c) = \begin{cases} -\infty & \text{if } a+b = 0.99 \\ \pm k & \text{if } a+b = 1.00 \\ \infty & \text{if } a+b = 1.01 \end{cases}$$

The Figures also suggest that as a, b, c, each increase, the distribution of r approaches normality. Notice that if in (3) we have $m=n$, the ratio of sums r^* is also a ratio of means, and

$$r^* \sim r(na,nb,nc).$$

Applying a general lemma of C. R. Rao [14], it easily follows that

$$\sqrt{n}\ [r(na,nb,nc)- \frac{a+b}{a+c}]\to N(0,\frac{(a+b)(b+c)}{(a+c)^3}).$$

This implies that the large sample distribution of $r(na,nb,nc)$ approaches normality as n increases and ultimately becomes degenerate at $(a+b)/(a+c)$.

For completeness we also wish to mention that the first two authors [5] have pointed out that $a+c$ is the quantity which determines whether or not moments of r are finite; in particular, it was found that

$E(r)<\infty$ iff $a+c>1$.

and

$Var(r)<\infty$ iff $a+c>2$.

The fact that some moments of r may be infinite is further evidence of the unusual nature of the distribution of r.

5. Conclusion

This paper has presented exact "closed form" distributional results (6), (7) for the ratio of two correlated gamma r.v.'s. These results also allow the statistician to obtain an efficient numerical approximation to the exact distribution with any degree of accuracy.

The results presented above for the ratio of correlated gamma variates suggest that the probability distributions of ratios of random variables, and of small-sample ratios of sums of random variables, often take unusual forms. In addition to the distribution theory presented here, the p.d.f.'s have an important implication for the practicing statistician: calculation of probability statements involving ratios of r.v.'s is often undertaken assuming normality, but as the above graphs have shown, these calculations must be treated with considerable caution. Furthermore, our results indicate that the accuracy of such probability calculations may rely heavily on precise estimation of the parameters of the ratio distribution. We hope to address this and other problems in future papers as part of our continuing study of ratios of random variables.

References

[1] Abramovitz, M. and Stegun, I. (1964). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Washington: U.S. Government Printing Office.

[2] Cramer, H. (1937). Random Variables and Probability Distributions, (Cambridge Tracts in Mathematics, No. 36), Cambridge University Press.

[3] David, F. N., and Fix, E. (1961). "Rank Correlation and Regression in a Nonnormal Surface," Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, 177-197.

[4] Fieller, E. C. (1932). "The Distribution of the Index in a Normal Bivariate Population," Biometrika, 24, 428-440.

[5] Flueck, J. A., and Holland, B. (1973). "Ratio Estimation Problems in Meterological Research," Proceedings, Third Conference on Probability and Statistics in Atmospheric Science, American Meterological Society, Boston, Mass.

[6] Flueck, J. A., and Holland, B. (1974). "The Moments and Distribution of the Ratio of Two Gamma Variables," 1973 Proceedings, Social Statistics Section, American Statistical Association, Washington, D.C.

[7] Geary, R. C. (1930). "The Frequency Distribution of the Quotient of Two Normal Variates," J. Roy. Statist. Soc., 93, 442-446.

[8] Gradshteyn, I. and Ryzhik, I. (1965). Table of Integrals, Series, and Products. New York: Academic Press.

[9] Gurland, J. (1948). "Inversion Formulae for the Distribution of Ratios," Ann. Math. Statist. 19, 228-237.

[10] Johnson, N. L, and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley and Sons, Inc.

[11] Kendall, M. G. and Stuart, A. (1963). The Advanced Theory of Statistics, Vol. I., Second Edition. London: Charles Griffin and Co., Ltd.

[12] Kullback, S. (1936). "The Distribution Laws of the Difference and Quotient of Variables Independently Distributed in Pearson Type III Laws," Ann. Math. Statist. 7, 51-53.

[13] Marsaglia, G. (1965). "Ratio of Normal Variables and Ratios of Sums of Uniform Variables," J. Amer. Statist. Assoc. 60, 193-204.

[14] Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, Inc.

[15] Rietz, H. L. (1936). "On the Frequency Distribution of Certain Ratios," Ann. Math. Statist. 7, 145-153.

Footnotes

* The authors' names are listed in alphabetical order.

Table 1. Summary of Behavior of $f(r) = f(r;a,b,c)$ for $r=0,1$.

		$a+b>2$	$a+b=2$	$1<a+b\leq 2$	$a+b=1$	$a+b<1$
$b+c>2$		$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1)=\pm k$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1)=\pm k$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1)=\pm k$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1)=\pm k$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=k^*$ $f'(1)=\pm k$
$1<b+c\leq 2$	$c>1,\ b<1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$
	$c=1,\ b=1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=k^*$ $f'(1^+)=-k^*$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=k^*$ $f'(1^+)=-k^*$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\pm k$ $f'(1^+)=-k^*$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=-k^*$ $f'(1^+)=-k^*$	Impossible
	$c=1,\ b<1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=k^*$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=k^*$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\pm k$ $f'(1^+)=-\infty$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=-k^*$ $f'(1^+)=-\infty$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=k^*$ $f'(1^-)=-k^*$ $f'(1^+)=-\infty$
	$c<1,\ b>1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=\infty$	Impossible	Impossible
	$c<1,\ b=1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-k^*$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-k^*$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-k^*$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-k^*$	Impossible
	$c<1,\ b<1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=+\infty$
$b+c\leq 1$		$f(0)=0$ $f'(0)=0$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$

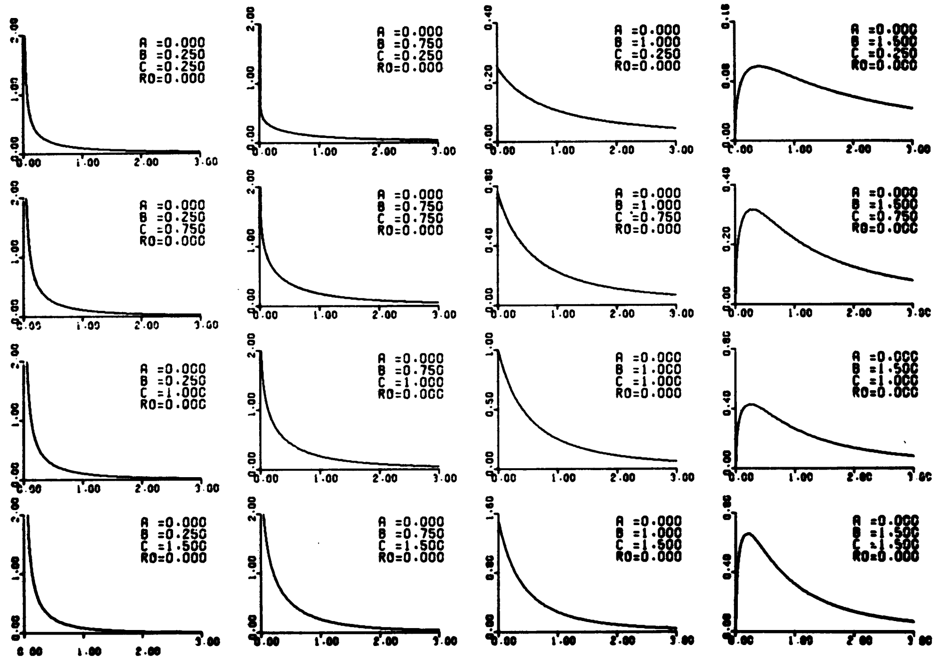


Figure 1. Plots of f(r;a,b,c) versus r.

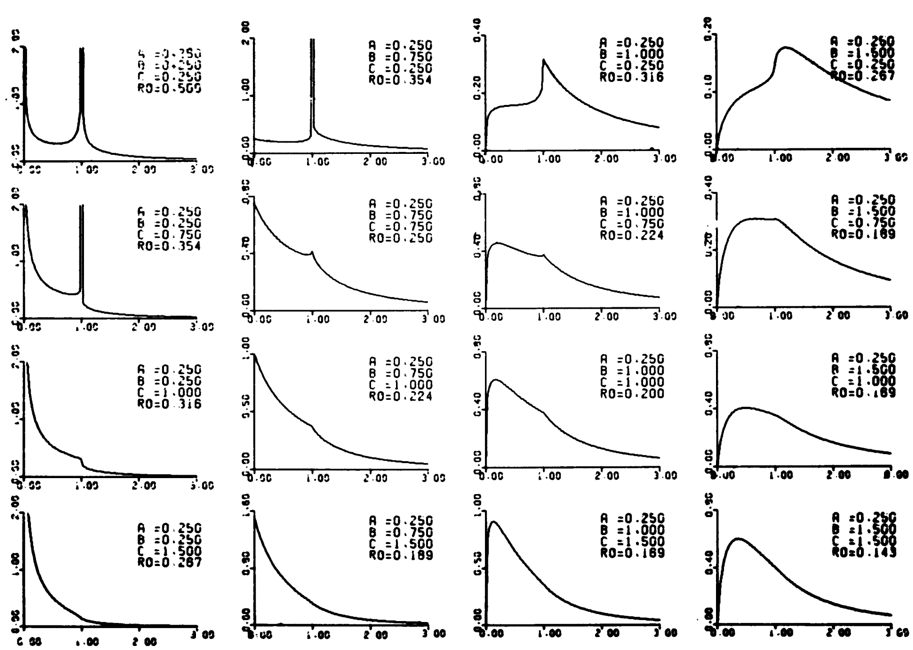


Figure 2. Plots of f(r;a,b,c) versus r.

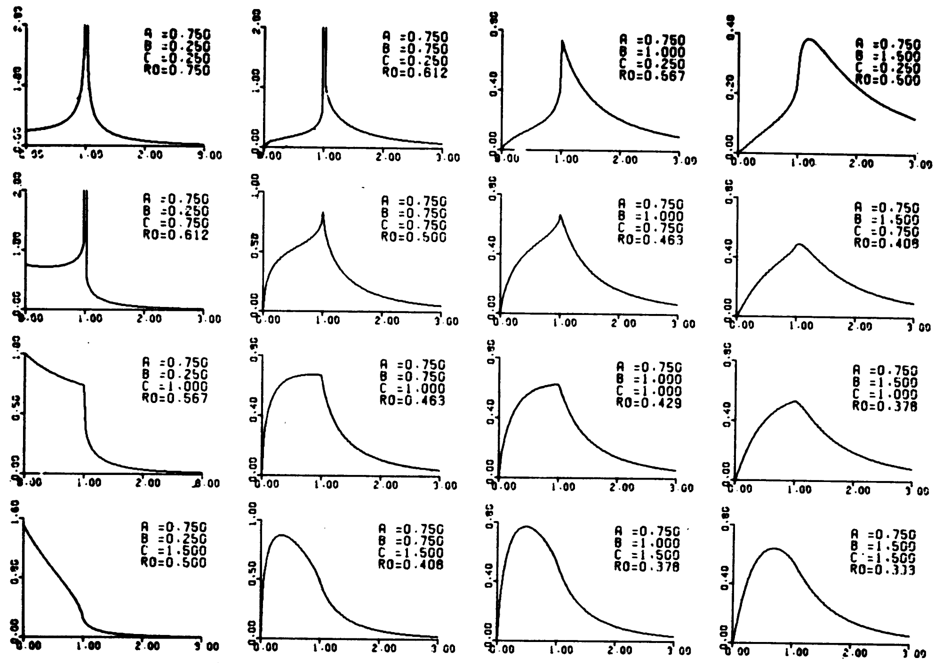


Figure 3. Plots of f(r;a,b,c) versus r.

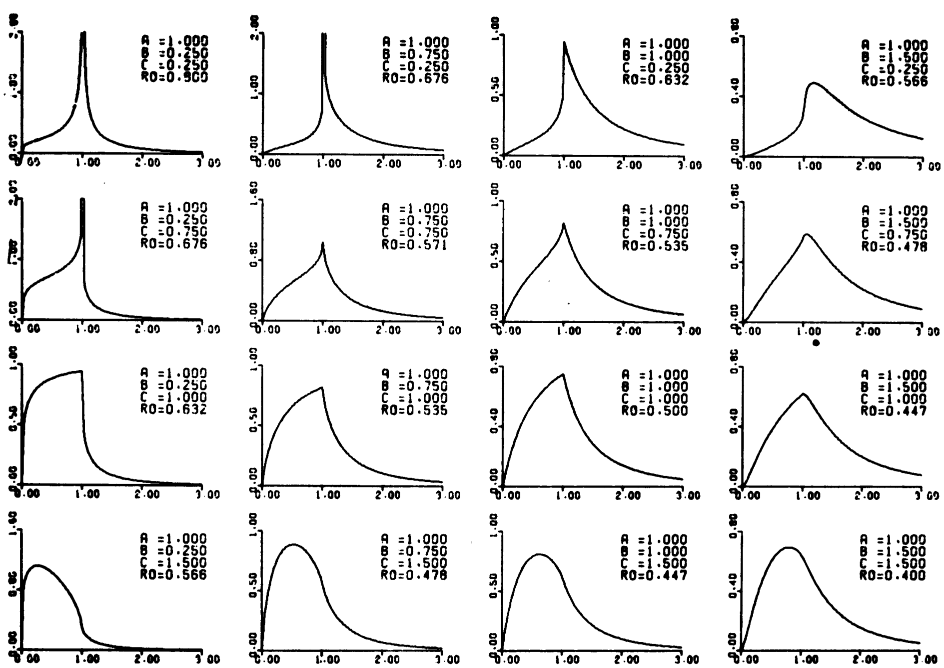


Figure 4. Plots of f(r;a,b,c) versus r.

THE EFFECTS OF SOCIAL SUBAREAS IN URBAN RESIDENTIAL MOVEMENT: AN ANALYSIS OF RESIDENCE HISTORIES[1]

William H. Frey, Center for Studies in Demography and Ecology
Department of Sociology, University of Washington

Areal factors have been used in various ways in studies of metropolitan residential movement. Some have used an ecological approach showing, for example, that those areas with high mobility rates can be characterized by certain social and demographic characteristics (Moore, 1969). Other studies have focused on the demographic effects of net migration on a region by looking at the net movement from various social and economic subareas (Goldstein, 1965). A third type of analysis looks at the characteristics of movers from social and demographic areas of various types (Rossi, 1955). Although the first two types of studies are useful in determining broad structural patterns of change in an urban area, they give little in the way of clues as to which individuals are doing the moving. The third type of study which looks not only at areal characteristics but at individual factors as well, allows one to differentiate between the effects of each on mobility. This type of study is least represented in the literature due to the difficulty in obtaining relevant data for both characteristics of the individual and place of origin.

Utilizing residence histories and census tract data, this analysis will investigate the influences of areal and individual characteristics as mobility determinants in the state of Rhode Island. The state generally represents the greater Providence metropolitan area--a medium-sized, mature northeast urban center which has undergone significant population decline in its central city area. Areal influences on movement will be assessed here in terms of small social subareas (census tracts).

AREAL INFLUENCES

Perhaps the most noteworthy physical characteristic in this regard is neighborhood density. The density criterion measured by "population per square mile" has proved successful in the explanation of mobility at the aggregate level particularly for a mono-nuclear city around which population density tends to exhibit a negative exponential relationship with the distance from the city center (Moore, 1971). This is in part explained since areas of greatest density are disproportionately composed of apartments, multiple family dwellings, and rented housing which generally attract the more mobile segments of the population. However, other studies suggest some very real environmental components associated with neighborhood density which could predispose one to movement (Lansing and Hendricks, 1967).

There is some agreement among writers on mobility that social areal characteristics are of more importance than either locational or physical characteristics (Abu-Lughod and Foley, 1960; Rossi, 1955). Shevky and Bell (1955) and more recent factorial ecologists (see Berry, 1970) have found urban regions to be differentiated according to dimensions of socio-economic status, family status, and minority status. This study will focus on the first two of these dimensions in addition to the past turnover status of urban subareas as social areal factors.

One might look at the household and housing characteristics usually associated with certain socio-economic status areas. In general upper socio-economic household heads are more likely to live in owned, single-family dwellings and are usually further along in the life cycle than are household heads from lower class areas. Since each of these characteristics are related to lower mobility levels, one might expect generally lower residential movement originating in higher socio-economic neighborhoods.

Aside from the compositional effects associated with the mobility of each status-linked area, one might look at mobility behavior of individual households because of their residence in these status areas. For example, it is reasonable to assume that lower and middle status neighborhoods are the source areas of those households which are upwardly mobile both socially and residentially.

One further consideration is the changing class status of areas within the metropolitan community. According to the filtering model (Grigsby, 1963), a few residents of medium or high income areas tend to vacate their dwellings due to obsolescence of the neighborhood location, As the process gets under way, other long-time residents react to the same stimulus or to the undesirability of new arrivals into the neighborhood.

There is also a segregation of residential areas by family status within the larger metropolis. A simplified model suggests that low family status areas will be inhabited by renters, apartment dwellers, and young people--all characteristics shown to be highly related to mobility. High family status areas will consist mostly of homeowners, single-family unit dwellers and families with children--characteristics associated with infrequent movers (Johnston, 1971). Aside from these compositional effects, the family status of a neighborhood has environmental influences on mobility incidence. In Rossi's (1955) study, the full families that resided in low family status areas showed higher mobility than any other household type.

The interest here in neighborhood turnover status stems again from Rossi's work which shows that when the socio-economic status of an area is taken into account, families in high turnover areas are more likely to move than those in low turnover areas.

Although Social Areal Analysis and other factorial ecologies isolate what appear to be independent social dimensions of urban areas, intercorrelations and interactions exist between them, many times in regularly predictable patterns. Also, it is possible that each of the social characteristics might have quite different mobility effects among various urban locations. In light of these issues, the following questions are relevant: (1) Which social areal factors are most important in the determination of the incidence of residential mobility, given the fact that various interdependencies exist? (2) To what extent do areal determinants operate differently in the explanation of mobility incidence for zones of various population densities? (3) What are the relative effects of individual level determinants, on the one hand, and areal determinants, on the other, in the explanation of mobility incidence?

DATA AND TABULATION TECHNIQUE

This study utilizes two data sources: residence histories as well as 1960 and 1970 census tract data. The residence histories were collected from a probability sample of Rhode Island adults who were interviewed in 1967 and 1968 as part of a survey undertaken by the Brown University Population Research Laboratory (see Organic & Goldstein, 1970).

A residence history was obtained from each person interviewed that included every place of residence since birth. For purposes of this study, the portion of the residence history used was only that between 1955 and 1967 since our measures of the areal characteristics associated with each residence were obtained from the 1960 and 1970 census and it was felt that these measures could not be inferred to residences before 1955. Portions of the residence histories were excluded in the following instances: when the individual was not residing in the state; and those years of residence before age eighteen. The former exclusion was made since the focus of the study is only on intra-urban mobility; the latter was made since most moves made prior to age 18 were probably decided by parents.

The data were tabulated utilizing a technique wherein the residence history of each respondent is divided into one-year life segments and these segments were the units of analysis. Since 13 years of a respondent's residence history can be included in the analysis, it is possible for a single respondent to contribute thirteen segments or analysis units to the study; hence although 2,233 survey respondents provide the source of information in the form of residence histories, the actual analyses are based on 22,644 one-year life segments. Although the original survey constituted a probability sample of Rhode Island adults in 1967-68, the population of one-year segments does not comprise a representative sample for the analogous population in 1955-67 because of death and out-movement of previous Rhode Island residents and due to memory failure associated with the residence histories themselves. Each segment was classed as mobile or non-mobile depending upon whether a move had taken place during the segment year examined. Tabulations in the analysis, then, will be presented in terms of "mobility rates"--(move segments/total segments) X 100--for categories associated with the independent variables.

This data source is unique in that census tract locations were recorded for each residence in the histories. Here it is possible to categorize the one-year segments by census tract characteristics and to analyze the effects of these characteristics on mobility incidence.

The social areal characteristics used to classify tracts represent measures of socio-economic status, change in socio-economic status, family status, change in family status, and turnover status. These measures are based on 1960 and 1970 census tract data for the state of Rhode Island (U.S. Bureau of the Census 1960; 1970). Further details on these measures are given in Appendix A.

In addition, the state has been partitioned into three zones based on levels of population density for towns in the state. In general, zone 1 refers to central city areas and industrial satellites; zone 2 includes immediate suburbs to a large extent; and zone 3 refers to towns of lowest population density which characterize areas on the periphery of the metropolitan area. In a crude sense, the density index can also be considered a level of urbanization index.

ANALYSIS

Overall Mobility Incidence Undertaken first will be analyses directed toward answering the question regarding which areal factors are most important in the determination of mobility incidence. The investigation will center around the tract characteristics just discussed. Table 1 shows unadjusted mobility rates associated with the various areal characteristics. Highest levels of movement are shown for areas of low socio-economic and low family status, high turnover status and in areas of downwardly changing family and socio-economic status. Higher rates are also shown for zones of greater density. The strongest relationships are seen for the areal factors: family status and socio-economic status.

The adjusted rates in Table 1 are presented in an attempt to control for intercorrelations among tract characteristics. This is a modification of the dummy variable regression technique proposed by Andrews, Morgan and Sonquist (1967) as multiple classification analysis.

Two sets of adjusted rates are shown in the Table. The first set includes the zone index as an independent variable and the second does not. The first set shows rather dramatically that when all social areal characteristics are taken into account, the zone of origin has virtually no independent effect on mobility incidence. In both multivariate analyses, the family status factor still predominates as the most important mobility determinant, but the adjusted rates for low family status areas are been reduced. Less importance is attributed to socio-economic status which is now on a par with turnover status. The bottom

Table 1

MOBILITY RATES AND CONTRIBUTIONS TO MOBILITY EXPLANATION
BY TRACT CHARACTERISTICS
Population: One-year Segments in Zones 1, 2, and 3.

Adjusted rates are adjusted for all tract characteristics

Tract Characteristic	Mobility Rates Unadjusted	Adjusted[a]	Adjusted[b]	Number
Socio-economic status				
High	8.2	8.4	8.4	6,968
Medium	8.4	9.4	9.4	8,359
Low	12.7	11.4	11.4	7,317
	Eta .0701	Beta .0411	Beta .0406	
Change in socio-economic status				
Upward	8.8	9.3	9.3	6,979
Moderate	9.3	9.7	9.8	7,333
Downward	10.9	10.1	10.1	8,332
	Eta .0303	Beta .0110	Beta .0111	
Family status				
High	7.2	7.9	7.9	7,301
Medium	9.2	9.6	9.6	8,362
Low	13.1	11.9	11.9	6,981
	Eta .0799	Beta .0546	Beta .0535	
Change in family status				
Upward	8.9	8.9	8.9	5,616
Moderate	9.0	9.5	9.5	9,126
Downward	11.1	10.6	10.6	7,902
	Eta .0344	Beta .0230	Beta .0231	
Turnover status				
High	11.9	11.6	11.6	5,905
Medium	10.1	9.7	9.7	8,042
Low	8.0	8.5	8.5	8,697
	Eta .0524	Beta .0419	Beta .0420	
Zone				
Zone 1	11.8	9.7		9,878
Zone 2	8.5	9.8		7,815
Zone 3	7.5	9.7		4,951
	Eta .0632	Beta .0025		
Overall	9.7			26,444

Tract Characteristic	Contributions to Mobility Explanation[c] Percentage of Total Variance	Percentage of Explained Variance
Socio-economic status	.14	13.6
Change in socio-economic status	.01	1.0
Family status	.64	62.1
Change in family status	.07	6.8
Turnover status	.17	13.6
Zone	.00	0.0
Total	1.03	100.0

[a]Includes tract characteristic, zone.

[b]Excludes tract characteristic, zone.

[c]entered in a stepwise manner as follows: family status, socio-economic status, turnover status, change in family status, change in socio-economic status, zone.

Table 2

ADJUSTED MOBILITY RATIOS FOR ZONES BY TRACT CHARACTERISTICS

Ratios are computed from mobility rates adjusted for all tract characteristics.

Tract Characteristic	Overall	Mobility Ratios[a] Zone 1	Zone 2	Zone 3
Socio-economic status				
High	.87	.92	.89	1.05
Medium	.97	.85	1.06	.92
Low	1.18	1.10	1.12	1.23
	Beta .0406	Beta .0418	Beta .0208	Beta .0287
Change in socio-economic status				
Upward	.96	.85	1.13	1.04
Moderate	1.01	1.04	1.04	.89[b]
Downward	1.04	.91	.89	2.16[b]
	Beta .0111	Beta .0303	Beta .0281	Beta .0454
Family status				
High	.81	1.19	.76	.97
Medium	.99	.89	1.27	.95[b]
Low	1.23	1.04	1.24	3.43[b]
	Beta .0535	Beta .0312	Beta .0757	Beta .0728
Change in family status				
Upward	.92	.95	.82	.99
Moderate	.98	.91	1.07	.91
Downward	1.09	1.08	1.01	1.50
	Beta .0231	Beta .0308	Beta .0271	Beta .0490
Turnover status				
High	1.20	1.28	1.02	1.07
Medium	1.00	.97	1.15	.89
Low	.88	.81	.88	1.03
	Beta .0420	Beta .0683	Beta .0340	Beta .0210

[a]The mobility ratio is defined as: $\frac{\text{The mobility rate for the category (e.g., high SES status)}}{\text{The mobility rate for the subpopulation (e.g., Zone 3)}}$

For example, the adjusted ratio for high SES status tracts in the overall population is equal to (from Table 1): $\frac{8.4}{9.7} = .87$

[b]Based on less than 100 cases

Table 3

CONTRIBUTIONS TO MOBILITY EXPLANATION BY TRACT CHARACTERISTICS AND BY TRACT CHARACTERISTICS WHEN INDIVIDUAL HOME OWNERSHIP STATUS AND AGE/MARITAL STATUS ARE CONTROLLED

Tract Characteristic	Percentage of Variance Explained by: Tract Characteristics Only	Individual Home Ownership Status and Age/Marital Status	Tract Characteristics in Addition to Home Ownership & Age/Marital Status
Socio-economic status	.49	11.38	.05
Change in socio-economic status	.09	11.38	.05
Family status	.64	11.38	.05
Change in family status	.12	11.38	.01
Turnover status	.27	11.38	.11

portion of the Table shows, however, the combined factors explain a low percentage of the total mobility variance.

Mobility Incidence for Metropolitan Zones It has been suggested that social areal factors will affect mobility differently in various urban locations. In Table 2, zone-standardized mobility ratios are presented which allow one to compare mobility differentials of each areal factor across zones.

Looking at turnover status, its overall direct relation to mobility can be seen particularly in segments located in zone 1. The pattern is quite different in zones 2 and 3, which display much smaller differentials. The overall inverse relation of family status with mobility is not exactly duplicated in each of the three zones but two general patterns are evident: the pattern of a U-shaped mobility relation in zone 1 and the pattern of an approximately inverse mobility relation in zones 2 and 3. The overall socio-economic status-mobility differential is not replicated in each of the three zones. However, in all instances, the lower socio-economic tracts display highest mobility. Differing patterns across zones are shown also for the factor, change in socioeconomic status; the zone 2 differential exhibits a pattern exactly opposite to the overall trend. Although there is deviation from the overall differential in the change in family status factor, the general association of mobility with declining area status can be seen in all three zones.

These findings might be summarized as follows: areal factors individually or as a group display a disappointingly low level of explanation toward mobility incidence. The family status of an area goes farthest toward explaining mobility variance in the overall analyses, and the other factors show differentials in expected directions. The overall rates mask quite different patterns across the three urban zones; turnover status is the most successful areal mobility determinant in zone 1 with family status dominating in zones 2 and 3. This suggests that whatever environmental effects social areal characteristics have on mobility, operate within specific zonal locations in the urban area.

Areal and Individual Influences on Mobility Individual level factors were not taken into account in the preceding analyses, and it is conceivable that the areal mobility differentials, though small, result more from individual related characteristics than from environmental effects. Further tabulations from these data (not shown) indicate that there is a tendency for the areal categories which scored high on mobility to be composed disproportionately of high mobility-related individual level characteristics.

Perhaps the clearest way to show the effects individual factors have on the areal differentials is to look at the relative percentages of mobility variance explained by areal factors when individual effects have been taken into account. In the first column of Table 3 is shown the percentage of variance explained by each areal factor alone; in the second column it can be seen that 11.38% of mobility variance is explained by two individual level factors--home ownership and age/marital status, a life-cycle factor.[2] The third column in the Table shows the percent of variation explained by areal factors in addition to that in the second column.

It is apparent that the individual level determinants contribute far more to the explanation of individual mobility incidence than do the areal determinants; moreover when areal effects are controlled for age/marital status and home ownership, the additional variance contributed by the areal factors is negligible. Tabulations (not shown) demonstrate this to be the case within each zone as well. Hence the areal mobility differentials shown earlier are in large part a function of individual and household determinants.

EVALUATION

A major finding in this investigation is that areal characteristics (at the census tract level) are weak mobility determinants. When individual and household determinants are taken into account, the areal effects on mobility are negligible. However, within the small percentage of variation explained by areal factors, the findings show that the effect of an area's turnover status is greater in densely populated central city locations and that the family status of an area predominates in the suburban zones.

This investigation has significant implications for those who wish to isolate relevant areal and environmental factors characteristic of high mobility areas. Our results suggest that at the minimum such investigations should stratify areas by aggregate measures of variables that are also successful individual mobility determinants (for example, home ownership and age/marital status) and, if possible, individual level data should be obtained. Although our data (not shown) indicate that social areas may have some contextual effects on individual and household mobility, such influences on aggregate movement are minimal in comparison to the compositions of areas with regard to the major individual level determinants.

In regard to ecological mobility patterns in the urban area, the results presented here suggest that aggregate net movement patterns observed for various socio-economic and family status areas are not a consequence of selective out-movement due to the social environment. Those aggregate differentials may in part result from the composition (for example: age, home ownership) associated with the socio-economic and family status areas or alternatively from differences in the "drawing power" of certain social areas in terms of selective in-movement.

Finally the analysis supports the contention that whatever influence the social environment does exert on mobility is differentiated by zonal location in the urban area. Further research with this data will focus on determinants of zonal destinations for intra-urban movers.

FOOTNOTES

1. This paper is part of a larger study of residential mobility and migration to be published in the forthcoming monograph Residential Mobility, Migration and Metropolitan Change by Alden Speare, Jr., Sidney Goldstein and William H. Frey (Ballinger Press). The work was supported by United States Public Health Grant HS-00246 from the National Center for Health Services Research and Development. The data were collected by the Population Research Laboratory of Brown University. The author is grateful to the collaborators in the larger study for their helpful comments on this paper.
2. The factor, home ownership consists of two categories: owners and renters. The life-cycle factor, age/marital status, consists of five categories: never married individuals and ever married household heads aged 18-29, 30-44, 45-64, and 65 and over. The explained variance attributed to these factors in Table 3 are based on dummy variable regression analyses.

REFERENCES

Abu-Lughod, Janet and Mary Mix Foley. 1960. Part Two - Consumer Strategies. In Foote, Nelson N. et. al., 1960.Housing Choices and Housing Constraints. New York: McGraw-Hill.

Andrews, Frank, James Morgan, and John Sonquist. 1967. Multiple Classification Analysis. Ann Arbor: Institute for Social Research

Berry, Brian J.L. and Frank E. Horton (editors). 1970. Geographic Perspectives on Urban Systems. Englewood Cliffs: Prentice-Hall.

Goldstein, Sidney. 1965. The Impact of Migration on the Socio-economic Structure of Cities and Suburbs. Sociology and Social Research, 50: 5-35.

Grigsby, William G. 1963. Housing Markets and Public Policy. Philadelphia: The University of Pennsylvania Press.

Johnston, R.J. 1971. Urban Residential Patterns. New York: Praeger Publishers, Inc.

Lansing, John B. and Gary Hendricks. 1967. Automobile Ownership and Residential Density. Ann Arbor: Institute for Social Research.

Moore, Eric G. 1969. The Structure of Intra-Urban Movement Rates: An Ecological Model. Urban Studies. 6: 17-73.

____________. 1971. Comments on the Use of Ecological Models in the Study of Residential Mobility in the City. Economic Geography. 47: 73-85.

Organic, Harold N. and Sidney Goldstein. 1970 in Kessler, Irving I. and Morton L. Levin. The Community as an Epidemiological Laboratory A Casebook of Community Studies. Baltimore: The Johns Hopkins Press.

Rossi, Peter H. 1955. Why Families Move. New York: The Free Press.

Shevky, Eshref F. and Wendell Bell. 1955. Social Area Analyses. Stanford: Stanford University Press.

U.S. Bureau of the Census. 1960. Censuses of Population and Housing: 1960 Census Tracts.

________________________. 1970. Censuses of Population and Housing: 1970 Census Tracts.

APPENDIX A: DESCRIPTION OF TRACT CHARACTERISTICS

The following are descriptions of the indices used as tract characteristics based on 1960 and 1970 census measures. In Tables 1,2 and 3 the units of analysis are one-year life segments created from residence histories which were collected from a probability sample of Rhode Island adults in 1967 and 1968. The segments represent single years of residence for these individuals between 1955 and 1967. Since census tract locations were recorded for each segment, it was possible to categorize each residence by the tract characteristics below. Tract characteristic scores for segments representing intercensal years (e.g., 1962) were linearly interpolated from the tracts' 1960 and 1970 scores. In all Tables each tract characteristic is recoded into three categories such that roughly one-third of the tracts in 1960 fall in each.

Socio-economic Status Index

Tract score is defined as the standard score of the following:

Average of (a) Standard score of percent families earning less than $4,000 a year in 1960 and $5,000 a year in 1970; (b) Standard score of percent males reporting occupations other than operatives, service workers, or laborers; (c) Standard score of percent persons 25 years and older who have completed high school. The 1960 standard scores are relative to other R.I. tracts in 1960. The 1970 standard scores are relative to other R.I. tracts in 1970.

Change in Socio-economic Status Index

Tract score is the difference of the 1960 score from the 1970 score on the Socio-economic Status Index.

Family Status Index

Tract score in 1960 is defined as the standard score (relative to other R.I. tracts in 1960) on the measure: married couples with children under 18 divided by all occupied dwelling units.

Tract score in 1970 is defined as the standard score (relative to other R.I. tracts in 1970) on the measure: Husband-wife families with children under 18 divided by all occupied dwelling units.

Change in Family Status Index

Tract score is the difference of the 1960 score from the 1970 score on the Familism Status Index.

Turnover Status Index

Tract score is defined as a number of persons not residing in the same house five years ago divided by population five years of age and over multiplied by 100.

ON THE DECREASE IN BLACK MALE LIFE EXPECTANCY

Kurt Gorwitz, Office of Health and Medical Affairs

Introduction

In the United States, age-adjusted mortality rates for nonwhite male and female residents have for some time been approximately 40 percent higher than comparable figures for their white counterparts. Life expectancy, as a result, has been about five years lower. This disparity has been masked in commonly published crude mortality statistics since these do not reflect the nonwhite population's higher percentage of younger residents and concomitant lower percentage of older residents.

Age-specific and age-adjusted mortality rates declined fairly steadily for some time for each of the four race and sex groups. This decrease has continued to the present time for both white and nonwhite females. However, in the past decade (between 1960 and 1970) rates remained essentially unchanged for white males and increased quite appreciably for nonwhite males. Presently (1970), life expectancy is approximately 61 years for nonwhite males, 68 years for white males, 69 years for nonwhite females and 75 years for white females.

We here present Michigan data for 1959-61 and 1969-71, focusing on the relative impact of various causes of death and on variations in mortality rates at specific age levels. Possible cultural, economic and social issues related to the trends noted will be discussed and the implications of the widening disparity in rates will be considered. Although similar detailed data for the United States for the years 1969-71 have not as yet been published, available statistics indicate comparable trends.

Analysis of Findings

Michigan's total population increased 13.4 percent, from 7,823,000 to 8,875,000, between 1960 and 1970. During the same time period, the number of nonwhite residents increased 41.4 percent, from 737,000 to 1,042,000. Approximately half of this higher growth rate reflects the continuing substantial in-migration of younger nonwhite males (and their families) seeking employment in automobile related professions, particularly in the first half of the decade. More than 95 percent of all nonwhite residents are black and the terms nonwhite and black are therefore used interchangably here. Approximately 75 percent of nonwhite residents, but only 10 percent of white residents, live in Detroit, Michigan's largest city. Most of the rest live in the state's other urban centers.

In 1959-61, annual age-adjusted mortality rates were 10.1 (per thousand) for white males, 6.6 for white females, 11.4 for nonwhite males, and 9.1 for nonwhite females. By 1969-71, these rates had remained unchanged for white males, decreased 10.6 percent (to 5.9) for white females, increased 15.8 percent (to 13.2) for nonwhite males, and decreased 6.6 percent (to 8.5) for nonwhite females. In 1959-61, black male mortality was 32.6 percent above the general state rate. By 1969-71, this excess was 53.5 percent. Conversely, white females (the group with the lowest rate) were 23.3 percent below the state average in 1959-61 and 31.4 percent in 1969-71. In the latter period, the black male age-adjusted mortality rate was more than twice the comparable figure for white females.

Approximately 2/3 of the increase in the black male age-adjusted mortality rate resulted from a dramatic rise in deaths in the age group 15-44. Between 1959-61 and 1969-71, rates increased 131.3 percent (from 1.6 to 3.7) at ages 15-24, 100.0 percent (from 2.8 to 5.6) at ages 25-34 and 71.7 percent (from 5.3 to 9.1) at ages 35-44. Concurrently, they remained essentially unchanged in this age group for the three other race-sex groups. In comparison with rates for white males, black male rates in 1959-61 were 14.3 percent higher at ages 15-24, 75.0 percent higher at ages 25-34, and 89.3 percent higher at ages 35-44. By 1969-71, they were 117.6, 250.0, and 184.4 percent higher, respectively. In comparison with similar figures for black females, these percentages were 60.0, 21.7, and 15.2 in 1959-61 and 236.4, 166.7, and 89.6 in 1969-71. That is, the most dramatic absolute and relative rise occurred in the age group 15-24.

Approximately 2/3 of this increase in black male deaths in the age group 15-44 was in turn caused by a major rise in mortality from two causes--accidents and homicides. Between 1959-61 and 1969-71, the death rate for accidents increased 59.8 percent (from 59.7 per 100,000 to 95.4) while the rate for homicides rose 236.5 percent (from 54.0 to 181.7). These two causes accounted for approximately 1/3 of all black male deaths in this age group in 1959-61 compared with 1/2 in 1969-71. In comparison with rates for white males, black male rates in 1959-61 were 11.2 percent lower for accidents and 1,442.8 percent higher for homicides. By 1969-71, the comparable figures were 13.3 and 2,012.8 percent higher, respectively, for black males. In comparison with black females, black male rates were 209.3 and 193.5 percent higher in 1959-61 and 250.7 and 544.3 percent higher in 1969-71. Concurrently, while black male death rates decreased from a number of other leading causes such as heart disease, cancer, vascular lesions, and diabetes, they generally remained higher than similar figures for each of the other three race and sex groups.

At age 15, black males in 1969-71 had a life expectancy of 49.1 years, a decrease of 3.7 years from the 52.8 noted in 1959-61. This former figure was 6.0 years less than the comparable 55.1 for white males, 8.1 years less than the 57.2 for black females and 13.0 years less than the 62.1 for white females. The gap gradually decreased with advancing age. Beginning at age 65, black male life expectancy exceeded white males'.

In 1959-61 and 1969-71, white males, white females, and nonwhite females at birth had a 70-75 percent probability of eventually dying from the three leading causes of death--heart disease, cancer, and vascular lesions. This probability for nonwhite males was 64.2 percent in 1959-61 and 58.3 percent in 1969-71. Compensating for this decrease was an increase in the probability of dying from accidents (from 4.7 to 5.8 percent) and from homicides (from 2.9 to 6.9 percent). Given current rates, one out of eight black males will ultimately die from an accident or from homicide. This probability is one out of 17 for white males, one out of 30 for white females and one out of 26 for nonwhite females.

Based on 1969-71 rates, elimination of heart disease deaths would increase the life expectancy in each of the four race-sex groups by 5-6 years with the least impact (5.1 years) on black males. Should no further cancer deaths occur, life expectancy would rise about 2.5 years. Homicides, which reduced life expectancy of black males by 0.8 years in 1959-61, were responsible for a reduction of 2.3 years in 1969-71 compared with 0.2 years for white males, less than 0.05 years for white females and 0.5 years for black females. That is, homicides and cancer have virtually an equal impact on reducing black male life expectancy. The impact of accidents as well as influenza and pneumonia in reducing black male life expectancy was also relatively large.

Discussion

The facts presented here surely should be a matter of major concern. Prime among the questions raised are the reasons for the reduced black male life expectancy, the impact of the relatively large black male in-migration, and whether or not remedial actions are available. While some definitive responses readily come to mind, others cannot be documented or may be open to question. It would certainly appear, however, that the trend noted resulted from the interaction of a number of factors or conditions.

As documented, most directly responsible for the decrease in black male life expectancy has been the dramatic increase in homicides, particularly in the age group 15-44. In fact, the number has continued to go up, virtually doubled in the last five years and is currently (1974) at a record level. More than three-quarters of all homicides of black males in 1973 were caused by handguns, primarily either in a crime of passion between relatives and/or acquaintances or as the end result of the commission of a crime. Handguns are readily available in our major cities and the number in the hands of private citizens is estimated to be in the hundreds of thousands or even millions. Surely the data presented here provide an indication of the need for vigorous enforcement of existing statutes as well as new legislation required to control and limit the manufacture, distribution, and possession of handguns. It may logically be argued that the rising and excessive homicide rate among black males is directly related to cultural, economic and social issues which must ultimately be reconciled and resolved. However, such effort would require a long-range, ongoing commitment and does not preclude the need for concurrent preventive actions related to manifestations of these issues.

The number of black residents increased 40 percent in Michigan between 1960 and 1970. This figure was even higher in the younger ages and was particularly pronounced in the city of Detroit. Unemployment and under-employment is appreciably above the national average and has resulted in a widening economic gap. It is our impression also that employment with an actual or potential health hazard is increasingly filled by black males. Rigorous enforcement of safety standards and pollution control requirements as well as expanded vocational training with the potential for meaningful upward mobility would all appear to be indicated.

In northern industrial states such as Michigan, black residents can be expected to continue living primarily in core areas of large cities. The ongoing drift of health manpower and facilities to the suburbs must be halted and reversed. Reasonable access to required services as well as some equitable third-party payment mechanism would appear to be essential to bridging presently existing gaps.

It is possible that mortality rates have been rising among black males because of the large scale migration from Southern states of individuals concentrated in the lowest socio-economic levels. This in-migration has now virtually ceased and there is, in fact, evidence of some out-migration. We cannot document the relative impact of this migration and do not know whether black male mortality rates differ between long-term residents and recent arrivals. A detailed analysis within the context of socio-economic and cultural variables is indicated.

TABLE 1

ANNUAL AGE-SPECIFIC MORTALITY RATES (PER 1,000) BY RACE AND SEX

MICHIGAN

1959-1961 AND 1969-1971

AGE (IN YEARS)	White Male 1960	White Male 1970	White Female 1960	White Female 1970	Nonwhite Male 1960	Nonwhite Male 1970	Nonwhite Female 1960	Nonwhite Female 1970
TOTAL (CRUDE)	10.2	10.1	7.2	7.3	8.7	10.6	6.8	7.0
(AGE ADJ.)	10.1	10.1	6.6	5.9	11.4	13.2	9.1	8.5
-1	26.2	20.7	19.5	15.5	43.4	38.2	34.1	30.7
1-4	0.9	0.8	0.8	0.7	1.3	1.3	1.0	1.2
5-14	0.5	0.5	0.3	0.3	0.5	0.6	0.4	0.4
15-24	1.4	1.7	0.5	0.6	1.6	3.7	1.0	1.1
25-34	1.6	1.6	0.8	0.8	2.8	5.6	2.3	2.1
35-44	2.8	3.2	1.8	2.0	5.3	9.1	4.6	4.8
45-54	8.8	8.4	4.7	4.5	12.6	15.4	10.2	9.6
55-64	21.6	21.6	11.1	10.4	26.6	29.1	22.0	18.9
65-74	49.5	49.4	28.9	25.9	53.5	53.9	39.1	36.3
75-84	106.6	106.5	78.9	68.7	78.6	90.7	63.1	63.0
85 +	220.2	194.4	195.4	162.0	136.7	108.8	127.2	96.2

TABLE 2

RACE AND SEX SPECIFIC MORTALITY RATES AS A PERCENT OF

THE TOTAL 1959-1961 MORTALITY RATE

MICHIGAN

YEAR	WHITE MALE	WHITE FEMALE	NONWHITE MALE	NONWHITE FEMALE
1959 - 1961	117.4	76.7	132.6	105.6
1969 - 1971	117.4	68.6	153.5	98.5
CHANGE	-	- 8.1	+ 20.9	- 7.1

TABLE 3

ANNUAL MORTALITY RATES (Per 100,000) IN THE AGE GROUP 15-44
BY RACE, SEX, AND LEADING CAUSES OF DEATH
MICHIGAN 1959-1961 AND 1969-1971

	White Male		White Female		Nonwhite Male		Nonwhite Female	
CAUSE OF DEATH	1960	1970	1960	1970	1960	1970	1960	1970
TOTAL	195.7	209.2	106.5	103.5	336.3	566.1	269.4	240.6
HEART DISEASE	42.2	29.1	11.8	8.8	58.4	58.0	45.2	28.6
CANCER	24.4	24.7	30.3	24.5	31.4	28.5	47.0	29.2
VASCULAR LESIONS	5.8	4.9	5.2	5.0	18.0	15.1	23.0	18.6
INFLUENZA & PNEUMONIA	3.3	3.8	2.8	3.1	11.1	24.3	12.6	10.2
DIABETES	3.0	2.4	2.4	2.0	7.6	7.0	4.8	6.5
ACCIDENTS	67.2	84.2	15.9	23.6	59.7	95.4	19.3	27.2
HOMICIDES	3.5	8.6	1.6	2.8	54.0	181.7	18.4	28.2
ALL OTHER	46.3	51.6	36.5	33.6	96.2	156.0	99.1	91.9

TABLE 4

LIFE EXPECTANCY AT SPECIFIED AGE LEVELS BY RACE AND SEX
MICHIGAN 1959-1961 AND 1969-1971

	White Male		White Female		Nonwhite Male		Nonwhite Female	
AGE	1960	1970	1960	1970	1960	1970	1960	1970
0	67.7	68.1	74.7	75.5	64.3	61.1	68.0	69.4
1	68.5	68.6	75.2	75.6	66.2	62.5	69.4	70.6
5	64.8	64.8	71.3	71.8	62.5	58.8	65.7	67.0
15	55.1	55.1	61.5	62.1	52.8	49.1	55.9	57.2
25	45.8	45.9	51.7	52.4	43.6	40.7	46.4	47.8
35	36.5	36.6	42.0	42.8	34.6	32.8	37.3	38.7
45	27.3	27.6	32.6	32.5	26.2	25.4	28.4	30.4
55	19.4	19.5	23.8	24.8	19.0	18.8	21.4	22.9
65	12.7	12.9	15.9	16.9	13.3	13.4	15.4	16.5
75	7.8	8.1	9.4	10.4	9.3	9.7	10.4	11.6
85	4.3	5.2	4.8	6.0	5.1	7.5	5.4	7.8

TABLE 5

PROBABILITY OF EVENTUALLY DYING FROM SPECIFIED CAUSES BY RACE AND SEX
MICHIGAN 1959-1961 AND 1969-1971

	White Male		White Female		Nonwhite Male		Nonwhite Female	
CAUSE OF DEATH	1960	1970	1960	1970	1960	1970	1960	1970
TOTAL	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
HEART DISEASE	43.2	43.0	41.0	43.0	34.4	32.5	37.4	38.4
CANCER	16.0	17.4	16.0	16.1	17.2	16.8	14.0	14.3
VASCULAR LESIONS	11.2	9.3	16.6	14.4	12.8	9.0	18.3	15.4
INFLUENZA & PNEUMONIA	2.8	2.7	2.3	2.8	4.6	4.3	4.4	4.0
DIABETES	1.9	2.2	3.6	3.6	1.7	2.3	4.1	4.8
ACCIDENTS	5.2	5.4	3.7	3.2	4.7	5.8	2.3	2.5
HOMICIDES	0.2	0.5	0.1	0.1	2.9	6.9	1.2	1.3
ALL OTHER	19.5	19.6	16.7	16.8	21.8	22.6	18.4	19.5

TABLE 6

REDUCTIONS IN LIFE EXPECTANCY DUE TO SPECIFIC CAUSES OF DEATH BY RACE AND SEX
MICHIGAN 1959-1961 AND 1969-1971

	White Male		White Female		Nonwhite Male		Nonwhite Female	
CAUSE OF DEATH	1960	1970	1960	1970	1960	1970	1960	1970
HEART DISEASE	6.5	6.6	5.3	5.7	5.4	5.1	5.8	6.0
CANCER	1.9	2.5	1.9	2.7	2.5	2.6	2.4	2.6
VASCULAR LESIONS	1.1	1.0	1.9	1.5	1.3	1.2	2.2	2.2
INFLUENZA & PNEUMONIA	0.5	0.4	0.3	0.4	0.9	0.8	0.8	0.8
DIABETES	0.2	0.7	0.4	0.4	0.2	0.3	0.6	0.8
ACCIDENTS	1.5	1.5	0.9	0.8	1.3	1.7	0.6	0.8
HOMICIDES	0.1	0.2	*	*	0.8	2.3	0.3	0.5

*Less than 0.05

A MODEL OF INCOME DIFFERENTIALS BETWEEN DEVELOPING COUNTRIES

By K. L. Gupta, University of Alberta

Recently, Arrow stated that, "Inequality in economic development among countries and among groups and regions within a country provides a second, somewhat complicated difficulty for neoclassical theory. A purely neoclassical answer would explain differences in per capita income by differences in physical and human assets per capita. This of course raises the further question, how this came to be, a question which would require a fully dynamic model to answer; but I think the more compelling problem is that the differences in income seem much too vast to be explained by factor differences".[1/] This paper is an attempt to examine the extent to which the neoclassical theory does provide an explanation of per capita income differentials between developing countries. A distinguishing feature of the paper lies in the fact that the single equation neoclassical model is treated as part of a set of a simultaneous equation model which explicitly allows for the role of demographic factors. The scheme of the paper is as follows. In Section I we present the model, Section II presents the results, and the last section briefly summarizes the findings.

I. The Model

The neoclassical explanation is a straight forward one. Consider the following production function:

$$Y = F\ (K,L) \qquad (1)$$

where Y, K, and L are GNP, capital, and labor respectively.

Assuming that this production function is linear and homogeneous, it may be written in per capita terms as

$$\frac{Y}{N} = F\ \left(\frac{K}{N}\ ,\ \frac{L}{N}\right) \qquad (2)$$

where N is population and (L/N) represents total labor force participation rate. Thus equation (2) states that, abstracting from technical change, per capita income depends on per capita capital and the labor force participation rate. In other words, countries which exhibit high k(=K/N) and PR(=L/N), should be expected to have high per capita income and vice-versa.

The above model has one major short-coming and that is the exogenous nature of the labor force participation rate. The correct procedure would be to explicitly allow for the role of demographic factors as determinants of PR and thus in the determination of differences in per capita income. This is the procedure we follow.

For our purpose, we face an empirical problem before this can be done. The data on international labor force participation rates are highly inadequate and imprecise. Thus the estimation of an equation like (2) is very difficult. In order to overcome this difficulty, following demographic theory, we postulate that participation rates depend on dependency rates - a variable for which reliable data are available - such that

$$PR = f(DR)\ f' < 0 \qquad (3)$$

where DR stands for dependency rates. Equation (3) implies that the higher are the dependency rates, the lower will be the participation rates. Hence, a priori, we should expect a negative partial relationship between per capita income and dependency rates.[2/]

Assuming that the empirical model is represented by a linear equation and further that quality differences in labor force are represented by differences in literacy rates and further that non-capital resources also affect per capita income, the equation to be estimated is given by

$$Y = a + \underset{>0}{bk} + \underset{<0}{cDR} + \underset{<0}{dDEN} + \underset{>0}{cLIT} \qquad (4)$$

where the expected signs of the coefficients are indicated below the coefficients and DEN stands for population density which is used to measure non-capital resources, an admittedly crude measure as pointed out by Adelman.[3/] LIT stands for literacy rate. Per capita energy consumption is used as a proxy for per capita fixed capital[4/] and is measured in kilograms per capita coal equivalents.

We now assume that the dependency rate is an endogenous variable. To take account of this, we specify the following additional equation:

$$DR = a_o + b_oBR \qquad (5)$$

$$BR = a_1 + b_1FPR + c_1y + d_1IMR + c_1LIT + f_1ALF \qquad (6)$$

$$FPR = a_2 + b_2BR + c_2AFL \qquad (7)$$

$$IMR = a_3 + b_3y + c_3ALF + d_4LIT + e_4HB \qquad (8)$$

where, in addition to the variables already defined,

- BR: birth rate per 1000
- FPR: female labor force partifipation rate
- IMR: infant mortality rate
- ALF: percentage of labor force in agriculture
- HB: number of persons per hospital bed

Detailed explanations for these specifications can be found in the author's other paper.[5/] Therefore only brief comments on these equations are offered here.

Dependency rate equation

This equation is straightforward and its rationale has been ably summarized by Leff. Thus, "demographic theory indicated that a prolonged high birth rate will affect a population's age composition, placing a relatively large percentage of population in the younger age bracket".[6/]

Birth rate equation

This equation embodies the hypotheses put forward by Adelman (1963), Becker (1960), Mincer (1963). Cain (1966), and Schultz (1969, 1973), among others. Very briefly, the inclusion of per capita income follows from the theory of consumer choice, as for example, argued by Becker. Female participation rate is included as a proxy for the, "opportunity income of women and their access to the labor market."[7/] The inclusion of infant mortality is justified in terms of the replacement needs of a family for children {Gregory et.al. (1972), Schultz (1973)}. The role of education has been discussed extensively and need not be elaborated here. The reason for the inclusion of the percentage of agricultural labour force is based on the well known argument that agricultural activity is more conducive to higher birth rates than non-agricultural activity.[8/]

Female participation rate equation

Birth rates affect the supply of labor and is therefore included as an argument of this equation.[9/] It is now well recognized that labor force participation rate is higher in largely agrarian economies[10/] and we therefore expect a positive relationship between female participation rate and the level of non-industrial development where the latter is measured by the percentage of the labor force in agriculture.

Infant mortality equation

This equation is similar to the one used by Adelman. The variable, the number of persons per hospital bed, is used as an index of the availability of health care services.

For estimation purposes it is not necessary to estimate the entire model. Instead, we use the two-stage least squares method. In the first stage, we use all the exogenous variables as instruments to estimate DR which is the endogenous variable in equation (4). In the second stage, using this estimated value of DR, we estimate equation (4) by the method of ordinary least squares. This provides consistent estimates of the parameters of equation (4). Two important points about our method of estimation should be noted. First, the instruments used are derived from an a priorily defined model and second, we use all the exogenous variables as instruments and not a subset of them, which is always an arbitrary procedure.

II. The Results

The model was applied to a cross-section of forty developing countries. The countries included are: Argentina, Bolivia, Brazil, Ceylon, Chile, Colombia, Cost-Rica, D. Republic, Ecuador, El Salvador, Greece, Guatemala, Honduras, India, Indonesia, Israel, Japan, Jamaica, Jordan, Malaysia, Malta, Mexico, Morocco, Nicaragua, Pakistan, Panama, Paraguay, Peru, Philippines, Portugal, S. Korea, Spain, Taiwan, Thailand, T. Tobago, Turkey, UAR, Uganda, Uraquay, and Venezuela.

The period covered was the 1960's. The data were collected from the unpublished World Tables of the World Bank and various publications of the United Nations.

The results of equation (4) are as follows:

$$y = 601.1077 + \underset{(7.805)}{0.20789k} - \underset{(1.838)}{12.61856DR} - \underset{(0.690)}{0.07932DEN} + \underset{(1.366)}{1.87395LIT}$$

$$\bar{R}^2 = 0.808$$

We can see from these results that our version of the neoclassical model explains about 81% of the per capita income differentials between the developing countries. Except population density, all the coefficients exceed their standard error. The sign and significance of the coefficient of dependency rate is particularly noteworthy. Recall that we were forced to abandon the proper neoclassical model due to lack of information on labor force participation rates. We can see that the use of dependency rates gives fairly reasonable rates. At the same time, it also highlights the importance of demographic factors.

In order to examine the relative impact of various factors on per capita incomes, we calculated elasticities at the mean. These are given below:

TABLE 1
Elasticities

With respect to	Elasticity
k	0.41
DR	-1.60
LIT	0.33

If we consider k and LIT to represent crude indicators of investment in physical and human capital, respectively, we can see that their significance is somewhat similar. The elasticity with respect to dependency rates on the other hand is significantly different - almost four times. It would thus appear that developing countries would get a greater pay-off by reducing dependency rates, which ultimately means reducing birth rates. Our findings thus once again highlight the importance of population control for adequate economic development.

III. Conclusions

In this paper we have made use of the well-known neoclassical model to explain per capita income differentials between developing countries. However, we modified this model in two important ways. First, we replaced labor force participation rate by the dependency rate which is more reliably measured. And second, we treated the modified single equation neoclassical model as part of a simultaneous equation model which explicitly allowed for the role of demographic factors as determinants of per capita income differentials. The results turned out to be highly satisfactory to the extent that the model explained about 80 percent of the variation.

Depending on the availability of data, it would be worthwhile refining our model. This could be done in a variety of ways. First, equation (4) should be specified so as to include both male and female labor force participation rates in view of the fact that these rates show considerable variation. Then include equations explaining these two participation rates. Second, more refined measures of non-capital and capital resources and literacy could be used. Finally, the model could be disaggregated, making allowance for the so called 'dual' economy characteristic of these countries.

Footnotes

1/ Arrow [1974], p. 2.

2/ See also Enke [1973].

3/ Adelman [1963].

4/ See Kim [1969].

5/ See Gupta (1973).

6/ See Leff [1969], p. 887.

7/ Schultz (1969), p. 155. See also Mincer (1963) and Cain (1966).

8/ Schultz (1969).

9/ See Benham (1971) and Cain (1966, 1973).

10/ Farooq (1972).

References

1. Irma Adelman, "An Econometric Analysis of Population Growth", American Economic Review, June 1963, 314-39.

2. K. J. Arrow, "Limited Knowledge and Economic Analysis", American Economic Review, March 1974, 1-10.

3. Gary S. Becker, "An Economic Analysis of Fertility" in Demographic and Economic Change in Developed Countries, Universities National Bureau Conference Series II Princeton, N.J.: Princeton University Press, 1960.

4. Glen C. Cain, Married Women in Labor Force: An Economic Analysis, Chicago: University of Chicago Press, 1966.

5. Stephen Enke, "Population Growth and Economic Growth", The Public Interest, No. 32 (Summer 1973), 86-96.

6. G. M. Farooq, "An Aggregative Model of Labor Force Participation in Pakistan", The Developing Economies, September 1972, 267-89.

7. Paul E. Gregory et.al., "A Cost-Inclusive Simultaneous Equation Model of Birth Rates", Econometrica, July 1972, 681-88.

8. K. L. Gupta, "A Model of Fertility Behavior in India", Paper presented at the American Statistical Association, December 1973, New York.

9. Y. C. Kim, "Sectoral Output-Capital Ratios and Levels of Economic Development: A Cross-Section Comparison of the Manufacturing Industry", Review of Economies and Statistics, 1969, 453-58.

10. N. H. Leff, "Dependency Rates and Savings Rates", American Economic Review, December 1969, 886-96.

11. Jacob Mincer, "Market Prices, Opportunity Costs and Income Effects", in Measurement in Economics, Standford: Standford University Press, 1963.

12. Paul T. Schultz, "An Economic Model of Family Planning and Fertility", Journal of Political Economy, March/April 1969, 153-80.

13. ______________, "A Preliminary Survey of Economic Analysis of Fertility", American Economic Review, May 1973, 71-78.

A STUDY OF HEALTH CARE PATTERNS OF PUBLIC ASSISTANCE RECIPIENTS[1]

Charles P. Hall, Jr., John A. Flueck, and William F. McKenna, Temple University

Introduction

The general objectives of the Medicaid program, initiated under Title XIX of the Social Security Act, are well known, i.e., to provide better health care to the poor. This program is also known to vary considerably from state to state under federal guidelines that mandate coverage of certain basic services and also allow for a broad range of optional services to be provided under the umbrella of federal cost-sharing with the states. As has been true under most similarly funded programs, the extent to which states are able to capture federal money to support services is, in part, directly related to the general affluence of the state. Under Medicaid, as with a number of other programs, the southern states have generally lagged considerably behind others in this regard.

A major, legitimate concern of federal policymakers concerns the impact that any restriction in coverage of optional services may have on the health of Medicaid eligibles as well as on the utilization and cost of services. In this regard, the specific objectives of this study sponsored by the Social and Rehabilitation Service of DHEW, were as follows:

Primary objectives. To determine (1) the proportion of Medicaid eligibles who are receiving health care services (both basic and optional services); (2) what specific services are used; (3) where services are received, i.e., from what type of provider; and (4) how and by whom they are paid.

Secondary objectives. To determine (1) the so-called "utilization effects," whether optional services were obtained through the substitution of one of the basic covered services, and to make comparisons between jurisdictions providing different coverages of medical services regarding the proportions of eligibles receiving various health care services; (2) the proportion of eligibles who have perceived a need for health care services but who did not receive them and to ascertain the reason(s) why they did not receive them.

Tertiary objective. To determine the health status of the eligibles.

Design of the Survey

From the outset, the project team felt it would be necessary to gather most of its data from a field survey. Secondary sources of Medicaid experience, including records of state agencies and the office records of providers, were considered to be unsatisfactory and incomplete in several respects in comparison to direct contact with eligibles. This would be particularly true with respect to the determination of unmet needs, since they would never appear in official records; furthermore, even where services had been used, it was found that paid claim tapes were often inaccurate for any one of several reasons, ranging from slow reporting to data processing errors. In addition, they could provide no information on utilization and expenditures by eligibles outside of the Medicaid system.

It was, therefore, determined that the major focus of investigation would be a survey of Medicaid eligibles in locations offering different programmatic benefits under Medicaid. It is important to note that Medicaid eligibles rather than recipients were surveyed, since this made it possible to also gather information on those poor who sought and received health care services outside the system, either because they were unaware of available benefits, their providers refused to participate, or any of several other possible reasons.

For purposes of the survey, Medicaid eligibles were defined as cash welfare recipients in the particular locations; the "medically indigent" or others who may have qualified in the absence of categorical public assistance benefits were not included. These latter groups were excluded for two major reasons: (1) it was impossible to identify them before they actually became users of the system; and (2) not all of the sites selected provided benefits for them.

In addition to the constraints listed above, institutionalized welfare recipients were also eliminated from the survey on the grounds that their health care needs and the health system which serves them are both likely to be different than for noninstitutionalized eligibles. Finally, welfare recipients under the Aid to the Blind program were also eliminated because they were small in number and are typically serviced by other specialized programs.

In summary, the survey population of interest, or target population, consisted of noninstitutionalized cash welfare recipients in the Aid to the Permanently and Totally Disabled (AD), Old Age Assistance (OAA) and Aid to Families with Dependent Children (AFDC) programs in each of the four local areas of Atlanta, Little Rock, Oklahoma City and Trenton.

The sites were chosen purposively, after considerable preliminary investigation, to ensure a spectrum of optional service coverage and local agency cooperation in the survey. The identified population was produced by using the county welfare categorical assistance eligibility file as of March 1973. This resulted in the sampling unit being the welfare case; the Atlanta LA was represented by Fulton County (which was estimated to contain 91 percent of the Atlanta welfare caseload); Pulaski County (including Little Rock, North Little Rock, College Station, and Jacksonville, Arkansas) represented the Little Rock LA; Oklahoma County (which includes Midwest City, and most of Oklahoma City) represented the Oklahoma City LA; and Mercer County

(Trenton, Mercerville, and Princeton) represented the Trenton LA.

The population of interest and the identified population, the population that was attempted to be sampled, have been described. However, a third population, the sampled population, is the population that was actually sampled. Like the initial population of interest, this is a prescribed population and is composed of all those sample units in the identified population who could have been successfully interviewed. In short, only those sample units who were willing and who could have been reached are members of the sampled population. The interest was in maximizing the intersection between the identified population and the sampled population. In this study, those welfare cases who were home during the hours of 8:00 A.M. to 10:00 P.M. in June, July, and August, 1973, who had not moved and had a correct address, who had moved and had established a new address, and who had or would have answered any (or the majority) of the questions were members of the sampled population.

An additional restriction on the sampled population was that only those individuals in these cases who were eligible for welfare at some period within the last three months were fully interviewed. Those individuals who stated that they were not eligible in the last three months or who did not know if they had been eligible, were declared ineligible and no interview data were collected beyond the initial demographic and eligibility data.

Questionnaire design is an art which is gradually becoming more of a science. In constructing the questionnaire for this survey, an attempt was made at obtaining questions that were similar to those successfully utilized by other "health" surveys, such as the National Center for Health Statistics, Health Interview Survey [2].

Considerable attention was given over a three-month period to the construction of the questionnaire. Numerous discussions occurred between the subject-matter members of the project and the project statistician. Topics discussed included the number of optional services to investigate, the selection of the appropriate respondent, the quantity and level of detail of the desired information, the recall period, procedures for maximizing the quality of the data, interviewer bias, the length of the questionnaire, the potential coding problems, etc. The literature on questionnaire construction and response errors were consulted, and rough drafts of the questionnaire were "hot-housed" on university employees and local welfare eligibles. During this period, the survey subcontractor, Marketing Information Service, Inc., Atlanta, Ga., worked on the formulation and flow of the questionnaire and, with the Temple project team, conducted a pretest of it using approximately twenty-five cases in each of the four local areas.

The final questionnaire contained a total of approximately 228 questions spread over eleven sections (A. Demographic, B. Eligibility and Health Insurance, C. Overnight Hospitalization, D. Dental Care, E. Eye Care, F. Hearing, G. Physician Services, H. Prescription and Nonprescription Drugs, I. Home Health Care, J. Family Growth and Planning Services, and K. General Health Status). Because of skip patterns and welfare program differences, however, no one was called upon to respond to all questions.

It was felt that the family growth and planning section of the questionnaire contained a number of sensitive and "threatening" questions. These were handled in the traditional manner by placing them at the end of the interview in the belief that this would give the interviewer an opportunity to establish rapport and would minimize any possible damage to the survey. It was somewhat surprising to find that answers to these questions appeared to be as complete as for other questions in the survey, and many interviewers commented on the willingness of respondents to provide information in this as well as other areas.

Both purposive and probability-based selections were utilized in this study. As mentioned above, the selection of the four sites was purposive.

The second stage of the sample design utilized probability-based selection of the welfare cases. Because previous studies [3] suggested that there tended to be different utilization rates between various subsets of the U.S. welfare population, and because SRS desired to gain knowledge about the health care and practices of various subsets of Medicaid users, stratified random sampling with proportional allocation was selected as the sampling technique [4].

The third stage employed 100 percent sampling of all eligible individuals within the welfare case.

Both a priori and a posteriori stratification were utilized, the former on the identified population, and the latter on the collected sample units.

The selection of the specific a priori stratification variables was largely dictated by the importance of such variables as indicated by past published and unpublished research, the desire to keep the number relatively small, and their availability. The results were that a set of four variables--type of aid program, race, age, and size of family--was selected. Their utilization in all four sites was as follows: (1) AD--classified by the racial categories of white (W), American Indian (AI), black (B), and unclassified (UC); (2) OAA--cross-classified by race (W, AI, B, UC) and age (65 through 74, 75 and above); and (3) AFDC--cross-classified by race (W, AI, B, UC) and family size (three or less, four or more).

For prestratification purposes, the above variables were based on welfare agency classification; the AI category was only used in Oklahoma City, and the UC category only in Little

Rock. The reason for the UC classification was that a federal directive had been issued that called for the elimination of classification by race. Little Rock was the only site that had implemented this new policy to any significant degree.

The racial and sex composition of the interviewers varied, with Trenton having black, white and Spanish-speaking interviewers, mostly college-aged males, while Oklahoma City had mainly white females of middle forties and fifties and college age. The supervisors were all experienced survey field personnel, and contrary to expectation, both supervisors and interviewers reported no unusual problems in interviewing members of this selected welfare population. This was supported by the fact that response rates by site ranged from about 75 percent to 86 percent [1].

The main survey was in the field for the months of June, July and August, with the majority of the interviewing conducted between 8:30 A.M. and 4:30 P.M. on weekdays. Interviewers were requested to seek an interview environment that permitted the respondent, or respondents, to give undivided attention to the questions of the interview. Up to three callbacks (subsequent visits to the interviewee's address) were utilized before a case was terminated, with evening (7:00 P.M. to 10:00 P.M.) callbacks being frequently used.

Initial demographic data (names, ages, relationships, sex, welfare eligibility, etc.) were collected for each member of the household (Sections A and B of the questionnaire); only data on "eligibles" was collected for all other sections of the questionnaire. "Eligibles" were defined as those listed on the Medicaid card, which was asked for by the interviewer, or those designated as eligible by the respondent.

The usual field problems of "no one at home," "vacant dwelling," "refusals," "incomplete interview," "invalid address," "deceased," etc., were encountered. Two of these problems, "invalid address" and "ineligible for welfare," had considerably higher frequencies in most of the locations than anticipated. Invalid addresses were either returned to the county welfare office for resolution or neighbors were solicited for the "new" or "correct" address. It was felt that the vigorous pursuit of invalid addresses was partially responsible for the relatively favorable response rates.

It should be noted that, contrary to normal personal interview procedures, no introductory letter was utilized. Indications from the pretest were that this letter created more problems than it solved by generating apprehensions among respondents.

After the interview was designated as completed by the interviewer, a number of checks and studies were made on the quality of the collected data, including a field edit, home-office edits, an acknowledgement of the interview, a 10 percent reinterview study, and a validity study.

The acknowledgement of an interview was a brief letter, with a postage guaranteed envelope, that also contained two questions from the main survey--one demographic, and one on doctor visits in 1973. The hope was that with these two questions one would get a more substantial indication of the presence of an interview than with the traditional yes-no question. Unfortunately, this approach was not particularly successful, for the return rates did not exceed 25 percent.

A reliability study was conducted by reinterviewing 10 percent of the respondents in each location using eleven important questions from the questionnaire. These reinterviews were typically conducted by the most "senior" interviewers within a week or two of the original interview. Results of this study indicated reliability of in excess of 90 percent with the exception of responses about prescriptions and certain aspects of dental care [1].

A validity study was conducted in the main survey relying largely on the "paid-claims tapes" secured from three of the sites (they were unavailable to the project team in Atlanta). Again, results indicated that information on prescriptions and dental care were weakest.

Some Preliminary Results

Survey data were collected for over 3,000 cases which included over 8,000 Medicaid eligible individuals. Needless to say, the resulting database is substantial and analysis will proceed for some time.

In this paper some preliminary findings will be reported with respect to the utilization of key health services, the proportion of eligibles who incurred out-of-pocket expenses for those services, the proportion of eligibles who perceived themselves to have unmet needs for the services, and the interviewee ratings of health status.

At the outset, it should be noted that the research group ranked the Medicaid benefits in the four survey locations on the basis of comprehensiveness. In general, there was agreement that Trenton had by far the best overall benefits, followed by Atlanta, Little Rock and Oklahoma City, in descending order. Trenton covered virtually every allowable optional service. Atlanta, though providing full coverage for prescriptions, did not cover dentures at all and gave only partial coverage--via a screening program for AFDC children to age six--for routine dental care, orthodontics, eyeglasses, optometric services and hearing aids. Little Rock did not cover prescriptions or hearing aids, but full coverage was afforded for routine dental care, orthodontics (prior approval required), and dentures. Eyeglasses and optometric services were provided only in the screening program for children to age twenty. In Oklahoma City, there was no coverage for prescriptions or dentures, and all of the other above-mentioned benefits were generally provided only for AFDC children through age twenty via a screening program. Oklahoma City also had a limit of only four covered physician visits per month. These coverage patterns should be kept in mind when reviewing the findings reported below, and displayed in Tables 1-4.

Table 1 presents data on the percent of Medicaid eligibles, classified by categorical assistance group, who made use of five selected health services during given time periods prior to the interview date. Note that the first two services listed are mandatory services, i.e., they must be covered by state Medicaid programs or no federal funds would be made available to help finance the state program. The last three services listed, however, are "optional" services, i.e., it is not required that they be included in a state plan, but if they are, the federal government will share in their cost.

Note that the data in Table 1 speak directly to two of the primary objectives of the study: the proportion of Medicaid eligibles who are receiving health care services and the utilization rates for these specific services.

As can be seen, Trenton had the highest or tied for the highest proportion of eligibles reporting utilization in thirteen of fifteen subpopulations (three aid categories; five services). Atlanta had the highest percentage of utilization for physician services among the OAA and for visits to the dentist among the AFDC. It also tied for the highest percentage of physician use in both the AD and AFDC subpopulations.

The comprehensive coverage of the New Jersey Medicaid program must certainly be a major factor in the persistently high utilization in Trenton. Not only were its benefits extensive, but the survey indicated a very high level of awareness of available benefits on the part of Medicaid eligibles. Yet there was no evidence to suggest that this greater awareness of benefits resulted directly from efforts by the Medicaid program staff to "market" their product more aggressively in New Jersey. In fact, based on the standard procedures for disseminating Medicaid information which were outlined to the researchers by program staff in each of the four locations during a presurvey site visit, Trenton may have been less zealous in this regard than some of the other cities. Naturally, there may have been some discrepancy between the standard procedures that were reported and what was actually done.

In any case, it seems plausible to suspect that at least some factors unrelated to Medicaid may also have been operative. This is especially true when Trenton's utilization level remains consistently higher even when compared with cities having similar benefits, e.g., Atlanta for prescriptions. Trenton was the only non-southern location in the survey, and some national health statistics have indicated higher levels of the use of health services in the Northeast than in the South. Further investigation may show that regional differences are as much a function of educational level (generally higher in the Northeast) and/or variations in regional medical practices as of program differences.

The relatively strong showing of Atlanta with reference to physician visits and dental care can be traced to the central role played by Grady Hospital and some other special public and private programs in that city. Grady is a large, centrally located municipal hospital which is readily reached by public transportation. Long before the enactment of Medicaid, it had served as the primary provider of health services to the poor, a role which it continues to play--even for extra-programmatic services--especially for black Medicaid eligibles. In addition, the OEO-sponsored Atlanta Southside Comprehensive Health Center and the Ben Massell Dental Clinic, among others, provided additional accessibility to these services.

In five of six subpopulations dealing with mandatory services, Little Rock showed the lowest reported utilization. This was consistent with a presurvey finding that Little Rock had the most serious problem of the four cities studied with a maldistribution of providers and facilities. Major poverty pockets were concentrated in the eastern section of the city, while all the hospitals and most physicians were located in the western part of town. The situation was further exacerbated by a poor public transportation system which did little to enhance the accessibility of services to the poor.

Consistent with the ranking assigned to it, Oklahoma City showed the lowest level of utilization in six of the nine optional service subpopulations.

Generally speaking, utilization levels for the three optional services shown followed a predictable pattern in light of Medicaid coverage and other known available services. Thus, while Atlanta, without Medicaid dental benefits (except for AFDC children to age six), showed higher utilization in two of three aid categories than did Little Rock, with full dental coverage, this was not surprising in light of the several public and private dental programs in Atlanta and the maldistribution of providers in Little Rock.

For vision care, only Trenton provided Medicaid benefits, but Atlanta did relatively better than the other noncovered cities in two of three aid categories, and once again, it had several extra-Medicaid vision care services available.

With respect to prescriptions, only Atlanta and Trenton provided coverage. The relatively good showing by Oklahoma City, despite no Medicaid coverage, probably results from an awareness on the part of physicians of the existence of the special drug allowance provided by the county welfare agency in connection with the cash welfare benefit and the availability of a free county dispensary after the drug allowance has been exhausted.

Rather substantial differences in utilization rates were found when age breakouts were examined. The contrasts were especially marked when the AFDC group was separated by adult/child category. In most instances, predictably higher levels of utilization were manifested by adults.

Racial differences tended to be even more pronounced, with proportionally more whites than blacks reporting utilization in most subpopulations.

Finally, the figures in Table 1 tend not to support the hypothesis that a "utilization effect" would reflect substitution of more costly basic services for uncovered optional services in those locations with very narrow program benefits (e.g., more doctors' visits in Little Rock, prescriptions not covered, than in Trenton --prescriptions covered--in order to permit injection of medication and/or dispensation of free medicine samples to persons who otherwise might never obtain needed medicines). There are, however, so many variables that influence the demand for health services, that some of the substitution may very well be present (indeed, in some subpopulations that have been reviewed this seems certain), though masked by other offsetting factors.

The percent of eligibles who incurred some out-of-pocket costs for various services are shown in Table 2. Trenton, again, showed the lowest figures in thirteen of fifteen subpopulations, missing its customary first ranking only with respect to OAA hospitalization and OAA physician visits. This would seem to further substantiate the earlier contention that Medicaid eligibles in Trenton were more aware of the benefits to which they were entitled than were the eligibles in the other locations. Obviously, one reason for incurring expenses with respect to any covered service would be that the eligible received the service outside the system--either involuntarily because of ignorance, or voluntarily because, e.g., of a desire to utilize a provider who had chosen not to participate in the program. An attempt was made, through the survey document, to ascertain why out-of-pocket expenses were incurred, but time has not yet permitted analysis of those responses.

Preliminary analysis has revealed very sharp racial differences with respect to out-of-pocket expenses, with proportionally more whites than blacks reporting these expenses in a large majority of cases. Again, it is not clear whether this reflects less awareness of available Medicaid benefits or a greater propensity to seek services outside the system, for any of a variety of reasons. Some evidence exists on both counts, but more analysis is needed before any final inferences may be drawn. It can be reported, though, that a fairly consistent pattern exists showing larger proportions of whites than blacks using private physicians and dentists. This may be suggestive of more treatment outside the Medicaid system, since a number of private practitioners have refused to participate in it.

Among the AFDC group, out-of-pocket expenses were reported far more frequently among adults than among children, a result that was hardly surprising in light of coverage patterns.

Note that the data in Table 2 speak directly to the primary objective of determining who pays for the health care of the poor.

A secondary objective of the study was to determine the proportion of eligibles who perceived a need for health care services that were not received. Information on this point is displayed in Table 3. Note that Trenton shows the lowest reported unmet needs in eleven of twelve subpopulations. The only other low was recorded by Atlanta with respect to prescriptions for the AFDC (Atlanta and Trenton both provided prescription benefits). Oklahoma City, on the other hand, had the highest or tied for the highest proportion of unmet needs in eight of the twelve categories listed.

Again, age and racial differences were significant. Predictably, a much smaller proportion of children than adults among the AFDC were reported to have unmet needs, probably reflecting the previously mentioned mandatory screening programs for children, increasing morbidity with age and, possibly a natural parental instinct to obtain care for the young at any cost.

It was, however, mildly surprising to find that when all subpopulations were compared, half showed greater unmet needs among blacks and half showed greater unmet needs among whites, despite the fact that in most subpopulations whites reported higher levels of utilization. This finding suggests that whites may have a greater propensity to seek and use medical services than blacks. While there are many possible reasons for this situation, and further analysis will be conducted, one plausible explanation is that poor whites seem less prone than poor blacks to be concentrated in narrowly defined "ghetto" areas. Because of the greater dispersion of their living areas, poor whites may be more exposed to the middle class health values of the community.

A tertiary objective of the study was to determine the health status of Medicaid eligibles. Evaluation of this point by the eligibles themselves is displayed in Table 4. In all three aid categories, Trenton had the lowest percentage of eligibles who reported their health as "poor." Not unexpectedly, the smallest proportion appeared among the AFDC; relatively few persons under age twenty (the preponderant group in the AFDC classification) would be expected to classify their health as poor. Also note that the intercity differences are smallest among the AFDC. This is consistent with the fact that benefits are more uniform across sites in this category for children. The high proportion of Atlanta ADs reporting poor health is probably accounted for by a substantially higher proportion of eligibles who reported chronic conditions.

It was interesting to note, however, that the eligibles appeared to have a remarkable degree of sophistication when asked to rate the quality of health services available to them; their ranking of services often differed sharply from their rating of health status. Thus, they apparently recognized that many factors other than the quality of available services influence an individual's health status.

Some Tentative Conclusions

It seems clear from the survey that the presence of Medicaid benefits does positively influence the use of health services, and it also reduces financial barriers to care and levels of perceived unmet needs. The influence of these benefits on health status are much less clear. Here, the presence of many other variables, including different patterns of medical practice, variations in the available mix of providers, the geographical spread of hospital facilities and a multitude of others, makes it difficult to reach definitive conclusions.

The data clearly show rather marked differences in racial patterns for the variables investigated. Additional analysis should lead to better explanations for these variations.

The project team intends to continue working with the survey data and hopes to report further on its findings in the near future.

ACKNOWLEDGEMENTS AND REFERENCES

[1] This paper is based on research conducted by the authors under contract with the Social and Rehabilitation Service of DHEW, contract #18P 56701/3-01. Readers wishing more detail should refer to the final report for that contract, dated October 1974.

[2] DHEW, Health Survey Procedure: Concepts, Questionnaire Development and Definitions in the Health Interview Survey, NCHS, Series 1, No. 2, May 1964.

[3] Effect of Medicaid on Health Care of Low-Income Persons, School of Public Health and Administrative Medicine, Columbia University, New York, contracts WA-406, SRS-ORDT-68-01 and SRS-69-50.

[4] Cochran, William G., Sampling Techniques, 2d ed., Wiley & Sons, 1963.

TABLE 1. UTILIZATION, PERCENT OF ELIGIBLES

	Trenton			Atlanta			Little Rock			Oklahoma City		
Service	OAA	AD	AFDC	OAA	AD	AFDC	OAA	AD	AFDC	OAA	AD	AFDC
1. Hospital 1972-73	34	45	25	26	42	20	29	29	16	34	43	24
2. Physician 1973	68	81	56	73	81	56	59	65	44	64	68	55
3. Dentist 1972-73	18	36	43	10	34	50	15	28	37	14	23	31
4. Eye doctor 1972-73	48	45	32	41	36	21	33	28	27	29	24	17
5. Prescription 1973	51	61	41	47	54	32	44	37	27	42	50	39

TABLE 2. OUT-OF-POCKET EXPENSE, PERCENT OF ELIGIBLES

	Trenton			Atlanta			Little Rock			Oklahoma City		
Service	OAA	AD	AFDC	OAA	AD	AFDC	OAA	AD	AFDC	OAA	AD	AFDC
1. Hospital 1972-73	14	3	7	22	11	10	20	10	17	12	11	16
2. Physician 1973	19	5	5	13	6	8	34	24	16	11	14	13
3. Dentist 1972-73	13	4	3	77	24	26	54	36	17	68	55	34
4. Eye doctor 1972-73	12	6	8	39	17	2	66	55	22	38	39	33
5. Prescription 1973	4	3	5	48	33	33	96	84	84	91	91	94

TABLE 3. UNMET NEEDS, PERCENT OF ELIGIBLES

	Trenton			Atlanta			Little Rock			Oklahoma City		
Service	OAA	AD	AFDC	OAA	AD	AFDC	OAA	AD	AFDC	OAA	AD	AFDC
1. Physician 1973	9	12	3	29	34	6	21	20	5	22	31	6
2. Dentist 1972-73	13	24	8	24	31	22	22	33	16	25	41	23
3. Eye doctor 1972-73	19	14	3	41	41	12	38	32	10	41	41	13
4. Prescription 1973	5	10	11	10	12	9	10	26	24	11	23	16

TABLE 4. HEALTH STATUS, PERCENT OF ELIGIBLES LISTING POOR

	Trenton			Atlanta			Little Rock			Oklahoma City		
Rating	OAA	AD	AFDC	OAA	AD	AFDC	OAA	AD	AFDC	OAA	AD	AFDC
Poor	15	41	5	36	62	6	37	46	6	30	45	7

MATHEMATICAL MODELS FOR HEALTH SURVEY DATA AND ASSOCIATED MEASURES OF HEALTH STATUS*

R. C. Hanumara, M. H. Branson, D. Shao and J. C. Chen
University of Rhode Island; O. Thornberry,
R. I. Health Services Research, Inc.

I. INTRODUCTION

This paper presents the results of model building of measures of use and need of health services and measures of health status for a defined population. The data presented were obtained through a statewide household survey conducted by Rhode Island Health Services Research, Inc. (SEARCH) during the early part of 1972.** The sample was a full probability sample with households randomly selected from within each of Rhode Island's thirty-nine cities and towns. Interviews were obtained for 93 percent of the families falling in the sample. Information on disability and utilization of health services was obtained for 3,086 families consisting of 9,383 individuals. More detailed information describing the survey methodology is available in Thornberry et. al. (13).

Mathematical models to describe the state of health of a population are given in a paper by Chiang (4). In the same paper, he also proposes measures of health. In addition to Chiang's index of health, Miller (8) evaluated several other health indicies. Much of this theoretical work does not appear to have been applied in practice primarily due to a lack of appropriate data. This paper is concerned with validating the mathematical model proposed by Chiang and computing the indices of health using the survey data. In addition, techniques of the analysis of multi-dimensional contingency tables to describe the relationships between variables are applied to the data.

II. MATHEMATICAL MODEL

In data analysis, mathematical models of phenomena are often developed for purposes of simplifying and summarizing raw data. Number of bed days per person per year (measuring need) and number of physician visits per person per year (use) are used in the study of mathematical models and in computing measures of health.

Chiang developed a model to predict the number of doctor's calls, clinic visits and the number of complaint periods that an individual had in a year. With N as the number of doctor's calls that an individual had in a year, Chiang has shown the probability distribution of N to be Poisson with expected value λ. Further, assuming λ is a random variable with the gamma density function, the model proposed by Chiang for each of the variables that he selected is negative bionomial, i.e.,

$$P(N=n)=\frac{(n+\alpha-1)!}{(\alpha-1)!n!}\beta^{\alpha}(1+\beta)^{-(n+\alpha)},\qquad n=0,1,2,\ldots$$

Chiang fitted the negative bionomial model to the data on the number of doctor's calls and the number of complaint periods for different age groups and applied Chi-square goodness of fit test to validate the results. He suggested further testing of the model.

In the present study, the number of physician visits and the number of bed days are variables which closely correspond to variables chosen by Chiang. There were 9383 individuals responding to those two questions. The number of physician visits ranges from 0 to 90, while the number of bed days ranges from 0 to 365. Using the method of moments, the two parameters, α and β are estimated. Expected frequencies are computed and are given in Tables 2.1 and 2.2.

It is apparent from Tables 2.1 and 2.2 that the observed and expected frequencies differ considerably and thus any goodness of fit test would reject the hypothesis of validity of the model. One might ask whether the model is appropriate or, if the model is appropriate, what might be the source for lack of fit. It is observed from the data that the distribution of N has several modes. For example, while the number of individuals with exactly 10 bed days was 175, only 38 and 23 individuals reported 9 and 11 bed days, respectively. Similarly, there were 141 people with 10 physician visits, while only 63 and 28 reported 9 and 11 physician visits, respectively.

The distribution of the number of bed days has peaks at 7, 10, 14, 20, 21, 28, 30 and troughs at other numbers. The individual's preference for these numbers is understandable for questions based primarily on memory recall. This phenomenon is also observed in another survey done by Brown University (2). In the SEARCH survey, the number of bed days within the last two weeks are also reported and this response is more likely to be accurate. More accurate response would be expected if an individual selected an interval for the number of bed

TABLE 2.1

Observed and Expected Number of Persons by the Number of Bed Days in a Year

Bed Day	Observed	Expected
0	4880	4135
1	505	1218
2	839	740
3	554	527
4	325	403
5	303	321
6	141	262
7	466	218
8	75	184
9	38	156
10	175	134
11	23	116
12	55	101
13	16	88
14	271	77
15	29	68
16	11	60
17	21	53
18	15	47
19	4	42
20	48	37
21	94	33
22	7	30
23	3	26
24	11	24
25	14	21
26	5	19
27	1	17
28	28	15
29+	359	14

SOURCE: SEARCH Household Survey

TABLE 2.2

Observed and Expected Number of Persons by the Number of Doctor's Visits in a Year

Doctor's Visits	Observed	Expected
0	2076	3129
1	2845	1632
2	1357	1094
3	726	789
4	603	588
5	287	447
6	326	344
7	108	268
8	129	210
9	63	165
10	141	130
11	28	103
12	302	82
13	20	66
14	17	52
15	59	42
16	16	34
17	12	27
18	13	22
19	3	17
20+	181	14

SOURCE: SEARCH Household Survey

TABLE 2.3

The Pooled Frequencies from Table 2.1

Bed Day	Code	Observed	Expected
0- 2	0	6224	5914
3- 5	1	1182	1395
6- 8	2	682	706
9-11	3	236	420
12-14	4	342	269
15-17	5	61	180
18-20	6	67	123
21-23	7	104	86
24-26	8	30	61
27-	9	388	44

days such as between 9 and 11 or between 8 and 12 in a year. With this in mind, the data was grouped into the intervals 0-2, 3-5, 6-8, etc. Expected frequencies in the intervals were then computed using negative bionomial model. The difference in observed and expected frequencies in Table 2.3 are much smaller compared with those in Table 2.1. However, one may not conclude from this the validity of the negative bionomial model because of the arbitrariness involved in the pooling of the data. Still, if one were to argue in favor of the negative bionomial model in view of the good fit obtained by Chiang for the data that he collected, the expected frequencies in Tables 2.1 and 2.2 may be viewed as adjusted for bias. Due to cost limitations, no follow-up studies to determine the extent or bias of response preference for certain numbers were conducted. The effect of bias on the mean may cancel out thus not affecting the measures of health which are discussed later. It is interesting to note that in demographic studies, the digit preference or age heaping is observed in reporting of certain ages. The causes and patterns of age digit preference vary from one culture to another, but preference to report ages ending in '0' and '5' and avoid certain numbers as 13 and 4 is noted. Some techniques (Shryock and Siegel, (10)) are available to adjust the data for bias in census reporting but these have not been useful in the present context. Other contagious distributions such as Neyman's Type A are appealing because they may have more than one mode but the problem of bias must be resolved before another model is proposed.

Techniques developed by Goodman, Bishop, etc., to analyse the data of multi-dimensional contingency tables are used (Hanumara and Branson, (7)) to fit models. Using bed days as dependent (response) variable, age, sex, socio-economic characteristic as independent (factors) variables an unsaturated logit model is fitted. The results from the model are as follows: Main effects of economic status, age, sex on bed days are significant, interaction effects of economic status and age, age and sex on bed days are significant, while the interaction effect of economic status and sex on bed days is not significant. The factors economic status and sex affect the number of bed days independently within the age level.

III. HEALTH STATUS INDEX

Traditionally death rate has been used as a crude health measure. However, in most industrialized countries, the death rate has been relatively stable over the past two decades. Thus measures of health incorporating not only mortality but also morbidity information are called for.

A health status index should fulfill two requirements: 1) it should be sensitive to significant changes in morbidity as well as mortality, 2) it should be subject to analysis into components providing a useful description of health problems underlying the index value. According to Goldsmith (6), a health status index serves three functions: Public information, administration, and medical science. Many of the recently proposed health indices have been evaluated by Miller (8). One of the criteria which Miller used for evaluation is the availability of input data. On a scale of 0 to 3, the index developed by Chiang (4) obtained a rating of 2, while the ratings for all other indices were 0 or 1. Chiang's index has a total rating of 10 out of a possible 13 in Miller's evaluation of thirteen indices. Miller's Q-index is the only other index which received a total rating of 10, while the rating for the rest ranged from 2 to 8. Miller's Q- index is much more program specific and the computation of this index is not feasible due to the unavailability of appropriate data. The interested reader will also find relevant discussions on health indices in Fanshel and Bush (5) and Sullivan (12).

Chiang's Index

Chiang's index is based on the probability distribution of three variables; the frequency of illness; the duration of illness in number of days, and the time lost due to death in a year measured in number of days. He writes for age group x:

$$H_x = 1 - \overline{N}_x\overline{T}_x - (½)M_x$$

where H_x is the mean duration of health or the fraction of a year in which an individual is living and free of illness, $\overline{N}_x$ is the observed average number of illnesses per person, $\overline{T}_x$ is the average duration of an illness in a year, M_x is the age-specific death rate for the year. The product $\overline{N}_x*\overline{T}_x$ is an estimate of the fraction of the expected duration that an individual is ill in a year. Now, H, the index of health for the entire population for the specified time period is defined to be

$$H = (I/P)\ \Sigma_x P_x H_x$$

where P_x is the age-specific population and P is the total population. The value of H ranges from zero to one with healthier populations having larger values of H. In the extreme case, if no illness and no death for the population; i.e., $\overline{N}_x\overline{T}_x=0$, $M_x=0$ for each x; $H_x=1$ and H=1. If everyone were ill with no deaths for the entire year; i.e., $\overline{N}_x\overline{T}_x=1$, $M_x=0$, then $H_x=0$ and H=0.

The Computation of Chiang's Index

In Chiang's model, illness is assumed to be a recognizable state of measurable duration, but how it is to be measured is not specified. The total duration of illness for an individual during a year has to be ascertained in order to compute Chiang's index. The number of bed days reported in the SEARCH survey data is only a part of illness period in a year for an individual. The total duration of illness for an individual is interpreted to be the sum of the number of bed days and the number of days with restricted activity but not in bed. Thus, the number of days of restricted activity apart from the days in bed had to be estimated. It should be noted that the term total duration of illness is consistent with the term "days of restricted activity" used in U. S. National Health Surveys. As the intent is to compute Chiang's index, his definitions are used here.

The number of bed days and the number of restricted activity days apart from the days in bed in a two week period are also available from the SEARCH survey data. Thus, (total duration of illness for two week period/bed days for two week period)*(bed days for twelve month period) is an estimate of the total duration of illness in a year. Data from two other surveys, Current Estimates (14) and Preliminary Report by the University of Chicago (15), are used to verify whether for specific-age group the ratios between total duration of

TABLE 3.1

Ratios from Current Estimates and SEARCH Household Survey

	Age			
	Under 6	6-16	17-44	45-
Ratio of Restricted Days and Bed Days (Current Estimates, 12 months)	2.358	2.083	2.249	2.409
Ratio of Total Duration of Illness and Bed Days (SEARCH survey, 2 weeks)	2.092	1.623	2.478	2.323

TABLE 3.2

Ratios (for 2 week period) from Preliminary Report and SEARCH Household Survey

	Age				
	1-5	6-17	18-44	45-64	65-
Ratio of Restricted Days and Bed Days (Preliminary Report)	1.671	1.565	1.955	1.947	1.767
Ratio of Total Duration of Illness and Bed Days (SEARCH survey)	2.220	1.660	2.478	2.382	2.248

TABLE 3.3

Statistics for R. I. 1971 from SEARCH Household Survey Used to Compute Total Duration of Illness

Age	Bed Days (2 weeks)	Restricted Days Exclusive of Bed Days (2 weeks)	Ratio of (2) + (3) and (2)	Bed Days (a year)
0- 4	0.515	0.602	2.169	4.461
5-14	0.554	0.352	1.635	3.980
15-24	0.342	0.432	2.257	4.504
25-44	0.386	0.564	2.461	4.796
45-64	0.629	0.869	2.382	8.261
65-	0.958	1.196	2.248	13.805

TABLE 3.4

Total Duration of Illness, Death Rate, and Population

Age	Total Duration of Illness	Death Rate (1971) per 1000	Population (1970)
0- 4	9.676	4.10	76035
5-14	6.507	0.37	174163
15-24	10.166	0.89	173643
25-44	11.803	1.66	210297
45-64	19.678	10.30	208655
65-	31.034	59.83	103932

illness and the number of bed days are in agreement with the ratio from the SEARCH survey data. This indicates the reliability of the ratio estimator, total duration of illness per person in a year.

The Current Estimates (14) gives restricted days and bed days for the year 1971. The ratios between them are computed and are shown in Table 3.1. This table also shows the ratios of total duration of illness and bed days for two week period, computed from SEARCH survey data.

In the Preliminary Report (15), the percentage of individuals who stayed in bed for one or more days during the last two weeks, and the average number of bed days experienced by those individuals who stayed in bed for one or more days during the last two weeks are given. The mean values of bed days of individuals can be computed by multiplying these two values. The mean values of restricted days can also be computed from this report. Then the ratios between bed days and restricted days are obtained. From SEARCH survey data, the mean values of bed days and restricted activity days not in bed as well as the ratios are computed. The results are shown in Table 3.2.

The National Health Surveys are carried out throughout the whole year and hence the estimates obtained from them are not influenced by the presence of an epidemic at a particular time of the year. There was an incidence of influenza at the time of SEARCH survey and this might have some small influence on the results relating to two week period. The difference in ratios in Table 3.1 are both positive and negative. On the other hand, the differences in ratios in Table 3.2 are all positive. It may be due to the fact that influenza mentioned before and also different geographic regions are involved.

For actual computation of the index, certain age groups are chosen for which the death rates are available. Table 3.4 gives the total duration of illness in a year as computed from the SEARCH survey data (Table 3.3) and the death rate for 1971 and the population for 1970 for the State of Rhode Island taken from Vital Statistics (9). Now, the health indices for different age groups are computed. For instance, the computations for age group zero to four are as follows:

$$H_x = 1 - \overline{N}_x\overline{T}_x - (½)M_x$$

$$= 1-(9.676/365)-(½)*.0041 = 0.971$$

Age	0-4	5-14	15-24	25-44	45-64	65-
H_x	.971	.982	.972	.967	.941	.885

Finally, Chiang's index H is obtained

$$H = (1/P)\Sigma_x P_x H_x$$

$$= 1/949725*(76035*0.971 + \ldots + 103932*0.885)$$

$$= 0.954$$

We have used certain age groups in computing H. The primary reason is the availability of data on death rates at these age groups. One can expect the value of H to be sensitive to different choices of age groups. Certainly, the result is improved with more age groups. However, it is not only the value of H that is interpreted but also individual values of H_x representing age groups. Therefore, too many age groups are not needed and six age groups which are used here are enough for interpretation.

Recalling that population proportions used in computing H are for the year 1970 instead of 1971, one might question its effect on H. This is done because the data on population is available only up to 1970 at the present time and practically no change is expected in population proportions from 1970 to 1971.

Chiang gives formulas for sample variance of H but we are unable to compute it. In Chiang's terminology, total duration of illness is the product of number of illnesses and average duration of illness and the formula for sample variance of H requires data on these two variables. We interpreted the total duration of illness in a way convenient to compute H using SEARCH survey data but this does not permit the computation of the sample variance of H.

The only index which we are able to compute still has some limitations. We say the health status of Rhode Islanders in numerical terms is 0.954. As the larger the value of H, the healthier population is, one might conclude that Rhode Islanders are in fact healthy. It is hard to imagine a situation where the value of H will be close to zero and lead one to conclude otherwise. Hence no conclusion can be drawn using a single value. It would have been interesting to compare health index of 1971 with previous years but data are not available to permit such study. Again, for lack of necessary data, no comparisons of health status of states are made.

Health indices for different age

groups have some use. The reliability of H may be seen by different values of H_x. H_x attains maximum value in the age group 5-14 and reaches the lowest value in the age group above 65 as one would expect. Health index might be used, for example, to study the health status of older people before and after medicare facilities.

At the time of conclusion of this study, Srivastava (11) is computing Chiang's index and also Chen's G- index (3) using different data base. Srivastava uses hospital days as a measure of illness as this is readily available for many sub-classifications of the population. Thus, health indices are being computed in his work for populations divided on the basis of census tracts or towns or social variables which lead to many comparative studies. Further, his studies show the G- index and Chiang's index are correlated.

FOOTNOTES

*This research was supported in part by Rhode Island Health Services Research, Inc. (SEARCH). The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of SEARCH.

**The SEARCH survey was conducted under a grant from the National Center for Health Services Research (#1-R18-HS-00720).

REFERENCES

1. Aday, L. A., "The Utilization of Health Services: Index and Correlates," Purdue University, 1965.

2. Brown University, "Brown 1970 Piggyback Follow-Up Questions and Frequency Distribution," Department of Sociology, Population Studies and Training Center, Providence.

3. Chen, M. K., "The G- Index for Program Priority," Proceedings of a Health Services Research conference held in 1972 at Tuscon, 1973.

4. Chiang, C. L., "An Index of Health: Mathematical Models," National Center for Health Statistics, Series 2, No. 5, 1965.

5. Fanshel, S. and Bush, J. W., "A Health Status Index and Its Application to Health Services Outcomes," Operations Research, 18, November-December 1970.

6. Goldsmith, S. B., "The Status of Health Status Indicators," Health Services Reports, 87, No. 3, March 1972.

7. Hanumara, R. C. and Branson, M. H., "Application of Log-Linear Modelling to Health Survey Data," University of Rhode Island, Kingston, 1974.

8. Miller, J. E., "Criteria for Evaluating the Application of Health Status Indices in a Management Context," Office of Research and Development, 1971.

9. Rhode Island Department of Health, "Vital Statistics," 1971.

10. Shryock, H. S., Siegel, J. and Associates, "Methods and Materials of Demography," U. S. Bureau of Census, 1971.

11. Srivastava, R. K., "Application of Health Status Indicators as Evaluatory Tools to Analyse Health Status for the State of Rhode Island," Thesis to be submitted to University of Rhode Island, Kingston, 1974.

12. Sullivan, D. F., "Conceptual Problems in Developing an Index of Health," National Center for Health Statistics, Series 2, No. 17, 1966.

13. Thornberry, O.; Scott, H. D.; and Branson, M. H.; "Methodology of a Health Interview Survey for a Population of One Million," paper presented at the 101st Annual APHA Meeting, San Francisco, November 1973.

14. U. S. Department of Health, Education and Welfare, "Current Estimates from the Health Interview Survey United States-1971," National Center for Health Statistics, Series 10.

15. University of Chicago, "Preliminary Report Number 1, Disability Days and Physician Contact for the Two Week Period Preceding the Interview Data by Age, Sex, Race, Residence, and Family Income." Center for Health Administration Studies, 1971.

PSYCHOLOGICAL INFLUENCES ON WHITE COLLAR PAY

Einar Hardin, Michigan State University

Many psychological characteristics of employees may have an impact on job performance: openmindedness, adaptability to new conditions, interpersonal competence, achievement motivation, trustworthiness, intelligence, knowledge, and so forth. Because the mix of psychological characteristics is likely to differ between the demand and supply sides of the labor market, pay differentials associated with these characteristics will tend to arise. Persons having bundles of characteristics generally desired by employers will tend to be paid more, given their other pay-determining attributes.

The voluminous research of the last decade on interpersonal differences in pay[1] gives little attention to psychological variables, and empirical analyses have made virtually no use of such variables other than intelligence. The present study is an attempt to reduce the gap. The roles of two psychological variables, dogmatism and readiness for change, are explored as additions to a human capital model - and alternately a demographic model - of pay, and the potential social economic benefits from alterations in dogmatism and readiness for change are calculated.

Psychological Variables

Dogmatism was defined by Rokeach in 1954 as "(a) a relatively closed cognitive organization of beliefs and disbeliefs about reality, (b) organized around a central set of beliefs about absolute authority which, in turn, (c) provide a framework of patterns of intolerance and qualified tolerance toward others." This concept was originally offered as a generalization of authoritarianism, because Rokeach believed the latter, especially when operationalized in test form, referred to fascist or right wing authoritarianism, although there was nothing to preclude the existence of an authoritarianism of the political left. A large pool of statements sampling the three basic aspects of dogmatism and useful in measurement was also presented.[2]

The subsequent twenty years have seen much analysis and use of the Rokeach concept and scale. Reliability appears to be reasonably high, as illustrated by r=.55 between test results separated by a five-year period, and seems fairly insensitive to the number of items. Dogmatism seems to have only a slight negative relationship to various intelligence test scores, including ACT, SAT, and CQT scores.[3] Persons differing in dogmatism have been compared on other characteristics, such as scores on standardized personality tests, interpersonal and group behavior, psychological disorders, perceptual functioning, time perspective, and problem solving. The emerging general picture of a dogmatic person is one who is not receptive to new ideas and beliefs, is intolerant, inflexible, and insecure, and does not function easily in unstructured settings, and whose parents had closed minds.[4]

It is evident that persons differing in dogmatism may differ in productivity and worth to an employer. Highly dogmatic employees can be expected to be less quick than others to offer or even accept proposals for problem solution and organizational change adequate to the needs of the firm, and they may be substantially less able to cooperate with other employees. Very high levels of dogmatism are likely to be strongly disabling. Being less valuable than otherwise similar persons, more dogmatic employees would be paid less.

The productivity effects of dogmatism may vary with the knowledge and skills of the employee. When the job is routine and predetermined and when careful and close compliance with rules and instructions is important, moderate dogmatism is perhaps of little consequence. In complex and relatively unstructured jobs which require personal initiative and a sensitive response to actions of others, dogmatism may impair productivity greatly and cause a severe economic penalty. Since those who hold complex jobs are likely to be the better educated, longer-service, male employees, any decline in pay which accompanies growing dogmatism may vary with education, length of service, and sex.

Whether an individual willingly accepts - or positively seeks - a change from one to another state of affairs depends in part on how the characteristics of the two states meet the needs of the individual, in part on the economic cost the individual incurs in going from one to the other. As economists have long argued in analysis of educational, occupational, and geographic mobility, the individual may be viewed as responding to change by weighing the difference in flows of benefits from the two states of affairs against the economic cost of the change. Readiness for change, the second psychological variable covered by the study, is not involved in this economic concept of response to change.

However, the change from one state of affairs to another may also entail a requirement to adjust to new circumstances, and the individual may not know with certainty that he will be able to make the required adjustment. As recognized in the concept of resistance to change, many persons may view the adjustment requirement and the self-doubts with distaste. Economists might say that the individual has a psychological cost of transition which must be outweighed by the difference between the net flow of benefits and the economic cost of transition, before a positive response occurs. Such a psychological cost represents low readiness for change.[5]

It is important to recognize, furthermore, that some persons may not view the change process with distaste or indifference but may welcome it. First, the change process may temporarily break up the monotony associated with any state of affairs. Second, it affords some persons a welcome opportunity to test their ability to overcome hazards and ambiguities and generally cope with change. For such persons, the process of change may have a positive value which in part or in whole offsets the economic cost, if any, of the transition. This positive psychological value represents high readiness for change.

The concept of readiness for change may be used in extending the economic analysis of change. In brief, given the usefulness of the potential as compared with the actual states of affairs and the economic cost of the transition, increased readiness means increased probability of a favorable response to a particular change.

Unfortunately, there is very much less empirical research on readiness for change than on dogmatism. Almost all is based on questionnaire data covering two insurance companies, one of which furnished the material for the present study. Trumbo reported a corrected odd-even reliability coefficient of .79 for the original nine-item scale of readiness for change and a .28 correlation with scores on the Wonderlich Personnel Test used in estimating the mental ability of job applicants; showed that high readiness for change was associated with positive attitudes toward past and prospective changes in technology and work environment but was virtually independent of job satisfaction; and demonstrated that high readiness for change among supervisors was associated with high readiness and low group cohesiveness among their direct subordinates. Using nine-item data from the second company, Nangle verified Trumbo's finding of a positive relationship with favorability toward changes in technology and work environment, and Faunce found higher readiness for change among employees who identified with high social class and had fathers with high occupation or income. After eliminating three items on the basis of logical heterogeneity and low item intercorrelations, Hardin showed that desire for changes in specific job aspects was associated not only with low satisfaction as to these job aspects but also with high readiness for change.[6] These findings suggest that the readiness for change scale does measure with some reliability an attitude which is reflected in observable behavior, is different from job dissatisfaction, has a bearing on response to particular technological and work changes, and does not merely summarize the comparative attractiveness of two states of affairs.

Readiness for change, as defined here, derives its economic importance from incessant changes in product and resource markets and in company technology, all of which call for continual and swift adaptation of the firm and its internal activities. Adaptation is swifter and more effortless, when the work force favors the very process of change, apart from the characteristics of the initial and terminal states of affairs, than when it takes a negative or hostile view of the change process. Other things being equal, high readiness for change among employees has a positive economic value for the employer, and he will pay a premium to secure and retain the services of highly ready employees.

The productivity effects and, hence, the pay premium of high readiness for change may vary with circumstances. Where much adjustment to change is needed, the pay premium is likely to be especially large. One may plausibly think that persons with much seniority and formal education are especially likely to hold complex jobs where opportunities for cooperation or obstruction are great; accordingly, the readiness for change premium would be higher among employees with much seniority and education. However, an interaction with education and seniority may be negative instead of positive, since adjustment to change originated by other persons, which is what the scale seems to measure, may be more needed in routine than complex jobs. This would make the pay premium for readiness for change larger among women, who tend to hold the routine and simple jobs, than among men.

Economic-Demographic Variables and Models

Many other variables are commonly thought to influence the rate of pay. According to human capital theory the pay of an individual depends not only on his basic ability and industry but also on the amount of human capital he has acquired through investment in formal schooling, in vocational training, in on-the-job training, and in learning by doing. When specific information is lacking concerning on-the-job training and learning by doing, research workers often substitute length of service with present and past employers. Following the general approach of Malkiel and Malkiel we use this basic model:

$$(1)\ \log_e PAY_i = \alpha_o + \alpha_1 ED_i + \alpha_2 SEN_i + \alpha_3 SEN_i^2 + \alpha_4 WORK_i + \alpha_5 WORK_i^2 + \varepsilon_i$$

where ED = years of formal schooling completed, SEN = years of service with present employer, WORK = years of service with prior employers, and α_o expresses the combined effect of all unspecified variables.

According to a demographic and theoretically less rigorous interpretation, pay depends on formal schooling, length of service with present employer, and chronological age. A distinct age-earnings profile having a maximum at some intermediate age is explained eclectically by reference to work experience with other employers, special training, obsolescence, degenerative phenomena, etc.[7]

From this interpretation we distill the following demographic alternate model of pay:

$$(2)\ PAY_i = \alpha_o + \alpha_1 ED_i + \alpha_2 ED_i AGE_i + \alpha_3 ED_i AGE_i^2 + \alpha_4 \log_e SEN_i + \alpha_5 ED_i \log_e SEN_i + \varepsilon_i$$

in which AGE = years of age at pay date. Years of work experience with other employers, WORK, is excluded because, among men more than women, it is highly correlated with AGE, once ED and SEN are held constant. If $\alpha_2 > 0$ and $\alpha_3 < 0$, as expected, the age-earnings profile reaches a maximum at AGE = $-\alpha_2/2\alpha_3$, regardless of the years of formal schooling, but the profile is at a higher level, especially in years close to those of maximum earnings, the higher is the level of formal schooling.[8] Finally, the model concedes that education and seniority may have an interaction effect upon pay in addition to their main effects.

Two variables normally associated with pronounced pay differences are absent from both models: sex and occupation. The sex differential in pay may not be merely a level or scale factor but may vary with personal characteristics, and even the disturbance term may have different variance. Therefore, it seemed preferable to analyze the data for men and women separately instead of building interaction terms for sex into a single model.

The exclusion of the occupation variable is based on different reasons. According to economic theory for a nonunion market, such as the one in which the present study was conducted, occupational pay differences reflect (a) the nonwage disadvantages (noise, dirt, hazard, work load, etc.) inherent in each of the occupations and (b) the prices payable for the grades of labor optimally suited to each occupation. When pay differences arise exclusively from nonwage disadvantages (so-called equalizing differentials), the model of pay should include variables explicitly representing these disadvantages. In the present study no reliable information was available about the presumably limited differences in nonwage disadvantages among occupations.

On the other hand, when pay differences reflect only differences in quality of labor, the model of pay used in estimating the prices paid for labor quality should give no recognition at all to occupation. This can be seen clearly when there is no variation in tasks and remuneration within occupations but when employee characteristics determine fully who gets access to the various occupations. A regression equation that contains the occupation variable along with the personal characteristics variables is likely to assign all influence upon pay to the occupation variable, the coefficient of which will represent a hopelessly confounded effect of all the personal characteristics.

Models Augmented with the Psychological Variables

The human capital model and the alternative demographic model formed the core of the models incorporating the two psychological variables: dogmatism (DOGM) and readiness for change (RFC). In the main analysis these additional variables were included only in their main-effects form, giving the equations:

(3) $\log_e PAY_i = Z_i' + \alpha_6 DOGM_i + \alpha_7 RFC_i$

(4) $PAY_i = Z_i'' + \alpha_6 DOGM_i + \alpha_7 RFC_i$

where Z_i' and Z_i'' represent the sums of all terms in equations (1) and (2). However, interactions of DOGM and RFC separately with ED and SEN in the human capital model and with ED and $\log_e SEN$ in the alternative demographic model were also explored by inclusion of the corresponding terms in expanded versions of (3) and (4). Thus, the analysis was designed to show what effects, if any, dogmatism and readiness for change exerted upon pay when several generally recognized pay determinants were held constant.

Procedure

While economic and demographic data sufficient for estimating the coefficients of models (1) and (2) can often be obtained from the files of employers, appropriate psychological data are seldom available. First, many organizations neither collect nor use psychological data on their employees, or they collect data on dimensions having little relevance for pay. Second, when such data have been collected and used and if they are found to have the expected statistical relationship to pay, they may be suspect in the eyes of some research workers on the premise that the relationship arose because the organization falsely thought the data had a bearing on the individual's worth as an employee and because it placed and rewarded the person accordingly.[9]

Collection of primary data for studies of psychological influences upon pay may involve bringing all employees together in a company cafeteria or auditorium to fill out psychological questionnaires and to take tests, with substantial cost in money and interruption for the employing organization. Alternatively, it may require an interview project or a mail questionnaire survey, in which strict control against interference is difficult to attain and the nonresponse rate may become uncomfortably high. At an exploratory stage, there is much to be said, then, for secondary use of research files even if these were created for other purposes and perhaps in other times.

This study made use of data from an earlier research project.[10] The files contained complete and usable information on sex, age, education, seniority, readiness for change, and dogmatism for 159 women and 49 men, who jointly represented about two-thirds of all office employment in a casualty insurance company and who included supervisory and nonsupervisory employees engaged in the various departments commonly found in such companies. Seniority (SEN) was calculated as years of employment with the company. Education (ED) represented years of formal schooling completed; "some college" and "college" were coded as 14 and 16 years of schooling. No usable direct information being available, years of work experience with previous employers was calculated as WORK = AGE - (ED + SEN + 6), in accordance with common usage among human capital researchers. This formula implies an assumption, more warranted for men than women, that each person is working from the end of formal schooling to the start of employment with the current employer.

Employee dogmatism was measured by a Likert score based on agreement-disagreement responses to these eleven statements: Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups; The worst crime a person could commit is to attack publicly the people who believe in the same thing he does; It is only natural that a person would have a much better acquaintance with ideas he believes in than with ideas he opposes; In this complicated world of ours the only way we can know what is going on is to rely on leaders or experts who can be trusted; The present is all too often full of unhappiness; it is only the future that counts; It is only when a person devotes himself to an ideal or cause that life becomes meaningful; There are two kinds of people in this world: those who are for the truth and those who are against the truth; Man on his own is a helpless and miserable creature; It is only natural for a person to be rather fearful of the future; It is better to be a dead hero than a live coward; and The main thing in life is for a person to want to do something important. Each statement had six responses with associated weights: 6 = I agree very much; 5 = I agree on the whole; 4 = I agree a little; 3 = I disagree a little; 2 = I disagree on the whole; and 1 = I disagree very much. A score with a possible range from 11 to 66 was computed as the sum of weights of the selected responses; persons failing to respond to one or more items were removed from the study. The statements were chosen by Kamenske from the Rokeach pool. The criteria of selection were that all three basic aspects of dogmatism must be reflected among the items, that there should be a high item reliability and discrimination, and that the items should be plausible as part of a questionnaire on job attitudes of clerical employees. In data from another company Trumbo found strong positive association between responses to individual statements and the eleven-item total score.[11]

Readiness for change was measured by a Likert score largely based on agreement-disagreement responses to six statements. Five of these were: One can never feel at ease on a job where the ways of doing things are always being changed; The trouble with most jobs is that you just get used to doing things in one way and then they want you to do them differently; I would prefer to stay with a job I know I can handle than to change to one where most things would be new to me; I like a job where I know that I will be doing my work about the same way from one week to the next; When I get used to doing things in one way, it is disturbing to have to change to a new method. Each of these statements had five response categories with associated weights: 1 = I strongly agree; 2 = I agree a little; 3 = I neither agree nor disagree; 4 = I disagree a little; and 5 = I strongly disagree. The sixth item was: The job that you would consider ideal for you would be one where the way you do your work: 1. is always the same; 2. changes very little; 3. changes somewhat; 4. changes quite a bit; and 5. changes a great deal. After persons not responding to all six items were dropped, a readiness-for-change score with a possible range from 6 to 30 was computed by summing the weights of the six selected responses. The items were selected from the nine-item scale on the basis of logical similarity and satisfactory item intercorrelations.[12]

The pay rates were fixed before the questionnaire survey was undertaken. Hence, any observed relationship of pay to dogmatism or readiness for change cannot be interpreted as a self-validation or artificial certification effect.

Data on weekly pay at survey time had been obtained from company records. A rate range existed for each job but permitted wide variation according to management judgment. No overtime or other premiums were paid. There existed no collective bargaining agreement and no union holding or seeking recognition as a bargaining agent.

Preliminary analysis showed that pay had different means and standard deviations for the two sexes. The two psychological variables, DOGM and RFC, were not strongly interrelated ($r = -.38$ among men and $-.20$ among women), and they were also rather weakly related to the constituent variables in the economic models (1) and (2), as shown by correlations up to .20 among women and .31 among men. However, interaction variables involving DOGM and RFC often were strongly correlated with some component variables. In particular, correlations between ED·PSY and PSY and between SEN·PSY and SEN (where PSY represents either DOGM or RFC) ranged from .76 to .93 among women and from .93 to .98 among men. Because the latter pattern made it difficult to distinguish between main-effects and interaction-effects models in sample data, the estimated coefficients are presented for main-effects models only.

The coefficients of models (3) and (4) as well as models expanded by interaction terms involving DOGM and RFC were estimated by the method of ordinary least squares. Errors in variables are likely to be fairly large in the case of psychological variables, such as Likert scores. The conservative nature of the estimates should be kept in mind when the results are interpreted.

Results

The regression equations corresponding to the human capital models (1) and (3) are shown in Table 1. In model (1), that is excluding the psychological variables, the coefficients had sensible signs and values, and they were statistically significant, except for SEN and SEN^2 among men, where $P = .07$ and $.31$, respectively. In all, the standard human-capital model covering ordinary economic variables appeared to be relevant to pay in this organization.

When DOGM and RFC were added as main effects and model (3) was estimated, the adjusted $\bar{R}^2$ rose; the increase was statistically significant in the women's sample ($F = 5.04$) and was near significance in the men's sample ($F = 3.02$). The coefficient of DOGM was clearly significant for both sexes, and that of RFC was of borderline significance ($P = .06$) for women. Inclusion of DOGM and RFC left most of the other coefficients unchanged, but it made the coefficient of ED nonsignificant ($P = .10$) among women and reduced and rendered clearly nonsignificant ($P = .41$) that of SEN among men.

According to the estimates of model (3) an increase in DOGM by one score point reduced the pay of women by about .3 per cent and that of men by about .9 per cent. Furthermore, an increase in RFC by one score point raised the pay of women by .3 per cent but left the pay of men unchanged.

In the terminal equations obtained when a stepwise deletion and addition procedure with a $P = .05$ stopping rule was applied, the estimated impacts of DOGM and RFC retained their signs and had almost the same values as in the preceding equation. Thus, the results were not unduly sensitive to the specification of the model.

Several analyses were conducted in a search for those interactions of DOGM and RFC with ED or SEN which seemed plausible in advance. Here, model (3) was augmented with terms representing either ED·DOGM and ED·RFC or SEN·DOGM and SEN·RFC. When added pairwise to the equation, the coefficients of these additional variables were found non-significant singly as well as jointly. However, because of the high correlations with certain interaction terms, the stepwise procedure as applied to the expanded equations frequently retained the interaction term while rejecting the corresponding main psychological variable. Nevertheless, the revised equations fit the data, in terms of $\bar{R}^2$, only about as well as did the equations derived from model (3).

The coefficients of the alternative demographic model (4) were also estimated. When all the variables were included, the coefficient of

Table 1. Estimates for the Human Capital Model of $\log_e$ Weekly Salary

Variable	Women (n = 159) Model (1)	Model (3)	Terminal	Men (n = 49) Model (1)	Model (3)	Terminal
Intercept	3.53 (.143)	3.67 (.157)	3.92 (.0706)	3.56 (.195)	4.07 (.284)	4.06 (.249)
ED	.0262 (.0116)	.0191[a] (.0116)	N.S. -	.0467 (.0141)	.0411 (.0138)	.0410 (.0131)
SEN	.0718 (.0119)	.0726 (.0116)	.0733 (.0111)	.0640[c] (.0346)	.0297[e] (.0359)	.0256 (.00789)
SEN^2	-.00318 (.00127)	-.00316 (.00124)	-.00323 (.00121)	-.00346[d] (.00334)	-.000400[f] (.00344)	N.S. -
WORK	.00770 (.00304)	.00610 (.00300)	N.S. -	.0321 (.0112)	.0302 (.0108)	.0298 (.0103)
$WORK^2$	-.000288 (.000108)	-.000217 (.000107)	N.S. -	-.000978 (.000428)	-.000861 (.000415)	-.000842 (.000393)
DOGM	- -	-.00294 (.00133)	-.00339 (.00132)	- -	-.00919 (.00362)	-.00912 (.00318)
RFC	- -	.00348[b] (.00185)	.00435 (.00183)	- -	-.000897[g] (.00488)	N.S. -
$\bar{R}^2$	.4725	.5022	.4914	.4553	.5089	.5313

[a]P = .10, [b]P = .06, [c]P = .07, [d]P = .31, [e]P = .41, [f]P = .91, [g]P = .86

N.S. indicates the variable did not survive in the stepwise regression procedure, P = .05.

DOGM was negative and significant for both women (a_6 = -.161, s(a) = .0783, P = .042) and men (a_6 = -.897, s(a) = .347, P = .013). The coefficient of RFC was positive for both sexes and significant for women (a_7 = .237, s(a) = .110, P = .033) but not for men (P = .466). Thus, the pay penalty for an additional score point of dogmatism was about 16¢ per week among women and about 90¢ per week among men, while a pay premium amounting to about 24¢ per week per score point of readiness for change was paid only to women. Very similar results were obtained, when the stepwise procedure was applied to model (4). Again, there was a substantial amount of exploration for interaction effects involving DOGM and RFC, and the results closely resembled those of the human capital model.

While no formal significance test was made, it seems obvious that the impact did differ between women and men. The evidence for the existence of interaction effects involving the two psychological variables with other variables was at best erratic and weak. It may be wise, at the current time, to regard dogmatism and readiness for change as having main effects impact upon pay and interaction effects with sex alone.

Social Economic Benefits from Psychological Change

The finding that dogmatism and readiness for change had an impact on the salaries of employees implies that these two psychological variables also affected the output of the nation. Indeed, one may estimate the output gains, often called social economic benefits, from decreased dogmatism and increased readiness for change on the basis of the results in Table 1. Three main assumptions are needed for such estimates. First, according to a premise commonly used by economists, the contribution of an employee to national output is assumed to be roughly equal to the employee's compensation, composed of direct salaries and the value of fringe benefits. Second, the value of fringe benefits is assumed to be unaffected by both dogmatism and readiness for change, because data were not collected. Third, although lowered dogmatism and increased readiness for change may alter the employee's behavior as a citizen, with attendant changes in government policy and actions, economic effects through other links than the person's employment are assumed to be absent.

Among women as well as men, the standard deviation of individual scores was about seven points of dogmatism and about five points of readiness for change. According to regression results, a decrease in dogmatism of both sexes by seven points and an increase in readiness for change of women by five points would raise the weekly pay, at 1957 levels, by $2.16 for women and by $6.31 for men. Expressed in annual rates at 1973 levels[13] the social economic benefits would amount to about $216 per woman and $631 per man.

It is not known whether educational and training programs can be redesigned or the environment of children in their formative years can be modified to bring about such a change in the two psychological variables, nor are the costs and side effects known. However, if upheld by replication work in progress, the estimated annual social economic benefits are large enough to warrant serious analysis of the factors determining dogmatism and readiness for change, with special attention to the potentials of programs for deliberate psychological change.

Concluding Remarks

The influences of dogmatism and readiness for change upon white collar pay were estimated by including these two variables in widely accepted models of pay determination where sex, education, seniority, and prior work experience were held constant in one and sex, education, seniority, and age were held constant in the other model. Clearly, both dogmatism and readiness for change deserve recognition in their own right.

The findings support the general thesis that a number of psychological variables should be included in the economists' models of pay determination. In expanding the models, it is important to avoid those psychological variables, such as job satisfaction and many work attitudes, which cannot be regarded as causes of pay. Advice from psychologists is needed in the choice of promising variables and of measurement methods.

Footnotes

[1] See Mincer, Jacob. "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach." Journal of Economic Literature, Vol. 8, No. 1 (March 1970), pp. 1-26.

[2] See Rokeach, Milton. "The Nature and Meaning of Dogmatism." Psychological Review, Vol. 63, No. 3 (May 1954), pp. 194-209, for concepts and his "Political and Religious Dogmatism: An Alternative to the Authoritarian Personality." Psychological Monographs, Vol. 70, No. 425 (1956) for item pool.

[3] See Rokeach, Milton. The Open and Closed Mind (New York: Basic Books, 1960), p. 190; Zagona, Salvatore V. and Zurcher, Louis A., Jr. "Notes on the Reliability and Validity of the Dogmatism Scale." Psychological Reports, Vol. 16, No. 3, Part 2, (June 1965), p. 1236; and Thompson, Robert C. and Michel, Jerry B. "Measuring Authoritarianism: A Comparison of the F and D Scales." Journal of Personality, Vol. 40, No. 2 (June 1972), p. 185.

[4] This account relies heavily on Vacchiano, Ralph B., Strauss, Paul S., and Hochman, Leonard. "The Open and Closed Mind: A Review of Dogmatism." Psychological Bulletin, Vol. 71, No. 4 (April 1968), pp. 261-273.

[5] Thus, resistance to change may be interpreted as low readiness for change, but in much of the literature there is a failure to distinguish between low readiness and either the basic unattractiveness of the potential as compared with the actual state of affairs or the individual's economic cost of going to a new state of affairs.

[6] Trumbo, Don A. "Individual and Group Correlates of Attitude toward Work-Related Change." Journal of Applied Psychology, Vol. 45, No. 5 (October 1961), pp. 338-344; Nangle, John E. A Study of Certain Aspects of Organizational Communications within a Medium-Sized Insurance Company Undergoing Technological Change (unpublished doctoral dissertation, Department of Psychology, Michigan State University, 1961); Faunce, William A. "Social Stratification and Attitude toward Change in Job Content." Social Forces, Vol. 39, No. 2 (December 1960), pp. 140-148; and Hardin, Einar. "Job Satisfaction and the Desire for Change." Journal of Applied Psychology, Vol. 51, No. 1 (February 1967), pp. 20-27.

[7] Malkiel, Burton G. and Malkiel, Judith A. "Male-Female Pay Differentials in Professional Employment." American Economic Review, Vol. 63, No. 4 (September 1973), pp. 693-705. For illustrations of the demographic approach see Rees, Albert and Shultz, George P. Workers and Wages in an Urban Labor Market (Chicago: University of Chicago Press, 1970); and Johnson, George E. and Youmans, Kenwood C. "Union Relative Wage Effects by Age and Education." Industrial and Labor Relations Review, Vol. 24, No. 2 (January 1971), pp. 171-179.

[8] When translating cross-sectional age-earnings relationships into patterns of longitudinal behavior one must allow for general price change, technological change, and accumulation of tangible capital. A cross-sectional downturn may represent an increasingly severe retardation in growth of money or real pay, rather than an actual decline over the course of time.

[9] Such a cloud of suspicion also hovers over generally accepted variables, such as age, sex, education, and race. For one view in this regard, see Berg, Ivar. Education and Jobs: The Great Training Robbery (Boston: Beacon Press, 1971).

[10] For additional background information and some substantive findings see Hardin, Einar. "The Reactions of Employees to Office Automation." Monthly Labor Review, Vol. 83, No. 9 (September 1960), pp. 925-932.

[11] See Kamenske, Gloria Cheek. Some Personality Factors in Attitude toward Technological Change in a Medium Sized Insurance Company (unpublished doctoral dissertation, Department of Psychology, Michigan State University, 1965); and Trumbo, "Individual and Group Correlates," op. cit.

[12] See Hardin, "Job Satisfaction," op. cit., for a further discussion and for a distribution of item intercorrelations.

[13] By assuming 50 paid weeks per year and adjusting the 1957 salary level for a doubling of average weekly earnings (up 98.3%) and average hourly earnings (up 96.2%) of nonsupervisory workers in finance, insurance, and real estate from 1957 to 1973; see the Manpower Report of the President, April 1974 (Washington: U.S. Government Printing Office, 1974), p. 315.

METRIC MULTIDIMENSIONAL MAPPING OF CANDIDATES IN AN ISSUE SPACE

Melvin J. Hinich, Virginia Polytechnic Institute and State University
Lawrence Cahoon, Carnegie-Mellon University

INTRODUCTION

The spatial theory of electorial competition provided the basis for the development of the methodology presented in this paper. Spatial analysis seeks, under a variety of assumptions to ascertain the policies candidates should adopt. The essential assumption of spatial theory is that each candidate's strategy and each citizen's preference can be represented in an Euclidean space, and that each citizen's loss function can be represented as a metric in this space. For the individual citizen this space is developed over time as he views the changing political situation. The U.S. has a representative form of government in which the candidates are judged to a large extent on their positions on the issues. The election process generates uncertainty concerning the nature of the many issues involved, and concerning the candidates' position on these issues. The theory assumes that the citizens, in order to more readily deal with this situation, condense the issues into a space in which the dimensions represent their underlying preferences. These dimensions then do not necessarily represent the issues directly. Since the citizens are concerned primarily with the future activity of the candidates it is assumed that they use this space to predict the future positions of the candidates. Thus a "spatial picture" at any point in time is a plot of the predicted positions of the candidates relative to the citizens underlying preferences. When asked how he feels about a given candidate the citizen responds in terms of a loss function relative to this "spatial picture".

With this basis we use the feeling thermometer data collected in the University of Michigan Survey Research Center's 1968 Presidential Survey to map the candidate positions and the respondent's ideal points into a joint two dimensional space. The scaling technique is based on the assumption that the thermometer scores given by the respondents for each candidate is a monotonic function of weighted Euclidean distance. The positions of the candidates, the respondents' ideal points, the issue weights, axis orientation, and the mean ideal point of the population are all identified in the model used to develop the methodology, provided some constraints are satisfied.

THE MODEL

The metric which specifies the utility function of the individual respondents is basic to any metric scaling technique. We assume that this utility function is of the form

$$(1)\quad -U(\underline{\theta}_j,\underline{X}_i) = [(\underline{\theta}_j-\underline{X}_i)'A(\underline{\theta}_j-\underline{X}_i)]^{1/K}$$

when $\underline{\theta}_j = (\theta_{j1},\theta_{2j})'$ is the position of the jth candidate, $\underline{X}_i = (X_{i1},X_{i2})$ is the ith citizen's ideal point, $\underline{A}$ is a diagonal matrix of positive issue weights a_1, a_2, and K is a positive integer. We have constrained ourselves to a two-dimensional space since that is the dimensionality of the space in which the candidate positions will be estimated in the application of the methodology to the 1968 SRC survey data. When K=1(1) is the quadratic loss function common in statistics. We have found that the date suggests a value of K greater the 1 to be more appropriate.

Suppose that each citizen rates each of p+1 candidates on a thermometer scale that varies from 0 to 100. If 100 is the response given by an individual to a candidate whom he likes extremely well the thermometer score takes the form

$$(2)\quad T(\underline{X}_i,\underline{\theta}_j) = 100 - [(\underline{\theta}_j-\underline{X}_i)'A(\underline{\theta}_j-\underline{X}_i)]^{1/K}$$

For factor analysis to be useful it is necessary to eliminate the nonlinerity in expression (2).

Before proceeding to those modifications, let us first make clear a few assumptions which are necessary for the procedure to be applicable to spatial analysis. First, we assume that all citizens have the same perceptions of the candidate. Secondly, the weight a_1 and a_2 are independent of $\underline{X}$. These two assumptions possess a long history in spatial analysis. It must also be admitted that they are restrictive. In an ideal situation it is desirable to allow different perceptives on the part of different people. Rather than continue into an extensive discussion of these assumptions let us simply note that we must begin somewhere so we begin with these assumptions. Our third assumption is that, the covariance matrix, $\underline{\Sigma}$, is known, and is a diagonal matrix. This assumption may seem restrictive. It is not. The reason for making the assumption is that $\underline{A}$ and $\underline{\Sigma}$ are not jointly identifiable. In practice $\underline{A}$ and $\underline{\Sigma}$ jointly provide the same type of information about the space. That is, they both provide information on the relative importance the citizens give to the different dimensions. Since this is true, when $\underline{\Sigma}$ is not known it is sufficient to assume that $\underline{\Sigma}=\underline{I}$ and obtain the information about the relative importance of the dimensions from the estimate of $\underline{A}$ which we will obtain. In requiring that $\underline{\Sigma}$ be diagonal we are simply stating that the axis of the space which will be recovered will be that for which $\underline{\Sigma}$ is diagonal. This is possible since for a general covariance matrix $\underline{\Sigma}$ it is possible to rotate the space such that relative to the rotated set of axis the new covariance matrix $\underline{\Gamma'\Sigma\Gamma}$ is diagonal. $\underline{\Gamma}'X$ is the new set of observations and $\underline{\Gamma}$ is an orthogonal matrix. With these assumptions it is possible to proceed to the discussion of the methodology.

THE PROCEDURE

Subtracting equation (2) from 100 and allowing for an additive error in the individual observation of the candidates we then have a set of observations of the form,

$$(3)\quad D_{ij} = [(\underline{\theta}_j-\underline{X}_i)'\underline{A}(\underline{\theta}_j-\underline{X}_i)]^{1/K} + E_{ij}$$

the errors, E_{ij}, are assumed to be independent with unknown variance σ^2.

Briefly, the procedure uses the observations D_{ij} to construct a new set of observations which are linear in $\underline{X}_i$ and $\underline{\theta}_j$. These observations, when viewed as a set of observations on the set of candidates will satisfy the assumptions necessary for factor analysis when $K = 1$ or 2. For $K>2$ some modifications are necessary. Thus the covariance matrix of the new set of observations will be constructed. Factor analysis will be applied to that matrix. As usual with factor analysis the factored matrix will be a rotation of the factorization we desire. While factor analysis must in general be satisfied with an arbitrary rotation of the true matrix, we will use the additional information available to us about the observations to construct a regression to identify jointly the rotation and the axis weights. This will in turn allow us to estimate the positions of the candidates. Once that has been done we may then return to the set of observations which we constructed to obtain estimates of the individual's ideal points.

The Procedure was first developed relative to squared Euclidean distance. Thus to eliminate the non-linearities in (2) we first raise the observations to the k^{th} power to obtain,

(4) $D_{ij}^K = \underline{\theta}_j'\underline{A}\theta_j - 2\underline{\theta}_j'\underline{AX}_i \pm X_i'\underline{AX}_i + \delta_{ij}$

where δ_{ij} is generic for the error term. From (4) the construction of a set of observations linear in $\underline{X}_i$ and $\underline{\theta}_j$ is straightforward. First we note that the entire problem is invariant with respect to the origin of the underlying coordinate system. For that reason it is possible to simplify the mathematics of the problem by defining the origin to be located at the position of one of the candidates. Specifically let $\theta_{p+1}=0$. Thus for $j=p+1$ (4) becomes,

(5) $D_{ip+1}^K = \underline{X}_i'\underline{AX}_i + \delta_{ip+1}$

The set of observations desired is formed by subtracting (5) from (4) and then subtracting the mean $\overline{D}_j^K - \overline{D}_{p+1}^K$ to obtain,

(6) $Y_{ij} = D_{ij}^K - D_{ip+1}^K - (\overline{D}_j^K - \overline{D}_{p+1}^K) = 2\underline{\theta}_j'\underline{AX}_i + \delta_{ij}$.

Let $\underline{Y}_i = (Y_{i1},\ldots,Y_{ip})'$ and $\underline{Y}$ be the nxp matrix whose jth element is Y_{ij}. The sample covariance matrix of the $\underline{Y}_i$ is $\frac{1}{n-1}\underline{Y}'\underline{Y}$. It is this matrix which we wish to factor. Since $\frac{1}{n-1}Y'Y$ is the covariance matrix of $\underline{Y}_i$ it estimates

(7) $4\underline{\theta}'\underline{A}^2\underline{\theta} + \underline{C} + \underline{\psi}$

where $\underline{\theta} = (\underline{\theta}_1,\ldots,\underline{\theta}_p)$, $\underline{\psi}$ is the covariance matrix of the vector of errors $(\delta_{i1},\ldots,\delta_{ip})$ and C is the matrix of covariance between $2\underline{\theta}_j'\underline{AX}_i$ and δ_{ij}, $i,j = 1,\ldots p$. For $K = 1,2$ $\underline{C}$ is a matrix of zeros and $\underline{\psi} = \underline{1}'\underline{1}d+\underline{D}$ where $\underline{1}=(1,\ldots,1)'$, d is a constant and $\underline{D}$ is a diagonal matrix. Thus when $K=1$ or 2 $\frac{1}{n-1}\underline{Y}'\underline{Y}$ is factored, by the method of factor analysis we obtain an estimate

(8) $\underline{\Lambda} = (2\underline{\theta}'\underline{A},\underline{1})\underline{P}$

where $\underline{P}$ is a 3x3 orthogonal matrix.

If $K>2$ some modifications are necessary. From equation (6) it is possible to specify the exact form of $\underline{C}+\underline{\psi}$ in (7). If we ignore moments greater than two this can be written as $\underline{E}\sigma^2$. A typical term in the matrix $\underline{E}$ is

$\text{cov}\{(\underline{\theta}_j-\underline{X}_i)'\underline{A}(\underline{\theta}_i-\underline{X}_i),[(\underline{\theta}_j-\underline{X}_i)'\underline{A}(\underline{\theta}_j-\underline{X}_i)]^{\frac{K-2}{K}}\}$.

This can be estimated from the observations as $\text{cov}(D^K,D^{K-2})$. Once the matrix $\underline{E}$ is estimated we use the fact that $4\underline{\theta}'\underline{A\theta}$ is of rank two to estimate σ^2. That is we search over σ^2 to find the value of σ^2 such that the smallest p-2 eigenvalues of

(9) $\frac{1}{n-1}\underline{Y}'\underline{Y} - \underline{E}\,\sigma^2$

are near zero. Specifically we choose σ^2 such that the mean of the last p-2 eigenvalues of (9) is zero. Then (9) estimates $4\underline{\theta}'\underline{A}^2\underline{\theta}$. This matrix is then factored by the principle components version of factor analysis to obtain an estimate of $2\underline{\theta}'\underline{A\Gamma}$ where $\underline{\Gamma}$ is $\underline{a}$ 2x2 orthogonal matrix.

Thus for $K=1,2$ we have obtained an estimate of $(2\underline{\theta}'\underline{A},\underline{1})\underline{P}$ and for $K>2$ we estimate $2\underline{\theta}'\underline{A\Gamma}$. When $K=1,2$ a regression is constructed using the fact that we have a column of 1's in the matrix $2\underline{\theta}'\underline{A},\underline{1}$ to identify two of the three rotational parameters in $\underline{P}$. Once these have been identified the matrix $(2\underline{\theta}'\underline{A},\underline{1})\underline{P}$ is rotated to form an estimate of $(2\underline{\theta}'\underline{A\Gamma},\underline{1})$.

The estimate of $\underline{\theta}'\underline{A\Gamma}$ is next used together with the means $\overline{D}_j^K-\overline{D}_{p+1}^K$ to obtain estimates of $\underline{\Gamma}$ and $\underline{A}$. We begin by noticing that

(10) $\overline{D}_j^K - \overline{D}_{p+1}^K = \underline{\theta}_j'\underline{A\theta}_j - 2\underline{\theta}_j'\underline{A\overline{X}} + \delta_{ij}$

If $K>2$ a minor correction is made to (10) to account for the bias introduced into the error term by raising the observations to the K^{th} power. Post multiplying our estimate of $\underline{\theta}'\underline{A\Gamma}$ by the unknown $\underline{\Gamma}'\underline{A}^{-1/2}$ and forming the inner product $(\underline{\theta}'\underline{A\Gamma\Gamma}'\underline{A}^{-1/2})(\underline{\theta}'\underline{A\Gamma\Gamma}'\underline{A}^{-1/2})'$ we obtain an estimate of $\underline{\theta}_j'\underline{A\theta}_j$. Post-multiplying the estimate of $\underline{\theta}_j'\underline{A\Gamma}$ by the unknown $\underline{\Gamma}'\underline{\overline{X}}$ we obtain an estimate of $\underline{\theta}_j'\underline{A\overline{X}}$ the sum of these two provides an estimate of $\overline{D}_j^K - \overline{D}_{p+1}^K$. Expanding this expression we have the regression model

(11) $\overline{D}_j^K-\overline{D}_{p+1}^K = \alpha_0\hat{M}_{j1}^2 + \alpha_1\hat{M}_{j2}^2 + \alpha_2\hat{M}_{j1}\hat{M}_{j2}$
$\quad + \alpha_3\hat{M}_{j1} + \alpha_4\hat{M}_{j2} + \delta_{ij}$

where our estimate of $\underline{\theta}'\underline{A\Gamma}$ is $\hat{M}=(\hat{M}_{ij})$. The coefficients $\alpha_0,\ldots,\alpha_4$ are nonlinear functions of $a_1,a_2,\overline{X}_1,\overline{X}_2$, and the rotation parameter in the matrix $\underline{\Gamma}$. $\underline{\overline{X}} = (\overline{X}_1,\overline{X}_2) = n^{-1}\Sigma_{i=1}^n \underline{X}_i$

$\alpha_0,\ldots,\alpha_4$ are then used to solve the system of equations for $a_1,a_2,\overline{X}_1,\overline{X}_2$ and the rotation parameter. It then becomes a single matter to use these estimates together with the estimates of $\underline{\theta}'\underline{A\Gamma}$ to estimate $\underline{\theta}$, the matrix of candidate positions.

The respondents' ideal points can also be estimated. This is done by using the estimates $\underline{\theta}'\underline{A\Gamma}$, and Γ together with the set of observations $Y_{ij} = 2\underline{\theta}_j\underline{AX}_i + \delta_{ij}$ to form a regression to identify the $\underline{X}_i$.

THE DATA

The procedure has been applied to the 1968 SRC thermeter data. Prior to a discussion of the results of that analysis several comments are necessary. We have assumed that all citizens have the same perceptions of the candidates. To provide for a greater certainty in the truth of this assumption the population was divided into three groups: Democrats, Independents, and Republicans. A respondent who leaned towards the Democrats or to the Republicans while belonging to neither group was nontheless included with that group, as well as with the Independents. Thus there is some overlap in the three groups. Also, while the survey question did provide for an answer in case a respondent did not know a candidate, it was also possible to assume a score of 50 could be given to a candidate whom the respondent was unfamiliar with. A quick review of the actual responses given indicates that this did in fact occur. To eliminate distortions caused by this any respondent who gave more than three of the eleven candidates a score of 50 was eliminated from the sample.

For the utility function $U(\theta_j,X_i)$ a value of K equal to 4 was chosen. Such a value provides for a myopic view of the candidates. That this is the case is evident from the data. This effect also shows up in the plots of the candidates positions. The Republicans (Fig. 3) distinguish more between Agnew and Reagon than do the Democrats (Fig. 1).

Figures 1, 2, and 3 show the positions of the candidates as estimated by the procedure. The distances in these plots have been changed to Euclidean distances for ease of interpretation. Wallace and LeMay were eliminated from the study because, after a review of the data, it was felt that many respondents wished to score them beyond the permissable range of the thermometer scale. Also included with these plots are the information concerning the relative axis weights, error variance, predicted and true votes, and the R^2 for the regression used to estimate the rotation and axis weights. a_1 is the weight on the horizontal dimension, a_2 the weight on the vertical dimension, a_3 represents an efficiency weight. That is, it was felt that certain candidates were considered by the citizen's to be clear possibilities as presidential candidate, the others were not, thus the citizens would tend to like (or dislike) those who were more likely to be president at some future date. This weight was specific to Nixon, Kennedy, Johnson and Humphry. Thus these four candidates are viewed as being at $(\underline{\theta}_j,0)$, while the remainder are at $(\theta_j,1)$. The ideal points are at $(\underline{X}_i,0)$. This effect can be considered by a simple modification of the regression. It cannot be treated in the covariance matrix of the $\underline{Y}_i$ because the ideal points do not vary over that dimension. That the weight a_3 is positive in all cases indicates a likeing for Nixon, Kennedy, Johnson, and Humphry which is distinct from how the citizens see the other candidate.

The line used to predict the votes for each candidate is not the bisector of the line joining Nixon and Humphry, but is a line parallel to the bisector. Moving the line was considered justified because the thermometer data was gathered shortly after the election, thus creating some possible post election bias. The line chosen is that which maximizes the sum of the percentages of correctly predicted votes for Nixon and Humphry.

One rather interesting result became evident when the ideal points of the citizens who claimed to support Nixon and Humphry but abstained from voting were plotted (see Fig. 4). The plot indicates that the reason for abstention was not alienation, but that these citizens were on the dividing line between voting for Nixon or Humphry.

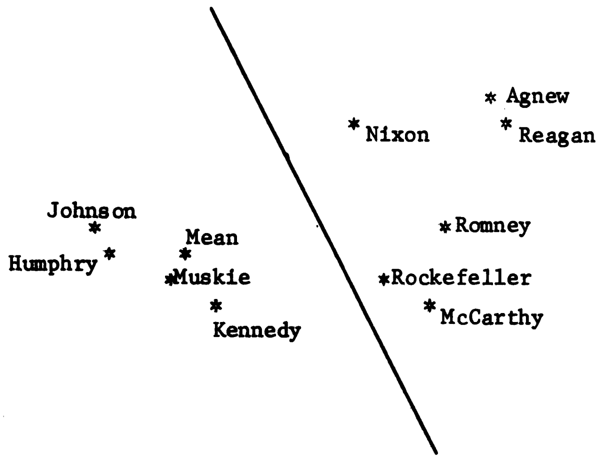
The interpretation of the meaning of the dimensions must be made from the positions of the candidates or from external information. In each of the three groupings it seems that the horizonal dimension is a party dimension. The vertical dimension is more difficult to interpret. Considering Kennedy and McCarthy's anti Viet Nam war stands, and Kennedy's position on the race issue, the vertical dimension could be regarded as a liberal-conservative social action dimension. One hesitates to use terms such as liberal and conservative here because of the many possible meanings these words may have in the political sphere, however it is felt that their use is clear in this instance. The position of Reagan relative to Nixon and Agnew in Fig. 3, (the Republican respondents), is consistant with such an interpretation, as is the position of Nixon relative to the democrats as an entire group. As mentioned before the Democrats, and Independents as well, have some difficulty separating Agnew and Reagan, thus since Agnew was Nixon's running-mate, their positions are determined, in the eyes of the voters, by Nixon's position.

The horizontal dimension was conjectured to be a party dimension. The extreme position of Johnson, and Nixon are factors indicating the truth of that conjecture. It might be asked, then, why are Agnew and Reagan to the right of Nixon, since this would indicate that they are more loyal to the Republican party than Nixon is. This fact is evident in each of the three groupings. It must be remembered that in 1968 Agnew was an unknown political figure, and for the Democrats this seems to be true of Reagan as well. With Nixon being the presidential candidate, the voters then would be very likely to place Nixon closer to their mean position than they would Agnew or Reagan.

Conclusions

We have given a method by which the positions of candidates may be estimated in an Euclidean space. This method has then been applied to real data from the 1968 presidential election. The principle objective here is not simply to secure by some means a spatial map of the candidates that satisfies some intuitive criteria. Rather it is to develop a multidimensional scaling procedure based on spatial theory's assumptions. In this way we hope to link theory to the empirical world.

Fig. 1 Predicted Position of Candidate Using Democratic Respondents



Estimated Error Variance $\hat{\sigma}^2 = 154.7$
Ratio of Axis Weights
$\hat{a}_2/\hat{a}_1 = 5.43$
$\hat{a}_3/\hat{a}_1 = 0.39$
Regression $R^2 = 0.928$

Predicted Vote Percentages
Nixon..........22.6%
Humphry........77.4%
True Vote Percentages
Nixon..........21.6%
Humphry........78.4%
Percentage of Votes Correctly
Predicted..........82.0%

Fig. 2 Predicted Position of Candidate Using Independent Respondents.

Estimated Error Variance $\hat{\sigma}^2 = 164.4$
Ratio of Axis Weights.
$\hat{a}_2/a_1 = 2.44$
$\hat{a}_3/\hat{a}_1 = 1.12$
Regression $R^2 = 0.744$

Predicted Vote Percentages
Nixon..........58.3%
Humphry........41.7%
True Vote Percentages
Nixon..........62.5%
Humphry........37.5%
Percentage of Votes Correctly
Predicted..........83.7%

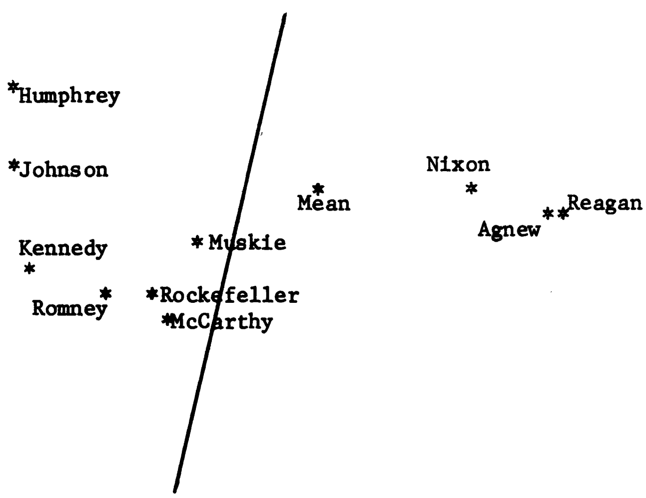
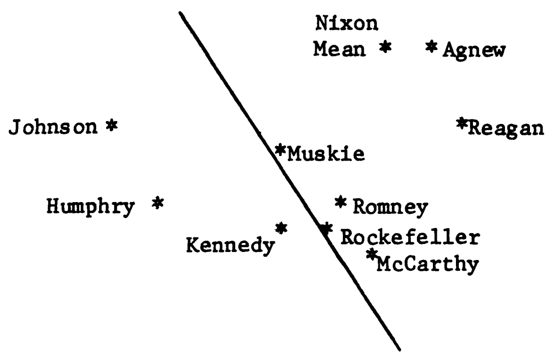


Fig. 3 Predicted Position of Candidates Using Republican Respondents



Estimated Error Variance $\hat{\sigma}^2 = 171.5$

Ratio Axis Weights

$\hat{a}_2/\hat{a}_1 = 1.74$

$\hat{a}_3/\hat{a}_1 = 0.79$

Regression $R^2 = 0.963$

Predicted Vote Percentages

Nixon..........90.4%

Humphry........ 9.6%

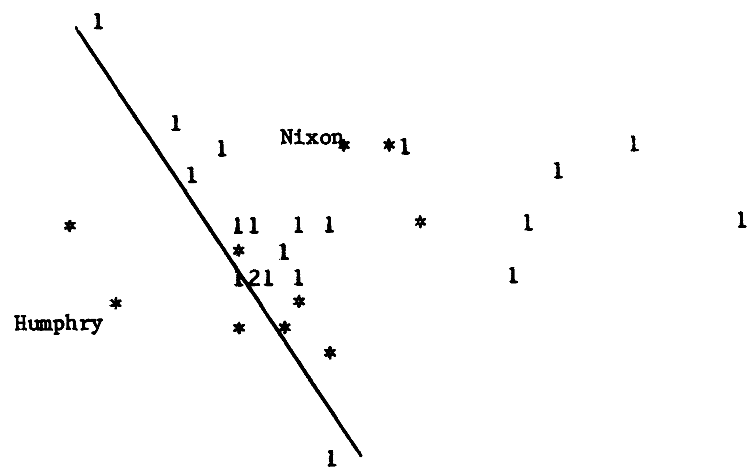
True Vote Percentages

Nixon..........92.6%

Humphry........ 7.4%

Percentage of Votes Correctly Predicted.........93.0%

Fig. 4 Republican Candidate and Position of Respondents who supported Nixon or Humphry but did not vote in the election. The numbers indicate the number of Respondents at any one position. 21 respondents are plotted, 5 more are located in the same position as the candidates, and 1 person is outside the area of the plot. Note the tendency for these respondents to be near the line separating Nixon Voters from Humphry Voters. (Candidates positions are the same as in Fig. 3 above.)



REFERENCES

Anderson, T.W., (1950) The Asymptotic Distribution of Certain Characteristic Roots and Vectors. Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley and Los Angeles, 103-130.

Anderson, T.W., and Rubin, Herman, (1956). Statistical Inference in Factor Analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V, University of California Press, Berkeley and Los Angeles, 111-150.

Anderson, T.W., (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Anderson, T.W., (1963). Asymptotic Theory for Principle Components Analysis. Ann. Math. Statist. 34 122-148.

Clarke, M.R.B., (1970). A Rapidly Convergent Method For Maximum-Likelihood Factor Analysis. British Journal of Math. Stat. Psychology. 25 43-52.

Daalder, Hans, and Rusk, Jerrold G. (1972). Perceptions of Party in the Dutch Parliaments. In S.C. Patterson and J.D. Wahlke, eds., Comparative Legislative Behavior. Wiley, New York.

Draper, N.R., Smith, H. (1966). Applied Regression Analysis. Wiley, New York.

Harman, (1967). Modern Factor Analysis. Univ. of Chicago Press. Chicago.

Johnston, J. (1960). Econometric Methods. McGraw Hill. St. Louis.

Jöreskog, K.G., (1967). Some Contributions to Maximum Likelihood Factor Analysis. Psychometrica. 32 443-482.

Karns, David A., (1974). Legislative Roll Call Analysis: The Individual Differences Scaling Approach. Presented at Midwest Political Science Meetings, Chicago 1974.

Lawley, D.N., (1967). Some New Results in Maximum-Likelihood Factor Analysis. Journal of the Royal Soc. of Edinburgh. 67 256-264.

Lawley, D.N., and Maxwell, A.E., (1971). Factor Analysis As A Statistical Method. 2nd ed. Butterworth, London.

Morrison, Donald F., (1967). Multivariate Statistical Methods. McGraw-Hill, New York.

Morrison, Richard, (1972). A Statistical Model of Legislative Roll Call Analysis. Journal of Mathematical Sociology 2 235-48.

Page, Benjamin I., and Brody, Richard A. (1973). Indifference Alienation and Rational Decisions: The Effects of Candidate Evaluations on Turnout and Vote. Public Choice.

Page Benjamin I. (1974). Presidential Campaigning, Party Cleavages, and Responsible Parties. Working Paper, Department of Political Science, University of Chicago.

Rae, Douglas, and Taylor, Michael, (1971). Decision Rules and Policy Outcomes. British Journal of Political Science. 1 71-90.

Rao, C.R., (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

Riker, William H., and Ordeshook, Peter C., (1973). An Introduction to Positive Political Theory. Prentice-Hall, Englewood Cliffs.

Rabinowitz, George B. (1973). Spatial Models of Electoral Choice: an Empirical Analysis. University of North Carolina. Chapel Hill.

Rosenthal, Howard, and Sen, Subrata, (1973). Electoral Participation in the French Fifth Republic. American Political Science Review. 67 29-54.

Ross, John, and Cliff, Norman, (1964) A Generalization of the Interpoint Distance Model. Psychometrica. 29 167-176.

Rusk, Jerrold G., and Weisberg, Herbert, (1970) Dimensions of Candidate Evaluation. American Political Science Review. 64 1167-85.

Rusk, Jerrold G., and Weisberg, Herbert. (1972) Perceptions of Presidential Candidates. Midwest Journal of Political Science. 16 388-410.

Shepard, Roger N., Romney, A. Kimbal, and Nerlov, Sara Beth, eds., (1972). Multidimensional Scaling. Seminar Press. New York.

Wendell, R.E., and Thorsen, S.J., (1973). Some Generalizations of Social Decisions Under Majority Rule. Presented at Midwest Political Science Meetings, Chicago 1973.

AN EMPIRICAL ANALYSIS OF INTERSTATE DIFFERENTIALS IN THE RATE OF FERTILITY DECLINE, 1970 - 1971

Timothy D. Hogan, Arizona State University*

I

Birth rates have declined to record low levels in the United States. Between 1970 and 1971, the last year for which final statistics on births are available, the total number of live births in the United States declined five percent, and the U.S. birth rate declined seven percent between 1970 and 1971 (22, p. 1). Preliminary data for 1972 and 1973 indicate that the total number of births declined 8.5 percent between 1971 and 1972 and fell an additional 3.7 percent between 1972 and 1973. The average birth rate for 1972 was indicated to have been 9.4 percent lower than 1971, and the 1973 birth rate to have been 4.3 percent below the average rate for 1972 (20).

The so-called "birth-death" phenomenon that has occurred since 1970 has been nationwide in scope. Provisional data for 1973 (by place of occurrence) show that the birth rate declined in all states but Georgia (20). But while this general pattern of fertility decline has been nationwide, the relative magnitudes of these declines and the levels of fertility themselves among the states have been subject to considerable variation -- for example, annual declines between 1970 and 1971 ranged from over ten percent in crude birth rates in states such as Connecticut, New York, Washington, and California to a slight increase in South Carolina.

Given this pattern of variation, this paper presents preliminary results from empirical analysis of the factors associated with interstate variations in the percentage decline in the numbers of births and in the birth rate between 1970 and 1971.

II

Previous studies have examined the determinants of differential fertility in the United States utilizing cross-section data (3,7,8,9,10, 24), but have concentrated on explanation of variation in the levels of fertility. Analyses have also examined the factors associated with U.S. fertility trends, but these have generally investigated the trends in national fertility measures rather than differential regional trends. (1,6)

We will formulate the following linear stochastic model to test the association of a combination of socioeconomic and demographic characteristics of states with (a) the percentage decline in births and (b) the percentage decline in birth rates during the 1970-1971 period:

$$F = F(Y,U,B,R,L,M,Ed,A,P,I,Ec)$$

where

F refers alternatively to (a) the percentage decline in the number of live births and (b) the percentage decline in the crude birth rate between 1970 and 1971.

Y refers to the median income of families in 1969. The relationship between income and fertility has been studied extensively.[1] Typically, analysis has been in terms of association between family size and income or wealth levels. In our context, we will investigate the relationship between differences in the magnitude of the fertility decline and the relative income levels of families among states.

U refers to the 1970 unemployment rate. This additional economic variable is included as a measure of the influence of short run fluctuations in economic conditions upon fertility decisions -- particularly upon the timing of births.[2]

B refers to the proportion of the total 1970 Census population of the state that was Negro. Previous analyses have identified race as a significant determinant of differential fertility (2,15). Additionally, black attitudes toward fertility control and population growth have been the subject of considerable recent discussion (23), so that this variable has been included to study the significance of race as a factor in recent declines in fertility.

R refers to the proportion classified as rural residents of the total 1970 Census population. Previous studies of urban-rural differentials show substantially higher fertility for rural areas, even when education, income and other socioeconomic factors are taken into account (2,5,8,15).

L refers to the female labor force participation rate. Theoretical analyses have identified the opportunity cost of the mother's time as a very important factor in the family's fertility decisions (4,24); and a significant negative relationship between female labor force participation and fertility levels has generally also been empirically identified (6,7,9,24).

M refers to the proportion of women over 14 in 1970 who were married. Since the fertility of married women is much higher than that of unmarried women, we would expect the proportion married of the female population of child-bearing age to be a significant factor determining differences in fertility rates among states.

Ed refers to the median number of school years completed by women 25 years and over in 1970. A measure of female educational attainment has been included in our analysis because previous research has identified a strong inverse relationship between education and fertility (2,8,10,15). Additionally, a positive relationship between educational attainment and the rate of fertility decline might be hypothesized on the basis of a positive relationship between education and the awareness of new economic and social trends and/or changes in contraceptive technology that would effect fertility.

A refers to the median age of the female population in 1970. Ceteris paribus, the age composition of the female population in the child-bearing ages has important effects upon aggregate fertility because age specific fertility rates vary tremendously among different age groups. Further, the rate of fertility declines has been much higher for older women than those in the prime child-bearing ages. Inclusion of the median age of the female population as an explanatory variable provides an indirect measure of differences in the age composition of the female population.[3]

P refers to the percentage increase in the population aged 18-44 between 1970 and 1971. Even if age specific fertility rates remained constant, the number of births and the birth rate would be expected to change if the size of the child-bearing population varied during the time period; this age specific population growth measure has been included to investigate the significance of this factor.

I refers to the percentage change in the infant mortality rate between 1969 and 1970. The infant mortality rate is generally considered a sensitive measure of the general level of health and medical care (2); it has been included in the analysis as a measure of the importance of such health factors upon the rate of fertility decline.[4] Alternatively, the inclusion of this variable might be considered a test of a theory of fertility decline that has received considerable study -- the direct impact of the decline in infant mortality upon the number of births per woman (7,13), in the context of the recent U.S. experience.

Ec refers to the number of members of the conservationist society, the Friends of the Earth, per 1000 population. This variable might be interpreted as a proxy of the degree of the environmental concern among the states' residents.

Further definitional details and data sources are presented in the appendix for all variables.

It should be noted that a priori hypotheses concerning the direction of the relationship between most of the independent variables and the measures of fertility decline were not stated. Previous analyses have analysed the relationship between income, education, age structure, female labor force participation, race, urban-rural residence and fertility levels; but little prior evidence exists regarding the association between the socioeconomic variables and the rates of fertility decline.

Theory may imply a positive relationship between fertility levels and a particular independent variable, and the theory may also be supported by empirical results indicating the existence of such a positive association. If a linear model is assumed, it is also possible to extend the theory to imply a positive relationship between changes in fertility and variations in the particular independent variable. Unfortunately, neither variant of such a theory concerning the relation between fertility - independent variable provides much insight into the expected relationship between the rate of change in the dependent variable and the level of the independent variable. Therefore, the test of significance employed in the analysis was a two-tailed t test -- that is, no assumption was made about the direction of the specified relationship between the rate of fertility decline and most of the independent variables prior to analysis.

III

Ordinary least squares regression analysis was utilized to estimate the alternate versions of the model described in the previous section. Table 1 presents separate regression results with the percentage decline in births and the percentage decline in birth rates alternatively employed as the dependent variable in the multivariate model. A constant term was computed for each regression but not reported. The table reports beta coefficients for each variable, rather than standard regression coefficients to allow direct comparison of the relative impact of each factor upon the rate of fertility decline.[5] The estimated t value for each coefficient is also reported in parentheses below the respective coefficient, and those coefficients which are indicated to be statistically significant at the ten percent, five percent, and one percent levels of significance are designated by the superscripts "a," "b," and "c" respectively.

Of the eleven explanatory variables included in the analysis, the regression results show seven to be statistically significant at the ten percent level of confidence or above (when two-tailed test is applied) in the aggregate birth equation, while six are indicated to be statistically significant in the birth rate equation.

Interpreting the beta coefficients as measures of the relative impact of each significant explanatory variable upon the rate of fertility decline, the results show the median income measure to have been the most important variable in explaining the decline. Differences in urban-rural residence of the states' populations were indicated to be the second most important factor for both fertility measures, and the racial composition was shown to be the third most important factor. The proportion of the female population married and the measure of decline in the infant mortality were also found to be significant variables for both measure of the rate of fertility decline. Declines in aggregate births were also significantly correlated (positively) with female median age and with the variable measuring growth of the population in the child-bearing ages (inversely); these factors were found to have the same signs in the birth rate equation, but were not indicated to be statistically significant. The measure of environmental concern was shown to be positively related to both measures of the rate of fertility decline, but significantly correlated only in the case of the birth rate equation. Differences in educational attainment, female labor force participation, and unemployment rates were not indicated to be significant in either equation.

One interesting aspect of these results is the relatively high proportion of the total variance in the fertility measures explained by the combination of eleven independent variables included in the regression equations. The R^2's of .75 in both equations compare very favorably with cross-section analyses of U.S. differential fertility -- for example, Heer and Boynton (7) were able to explain only 35 percent of the total variance in birth rates among a sample of U.S. counties, with a model including seven explanatory variables.

The positive impact of median income levels was identified as the most important relationship by the regression results. These results could be interpreted to support several alternative hypotheses. One possible interpretation would be that children are an inferior good, so that higher income levels are associated with low levels of fertility. Alternately, the observed positive relation between income levels and the rate of fertility decline may be indirect evidence of the relationship between the opportunity cost of wives' time and fertility, since families' income levels and the value of time would be expected to be positively related. Finally, even after standardization for educational, racial, and residential differences, there may be a positive relationship between income levels and adoption of new contraceptive technology, so that the elimination of unwanted fertility among higher income individuals has become increasingly successful compared with the experience of the low income population.

The regression results also indicated that the rate of fertility decline was significantly larger in states' with relatively large rural populations. This observed relationship supports other evidence that rural fertility has been declining more rapidly than urban fertility -- that is, the differential between rural and urban fertility levels has been narrowing over time (8,15).

Similarly, the positive relationship indicated between the proportion of the population which was Negro and the rate of fertility decline can also be interpreted as the results of the lessening of the gap between higher Negro and lower white fertility rates. Alternatively, the observed relation might also be explained in terms of the effects of underlying differences in socioeconomic variables that were not adequately measured by the present aggregate analysis. For example, Negro income levels could have risen faster than white income levels, so that those states with relatively high proportion of Negroes would have larger fertility declines.

The regression results with respect to the effects of improvements in infant mortality can also be interpreted in at least two ways. If the infant mortality measure is assumed to be a proxy for improvements in health care (perhaps including the provision of abortion and family planning programs), then the positive relation discovered could be interpreted in terms of the positive influence of these health-related factors upon the rate of fertility decline. On the other hand, it is also possible to explain this evidence in terms of a direct relationship between changes in infant mortality and changes in fertility. Previous analyses have usually examined the infant survival factor in the context of developing societies, but Heer and Boyton (7) found a significant and positive relationship between infant mortality rates and birth rates for U.S. counties. Our evidence could also be interpreted as confirming this relationship in the context of the recent U.S. experience.

In the case of the birth rate equation, the significant positive relationship between the environmentalist measure and birth rate declines support the hypothesis that one of the important factors leading to the unexpectedly rapid decline in fertility of the 1970's has been the increasing concern with the negative impact of population growth upon the quality of life. A recent national opinion poll indicated that two-thirds of the respondents regarded U.S. population growth as a serious problem; and data collected in the 1970 National Fertility Study indicated that, among married women under 30 with less than two children, those concerned with population growth intend to have only half as many additional children as those not concerned (11).

For those two demographic variables for which a priori hypotheses concerning the nature of their impacts upon fertility decline were obvious -- the proportion of the female population married and the growth rate of the population of child-bearing ages during the 1970-1971 period -- the theoretically expected negative relationship was, in fact, observed in both regression equations.

Finally, the results did not demonstrate significant relationships, when the effects of the other factors are taken into account, between fertility declines and female educational attainment, female labor force participation, or the unemployment rate. The lack of significant relationship between the rate of fertility decline and female educational levels is somewhat surprising, given the positive associations that have been observed between educational attainment and contraceptive use and attitudes relating to abortion and fertility control (2,12). Similarly, substantial evidence exists of a close association between fertility and labor force participation by women, so a lack of evidence reaffirming such a relationship with respect to the rate of fertility decline was also somewhat unexpected. Failure to observe a significant relation between the rate of fertility decline and the unemployment rate was not as surprising, given the controversy concerning the importance of business cycle effects upon fertility rates (16). These results may be due, however, to shortcomings in the actual variables utilized in the empirical analysis as measures of the theoretical relationships under study. The existence of multicollinearity among the independent variables may also have resulted in the estimated coefficients of these variables to be nonsignificant.

IV

In summary, these results have emphasized some interesting aspects of the recent fertility experience. The evidence demonstrates that a high proportion of the differentials in the rate of fertility decline among the states can be explained on the basis of a combination of several socioeconomic and demographic characteristics of the states' populations. These results may also be interpreted as indirect evidence of the role of certain factors in the unexpectedly rapid decline in fertility experienced by the United States during the early 1970's. The regression analysis indicates that (1) ceteris paribus, the downward trend was most pronounced in states with relatively high median income levels; (2) a process of fertility differentials between rural and urban residents and between Negroes and Whites narrowing during the period also contributed to the fertility decline; and (3) empirical support exists for the hypothesis that growing the environmentalist sentiment had an impact upon fertility during the recent period.

FOOTNOTES

[1]For summaries of previous studies refer to Easterlin (4) or Simon (14).

[2]The relationship between the business cycle and short-run fluctuations in birth rates has received considerable analysis (1,8,16). Easterlin has also related changes in economic conditions to longer-run trends in fertility (4).

[3]The median age is an imperfect measure of such age compositional effects, since it is many different age compositions could have the same median age, but it is a single, available proxy measure which does provide some indication of age structure difference.

[4]The level of infant mortality rates in 1970 was also employed as an alternate variable, but was found not to be correlated with fertility declines.

[5]A "beta coefficient" is equal to the estimated coefficient times the ratio of the standard deviation of the independent variable divided by the standard deviation of the dependent variable.

*Associate Professor of Economics and Research Associate, Bureau of Business and Economic Research, College of Business Administration. The author would like to acknowledge support from the Faculty Research Grant Program at Arizona State University during the summer of 1974 which allowed time to pursue this research.

REFERENCES

1. G. Becker, "An Economic Analysis of Fertility," in National Bureau of Economic Research, Demographic and Economic Changes in Developed Countries, Princeton, N.J., 1960, pp. 209-240.

2. D.J. Bogue, Principles of Demography, New York, 1969.

3. P.N. De Tray, "Child Quality and the Demand for Children," Journal of Political Economy, 81 (March-April, 1973), Part II, S70-S95.

4. R.A. Easterlin, "Towards a Socio-Economic Theory of Fertility: A Survey of Recent Research on Economic Factors in American Fertility," in Fertility and Family Planning: A World View, S.J. Behrman, et. al., eds., Ann Arbor, Michigan, 1969.

5. B. Gardner, "Economic Aspects of the Fertility of Rural-Farm and Urban Women," Southern Economic Journal, 38 (April, 1972), 518-524.

6. P.R. Gregory, J.M. Campbell and B.S. Chang, "A Simultaneous Model of Birth Rates in the United States," Review of Economics and Statistics, 54 (November, 1972), 374-380.

7. D.M. Heer and J.W. Boynton, "A Multi-variate Regression Analysis of Differences in Fertility of United States Counties," Social Biology, 17 (September, 1970), 180-194.

8. C.V. Kiser, W.H. Grabill, and A.A. Campbell, Trends and Variations in Fertility in the United States, Cambridge, Mass., 1968.

9. R.T. Michael, "Dimensions of Household Fertility: An Economic Analysis," 1971 Proceedings of the American Statistical Association, Social Statistics Sections, Washington, D.C., 1971, pp. 126-136.

10. R.T. Michael, "Education and the Derived Demand for Children," Journal of Political Economy, 81 (March-April, 1973), Part II, S128-S164.

11. R.R. Rindfuss, "Recent Trends in Population Attitudes," in Aspects of Population Growth Policy, R. Parke, Jr. and C.F. Westoff, eds., Research Reports of the Commission on Population Growth and the American Future, Volume 6, Washington, 1973, pp. 17-34.

12. N.B. Ryder, "Recent Trends and Group Differences in Fertility," in Toward The End of Growth, C.F. Westoff, et. al., eds., Englewood Cliffs, N.J., 1973, pp. 57-68.

13. T.P. Schultz, "An Economic Model of Family Planning and Fertility," Journal of Political Economy, 77 (March-April, 1969), 153-180.

14. J.L. Simon, The Effects of Income on Fertility, Monograph No. 19, Carolina Population Center, University of North Carolina, 1974.

15. T.L. Smith and P.E. Zopf, Jr., Demography: Principles and Methods, Philadelphia, 1970.

16. A. Sweezy, "The Economic Explanation of Fertility Changes in the United States," Population Studies, 25 (July, 1971), 255-267.

17. U.S. Bureau of the Census, Census of the Population: 1970, Volume I, Washington, D.C., 1972.

18. U.S. Bureau of the Census," Estimated Population of States by Age, July 1, 1971 and 1972," Current Population Reports, Series P-25, No. 500, May, 1973.

19. U.S. Bureau of the Census, Statistical Abstract of the United States: 1973, Washington, D.C., 1973.

20. U.S. Public Health Service, National Center for Health Statistics, "Provisional Statistics: Annual Summary for the United States, 1973," Monthly Vital Statistics Report, 22, No. 13 (June 27, 1974).

21. U.S. Public Health Service, National Center for Health Statistics, "Summary Report: Final Mortality Statistics, 1970," Monthly Vital Statistics Report, 22, No. 11 (February, 22, 1974).

22. U.S. Public Health Service, National Center for Health Statistics, "Summary Report: Final Natality Statistics, 1971," Monthly Vital Statistics Report, 23, No. 3 (June 7, 1974), Supplement.

23. R.G. Weisbrod, "Birth Control and the Black American: A Matter of Genocide," Demography, 10 (November, 1973), 571-590.

24. R.J. Willis, "A New Approach to the Economic Theory of Fertility Behavior," Journal of Political Economy, 81 (March-April, 1973), Part II, S14-S64.

DATA APPENDIX

The percentage declines in births and in birth rates between 1970 and 1971 were computed from data on live births and birth rates (births per 1000 population) recorded by place of residence (22). Unfortunately, final fertility data for the years after 1971 has not yet been published.

Data series for 1969 median family income, average 1970 unemployment rate for persons 16 and over, and the proportion of the total 1970 Census population that was Negro were compiled from the 1973 Statistical Abstract of the United States (19).

Information by state on the proportion classified as rural-farm of the total 1970 population, the proportion of females 16 and over in the labor force, the proportion of females 14 and over married, the median number of school years completed by women 25 and over, and the median age of the female population was collected from the 1970 Census of Population (17).

The percentage changes in the population 18-44 between 1970 and 1971 for each state were computed from age specific population data (18).

The percentage declines in infant mortality rates (the number of deaths to infants less than one per 1000 live births) between 1970 and 1971 by state were computed from data on infant mortality in 1970 and 1971 (21) and on the number of live births reported in (22).

The number of members of Friends of the Earth per 1000 population in each state was computed from data on FOE membership by state in 1972 reported in a private letter from the national office of the organization and population estimates of state populations as of July 1, 1972 presented in (18). It would have been preferable to have such membership rates for the 1970-71 period but membership data for years before 1972 were not available.

TABLE I

REGRESSION RESULTS: PERCENTAGE DECLINES IN BIRTHS AND IN BIRTH RATES AS FUNCTIONS OF SELECTED SOCIOECONOMIC AND DEMOGRAPHIC VARIABLES, FOR THE FIFTY STATES, 1970 - 1971.

Independent Variables	Dependent Variable: Births (Beta Coefficient)	Percentage Decline in Births/1000 Population (Beta Coefficient)
Y	.6449[c] (3.30)	.7367[c] (3.79)
U	.1540 (1.21)	.1421 (1.12)
B	.3274[b] (2.33)	.3167[b] (2.27)
R	.4527[c] (3.76)	.3845[c] (3.20)
L	-.0041 (0.03)	-.0143 (0.11)
M	-.2744[c] (2.52)	-.2005[a] (1.85)
Ed	.1260 (0.77)	.1346 (0.83)
A	.2148[a] (1.98)	.1660 (1.53)
P	-.2883[c] (2.53)	-.1359 (1.20)
I	.2451[c] (2.38)	.2035 (1.98)[a]
Ec	.1885 (1.46)	.2653[b] (2.07)
R^2	.752	.753

THREE MULTIPLE COMPARISON PROCEDURES FOR CONTRASTS AMONG CORRELATION COEFFICIENTS

Bradley E. Huitema, Western Michigan University

Researchers occasionally encounter situations in which it is appropriate to test the hypothesis $H_0: \rho_1 = \rho_2 = \ldots = \rho_J$ (where J is the number of independent groups and ρ is the population correlation coefficient) against the alternative hypothesis that H_0 is false. A fairly well known chi square test of this hypothesis is presented in several texts (e.g., Rao, 1965; Snedecor and Cochran, 1967). The problem of multiple comparisons among J coefficients has not, however, been given much attention. The only procedure formally described for this purpose appears to be one presented by Marascuilo (1966, 1971). Marascuilo's procedure is essentially the correlation coefficient analog to Scheffe's widely known procedure for multiple comparisons among means. The purpose of this paper is to (a) describe three alternate multiple comparison methods and to (b) discuss the relative merits of these procedures.

The basic advantage of the Marascuilo procedure is that it can be applied *a posteriori* to any number or type of contrasts of interest. This is a very impressive characteristic because the total possible number of contrasts among three or more groups is infinite. As might be expected, a disadvantage associated with this flexibility is low precision. If the contrasts of interest are planned and the number is not large (a table for defining "large" is presented later) the procedures described in the remainder of this paper are more powerful.

Method I

The most frequently encountered multiple comparison situation appears to be the one in which all pairwise (simple) contrasts are of interest. A powerful test in this situation is an analog to the Protected LSD test on means.

Method I is a two stage procedure:

Stage 1: Perform the overall chi square test at level α. If this test is nonsignificant retain the hypothesis that $\rho_1 = \rho_2 = \ldots = \rho_J$. If this test is significant proceed to stage 2.

Stage 2: Compute tests on all pairwise contrasts using the conventional formula for testing the equality of two population correlation coefficients i.e.,

$$\frac{Z_i - Z_j}{\sqrt{\frac{1}{n_i-3} + \frac{1}{n_j-3}}} = \underline{z}\text{ obt}$$

Z_i and Z_j are Fisher's Z transformations of the Pearson correlations based on samples i and j,

n_i and n_j are the sample sizes associated with the i^{th} and j^{th} samples and

$\underline{z}$ obt (disregarding sign) is the test statistic which is evaluated with the $100(\alpha/2)$ percentile of the unit normal distribution. The stage 2 tests should be carried out at the same α as was employed during stage 1.

Example:

Data from Marascuilo's article appear below:

School	1	2	3	4	5
r_j	.66	.70	.68	.92	.44
n_j	58	68	113	37	91

Stage 1 The test of the equality of the five population correlation coefficients is based on the following computational scheme.

Sample (School)	n	r	Z	n-3	(n-3)Z	(n-3)Z²
1	58	.66	.793	55	43.615	34.587
2	68	.70	.867	65	56.355	48.860
3	113	.68	.829	110	91.190	75.597
4	37	.92	1.589	34	54.026	85.847
5	91	.44	.472	88	41.536	19.605
				352	286.722	264.495

$$\sum_1^J\left[(n_j-3)Z_j^2\right] - \frac{\left[\sum_1^J (n_j-3)Z_j\right]^2}{\sum_1^J (n_j-3)} = 264.495 - \frac{(286.722)^2}{352}$$

$$= 30.94 = \chi^2_{obt}$$

The obtained χ^2 is evaluated with $\chi^2_{(\alpha, J-1)}$ which is 9.49 for $\alpha = .05$. The hypothesis $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5$ is rejected; hence we proceed to stage 2.

Stage 2 All pairwise contrasts are tested.

(1) Samples	(2) Est. Z Contrast	(3) Std. Error of Contrast	(4) $\lvert z_{obt} \rvert$
1-2	-.074	.183	.404
1-3	-.036	.165	.218
1-4	-.796	.218	3.649
1-5	.321	.172	1.868
2-3	.038	.157	.243
2-4	-.722	.212	3.411
2-5	.395	.164	2.415
3-4	-.760	.196	3.873
3-5	.357	.143	2.496
4-5	1.117	.202	5.532

Each test statistic $\lvert z_{obt} \rvert$ is evaluated with the critical value 1.96 found in the conventional normal curve table. It can be seen that six of the 10 pairwise contrasts are significant using $\alpha = .05$.

Method II

If a researcher is interested *a priori* in comparing the correlations from each of several groups against one control group correlation there will be only J-1 contrasts. The J-1 contrasts are tested as shown in stage 2 of Method I except that the test statistic is now

$$\frac{Z_{control} - Z_j}{\sqrt{\frac{1}{n_{control}-3} \quad \frac{1}{n_j-3}}} = \underline{t}D.$$

The obtained $\underline{t}D$ value is compared with the critical value of Dunnett's $\underline{t}D$ distribution with infinite degrees freedom. The one and two tailed critical values can be found in Dunnett (1955 and 1964).

Notice that the chi square test described as the first stage of Method I is not employed with Method II.

Simultaneous confidence intervals for the J-1 contrasts are computed using

$$Z_{control} - Z_j \pm \underline{t}D(\alpha, \infty)\sqrt{\frac{1}{n_{control}-3} + \frac{1}{n_j-3}}$$

for the upper and lower limits of Z which are then transformed to correlations.

Example:

Suppose school 1 of the Marascuilo data had been designated as the control group *a priori*. The hypotheses of interest are:

H_o: $\rho_1 = \rho_2$,
H_o: $\rho_1 = \rho_3$,
H_o: $\rho_1 = \rho_4$ and
H_o: $\rho_1 = \rho_5$.

The test statistics associated with these hypotheses were computed under the second stage of Method I. They are:

$\underline{z}obt_1 = .404$,
$\underline{z}obt_2 = .218$,
$\underline{z}obt_3 = 3.649$ and
$\underline{z}obt_4 = 1.868$.

The critical value against which these obtained values are compared is 2.44 for $\alpha = .05$, two-tailed. The H_o: $\rho_1 = \rho_4$ is rejected.

The 95% simultaneous confidence intervals for the J-1 contrasts are:

Estimated Contrast	95% Simultaneous Confidence Interval for True Difference Between ρ's
-.074	-.48, .35
-.036	-.41, .35
-.796	-.87, -.25
.321	-.10, .63

Method III

Method III is the correlation coeficient analog to the Dunn-Bonferroni procedure for contrasts among means. Whenever the contrasts among the correlations of interest are planned (*a priori*) Method III is appropriate. Both simple and complex contrasts can be analyzed using this procedure.

The general formula for testing any simple or complex contrast is:

$$\frac{c_1Z_1 + c_2Z_2 + \ldots + c_JZ_J}{\sqrt{\frac{c_1^2}{n_1-3} + \frac{c_2^2}{n_2-3} + \ldots + \frac{c_J^2}{n_J-3}}} = \underline{t}DB.$$

The test statistic $\underline{t}DB$ is evaluated using the critical value of the Dunn-Bonferroni $\underline{t}DB$ distribution associated with the number of planned contrasts between correlations (C), and infinite degrees of freedom. The critical values are available in Dunn (1961).

Simultaneous confidence intervals for planned contrasts are constructed using

$$c_1Z_1 + c_2Z_2 + \ldots + c_JZ_J \pm \underline{t}DB(\alpha, C, \infty)\sqrt{\frac{(c_1)^2}{n_1-3} + \frac{(c_2)^2}{n_2-3} + \cdots + \frac{(c_J)^2}{n_J-3}}$$

Example:

Suppose the three following complex contrasts among the five schools of the Marascuilo example had been planned before the data were collected:

$$\frac{\rho_1 + \rho_2}{2} - \frac{\rho_3 + \rho_4}{2},$$

$$\frac{\rho_2 + \rho_3}{2} - \rho_4 \text{ and}$$

$$\frac{\rho_1 + \rho_2 + \rho_3}{3} - \rho_5.$$

The corresponding sample contrasts (in terms of Fisher's Z), standard errors of contrasts and obtained $\underline{t}DB$ statistics are:

Est. Z Contrast	Est. Std. Error of Contrast	$\lvert \underline{t}DB_{obt} \rvert$
-.379	.1342	2.82
-.741	.1885	3.93
.358	.1269	2.82

The absolute value of each obtained $\underline{t}DB$ statistic is evaluated using the tabled value of $\underline{t}DB$ for C=3. Since the critical value for α=.05 is 2.39 all three complex contrasts are declared significant.

The 95% simultaneous confidence intervals for the three complex contrasts are:

Estimated Contrast	95% Simultaneous Confidence Interval for True Difference Between ρ's
-.379	-.60, -.06
-.741	-.83, -.28
.358	.06, .58

Comparison of Methods I, II and III with Marascuilo's Procedure

Marascuilo's procedure was applied to the contrasts used to illustrate Methods I, II and III. The results are presented in Table 1.

It can be seen that six pairwise contrasts are declared significant with Method I whereas Marascuilo's procedure detects four significant contrasts. Simultaneous confidence intervals are not shown in the comparison of these two procedures because Method I is a hypothesis testing procedure only.

When Method II is compared with Marascuilo's procedure we find that the same contrasts are declared significant. A comparison of simultaneous confidence intervals, however, reveals narrower intervals for Method II.

Method III declares all three planned complex contrasts significant whereas the Marascuilo procedure declares only one significant. Correspondingly, the simultaneous confidence intervals are narrower for Method III.

Table 1

Comparison of Marascuilo's Procedure with Methods I, II and III.

Pairwise Contrasts (Tests Only) $\alpha = .05$

Z Contrast	Marascuilo Procedure	Method I
$Z_1 - Z_2$	Not Significant	Not Significant
$Z_1 - Z_3$	Not Significant	Not Significant
$Z_1 - Z_4$	Significant	Significant
$Z_1 - Z_5$	Not Significant	Not Significant
$Z_2 - Z_3$	Not Significant	Not Significant
$Z_2 - Z_4$	Significant	Significant
$Z_2 - Z_5$	Not Significant	Significant
$Z_3 - Z_4$	Significant	Significant
$Z_3 - Z_5$	Not Significant	Significant
$Z_4 - Z_5$	Significant	Significant

J-1 Control vs. Other Group Contrasts

Z Contrast	Marascuilo Procedure Simultaneous Confidence Interval		Method II Simultaneous Confidence Interval	
$Z_1 - Z_2$	-.64, .49	Not Significant	-.521, .373	Not Significant
$Z_1 - Z_3$	-.54, .47	Not Significant	-.439, .367	Not Significant
$Z_1 - Z_4$	-1.47, -.12	Significant	-1.328, -.264	Significant
$Z_1 - Z_5$	-.21, .85	Not Significant	-.099, .741	Not Significant

Three Planned Complex Contrasts

Z Contrast	Marascuilo Procedure Simultaneous Confidence Interval		Method III Simultaneous Confidence Interval	
$\frac{Z_1+Z_2}{2} - \frac{Z_3+Z_4}{2}$	.79, .03	Not Significant	-.70, -.06	Significant
$\frac{Z_2+Z_3}{2} - Z_4$	-1.32, -.16	Significant	-1.19, -.29	Significant
$\frac{Z_1+Z_2+Z_3}{3} - Z_5$	-.03, .75	Not Significant	.06, .66	Significant

Table 2

The Procedure Providing the Narrowest Confidence Intervals for Various Combinations of C and J

Number of Planned Contrasts (C)	Number of Groups (J) 3	4	5	6	7	8	9	10
2	III	III	III	III	III	III	III	III
3	III	III	III	III	III	III	III	III
4	M	III	III	III	III	III	III	III
5	M	III	III	III	III	III	III	III
6	M	III	III	III	III	III	III	III
7	M	III	III	III	III	III	III	III
8	M	III	III	III	III	III	III	III
9	M	III	III	III	III	III	III	III
10	M	M	III	III	III	III	III	III
15	M	M	III	III	III	III	III	III
20	M	M	III	III	III	III	III	III
25	M	M	M	III	III	III	III	III
30	M	M	M	III	III	III	III	III
35	M	M	M	III	III	III	III	III
40	M	M	M	III	III	III	III	III
45	M	M	M	III	III	III	III	III
50	M	M	M	III	III	III	III	III
100	M	M	M	M	III	III	III	III
250	M	M	M	M	M	III	III	III

Discussion

The optimum choice among Methods I, II, III and Marascuilo's procedure depends upon whether or not the contrasts have been planned in advance of data collection and the number and type (simple or complex) of interest. If all pairwise comparisons are of interest, as is usually the case, Method I appears to be the most appropriate. Many researchers will feel uneasy about using this procedure because it is analygous to performing an overall *F* test on several means and then, if *F* is significant, computing LSD tests on all pairwise differences between means. Several writers (e.g., Smawley, 1969) have warned against the use of this procedure but Carmer and Swanson (1973) have recently provided very convincing Monte Carlo data to support the use of this approach. If Method I is carried out using $\alpha = .05$ at both stages, the experimentwise error rate should be approximately .05 if Carmer and Swanson's results hold. The experimentwise error rate may differ slightly from α in the small sample case because Fisher's Z transformation is not exactly normally distributed in this situation. Another reason for expecting a slight departure is that it is possible to obtain no significant pairwise contrasts at stage 2 even though the stage 1 test is significant. Hence, if we define the experimentwise error rate as the ratio

Number of experiments in which one or more pairwise contrasts are falsely declared significant using Method I / total number of experiments,

the actual rate should be just slightly less than α.

Method II is more powerful for tests on J-1 simple contrasts than the Marascuilo procedure, but, it is not more powerful than Method I. The major advantage of Method II over Method I would seem to be the availability of simultaneous confidence intervals for the J-1 contrasts. The error rate associated with Method II is experimentwise.

The experimentwise error rate is given a different definition under Method II than under Method I. The Method II experimentwise error rate is defined as the ratio

number of experiments in which one or more of the control versus "other" group contrasts are falsely declared significant using Method II. / total number of experiments.

Method III will generally be preferred to Marascuilo's procedure when the researcher plans the contrasts. If the number of planned comparisons is too large, however, the Marascuilo procedure is superior in terms of the narrowness of the simultaneous confidence intervals. Table 2 can be used to make a choice between Method III and Marascuilo's procedure.

It can be seen that Marascuilo procedure is superior only when the number of groups is small and the number of planned comparisons is large. The situation in which the Marascuilo procedure is the only appropriate one is when no contrasts are planned and the investigator decides to test all contrasts that appear to be interesting.

REFERENCES

Carmer, S. G., & Swanson, M.A. An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. Journal of the American Statistical Association, 1973, 68, 66-74.

Dunn, O. J. Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56, 52-64.

Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 1955, 50, 1096-1121.

Marascuilo, L. A. Large-sample multiple comparisons with a control. Biometrics, 1964, 20, 482-491.

Marascuilo, L. A. Statistical methods for behavioral science research. New York: McGraw-Hill, 1971.

Rao, C. R. Linear statistical inference and its applications. New York: Wiley, 1965.

Snedecor, G. W., and Cochran, W. G. Statistical methods. Ames, Iowa: Iowa State University Press, 1967.

A WORKING LIFE TABLE FOR MALES OF RURAL NIGERIA, 1965*

M. I. IRO, UNIVERSITY OF LAGOS

1. 1. Introduction

A working life table shows, among other things, the joint effects of age-specific mortality rates and age-specific participation rates on the length of working life. The various causes of withdrawals from participation in gainful employment can be assessed by a careful examination of the internal structure of a working life table.

Most extant working life tables relate to European, North American and Japanese working populations (Azumi, 1958; Garfinkle, 1963; Great Britain, 1959; Wolfbein, 1949). Characteristics of the working population in the developing countries have been investigated in India (Gnanasekaran, 1960), Malaya (Saw Swee Hock, 1965) and Ghana (Kpedekpo, 1969). Tables of working life have been constructed for these countries. Urban and rural labour force were combined to construct an overall working life table.

The major difference between these tables referred to here and the Nigerian working life table is that the latter is constructed from data on rural employment patterns. One reason for this procedure is the lack of data on age-specific mortality rates in the urban areas; estimates of mortality rates from model life tables or from other similar countries would be so unreliable and subjective that they would introduce substantial bias to the analysis. The other factor is the predominance of the rural sector in the Nigerian economy. According to the 1963 census, 84.2% of the population of Nigeria were classified as rural dwellers; 80.2% of all employed persons aged 15 and above and 76.7% of employed males were also classified in the rural sector. The rural sector of the economy becomes a crucial subject for investigation.

II. Sources and Limitations of Data

The mortality data used in the construction of the tables come from a nation-wide survey conducted from April 1965 to May 1966. This provided life table survivors, l_x at exact ages. Data on the age distribution of the labour force come from the population census conducted in November, 1963. The Labour Force was defined to include employed persons aged 6 and above, who were at work or had jobs, whether they were full-time or part-time workers. "Work" was defined as 'economic activity' which contributed to the production of economic goods and services. Unemployed persons consisted of all persons aged 6 and over who, during the week preceding the census enumeration period, were not working but were actively seeking for work.

A major difficulty in defining economic activity in a developing country lies in the determination of what constitutes full-time employment in the rural sector. There is usually full-employment in agriculture which lasts for approximately three-fourths of the year: at the time of clearing the farm land, planting the crops and harvesting season; the other one-fourth of the year may be devoted to hunting, fishing, crafting and moving livestock from the savanah regions to the forest belts. If data on seasonal distribution of hours worked in any given year were available, it would be possible to construct different working life tables based on those who worked continuously throughout the year and those who worked for short periods of time. Any method of adjustment for seasonal unemployment and under-employment would result in dubious estimates of rates of employment. Suffice it to say that rural employment is interspered with considerable under-employment. (Jaffee and Berdecia, 1964).

The second problem encountered was in the census tabulation of employed persons, which was made in respect of all persons aged 15 and above. Such a tabulation is somewhat artificial and unrelated to the rural pattern of employment which starts at a much younger age.

The third problem related to the dating of the events which did not refer to the same calendar year (census year, 1963; mortality data, 1965). It is, however, not considered a very crucial defect that could bias the results. The proportional decrease of the Nigerian rural labour force between 1953 and 1963 was of the order of 6% annually due to structural shift from the rural to the modern sector but this loss of rural manpower was offset by a high birth rate estimated at about 50 per 1000 population.

Finally, the rural labour force is a residual after discounting for migration; the effect of rural to urban migration is to further reduce the available rural manpower.

III. Occupational Distribution of the Rural Labour Force

The occupational distribution of the male rural labour force for 1963 is shown below in Table I.

Table I: Occupational Distribution of Male Rural Labour Force

Occupational Code	Occupational Group	Number in Labour Force	Percentage
0	Professional, Technical and related workers	259,786	2.31
1	Administrative, Executive and Managerial Workers	17,361	0.16
2	Clerical Workers	60,188	0.54
3	Sales Workers including Petty Traders	699,644	6.22
4	Farmers, Fishermen, Hunters and Loggers	8,458,312	75.27
5	Miners, Quarrymen, etc.	10,770	0.10
6	Transport and Communication Workers	108,689	0.96
7-8	Craftsmen, Production Process Workers	945,769	8.41
9	Service, Sports, Recreation Workers	464,122	4.13
10	Unspecified Workers	31,061	0.28
11	Unemployed Persons	181,738	1.62
	Total Labour Force	11,237,440	100.00

More than 75% of the labour force were engaged in agriculture and related occupations and 7.8 million persons out of 9.4 million employed persons in agriculture were classified as crop farmers. Handicraft (8.4%) was particularly important. The main categories of handicrafts were tailors, weavers and vehicle mechanics. The rural professional group consisted of school teachers, herbalists and musicians.

IV. Adjustment of Census Labour Force Data

Table 2 below shows the total male population aged 15 and over and the number in the rural labour force.

Table 2: Rural Population and Rural Labour Force of Nigeria

Age Groups	Male Population	In Labour Force	Proportion in Labour Force Per 100
15-19	2,020,050	1,165,242	57.1
20-24	2,433,025	2,099,809	86.3
25-34	3,721,549	3,580,465	96.2
35-44	2,163,010	2,123,393	98.2
45-54	1,142,770	1,120,755	98.1
55-64	623,186	605,139	97.1
65-74	299,755	284,472	94.9
75+	298,078	258,165	86.6
Total	12,701,423	11,237,440	88.5

The proportion in the labour force rises from a minimum in the age-group 15-19 when the young are in full-time attendance at educational institutions or learning trades and reaches a maximum in the age group 35-44. In order to have a greater insight into the age distribution of the population in the labour force, data in Table 2 are split into quinquennial age groups by the use of Newton's formula:

$$P_x^1 = \tfrac{1}{2}[P_x + 1/8(P_{x-1} - P_{x+1})]$$

where P_x^1 is the population in quinquennial age group. The resulting population and numbers in the labour force are shown in Table 3.

The adjusted data in the table seem defective in some respects particularly in the age group 30-34, where the proportion was 100.00. To adjust this defect, the proportion working is derived by a different procedure rather than using the maximum value. As the population under study is entirely rural and agriculture, fishing, hunting and handicrafts are the principal occupations, little use will be made of the concept of "those in training" preparatory to entering the labour force such that ${}_5W_x = \tfrac{1}{2}({}_5W_x + {}_5W_{x+5})$. This will not yield a true value of ${}_5W_x$.

Table 3: Rural Labour Force and the Proportions in the Labour Force

Age Groups	In Population	In Labour Force	Proportions in Labour Force Per 100
15-19	2,020,050	1,165,242	57.70
20-24	2,433,025	2,099,809	86.30
25-29	2,003,904	1,861,586	92,90
30-34	1,717,645	1,718,879	100.00
34-39	1,242,678	1,215,428	97.80
40-44	920,332	907,965	98.70
45-49	667,624	655,268	98.10
50-54	475,146	465,487	97.80
55-59	364,282	354,837	97.40
60-64	258,904	250,302	96.70
65+	597,833	542,637	90.80
Total	12,701,423	11,237,440	

Two alternative procedures were considered: one was to assign the mean proportion $_5W_x$ in the labour force (92.2%) to all the age groups as the most likely $_5W_x$; the second alternative was to take the mean of the $_5W_x$ values from the age group 30-34 to the age group 50-54 as the most probable $_5W_x$. The latter method was preferred because there is technically no unemployment in the rural labour force.

The adjusted proportions working $_5W_x$ are tabulated in Table 4 below and shown graphically in Graph I. These rates compare very favourably with those of urban and rural Ghana for about the same period.

V. Table of Working Life

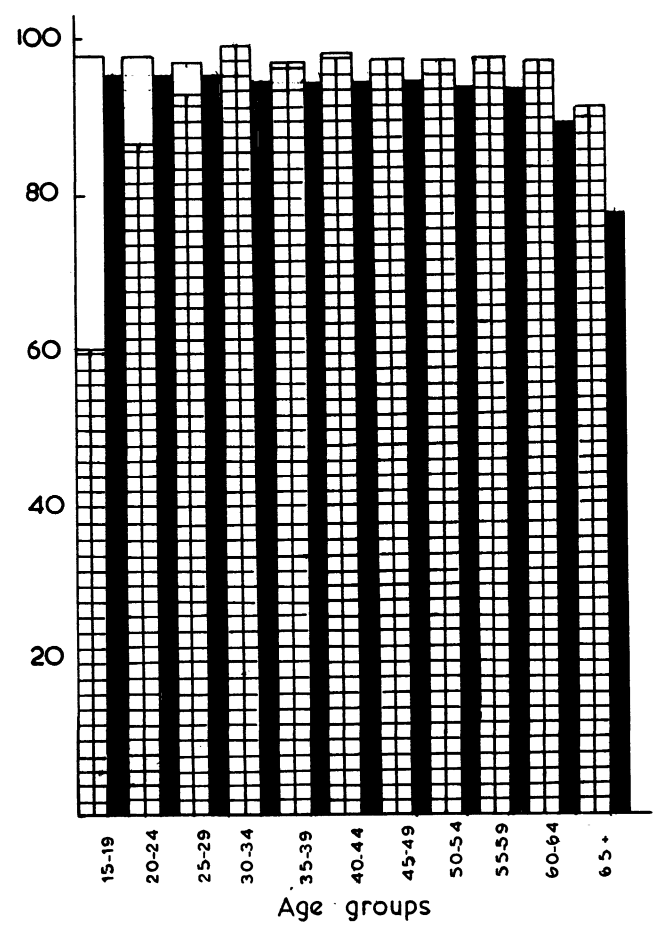
The various components of the working life table are displayed in Table 5. The procedure used in its construction is briefly explained in footnotes. The difference between the average expectation of life and the average expectation of working life increases with age. (See Graph II.)

Table 4: Proportions Working

Age Groups	Proportion Working at age x to x + n Unadjusted $_5W_x$	Adjusted Proportion Working at age x to age x + 5; $_5W_x$	All Ghana (Urban and Rural Proportions Working $_5W_x$
15-19	57.70	98.0	95.00
20-24	86.30	98.0	95.00
25-29	92.91	98.0	95.00
30-34	100.00	98.0	95.00
35-39	97.80	98.0	95.00
40-44	98.70	98.0	95.00
45-49	98.10	98.0	94.80
50.54	98.10	98.0	94.21
55-59	97.40	92.0	93.05
60-64	96.70	92.0	90.05
65+	90.80	92.0	77.88

Graph I

MALES: PROPORTIONS WORKING



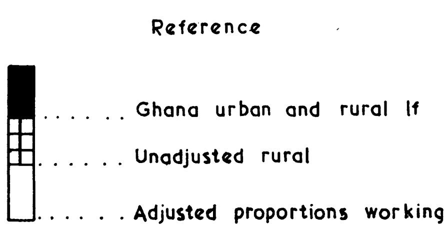
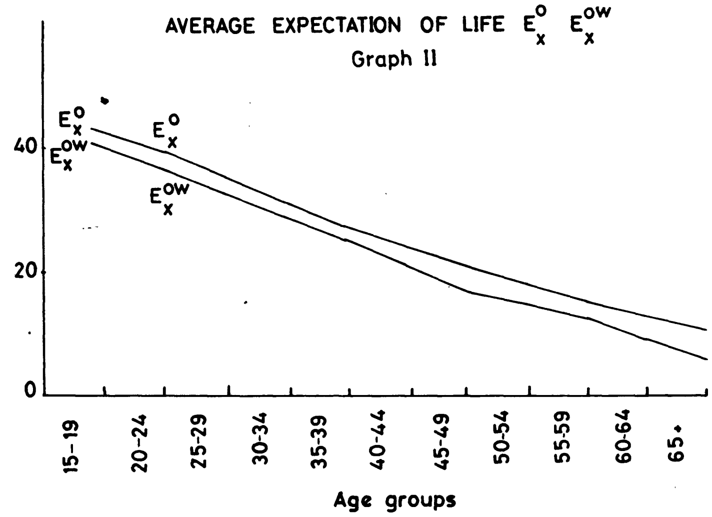


Table 5: Working Life Table for Males

Age Groups	Survivors at exact age x l_x	Years Lived in interval $_5L_x$	Proportion Working $_5W_x$	Number Working at exact age x l_x^W	Number in the Labour Force in age interval $_5L_x^W$	Total Years in interval T_x^W	Average Years of Working Life e_x^{oW}	Average Years of life e_x^o
15-19	629	3,095	0.98	616	3,032	24,926	40.5	42.9
20-24	609	2,995	0.98	597	2,935	21,894	36.7	39.2
25-29	588	2,895	0.98	577	2,838	18,959	32.8	35.5
30-34	569	2,785	0.98	558	2,730	16,121	28.9	31.6
35-39	545	2,660	0.98	534	2,608	13,391	25.1	27.8
40-44	519	2,497	0.98	509	2,448	10,783	21.2	24.1
45-49	480	2,265	0.98	470	2,218	8,335	17.1	20.8
50-54	426	2,000	0.98	417	1,903	6,117	14.7	18.1
55-59	374	1,732	0.92	344	1,593	4,214	12.3	15.3
60-64	319	1,425	0.92	293	1,310	2,621	8.9	12.5
65+	251	-	0.92	231	1,311	1,311	5.7	10.0

1. Column 2: Survivors at exact age taken from 1966 National Abridged life table for males.
2. Column 3: Computed from Column 2 by formula: $_5L_x = \frac{5}{2}(l_x + l_{x+5})$.
3. Column 4: Taken from Column 2 of Table 4.
4. Column 5: Column 2 x Column 4.
5. Column 6: As in Column 3 or Column 3 x Column 4.
6. Column 7: $\sum_{15-19}^{65} L_x^W$
7. Column 8: T_x^W / l_x^W
8. Column 9: Taken from 1966 Abridged life tables for males.

Graph II



VI. Losses from the Working Life

The major cause of loss (see Table 6) during working life is mortality in ages 15 to 65 which accounted for 79.4% of all losses. Old age, infirmity, permanent disability and retirement accounted for 20.6% of all losses.

In the rural labour force there is no clear dividing line between working and retirement. Permanent disability and old age are more meaningful in considering withdrawal from the rural labour force than retirement.

Of the total rural labour force estimated at over 11,237,440 in 1965, the total loss rate was 16.1 per 1000. Losses due to retirement were 2.6 per 1000. These compare with Ghana 2.4 per 1000 and Britain (1955) 9.7 per 1000.

Very little significance should be attached to retirement as a measurable concept in the rural economy. There is no system of retirement benefit or old age pension scheme. The extended family networks meet the social and financial obligations of the old. The tendency is for those who are physically strong to continue to participate in some sort of economic activity rather than contemplate complete withdrawal from the labour force.

Table 6: Losses from the Working Life

Age Groups	Total Losses due to all causes $_5S^w_x$	Losses due to deaths $_5d^w_x$	Losses due to deaths per 1000 $_5m^d_x$	Losses due to other Causes $_5r^w_x$	Losses due to other causes per 1000 $_5m^r_x$
15-19	19	19	6.3	-	-
20-24	20	19	6.5	1	-
25-29	19	19	6.7	-	-
30-34	24	23	8.4	1	-
35-39	25	25	9.6	-	-
40-44	39	38	15.5	1	-
45-49	53	53	23.9	-	-
50.54	73	38	20.0	35	18.4
55-59	51	31	19.5	20	12.6
60-64	62	42	32.1	20	15.3
65+					
Total	385	307	-	78	

Column 1: $_5S^w_x = l^w_x - l^w_{x+5}$; $_5S^w_x = {_5r^w_x} + {_5d^w_x}$

Column 2: $_5L^w_x \cdot {_5m_x}$

Column 4: $_5r^w_x = {_5S^w_x} - {_5d^w_x}$

References

1. Azumi, Koya:(1958): Length of work life of Japanese men, 1930 and 1955: Monthly Labour Review 81, pp. 1366-1368.

2. Fullerton, H.N. (1971): A table of expected working life for men, 1968: Monthly Labour Review, June 1971: pp. 49-55.

——— (1972): Generation tables of working life for men, 1940, 1950 and 1960: A paper presented at the P.A.A. meeting in Toronto, Canada, April 1972.

3. Garfinkle Stuart: (1963): The length of working life for males 1900-1960. U.S. Department of Labour, Report No. 8

——— (1963): Table of working life for men, 1960: Monthly Review, 86, pp. 820-823.

——— (1955): Changes in working life of men, 1900-2000: Monthly Labour Review, 1955, pp. 297-300.

4. Gnanasekaran, K.S. (1960): Increasing length of working life and its implications: Indian Economic Journal (Bombay): 7, pp. 405-414.

5. Great Britain (1959): Ministry of Labour: The length of working life of males in Great Britain: Studies in Official Statistics No. 4: H.M. Stationery Office, London.

6. Jaffee, A.J. & Berdecia, F.S. (1955): The concept and measurement of under-employment: Monthly Labour Review, March 1955, pp. 283-287.

7. Kpedekpo, G.M.K. (1969): Working life tables for males in Ghana, 1960: Journal of the American Statistical Association 64, pp. 102-110.

Kpedekpo, G.M.K. (1969): On working life tables in Ghana with particular reference to the female working population: Journal of the Royal Statistical Society Series A: 132, pp. 431-441.

8. Nigeria: Federal Office of Statistics (1968): Rural Demographic Sample Survey 1965/66.

——— 1963: (1963) Population Census Reports Vol. III: Combined National Figures.

9. Saw Swee Hock (1965): Malaya: Tables of male working life, 1957: Journal of the Royal Statistical Society Series (A): 128, pp. 421-438.

——— 1965): Uses of working life tables in Malaya: Proceedings of the World Population Conference, Belgrade.

10. Wolfbein, S.L. (1949): Length of working life: Population Studies, Vol. 3, pp. 286-294.

LOCAL AREA ESTIMATES OF UNEMPLOYMENT LEVELS AND RATES -- CONCEPTUAL ISSUES *

Hyman B. Kaitz, CSR Associates

Introduction

Interest in and use of estimates of unemployment levels and rates for individual areas of the United States has existed since World War II at least. Area classification of major labor market areas based on such estimates began after the 1948-49 recession. With the passage of the Area Redevelopment Act in 1960, major labor market areas, cities and rural areas were deemed eligible for Federal grants and loans on the basis of these same unemployment figures.

Subsequent legislation on area assistance eligibility extended the coverage of the program to smaller and smaller areas, down finally to individual Census tracts. The coverage of small areas under the Public Employment Program in 1971 extended to cities, counties of 75,000 population, balance of States, and separately to cities and counties of 7,500 or more population, with unemployment rates of 6 percent or more for three consecutive months.

More recently and importantly, the Comprehensive Employment Training Act of 1973 (CETA) uses a system of weighted elements under Title I for allocation of funds to areas; 37½ percent of this weight is based on the relative number of unemployed in the State as compared with the total number in all States. Prime sponsors under CETA may be entire States, or as small as any unit of general local government without regard to population criteria, but which meets certain other conditions. Eighty percent of Title II funds are allocated among eligible areas in accordance with the number of unemployed in an area of substantial unemployment relative to the number of unemployed in all such areas. An area of substantial unemployment is defined as one with an unemployment rate equal to or in excess of 6.5 percent for three consecutive months.

There can be no question, then, that area estimates of unemployment levels and rates have played an important role in financial assistance to communities, and that the importance of these estimates has increased with the enactment of CETA in 1973. Much discussion has surfaced lately (third quarter of 1974) about the possibility of reinstating the Public Employment Program in the near future if unemployment rates should continue to rise. In such a case, local area measures of unemployment would undoubtedly be considered again for the allocation of funds and jobs.

How good are these estimates of unemployment levels and rates? How appropriate are the concepts on which these estimates are based? These questions are taken up in the full report. In this paper, which constitutes the first part of the report, only the unemployment concepts are discussed.

Unemployment Concepts

Unemployment rates have generally been used as indexes of economic health. For the U.S. as a whole there are also many other indexes available such as real GNP growth or changes in the industrial production index. It has long been apparent that the relationship among these various indexes has been complex and puzzling, and that they have not always supported each other, both in intensity and in timing. Most recently, for example, we have had declines in real GNP in both the first and second quarters of 1974, but the unemployment rate appears to have been affected by these declines only marginally at best, a pattern which has been judged inconsistent with earlier experience and one which has confounded all forecasters 1/. The aggregation in these measures at the U.S. level presumably nets out large amounts of statistical error, and other sources of more or less random variation, so that they each exhibit acceptable patterns of regularity. These advantages do not quite obtain for individual geographic areas, so that it would not be surprising if a number of separate indexes of economic conditions for a single area tended to be in less agreement on some occasions as to what was actually happening than we would find to be true for the U.S. as a qhole. This is not necessarily a matter for concern, however, since single indexes cannot fully reveal the complexity of economic activity. The use of unemployment levels and rates as indexes must therefore represent a compromise with our needs. 2/. More narrowly, are the best measures of labor force slack to be found in unemployment rates and levels? The conventional unemployment concept will be examined below to some extent. We note that the CETA criteria accept the concept that persistence of high unemployment rates in an area equal to or exceeding 6.5 percent for at least three months is more significant than short term fluctuations above this level 3/. However there is a basic weakness in the use of triggers of this kind which set up an artificial dichotomy: that areas with unemployment rates at or above 6.5 percent differ in kind from those with an unemployment rate of 6.4 percent, or of 6.5 percent in only two months. This device appears in a variety of legislation and deserves detailed study before it is incorporated in future legislation.

Title I under CETA, which does not include a trigger (or selection) provision relies on more than just the unemployment levels of the areas, by including a weight for the number of adults in families below the low income level in the State relative to the total number of such adults in all States; this is a longer term, structural measure of poor economic health than the unemployment rate. A final weight under Title I is based on the allocation of funds in the preceding year, a device which maintains some continuity with the past but which diminishes the importance of current indicators (the unemployment rate).

For some years now the unemployment rate itself has been under attack from various sources as an inadequate or incomplete measure of economic deprivation, or of the extent of the labor surplus.

The official definition of unemployment in the national household survey covers those members of the civilian noninstitutional population 16 years of age and over, who:

1) had done no work for pay in the survey week, and
2) had been available for work, and
3) had made some explicit effort to look for work in the preceding month.

(There are some relatively minor additions and modifications to this concept which are not described here.)

The Gordon Committee reviewing the employment and unemployment data in 1962 was troubled by the fact that respondents to the household questionnaire of that time were counted among the unemployed if they volunteered the information that they had not looked for work because they didn't think they could find a job. Others who may have felt this way, but who didn't speak up were classified as being out of the labor force. Both of these groups were the so-called "discouraged" workers.

In response to the Gordon Committee recommendations, the 1967 revision of the household questionnaire added some supplementary questions to be asked of one-fourth of the full household sample of people who had not been working but who had not looked for work in the preceding four weeks. The data collected for these groups reveal that in the first quarter of 1974, for example, when there were 4.7 million unemployed people on the average, there were also about 0.7 million discouraged workers. For manpower program purposes, this added group should not be ignored. On the other hand, we know little about this group at the present time other than their demographic characteristics. We do not know how long they have been without work, and whether they have looked for work prior to the four weeks preceding the survey week when they indicated they had not looked for work. For some people, the desire or need for work may be sufficiently week so as to explain why they didn't look for work, if in their judgment jobs may have been available but hard to find, or of low quality and not worth seeking. Technically, they would be "discouraged" workers but our inferences about them would probably be wrong. A different group of discouraged workers live in large low income areas and realistically do not look for work for the same reason that they do not anticipate visits from foreign ambassadors. There are people in these low income areas who neither looked for work, or said they would have looked if they thought work was available, but who would take a job if community conditions were more favorable. We can only guess at the size of this latter group(which is probably not small) and it may be possible, under certain assumptions, to make estimates of its size. The net effect of excluding the discouraged workers, and the others who would not even be counted as discouraged under present definitions is to make inner cities appear to have less of a labor surplus than the actual conditions in these cities would suggest. 4/

An indirect econometric method for estimating the size of the labor surplus is one which relates the labor force participation rate to the employment rate(employment as a percent of population):

$$L/P = a + b\,E/P\ .$$

This equation has received much attention when applied to time series . In that application it usually included other variables, such as time, and lagged values of at least one of the two variables already present. The applicability of this formulation to the present paper arises from the following reasoning. When economic conditions worsen, and unemployment levels rise, the labor force tends to shrink (in the above equation, $b>1$). This has been interpreted as the net dominance of the "discouraged" worker effect over the "additional" worker effect (a family member who starts to look for work because the family breadwinner has been laid off 5/).

The equation above , after calculation of the coefficients, a and b, may be recast in the form:

$$L/P = a/(1 - b + ub)$$

In this form, the labor force participation rate is made an explicit function of the unemployment rate ($u = U/L$, where $U = L - E$). For simplicity of exposition consider that this model has been developed for the entire labor force, although in practice separate equations are usually calculated by age, sex and color. The empirically determined values of a and b are such that L/P rises as u falls. For "full employment" ($u = 0.04$), $L/P = (L/P)_F = a/(1- 0.96\ b)$. When unemployment is higher than four percent, the observed value of L/P should be lower than $(L/P)_F$. The difference, $(L/P)_F - L/P$ has been called "hidden" or "disguised" unemployment--those who have withdrawn from the labor force because of the decline in job opportunities. If this group is added to the unemployed (when $u>0.04$) we will get a new and higher unemployment rate.

The same approach, with some modifications, can be applied to cross-section data such as geographic areas. In this way, unemployment rates over four percent (or any other selected "norm") will be increased, while those below four percent will be reduced (whatever implications that may have!). The replacement of the original unemployment rates by these corrected measures may result in some reranking of the areas, since the equation does not have all the observations on the line, and an increase in the dispersion of unemployment rates. Inner cities would generally benefit from the use of this technique. While this approach has some analytic interest, it is not preferable at this time to other measures which may be considered. It should be noted that the measure of "hidden" unemployment developed via this model bears little relationship to the counts of "discouraged" workers which come from the household survey responses. For one thing, "hidden" unemployment vanishes when the observed unemployment rate is four percent, but the count of "discouraged" workers does not.

Labor surplus may be indicated not only by unemployment but also underemployment. People working less than full time but who would prefer to work full time are clearly underemployed. The shortfall in their hours can be converted to a full time equivalent unemployment measure; to balance this adjustment, unemployed people looking for part time work can have their numbers adjusted downward also to a full time equivalent basis. Both of these adjustments have been made to the regular count of unemployment and employment for some years now by the Bureau of Labor Statistics, to obtain a measure called "Labor force time lost." For the first quarter of 1974, for example, when the regular unemployment rate averaged 4.7 percent, the labor force time lost rate was 5.7 percent. It is reasonable to think that these would be superior indexes of labor surplus for geographic areas to those in use at present. The data requirements would be formidable, however.

Some manpower experts, Sar Levitan and Robert Taggart, have suggested in a recent article in the Monthly Labor Review that the number of unemployed should be augmented by the number of working poor. These and other measures which take income levels into account as well as unemployment are interesting and deserve a detailed examination in their own right. However,such an examination will not be attempted here.

In recent years a number of people have expressed dissatisfaction with the traditional measures of unemployment and labor force attachment. The January 31, 1974 policy statement of the National Manpower Policy Task Force says that current labor market statistics need to be supplemented or modified in order to improve our understanding of the labor market and its behavior. CETA, Section 312(c) says that "the Secretary /of Labor / shall develop preliminary data for an annual statistical measure of labor market related economic hardship in the nation. Among the factors to be considered in developing such a measure are unemployment, labor force participation, involuntary part time employment, and full time employment at less than poverty wages."

This resurgence of discussion of the complexities of labor market behavior and the inadequacies of presently available measures bodes well for future improvements. There are many questions to be asked in this connection, in addition to those which have been considered above. Some of these are:

What fraction of the discouraged workers want only part time jobs?

What are the job and wage requirements which would attract people from out of the labor force (who might thus be counted as part of a labor surplus) ?

How many people are working at more than poverty wages but are nevertheless in jobs below their capacity or potential?

The following subject has come in for little discussion elsewhere, for the most part, and is therefore given some attention here. Even if we were to restrict our attention to unemployment as conventionally measured, we would still have to recognize its dimension of differential persistence for individuals. CETA accepts the notion of persistence of unemployment for an area, but does not recognize the persistence of unemployment for individuals. Unemployment turbulence is a complex phenomenon which receives as yet little attention 6/ . For example, it is estimated that in 1968, a boom year, about 30 million spells of unemployment were begun by some 20 million different people. In that same year, unemployment averaged only 2.8 million per week. Every week, between 15 and 25 percent of the unemployed had just entered, and about the same proportion dropped out of the ranks of the unemployed. The total count of unemployment in a given week gives equal weight to the respondent who says he has been unemployed two weeks, and the one who says he has been out for three months. An individual who has a completed unemployment spell of one week is presumably not the target of manpower programs (except on grounds other than just the fact that he is unemployed). Many State unemployment insurance laws count the first week as a "waiting period" week, not subject to compensation, presumably on the grounds that one week can be tolerated by the worker, and at the same time helps to reduce State benefit expenditures. It has been suggested that the first several weeks of unemployment for a worker might be considered as frictional unemployment, and thereby excluded from measures of labor surplus either at the U.S. level, or across geographic areas. This is the same as restricting the measure to long term unemployment. Further thought should be given to this idea.

A related idea has been suggested by Geoffrey H. Moore: taking the product of the unemployment rate and the average duration of unemployment in the survey week, and calling it an index of "unemployment severity." What would this measure do for area classification? At present we equate two areas with the same unemployment rate, say, 5.0 percent. Suppose that in one area these people have been unemployed an average of 6 weeks, while in the other they have been unemployed for 12 weeks on the average at the time of the survey. It is a reasonable proposition that unemployment is more severe in the second area. The Moore index of severity would be almost twice as high in the second area as in the first(if the labor force denominator had also been adjusted to reflect this new measure). Whether or not there would be general agreement that conditions are twice as bad in the second area as in the first, no one would deny that unemployment in the second area is more severe and also different in some important respects from that in the first, with possibly different implications for manpower programs and fund allocation.

Even if we had a national Census every year, with reliable statistical measures of unemployment levels and rates by area, there would still be these conceptual problems to resolve. However, we don't have uniformly reliable statistical measures -- this is discussed in the larger report of which this paper is a part.

One final topic which has come up for discussion as a result of the dependence of unemployment rate calculations on employment obtained through establishment surveys, or alternatively through household surveys, is the measurement of such employment (a part of the labor force base for the unemployment rate) by place of work or place of residence of the worker. In this paper, we can consider a conceptual aspect of this issue apart from the measurement problem.

Take the following synthetic example:

In area A there are 100 people working, and 10 unemployed people living. In area B there are 100 people working and 5 unemployed people living. However 50 of the 100 people employed in area B live in A, while all of the people working in A also live in A.

If we take the unemployed where they live and the employed where they work (essentially the Manpower Administration procedure until 1974), we get these results:

Labor Force Status	Area A	Area B
Unemployed	10	5
Employed	100	100
Work force	110	105
Unemployment rate	9.1 %	4.8 %

On the other hand, if we take all of these people on the basis of their place of residence, we get:

Labor Force Status	Area A	Area B
Unemployed	10	5
Employed	150	50
Labor force	160	55
Unemployment rate	6.2 %	9.1 %

On a residence basis, the unemployment rates of A and B are almost the reverse of those calculated using employment on a place of work basis. The first set of calculations admittedly represent an unsatisfactory hybrid of unemployment by place of residence and employment by place of work with consequently ambiguous and hard to interpret results, and make for misleading comparisons among areas. The second calculation is consistent and readily understood -- all counts are on a residence basis. Nevertheless, if area unemployment rates are used for a specific purpose -- to allocate manpower program funds -- the residence concept may not be fully satisfactory either. Area A appears to be a bedroom community for a considerable proportion of the people working in B. But the residence approach underestimates the net number of jobs in B, and consequently the job creation potential. Where should the job training take place -- where people live, or where they customarily work? Or should the training programs be allocated between A and B on some other basis, such as the ability of these communities to run such programs? It is not unreasonable to put the manpower programs where the jobs are, other things being equal. A consistent conceptual basis for a calculation which differs from both of those illustrated above is to use the employment on a place of work basis, and unemployment on a "place of jobseeking" basis. Both sets of data can be collected in a household survey by asking employed people where they work, and unemployed people where they look for work.

In the above example, suppose we find that of the 10 unemployed people who live in A, 8 look for work usually in A, while 2 look in B. Of the 5 unemployed people living in B, 4 look in A, while 1 looks in B. On this basis we would then have the following:

Labor Force Status	Area A	Area B
Unemployed	12	3
Employed	100	100
Labor force	112	103
Unemployment rate	10.7 %	2.9 %

Does this calculation make more sense? In terms of local area politics and competition, it may not be particularly attractive. However it does have the virtue of excluding 100 percent bedroom communities in the allocation of manpower funds, which is a reasonable result. The calculation on a place of residence basis is probably more useful for welfare or poverty program fund allocation. Only the first calculation on a mixed basis lacks any good rationale.

As long as the areas which are considered are self-contained labor market areas, there is no problem since all three approaches will yield essentially the same results. If the area considered is adjacent to one or more other areas and there is a significant amount of interdependence among them evidenced by the commuting patterns of both their employed and unemployed workers, the issue discussed above arises. It may be that in such a case, the use of separate area indexes is not the ultimate answer, regardless of how these indexes are constructed, as long as they ignore the dynamic interaction among the areas. Some further thought and study is needed here.

The intent of this paper is to indicate that getting the most accurate measures of unemployment levels and rates on a local area basis may still leave us some distance from a fully satisfactory system of estimates for specific program purposes. And it may well be that different kinds of programs will require different kinds of estimates, based, perhaps, on different concepts. We must keep in mind that what may serve us well in terms of national ,general purpose estimates will not necessarily be good models for specific purpose local area estimates.

FOOTNOTES

1) According to Michael Evans of Chase Econometrics, the reason the unemployment rate is less than 5.5 percent rather than 6 percent or more anticipated on the basis of past relationships is that the unusually high corporate profits coupled with declining real wages have led to labor hoarding and lower layoffs. On the other hand, George Terborgh, Consultant to the Machinery and Allied Products Institute, estimates that over $25 billion of 1973 profits of nonfinancial corporations were "phantom" due to the undercosting of fixed assets and inventory consumption in a period of rapid inflation. Can it be that workers have benefited to the extent that corporate managers have suffered from a rather complex money illusion?

2) One may ask whether manpower program fund allocation should be tied to any index of labor surplus -- the unemployment rate or any alternative measure. I suggest that this cannot be a fully settled issue. While people may be helped directly by income subsidies, however tied to manpower programs, we have no good evidence that this is the appropriate way to bolster area economies and to get the maximum multiplier effect through investment of Federal funds. Policy makers clearly desire to attain more than the direct, short-term impact of the Federal programs. (For another approach to allocation criteria, see D. Hammermesh and H. Pitcher, "Economic Considerations for Manpower Revenue Sharing," Industrial and Labor Relations Review , July, 1974.)

3) However this rationale is weakened by the use of the monthly unemployment rates before seasonal adjustment. At the same time this procedure makes Title II of CETA more liberal in practice, and would tend to make more areas eligible under this provision. This may be desirable.

4) See: Herman Miller, "Subemployment in Poverty Areas of Large U.S. Cities," Monthly Labor Review , October, 1973.

5) In view of the studies which have been made of labor force flows (50 percent turnover in unemployment per month -- see Kaitz, Perry, Smith) these earlier concepts now appear somewhat simplistic.

6) Note the earlier discussion of this phenomenon.

* This paper is part of a larger report prepared for the Department of Economic and Community Development of the State of Maryland. Any opinions expressed here are those of the author and not necessarily those of the Department.

REFERENCES

Gastwirth, Joseph, "Estimating the Number of 'Hidden Unemployed', Monthly Labor Review, March, 1973

Kaitz, Hyman B.,"Analyzing the Length of Spells of Unemployment," Monthly Labor Review, November, 1970

Levitan, Sar and Taggart, Robert, "Employment-Earnings Inadequacy: A Measure of Welfare," Monthly Labor Review, October, 1973

President's Committee to Appraise Employment and Unemployment Statistics, Measuring Employment and Unemployment, G.P.O., 1962

Perry, George, "Unemployment Flows in the U.S. Labor Market," Brookings Papers on Economic Activity, II (1972)

Smith, Ralph E., "The Discouraged Worker in a Full Employment Economy," 1973 Proceedings of the Business and Economics Section, American Statistical Association

Tella, Alfred and Simpler, Norman, "Labor Reserve and the Phillips Curve," Review of Economics and Statistics, February, 1968

SURVEYS COMBINING PROBABILITY AND QUOTA METHODS OF SAMPLING

Benjamin F. King, University of Washington

INTRODUCTION

This paper discusses a Bayesian analysis of a model in which it is possible to have *full probability and quota sampling* within the same set of primary sampling units for the same survey.

The study was inspired by the analysis of biased measurement in the textbook of Pratt, Raiffa and Schlaifer (1965), but in order to better approximate the sampling procedures actually employed in survey research, that simple model has been considerably elaborated upon, resulting in optimization formulas that are more complicated.

We consider the following specification:

$$\tilde{x}_{uij} = \mu_u + \tilde{\gamma}_i + \tilde{\varepsilon}_{uij} \qquad (2.1)$$

$$\tilde{x}_{bik} = \mu_b + \tilde{\gamma}_i + \tilde{\varepsilon}_{bik}$$

$$j = 1,2,\ldots,n_u$$
$$k = 1,2,\ldots,n_b$$
$$i = 1,2,\ldots,p$$

$$E(\tilde{\gamma}_i) = E(\tilde{\varepsilon}_{uij}) = E(\tilde{\varepsilon}_{bik}) = 0 \qquad (2.2)$$

$$E(\tilde{\gamma}_i\tilde{\varepsilon}_{uij}) = E(\tilde{\gamma}_i\tilde{\varepsilon}_{bik}) = E(\tilde{\varepsilon}_{uij}\tilde{\varepsilon}_{bik}) = 0 \qquad (2.3)$$

$\tilde{x}_{uij}$ and $\tilde{x}_{bik}$ are observations from unbiased (probability) and biased (quota) measurement processes, respectively.

$\tilde{\gamma}_i$ is a random effect that is held in common by all observations, biased or not, in the i^{th} PSU.

$\tilde{\varepsilon}_{uij}$ is a second-stage disturbance term unique to the j^{th} observation in the full probability sample within the i^{th} PSU.

$\tilde{\varepsilon}_{bik}$ is correspondingly a second-stage random disturbance in quota sampling.

μ_u and μ_b are respectively the expected values of the $\tilde{x}_{uij}$ and $\tilde{x}_{bik}$, and $\beta = \mu_u - \mu_b$ is the bias.

Define

$$v_u = E(\tilde{\varepsilon}^2_{uij})$$
$$v_b = E(\tilde{\varepsilon}^2_{bik}) \qquad (2.4)$$
$$v_c = E(\tilde{\gamma}^2_i),$$

the variances of the random components on the right hand side of (2.1). It follows that

$$V(\tilde{x}_{uij}) = v_c + v_u$$
$$V(\tilde{x}_{bik}) = v_c + v_b \qquad (2.5)$$

and $\mathrm{Cov}(\tilde{x}_{uij}, \tilde{x}_{bik}) = v_c$.

Summarizing, our model assumes that p PSUs are randomly selected and that either full probability or quota sampling *or both methods* are applied to the units within each PSU. The within-PSU sample sizes n_u and n_b are constant across PSUs. Furthermore, it is assumed that both probability and quota observations are subject to the same random PSU effect; i.e., the bias conditional on PSU i is equal to the unconditional bias, $\mu_b - \mu_u$. The latter assumption implies that if, for example, a certain PSU has mean income that is higher than the national average, μ_u, the same random effect is present among the responses in quota sampling, although the within PSU average incomes for probability and quota sampling may differ substantially because of the fundamental bias, $\mu_b - \mu_u$.

Finally, we realize that actual survey practice involves sampling from finite populations, but for simplicity, we treat the selected PSUs as independent. Further, we assume that the within PSU observations in the samples of sizes n_u and n_b are independent conditional on γ_i.

PRIOR-TO-POSTERIOR ANALYSIS

Assume that the prior distribution of $(\mu_u, \mu_b)^t$ is bivariate normal with mean and variance

$$E\begin{bmatrix}\tilde{\mu}_u \\ \tilde{\mu}_b\end{bmatrix} = \begin{bmatrix}\bar{x}'_u \\ \bar{x}'_b\end{bmatrix},$$

$$V\begin{bmatrix}\tilde{\mu}_u \\ \mu_b\end{bmatrix} = \begin{bmatrix}V'_{uu} & V'_{ub} \\ V'_{ub} & V'_{bb}\end{bmatrix}, \qquad (3.1)$$

where V'_{ub} is the prior covariance of $\tilde{\mu}_u$ and $\tilde{\mu}_b$. Define

$$\tilde{\bar{x}}_u = \sum_{i=1}^{p}\sum_{j=1}^{n_u} \tilde{x}_{uij}/pn_u, \qquad (3.2)$$

the arithmetic mean of the unbiased observations, and

$$\tilde{\bar{x}}_b = \sum_{i=1}^{p}\sum_{k=1}^{n_b} \tilde{x}_{bik}/pn_b, \qquad (3.3)$$

the arithmetic mean of the quota sample.

Then the mean and variance of $(\tilde{\bar{x}}_u, \tilde{\bar{x}}_b)^t$, conditional on $(\mu_u, \mu_b)^t$ is

$$E\begin{bmatrix}\tilde{\bar{x}}_u & \mu_u \\ \tilde{\bar{x}}_b & \mu_b\end{bmatrix} = \begin{bmatrix}\mu_u \\ \mu_b\end{bmatrix},$$

$$V\begin{bmatrix}\tilde{\bar{x}}_u & \mu_u \\ \tilde{\bar{x}}_b & \mu_b\end{bmatrix} = \begin{bmatrix}v_c/p + v_u/pn_u & v_c/p \\ v_c/p & v_c/p+v_b/pn_b\end{bmatrix}. \qquad (3.4)$$

If we assume that the $\tilde{\gamma}_i$, $\tilde{\varepsilon}_{uij}$, and $\tilde{\varepsilon}_{bik}$ are normally distributed with p, n_u, and n_b, all > 0, then $(\tilde{\bar{x}}_u, \tilde{\bar{x}}_b)^t$ is conditionally bivariate normal. With nonnormal random components, the central limit theorem can be invoked to treat the sample means as approximately normal and approximately sufficient for $(\mu_u, \mu_b)^t$.

Define

$$H' = \begin{bmatrix} H'_{uu} & H'_{ub} \\ H'_{ub} & H'_{bb} \end{bmatrix} = \begin{bmatrix} V'_{uu} & V'_{ub} \\ V'_{ub} & V'_{bb} \end{bmatrix}^{-1}, \quad (3.5)$$

$$H = (n_u v_c v_b + n_b v_c v_u + v_u v_b)^{-1} \begin{bmatrix} pn_u(n_b v_c + v_b) & -pn_u n_b v_c \\ -pn_u n_b v_c & pn_b(n_u v_c + v_u) \end{bmatrix}, \quad (3.6)$$

and

$$H'' = H' + H. \quad (3.7)$$

It follows from the standard Bayesian development for the bivariate normal case that $\tilde{\mu}_u$ and $\tilde{\mu}_b$ have a joint posterior distribution that is normal, with mean and variance

$$\begin{bmatrix} \bar{x}''_u \\ \bar{x}''_b \end{bmatrix} = H''^{-1} \left[H' \begin{bmatrix} \bar{x}'_u \\ \bar{x}'_b \end{bmatrix} + H \begin{bmatrix} \bar{x}_u \\ \bar{x}_b \end{bmatrix} \right],$$

$$V'' = \begin{bmatrix} V''_{uu} & V''_{ub} \\ V''_{ub} & V''_{bb} \end{bmatrix} = H''^{-1}. \quad (3.8)$$

The posterior expectation of $\tilde{\mu}_u$ is of particular interest to us, and straightforward algebra yields

$$\bar{x}''_u = \bar{x}'_u + w_u(\bar{x}_u - \bar{x}'_u) + w_b(\bar{x}_b - \bar{x}'_b), \quad (3.9)$$

where

$$w_u = (n_u v_c v_b + n_b v_c v_u + v_u v_b)^{-1} [V''_{uu} pn_u(n_b v_c + v_b) - V''_{ub} pn_u n_b v_c]$$

and (3.10)

$$w_b = (n_u v_c v_b + n_b v_c v_u + v_u v_b)^{-1} [V''_{ub} pn_b(n_u v_c + v_u) - V''_{uu} pn_u n_b v_c].$$

Thus, depending on the posterior covariance structure, the posterior expectation of the mean of the unbiased process is affected by sample deviations of both the probability and quota means from their respective prior expectations.

In subsequent developments we shall require the posterior variance of $\tilde{\mu}_u$,

$$V''_{uu} = (H''_{uu} - H''^2_{ub} H''^{-1}_{bb})^{-1}. \quad (3.11)$$

It will be more convenient to work with the posterior precision, which, substituting from (3.5) and (3.6), can be written in the form

$$V''^{-1}_{uu} = u/v, \quad (3.12)$$

where

$$u = (H'_{uu}H'_{bb} - H'^2_{ub})(n_u v_c v_b + n_b v_c v_u + v_u v_b) + p^2 n_u n_b + H'_{uu} pn_b(v_u + n_u v_c) \quad (3.13)$$
$$+ H'_{bb} pn_u(v_b + n_b v_c) + 2H'_{ub} pn_u n_b v_c,$$

and

$$v = H'_{bb}(n_u v_c v_b + n_b v_c v_u + v_u v_b) + pn_u n_b v_c + pn_b v_u \quad (3.14)$$

It can be shown that if n_b is zero, implying a design that is totally devoted to unbiased sample selection, expressions (3.9) through (3.14) collapse to the usual univariate normal result.

OPTIMUM ALLOCATION OF A FIXED BUDGET

Consider a survey in which the principal task is point estimation of $\tilde{\mu}_u$, but for which the economic resources available are a fixed dollar amount k*. (This is not an uncommon situation in actual practice.)

Defining

k_c = the cost of "setting up" a staff for this survey in a single PSU,

k_u = the per unit cost of a within-PSU probability observation,

and k_b = the per unit cost of a quota observation, (4.1)

we write

$$k^* = k_c p + k_u pn_u + k_b pn_b. \quad (4.2)$$

It is assumed that elements of fixed cost for planning, overall supervision, etc., have been subtracted out of k*.

With a quadratic loss function of the error in estimating $\tilde{\mu}_u$, the optimum Bayesian estimator is $\bar{x}''_u$, the posterior mean. The posterior expected loss is proportional to the posterior variance of $\tilde{\mu}_u$. We shall consider the case where v_c, v_u and v_b are known, and $\bar{x}''_u$ will be used to estimate $\tilde{\mu}_u$. Given the fixed amount k*, we wish to allocate it between probability and quota sample observations in such a way as to maximize the preposterior variance of $\tilde{\bar{x}}''_u$, the controllable factor in the expected net gain from sampling (ENGS). Since

$$V(\tilde{\bar{x}}''_u) = V'_{uu} - V''_{uu}, \quad (4.3)$$

we must choose p, the number of PSUs, n_u and n_b in order to minimize V''_{uu}, given in (3.11) above, within the fixed budget k*.

Using (4.2) to express n_b in terms of p and n_u, we substitute in (3.13) and (3.14) and rearrange to get

$$\begin{aligned} u = {} & pv_u(|H'|v_b k_b - |H'|v_c k_c + H'_{uu}k^*) \\ & - p^2 H'_{uu} v_u k_c + |H'| v_u v_c k^* \quad (4.4) \\ & + n_u [p|H'|v_c(v_b k_b - v_u k_u) \\ & + pv_c k^*(H'_{uu} + H'_{bb} + 2H'_{ub}) \\ & + p^2(H'_{bb} v_b k_b - H'_{uu} v_u k_u + k^*) \\ & - p^2 v_c k_c(H'_{uu} + H'_{bb} + 2H'_{ub}) - p^3 k_c] \\ & - n_u^2 [p^2 k_u v_c(H'_{uu} + H'_{bb} + 2H'_{ub}) + p^3 k_u], \end{aligned}$$

and

$$\begin{aligned} v = {} & pv_u[H'_{bb}(v_b k_b - v_c k_c) + k^*] \\ & - p^2 v_u k_c + H'_{bb} v_u v_c k^* \\ & + n_u [pv_c(H'_{bb}[v_b k_b - v_u k_u] + k^*) \end{aligned}$$

$$- p^2(v_c k_c + v_u k_u)] - n_u^2 p^2 v_c k_u. \quad (4.5)$$

The desired approach is to maximize the posterior precision (3.12) by differentiating with respect to p and n_u, and setting the partial derivatives equal to zero. The resulting equations are very complicated, however, and a simultaneous solution for n_u and p will require numerical methods of analysis. In the following, we shall concentrate on the optimal choice of n_u for a fixed value of p, the number of PSUs, and then explore through examples, the effect of changes in p. We shall be especially interested in a p of about 100 because in typical surveys, as alluded to above, this may be a practical maximum that an organization can support in ongoing sampling operations. From the other side, there may be nonstatistical pressure to have at least that number of PSUs in order to satisfy demands for information from regional interest groups. In certain other survey situations, there may be reasons for keeping p small. For example, in the evaluation of a government poverty program with serious political implications, it may be desirable to keep the study down to 5 or 10 metropolitan areas for reasons outside of the statistician's control.

Differentiating (3.12) with respect to n_u, and using (4.4) and (4.5) to define u and v, we set

$$vdu - udv = 0 \quad (4.6)$$

The result is a quadratic equation in n_u,

$$an_u^2 + bn_u + c = 0 \quad (4.7)$$

where

$$a = p^4 + 2p^3 v_c(H'_{bb} + H'_{ub}) + p^2 v_c^2 (H'_{bb}+H'_{ub})^2 (1 - [v_b/v_u] q_b), \quad (4.8)$$

$$b = 2p^4 q_c + 2p^3 [2v_c q_c (H'_{bb}+H'_{ub}) - v_b q_b H'_{bb} - q^*] + 2p^2 v_c [(v_c q_c - v_b q_b)(H'_{bb}+H'_{ub})^2 - 2q^*(H'_{bb}+H'_{ub})] - 2pv_c^2 q^* (H'_{bb}+H'_{ub})^2, \quad (4.9)$$

and

$$c = p^4 q_c^2 + 2p^3 q_c [v_c q_c (H'_{bb}+H'_{ub}) - v_b q_b H'_{bb} - q^*] + p^2 [v_b q_b (2q^* H'_{bb} - v_u H'^2_{ub}) + q^{*2} - 4v_c q_c q^* (H'_{bb}+H'_{ub}) + (v_c q_c [H'_{bb}+H'_{ub}] - v_b q_b H'_{bb})^2] + 2pv_c q^* [(H'_{bb}+H'_{ub})(q^* + v_b q_b H'_{bb}) - v_c q_c (H'_{bb}+H'_{ub})^2] + v_c^2 q^{*2} (H'_{bb}+H'_{ub})^2. \quad (4.10)$$

In (4.8), (4.9) and (4.10) the costs are expressed relative to k_u, the per unit cost of an unbiased observation:

$$q_b = k_b/k_u,$$
$$q_c = k_c/k_u, \quad (4.11)$$
and $$q^* = k^*/k_u.$$

In analyzing the effect of prior parameters on the optimal allocation between full probability and quota sampling for a fixed value of p, there is no loss in generality from assuming that $q_c = 0$. Thus, (4.9) and (4.10) are simplified to

$$b = -2p^3 (v_b q_b H'_{bb} + q^*) - 2p^2 v_c (H'_{bb}+H'_{ub}) [v_b q_b (H'_{bb}+H'_{ub}) + q^*] - 2pv_c^2 q^* (H'_{bb}+H'_{ub})^2, \quad (4.12)$$

and

$$c = p^2 v_b q_b (2q^* H'_{bb} - v_u H'^2_{ub} + v_b q_b H'^2_{bb}) + 2pv_c q^* (H'_{bb}+H'_{ub})(q^* + v_b q_b H'_{bb}) + v_c^2 q^{*2} (H'_{bb}+H'_{ub})^2. \quad (4.13)$$

Using (4.8), (4.12) and (4.13), it can be shown through considerable algebraic manipulation that $b^2 - 4ac$ is a perfect square, x^2, such that

$$x = 2p(v_b/v_u)^{1/2} q_b^{1/2} [p^2 v_u H'_{ub} + pv_c (H'_{bb}+H'_{ub}) (v_b q_b H'_{bb} + v_u H'_{ub} + q^*) + v_c q^* (H'_{bb}+H'_{ub})^2]. \quad (4.14)$$

Thus the two roots of (4.7), when $q_c = 0$, are given by

$$n_u = (-b \pm x)/2a, \text{ or} \quad (4.15)$$

$$n_u = [p^2 (v_b q_b H'_{bb} + q^* \pm v_b^{1/2} q_b^{1/2} v_u^{1/2} H'_{ub}) + pv_c (H'_{bb}+H'_{ub}) \{v_b q_b [(1 \pm v_b^{1/2} q_b^{1/2} v_u^{-1/2}) H'_{bb} + (1 \pm v_u^{1/2} v_b^{-1/2} q_b^{-1/2}) H'_{ub}] + (2 \pm v_b^{1/2} q_b^{1/2} v_u^{-1/2}) q^*\} + v_c^2 q^* (1 \pm v_b^{1/2} q_b^{1/2} v_u^{-1/2}) (H'_{bb}+H'_{ub})^2]/a^*, \quad (4.16)$$

where

$$a^* = p^3 + 2p^2 v_c (H'_{bb}+H'_{ub}) + pv_c^2 (H'_{bb}+H'_{ub})^2 [1 - (v_b/v_u) q_b]. \quad (4.17)$$

We shall call a solution to (4.15) "feasible" if it falls in the range [0, q*/p], where q*/p is the maximum allowed by the budget constraint. When one root is in the feasible range the other is either negative or exceeds q*/p, and the precision is maximized when n equals the feasible n_u. If both roots are greater than q*/p, the optimal sample allocation is $n_u = q^*/p$. If one root is negative and the other greater than q*/p, the best allocation is $n_u = 0$. The number of observations to be selected by quota methods in each PSU is from (4.2) and (4.11)

$$n_b = (q^* - n_u p)/(pq_b). \quad (4.18)$$

SENSITIVITY ANALYSIS

The complexity of expressions (4.16) and (4.17) makes any formal analysis of the effects of prior parameters on n_u a near impossible task. We shall, however, examine some interesting special cases, and consider a few examples:

a. The case of $p = 1$, $v_c = 0$.

This is the model of Pratt, Raiffa and Schlaifer.

From (4.16) and (4.17) we have

$$n_u = q^* + v_b q_b (H'_{bb} \pm v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}} H'_{ub}). \quad (6.1)$$

We must consider the following possible situations:

(1) $H'_{bb} > v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}} |H'_{ub}|$, in which case both roots of (4.7) are greater than q^*, the total budget. The optimal allocation is to put all resources into full probability sampling, $n_u = q^*$, $n_b = 0$.

(2) $v_b q_b (v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}} |H'_{ub}| - H'_{bb}) > q^*$, in which case one root of (4.7) is greater than q^* and the other is negative. The optimal allocation is to use the total budget for quota sampling;

$n_b = q^*/q_b$, $n_u = 0$.

(3) $0 < v_b q_b (v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}} |H'_{ub}| - H'_{bb}) < q^*$. The optimal allocation is

$$n_u = q^* - v_b q_b (v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}} |H'_{ub}| - H'_{bb})$$

and

$$n_b = v_b (v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}} |H'_{ub}| - H'_{bb}). \quad (6.2)$$

b. The case of $\tilde{\mu}_u$ and $\tilde{\beta}$ independent a priori.

$\tilde{\mu}_u$ and $\tilde{\beta}$ uncorrelated implies that $V'_{ub} = V'_{uu}$. Then from substitution of $(H'_{bb} + H'_{ub}) = 0$ in (4.16),

$$n_u = [v_b q_b H'_{bb} (1 \pm v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}}) + q^*]/p \quad (6.3)$$

Note that the PSU component of the process variance, v_c, has no effect on the sample allocation. (It does, however, have a marked effect on the posterior variance.) It is clear that

$$n_u = [v_b q_b H'_{bb} (1 - v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}}) + q^*]/p \quad (6.4)$$

is the only feasible root within the budget constraint, and that $n_u < q^*/p$ requires

$$v_u^{\frac{1}{2}} v_b^{-\frac{1}{2}} q_b^{-\frac{1}{2}} > 1. \quad (6.5)$$

It is instructive to assume that $v_u = v_b$ and write (6.4) as

$$n_u = (q^*/p) + [V'_{uu} v_b q_b^{\frac{1}{2}} (q_b^{\frac{1}{2}} - 1)]/(\det V')p. \quad (6.6)$$

We see that n_u is only slightly decreased through q_b by a reduction in the relative cost of quota sampling, whereas a large v_b or high prior correlation between $\tilde{\mu}_u$ and $\tilde{\mu}_b$ can have more direct impact. With the assumption $v_u = v_b$, we can express v_b and V'_{bb} in terms of V'_{uu}:

$$v_b = c_1 V'_{uu}$$

$$V'_{bb} = c_2 V'_{uu}. \quad (6.7)$$

It follows from the case specification, $V'_{uu} = V'_{ub}$, that

$$\rho_{ub} = c_2^{-\frac{1}{2}}, \quad (6.8)$$

and from (6.6)

$$n_u = (q^*/p) + \rho_{ub}^2 c_1 q_b^{\frac{1}{2}} (q_b^{\frac{1}{2}} - 1)/(1-\rho_{ub}^2)p. \quad (6.9)$$

Hence the optimal allocation is invariant given q^*, p, q_b, the ratio of $v_b = v_u$ to V'_{uu}, and the prior correlation, ρ_{ub}, between $\tilde{\mu}_u$ and $\tilde{\mu}_b$.

We also see that the optimal n_u, relative to its maximum possible value, is

$$1 + \rho_{ub}^2 c_1 q_b^{\frac{1}{2}} (q_b^{\frac{1}{2}} - 1)/(1-\rho_{ub}^2) q^*, \quad (6.10)$$

which does not depend on p.

In order to examine this case further, in Table 2 we display the values of n_u that are optimal for various combinations of c_1 and ρ_{ub}, given

$$p = 100, \quad q_b = .1, \quad q^* = 1600,$$

$$V'_{uu} = V'_{ub} = 1. \quad (6.11)$$

$$\rho_c = v_c/(v_c + v_b) = 0.$$

The magnitudes of V'_{uu} and ρ_c do not affect the allocation given by (6.9) but they are necessary in order to evaluate the posterior variance. For the problem of estimating the proportion of a dichotomous population possessing one of the attributes $V'_{uu} = 1$ could well represent an approximate "vague" or diffuse prior for $\tilde{\mu}_u$ known to lie in [0, 1].

The $q^* = 1600$ is a typical total sample size for a household survey, spread over 100 PSUs.

Table 2 shows that in the example of the dichotomous population, where it is unlikely that $v_b = v_u$ would be much greater than .1, one must have a prior ρ_{ub} beyond the range of the table for any quota sampling to be desirable.

Table 3 shows the posterior precision relative to the optimum if the survey planner were to ignore Table 2 and devote all of his resources to quota sampling, with $n_b = 160$ ($q_b = .1$).

Consider the example of $c_1 = 0.1$ and $\rho_{ub} = 0.95$. $V'_{uu} = 1$ implies $v_b = v_u = 0.1$ and $V'_{bb} = 1.10803$. It can be shown in (3.11) - (3.14) that with the optimal allocation, $n_u = 16$, the posterior standard deviation of $\tilde{\mu}_u$ is

$$V''^{\frac{1}{2}}_{uu} = 0.0079.$$

The relative precision of full quota sampling, shown in Table 3 as 0.001 is, more exactly, equal to

0.000641.

2. VALUES OF n_u DETERMINED BY $c_1 = v_b/V'_{uu}$ AND ρ_{ub}.
($p=100$, $q_b=.1$, $q^*=1600$, $v_b=v_u$, $\rho_c=0$, $V'_{uu}=V'_{ub}=1$, i.e. $\rho_{u\beta}=0$.

c_1	ρ_{ub} .100	.300	.500	.700	.800	.900	.950	.990	.995	.999
.01	16	16	16	16	16	16	16	16	16	16
.1	16	16	16	16	16	16	16	16	16	16
.3	16	16	16	16	16	16	16	16	16	16
5	16	16	16	16	16	16	16	16	16	15
.7	16	16	16	16	16	16	16	16	16	15
1	16	16	16	16	16	16	16	16	16	15
10	16	16	16	16	16	16	16	15	14	5
25	16	16	16	16	16	16	15	13	11	0
100	16	16	16	16	16	15	14	5	0	0
1000	16	16	15	14	12	7	0	0	0	0

3. PRECISION OF TOTAL QUOTA SAMPLING RELATIVE TO OPTIMUM ALLOCATION SHOWN IN TABLE 2

c_1	ρ_{ub} .100	.300	.500	.700	.800	.900	.950	.990	.995	.999
.01	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
.1	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.006	0.031
.3	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.009	0.019	0.093
.5	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.016	0.031	0.145
.7	0.000	0.000	0.001	0.001	0.001	0.002	0.004	0.022	0.044	0.195
1	0.001	0.001	0.001	0.001	0.002	0.003	0.006	0.031	0.062	0.265
10	0.006	0.007	0.008	0.012	0.017	0.033	0.063	0.265	0.456	0.967
25	0.016	0.017	0.021	0.030	0.043	0.081	0.148	0.532	0.781	1.000
100	0.059	0.065	0.078	0.115	0.162	0.271	0.457	0.966	1.000	1.000
1000	0.388	0.422	0.482	0.625	0.758	0.951	1.000	1.000	1.000	1.000

Thus, if the optimal allocation is not obeyed, and maximum quota sampling blindly applied, the posterior standard deviation of $\tilde{\mu}_u$ will be 0.312 in spite of the seemingly high prior correlation, $\rho_{ub} = 0.95$. Positive intra-PSU correlation, i.e., $\rho_c>0$, will increase the relative efficiency of total quota sampling; but does not reduce the posterior standard deviation. For example, with $\rho_c = .5$, for the $c_1 = 0.1$, $\rho_{ub} = 0.95$ combination with optimum n_u,

$$V''^{\frac{1}{2}}_{uu} = 0.0326,$$

whereas if full quota sampling were employed the efficiency would be 0.0108 with

$$V''^{\frac{1}{2}}_{uu} = 0.313,$$

a negligible change from the case of $\rho_c=0$.

In Tables 2 and 3, consider an example where some quota sampling is desirable, as with $c_1= 10$ and $\rho_{ub}= 0.999$. Table 3 shows that for that combination (with $\rho_c = 0$) total quota sampling has an efficiency of 0.967. But the main idea in quota sampling is to save money, hence we might ask, "What if the budget q^* were cut in half, making possible only 8 full probability or 80 quota observations per PSU?" Comparison of variances shows that with total quota sampling at half the budget, the precision will be 0.782 times the optimum at full budget. If, however, the new reduced budget were to be totally devoted to full probability sampling, the efficiency would be only 0.205. Since 8000 observations is a very large sample, we might consider working with one-eighth of the original budget or $q^* = 200$, yielding 20 quota observations per PSU. In that case the efficiency drops to 0.365, but the posterior standard deviation of $\tilde{\mu}_u$ is still as small as 0.083.

c. $\tilde{\mu}_u$ and $\tilde{\beta}$ correlated, $\rho_c > 0$.

In the more general situation of $|\rho_{u\beta}| > 0$, (i.e., $V'_{uu} \neq V'_{ub}$), and $\rho_c > 0$, we must use formulas (4.16) and (4.17) for n_u, and the simplicity and invariance of the previous case are lost. In the full report of this work we examine several examples that illustrate this case, but restrictions of space do not permit discussion here.

CONCLUSION

The examples in this paper, as well as many others examined by the author by means of the time-sharing computer, indicate that the optimal use of quota sampling, even with a small number of PSUs, requires very high prior correlations between $\tilde{\mu}_u$ and $\tilde{\mu}_b$ and large variances of the measurement processes. Relative cost appears to have little impact on the results. It should come as no great surprise that in order to justify quota sampling, one ought to believe the process means to be correlated, but just how high a correlation is necessary may be of some interest.

Two final remarks:

1. First, we have discussed a model in which one of the measurement processes is unbiased, i.e. μ_u can be estimated without systematic error. In the light of increasing difficulties in eliciting response in household surveys, even under the best conditions, one wonders how many survey researchers would accept the realism of the model.

2. Secondly, our model has not really encompassed the case that leads to certain uses of quota sampling, especially in the surveying of current public opinion. Here quota methods are employed because there simply is not enough calendar time to get high response through follow-ups. In terms of our model, k_u is infinite (or at least greater than $k^*/_p$). Political pollsters are able to demonstrate the accuracy of their estimates by comparison with actual election returns, and their records are impressive. Can sociologists, urban planners, and market researchers who do not face equally severe timetables be sure of the magnitudes and directions of the biases in their use of quota sampling?

Bayesian methods are valuable, not because anyone seriously believes that a prior distribution is easy to assess, nor that it is psychologically stationary,but rather because these methods help the decision maker to better understand the implications of actions that he might otherwise choose out of habit, convenience, or questionable advice.

A METHOD FOR ESTIMATING NET MIGRATION AND NET EMPLOYMENT EFFECTS OF PROJECT LOCATION USING POPULATION CENSUS DATA

G. K. Kripalani, Western Michigan University

INTRODUCTION

The location of a project in an area affects job opportunities facing different sex and occupational groups differently, depending upon the nature of industries affected by project location and the sex/skill composition of labor demanded in the affected industries. Change in job opportunities affects net migration to or from the area and consequently the population in the area. Such changes have an impact on the prospective or potential unemployment in the area. This impact is, however, different from the total change in employment due to project location on account of leakage of new job opportunities to potential or prospective net migrants. Further, net migration response to changes in job availability varies according to age, sex and type of labor involved. Hence, for studying employment and population changes as a consequence of the location of a project in an area, labor must be subdivided into reasonably broad homogeneous groups by sex, occupation and age.

The estimation procedure is centered on the assumptions that (a) human resources are less than perfectly mobile, (b) there exist net migration response differentials by age and sex, (c) migration response to changes in jobs is restricted to labor force members (job seekers) only, (d) non-labor-force members of the population do not migrate except as spouses and dependent children of migrating labor force members, (e) desired employment participation is a parameter which is unaffected by changes in job opportunities, and (f) the probability of a prospective unemployed person to become employed as job opportunities expand is a function of sex and age. The procedure analyzes actual employment data for two adjacent census points to study area's labor market's past allocation patterns of new job vacancies to determine relative preferences for various age and sex groups. These relative preferences and estimates of job seekers by age and sex provide a basis for deriving relative probability coefficients. Estimates of net migration response differentials are then used to estimate the 'leakage' of jobs to prospective net migrants and, hence, to determine net employment, net migration and net population effects of project location or withdrawal.

It is hypothesized that net migration of a group is related to the excess population of the group, where excess population is defined as the difference between the actual population and the desired population relative to the actual level of employment available to the group. The desired population is defined as that population which, relative to the actual jobs available to the group, would have the desired employment participation. Net migration response coefficient of the group is equal to the proportion of the excess population that will net migrate. The concept of desired employment participation corresponds to labor force participation rate but is considered preferable to the latter on account of the subjective character of the latter and for other reasons not fully elaborated here.[1]

Consider a group i and a time interval (0, 1). If the group has survived population P_i^s at the end of the time interval at t = 1, employment E_i, the desired employment participation rate λ_i^*, then according to the above hypothesis, the number of net migrants of the group M_i will be given by

$$M_i = g_i(E_i/\lambda_i^* - P_i^s) \quad (1)$$

$$= g_i(E_i - E_i^*)/\lambda_i^* \quad (2)$$

where g_i denotes net migration response of the group and is equal to the proportion of the surplus population that will net migrate and $E_i^* = P_i^s \cdot \lambda_i^*$ denotes desired jobs relative to the group population P_i^s.

(1) may also be written as

$$M_i/P_i^s = g_i(E_i/P_i^s \lambda_i^* - 1) \quad (3)$$

or

$$\mu_i = g_i\left(\frac{E_i - E_i^*}{E_i^*}\right) = g_i\left(\frac{E_i}{E_i^*} - 1\right)$$

where $M_i/P_i^s = \mu_i$, the rate of net migration and $E_i^* = P_i^s \lambda_i^*$ = desired jobs. The real significance of this form of the model is that the net migration rate is proportional to the percentage deficit in jobs, a measure corresponding to 'potential' rate of unemployment in the without project situation.

If P_i is the actual population of the group at t = 1, the relationship between the actual population P_i and survived population P_i^s is given by

$$P_i = P_i^s(1 + \mu_i). \quad (4)$$

Substituting (3) into (4) we have

$$P_i = P_i^s(1 - g_i) + g_i E_i/\lambda_i^*. \quad (5)$$

g_i is a measure of the mobility of the group in question. $g_i = 1$ for a perfectly mobile group while $g_i = 0$ for a perfectly immobile group. When $g_i = 1$, $P_i = E_i/\lambda_i^*$ or the population at the end of the time interval at t = 1, is the desired population corresponding to the level of jobs at that time. For a perfectly immobile group $g_i = 0$, and the population at the end of the time interval at t = 1 is $P_i = P_i^s$ or the survived population.

A Property of the Proposed Model

A property of the proposed model relationship between net migration and surplus population (equation 1) or between net migration and job deficit (equation 2) is that when job opportunities available to a group increase as a result of the location of the project, net out-migration of the group is reduced or net inmigration increased; and hence in a net out-migration area a proportion of the new jobs is appropriated by those net migrants whose net out-migration has been withheld due to increased job opportunities in the area. An increase in job opportunities by ΔE_i would mean that the resultant net inmigration or withheld net out-migration is equal to

$$(6) \qquad \Delta M_i = g_i \Delta E_i/\lambda_i^*.$$

It is reasonable to assume that these net out-migrants who would have net out-migrated from the area in the without-project situation but who now stay in the area in the with-project situation would do so only when they have the desired employment participation in the area. In other words, the jobs appropriated by these net migrants out of the ΔE_i addition jobs are given by

$$(7) \qquad \Delta M_i \lambda_i^* = g_i \Delta E_i.$$

This means that the net reduction in the number of the unemployed in the area (assuming it to be a net out-migration, i.e., a job deficit area) is given by $(1 - g_i)\Delta E_i$ or the net reduction in the number of the unemployed per unit new job is given by

$$(8) \qquad 1 - g_i .$$

Now g_i is the coefficient of net migration response of a group. A perfectly mobile group has $g_i = 1$ and a perfectly immobile group has $g_i = 0$. The impact on the unemployed per unit job is thus greater for relatively more immobile groups than for these groups which are relatively more mobile.

Basis For Allocating Additional Jobs Due To The Project Among Age Groups

The preceding discussion on estimating the impact on the area's unemployed measured by the reduction in the number of the unemployed due to increase in job opportunities assumes that the additional jobs for the group as a result of the project is known. It is assumed here that total additional employment ascribable to the project (defined as the difference between employment in the with and without-project situations) is given in terms of sex separately for male and female categories.

Net migration response to job deficit or surplus population varies by age and sex and hence net employment effects of additional employment due to project will depend on how additional jobs for males and for females are distributed among age groups. The problem, therefore, is to find a suitable basis for allocating total additional jobs among age groups for each of these categories of labor.

The procedure outlined hereunder assumes that the employment, population, etc., situation in the project area in the without-project situation is given and known. The project is then superimposed on the area and its impact on variables of interest estimated. Further, the procedure outlined hereunder relates to an area which is assumed to be an area of net job deficit in both without-project and with-project situations. Appropriate procedures applicable to an area of net job surplus in the without-project situation or to an area of net job deficit in the without-project situation but which is expected to become an area of net job surplus in the with-project situation are not discussed here.

A critical question is involved in the choice of an appropriate method for allocating additional jobs among age groups. Does the labor market show differential preferences in allocating jobs by age or is the chance of a 'potential' (or prospective) unemployed person becoming employed the same for all age groups? Illustrative calculations with respect to some randomly chosen areas showed that the labor market displays significant selectivity differentials between age groups. Hence, the equal probability assumption was not considered reasonable. To estimate relative probabilities, a relative probability coefficient P_i is associated with each age i, so that the ratio of the probability of becoming employed of a job seeker in age group i to that of a job seeker in age group j is given by P_i/P_j. On this basis, the distribution of additional jobs among age groups will be based on the quantities $P_i S_i$ where

S_i = desired job vacancies or the number of job seekers and is equal to $E_i^* - E_i^h$ where

E_i^* = desired jobs by the age group i in the without-project situation:

E_i^h = number of persons in the age group i in continuous employment over the decade.

^oE_i = actual employment of the age group in the without-project situation.

In order to estimate P_i's, we may make use of available information on the actual distribution of job vacancies among age groups during the most recent decade, say (-1, 0) in the without-project situation. If ^oE_i and E_i^h are known, an appropriate basis for allocating additional jobs by age is given by the quantity

$$^oE_i^v = {}^oE_i - E_i^h$$

Since P_i's are relative, they may be so defined that

$$\sum_i p_i S_i = \sum_i S_i. \tag{9}$$

We have

$$^oE_i^v \Big/ \sum_i {}^oE_i^v = p_i S_i \Big/ \sum_i p_i S_i. \tag{10}$$

Hence since by definition $\sum_i p_i S_i = \sum_i S_i$, we have

$$p_i S_i = \sum_i S_i \cdot {}^oE_i^v / \sum_i {}^oE_i^v. \tag{11}$$

P_i's can now be calculated since all the quantities in the above relationship are known.

A look at equation (11) will show that distribution of jobs on the basis of $P_i S_i$ is the same as the distribution on the basis of job vacancies filled in decade (-1, 0) in the without-project situation viz. $^oE_i^v$ since the quantity $\sum_i S_i / \sum_i {}^oE_i^v$ is the same for all age groups and does not affect the distribution over age groups. Thus, a basis which drops the equal probability assumption leads to a basis which provides that the distribution of additional jobs due to the project over age groups should be on the basis of the actual experience of decade (-1, 0) in the matter of distribution of job vacancies.

The total number of additional jobs taken up in the with-project situation by prospective net migrants who otherwise, in the without-project situation, would have "net out-migrated" from the area is given by

$$\sum_s \sum_i \left\{ \Delta E_i\,(f) \cdot g_i\,(f) \right\} \tag{12}$$

where the double summation refers to sex and age.

The quantity (12) represents net migration effects in terms of leakage of additional jobs to prospective migrants. Thus, the net employment effect is equal to the difference between total additional jobs due to the project and the jobs lost by leakage to prospective net migrants.

The net migration and net population effects of project location in terms of numbers of people consist of two components viz (1) prospective migrants who take up some of the additional jobs and (2) their dependent migrants. This quantity (2) represents dependent net migration of spouses and children of associated withheld prospective primary net migrants. The procedure for estimating this component of dependent net migration is not being discussed in this paper.

The procedure outlined above is applied as an illustration to the State of Wisconsin for the year 1970. Since a high proportion of an area's job seekers is among the young age groups and since relative net migration response coefficients for young age groups are high, generally a very significant proportion of new jobs will be appropriated by potential (now withheld) net outmigrants and a relatively small proportion of new jobs will accrue to the potential unemployed of the area. Calculations based on 100 additional jobs in each sex category show the following results:

Based on Wisconsin State (Per 100 additional jobs in the relevant sex)

Jobs Accruing to:	Males	Females
(a) "Potential" Net Migrants (Leakage of Jobs)	91.7	83.3
(b) "Potential" unemployed of the area who would have remained in the area as unemployed	8.3	16.7

For reasons of space, actual application of the procedure is done with respect to Wisconsin Males only. See Tables 1 through 3.

Footnotes

[1] The concept of desired employment participation was developed by George S. Tolley in a Ph. D. dissertation (unpublished). See Jansen (1966). In the present study, desired employment participation by age and sex is assumed to be the corresponding U.S. national labor force participation rate.

References

Battelle Memorial Institute, Columbus Laboratories. 1966. Final Report on a Dynamic Model of the Economy of the Susquehanna River Basin.

Industrial Relations Research Institute. 1966. Retraining and Migration as Factors in Regional Economic Development, (September), Madison, University of Wisconsin.

Jansen, E. F., Jr. 1966. Employment participation Behavioral Relationships. Unpublished Ph. D. thesis, Department of Economics, North Carolina State University, Raleigh, N. C.

Kripalani, G. K., 1970. Structural Unemployment in the Evaluation of Natural Resource Projects, pp. 85-120, in G. S. Tolley (Ed.), "Estimation of First Round and Selected Subsequent Income Effects of Water Resource Investments." Published as a Report, IWR Report 70-1, under Contract No. DA49-129-CIVENG-65-11 by the U.S. Army Engineer Institute for Water Resources, Department of the Army Corps of Engineers, Springfield, Virginia, 1970.

Mazek, W. F., 1966. The Efficacy of Labor Migration With Special Emphasis on Depressed Areas. Working Paper CWR 2, June. Institute for Urban and Regional Studies, Washington University, St. Louis, Missouri.

Somers, Gerald G., 1954. Labor Supply for Manufacturing in a Coal Area. Monthly Labor Review, (December), 1327-1330.

Code to Expanded Symbols -- Notation Used

Column	Symbol	Significance
Table 1.1		
(1)	i	Stands for age group.
(2)	$P(i, m, 70)$	Population of males aged i in 1970.
(3)	$SR_{10}(i, m, 70)$	10-year survival rate applicable to males aged i in 1970
(4)	$\lambda^*(i + 10, m, 80)$	Desired labor force participation rate assumed for males aged (i + 10) in 1980.
(5)	$E^*(i + 10, m, 80)$	Desired jobs for males aged (i + 10) in 1980. = (2) x (3) x (4).
Table 1.2		
(2)	$E (i, m, 70)$	Employed males aged i in 1970.
(3)	$SR_{10}(i, m, 70)$	10-year survival rate (same as in col. (3) of Table 1.1).
(4)	$E^s_{10}(i, m, 70)$	1980 survivors of employed males aged i in 1970(assumed to be persons in continuous employment over the 1970-80 decade as also in 1980) = (2) x (3).
Table 1.3		
(2)	$E^* (i + 10, m, 80)$	Desired jobs for males aged (i +10) in 1980 (same as col. (5) Table 1.1)
(3)	$E^s_{10} (i, m, 70)$	Persons in continuous employment over the 1970-80 decade (Col. (4) of Table 1.2)
(4)	$JS (i + 10, m, 1980)$	Job seekers, males aged (i + 10) in 1980 (Col. (2) - Col. (3))
(5)	$P (i + 10, m, 1980)$	Relative probability coefficient (Relative probability of an unemployed male aged (i + 10) in 1980, becoming employed if additional male job opportunities arose in the area).
(8)	$g (i, m)$	Net migration response coefficient for males aged i (These coefficients are calculated by race, sex and age in other studies and have been commented upon in Kripalani, G. K. (1970) See References).

TABLE 1.1

WISCONSIN - MALES

Estimation of Desired Jobs in 1980, Assuming No Outmigration During 1970-80 Decade--

Age Group in 1970	Population 1970 (000's)	10-Year Survival Rate	Labor Force Participation Rate	Desired Jobs 1980 (000's)	Age in 1980
i	P (i, m, 70)	SR_{10}(i, m, 70)	λ^*(i + 10, m, 80)	(2) (3) (4)	i + 10
(1)	(2)	(3)	(4)	(5)	(6)
0- 4	195.4	.995	-	-	10-14
5- 9	235.1	.994	-	-	15-19
10-14	242.4	.987	.800	191.4	20-24
15-19	218.1	.981	.942	201.6	25-29
20-24	158.1	.980	.966	149.6	30-34
25-29	135.1	.978	.964	127.3	35-39
30-34	115.8	.970	.964	108.3	40-44
35-39	111.6	.954	.964	101.6	45-49
40-44	121.2	.929	.940	105.8	50-54
45-49	119.5	.889	.903	95.9	55-59
50-54	112.7	.832	.768	72.0	60-64
55-59	104.8	.759	.405	32.2	65-69
60-64	90.4	.663	.239	14.3	70-74
65-69	72.7	.548	.145	5.8	75-79
70-74	56.9	.431	-	-	80-84
75-84	63.1	-	-	-	85-94
85+	14.0	-	-	-	95+
4-5	86.3	.995	.181	16.1	14-15
6-7	93.2	.994	.436	40.4	16-17
8-9	97.6	.992	.539	52.2	18-19

Sources: Col. (2) and Col. (4): 1970 Census of Population. Col. 13: Vital Statistics of U.S., Annual 1967.

TABLE 1.2

WISCONSIN - MALES

Estimation of Survivors in 1980 of Those Employed in 1970 (= Persons Assumed to be in Continuous Employment Over 1970-80 Decade--

Age Group	Employed in 1970 (000's)	10-Year Survival Rate	1980 Survivors of Those Employed in 1970 (000's)	Age in 1980
i	E(i, m, 70)	SR_{10}(i, m, 70)	(2) . (3)	i + 10
(1)	(2)	(3)	(4)	(5)
14-15	16.4	.981	16.1	24-25
16-17	34.2	.981	33.6	26-27
18-19	46.1	.981	45.2	28-29
20-24	114.5	.980	112.2	30-34
25-29	122.5	.978	119.8	35-39
30-34	108.1	.970	104.9	40-44
35-39	105.5	.954	100.6	45-49
40-44	114.2	.929	106.1	50-54
45-49	110.6	.889	98.3	55-59
50-54	103.5	.832	86.1	60-64
55-59	92.1	.759	69.9	65-69
60-64	67.8	.663	45.0	70-74

(1)	(2)	(3)	(4)	(5)
65-69	28.2	.548	15.5	75-79
70-74	13.2	.431	5.7	80-84
75-84	7.6	-	-	85-94
85+	1.1	-	-	95+
4-5	0	.995	0	14-15
6-7	0	.994	0	16-17
8-9	0	.992	0	18-19

Sources: Col. (2) 1970 Census of Population. Col. (3) Vital Statistics of U.S., Annual 1967.

TABLE 1.3

WISCONSIN - MALES

Estimation of (a) Percent Distribution of 100 Additional Jobs by Age and (b) Jobs Taken up by Prospective Net Outmigrants and (c) Net Employment Effects. (Per 100 Additional Male Jobs.)

(Per 100 Additional Male Jobs)

Age in 1980	Desired Jobs 1980 (000's)	Persons in Continuous Employment 1970-80	Job Seekers 1980	Relative Probability Coefficient	Relative Weight for 100 Additional Jobs by Age	Percent Distribution of 100 Male Additional Jobs	Net Migration Response Coefficient	Jobs Taken by Potential Withheld Net Migrants (Labor Force)	Net Employment Effects
(i + 10)	@	@	@	@	(4).(5)	(6)/2.38	g(i,m)	(7) (8)	(7) -(9)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
14-15	16.1	0	16.1	.80	12.9	5.00	.9249	4.92	0.08
16-17	40.4	0	40.9	.89	36.0	13.95	.9249	12.90	1.05
18-19	52.2	0	52.2	.60	31.3	12.13	.9249	11.22	0.91
20-24	191.4	8.1	183.3	.58	106.3	41.20	.9249	38.11	3.09
25-29	201.6	86.9	114.7	.45	51.6	20.00	.9027	18.05	1.95
30-34	149.6	108.4	41.2	.34	14.0	5.43	.8806	4.78	0.65
35-39	127.3	115.5	11.8	.26	3.1	1.20	.8204	0.98	0.22
40-44	108.3	101.1	7.2	.18	1.3	0.50	.8204	0.41	0.04
45-49	101.6	96.0	5.6	.13	0.7	0.27	.5886[a]	0.16	0.11
50-54	105.8	99.7	6.1	.07	0.4	0.16	.5886[a]	0.09	0.07
55-59	95.9	88.8	7.1	.04	0.3	0.12	.4716[b]	0.06	0.06
60-64	72.0	66.1	5.9	.01	0.1	0.04	.4716[b]	0.02	0.02
65+	52.3	41.3	11.0	-	0	0	0	0	0
Total	1,314.5	811.9	502.6	-	238.0	100.0	-	91.70	8.30

[a] Based on age group 45-54.

[b] Based on age group 55-64.

Sources: Col. (5): Estimated by graphical curve fitting to relevant observations by a procedure outlined in Corps of Engineers Study edited by Dr. G. S. Tolley. See G. K. Kripalani (1970) in "References."

Col. (8): See explanation under "Code."

@ For Cols. (2), (3), (4), (5) See "Code."

STOCHASTIC INDICATORS OF OCCUPATIONAL MOBILITY, CANADA, 1951 - 1961

P. Krishnan and A. Sangadasa, University of Alberta

1. Introduction

A society is usually characterized as 'open' or 'closed' depending on the scope for mobility for every member of its population. Social mobility is a complex phenomenon constituted of a variety of interdependent factors. Occupational mobility is considered, by and large, as a first approximation of social mobility. In view of this practice, the present paper confines its attention to occupational mobility alone.

Various studies on occupational mobility in the western countries have been reviewed by Lipset and Bendix (1962). One of the generalizations arrived at by these authors is that the industrialized nations of the West are marked by a high degree of mobility as measured by the shift across the manual-non manual line and furthermore, there is relatively little difference in rates of mobility. It is also shown that the mobility patterns in western industrialized societies are determined by the occupational structure. A recent study on the United States by Blau and Duncan (1967) tries to look at the various determinants of mobility with data gathered at the time of 1960 Census. As far as Canada is concerned, there has been a dearth of studies in this area. Porter (1965) in his classic work The Vertical Mosaic analyzes occupational mobility in Canada as could be inferred from the data provided by the censuses.

> We might conclude that although there has been a transferring of workers from manual to non-manual occupations, it is questionable that all of this shift represents upward mobility from "lower-level" manual occupations. It seems that these lower white collar occupations have been filled more by the native labor force than by immigrants and that the shift has provided, at best, a questionable mobility for the native born.
>
> (Porter, p. 52)

Elsewhere the author writes:

> At each period of industrial growth, as new opportunities for upward mobility appear, each increment of skilled and professional roles is filled in part by immigration, in part by the Canadian-trained, and in part by upgrading. The same sources must replace those who leave the labour force for various reasons. There obviously has been some mobility for Canadian industrial workers, but there is little doubt that there could have been much more ....
>
> (Porter, p. 56)

Because of lack of sophistication in his methodology, Porter is not able to provide the readers with an estimate of the degree of mobility in the Canadian Society. This paper is directed toward that end by employing a stochastic process analysis of the mobility process.

2. Mobility as a Stochastic Process

Various attempts have been made in the last two decades at stochastic process modelling of mobility (Blumen, Kogan, and McCarthy, 1955; Joshi, 1956; Kemney and Snell, 1960, Hodge, 1966). An overview of the work done so far can be found in Bartholomew (1973). Since finite Markov chain approximation of the mobility process was not satisfactory, McGinnis (1968) re-examined the substantive basis of the models and has suggested the incorporation of the principle of cumulative inertia. Verification of this principle has been done in Land (1969), Morrison (1967) and Myers, McGinnis, and Masnick (1967). Models of social mobility incorporating this principle have been discussed by Henry, McGinnis, and Tegetmeyer (1972) and McGinnis and Henry (1973). Recently a semi-Markov approximation of the mobility process has been suggested (Ginsberg, 1972).

One has to collect very detailed data in order to test the goodness of fit of the above modified models. Very often we cannot do it for economic or for administration reasons. Insofar as one has to work with non-panel data, it is feasible to develop only simple models.

3. Stochastic Indicators of Occupational Mobility

Consider S_i (i = 1, 1,2...h) - the non-overlapping occupational categories - to constitute the state space. Let the mobility process be a Markov process in discrete time. If stationarity of the process is assumed, then the model is easy to handle. The stationarity assumption may be easily relaxed, if deemed necessary. Let $P = (P_{ij})$ be the matrix of transition probabilities. The elements of P are the stochastic indicators of occupational mobility. The major problem with which we are faced is the estimation of these probabilities when panel data are not available and one has access to only marginal distributions (macro data).

4. Estimation Procedure of the Stochastic Indicators

Since only macro data from the censuses are available, we are forced to make a few assumptions. The Miller OLS procedure (1952) leads to inadmissible estimators of transition probabilities. A way out is to impose constraints on parameters. These are:

a) non-negativity condition

$$P_{ij} \geq 0 \quad (i, j = 1,2...h)$$

b) row condition

$$\sum_{j=1}^{h} P_{ij} = 1 \quad (i = 1,2...h)$$

The criterion of minimization employed leads to two different types of estimators. If the criterion is minimizing the sum of absolute deviations (MAD), we get a Linear Programming Problem (LPP) estimator of the transition matrix (Rogers, 1968; Lee, Judge and Zellner, 1970). If the restricted least squares (RLS) technique is employed, we obtain the Quadratic Programming Problem (QPP) estimator (Lee, Judge, and Zellner, 1970). The QPP estimator is shown to be more efficient than the LPP solution (Lee, Judge and Zellner, 1970). In this paper we employed the LPP solution only for developing estimates of occupational mobility indicators for Canada.

Usually the marginal distributions utilized in estimation procedures are temporal in nature.

We do not have enough time series data on occupational distributions in Canada and, furthermore if the available data are put to use, the stationarity assumption is likely to be violated, for the occupational change patterns before and after 1951 are, for certain, not the same. So we have used the cross-section data on the regions for the census years 1951 and 1961 for estimation purposes. The rationale underlying this manipulation is that the different provinces reflect the occupational transitions in Canada over a period of time with the transitions governed by the same rule.

5. Data and Findings

The percent distributions by occupational categories, suitably collapsed, employed for this exercise are shown in Appendix Tables 1 and 2. The male and female distributions are treated separately. The estimated ten-year transition probabilities are shown in Tables 1 and 2. In each case the functional values are the minimum possible as the programming technique is employed.

6. Male Mobility Patterns

Now we interpret the results from a substantive point of view. As far as males are concerned, the greatest movement is noticed in the primary sectior. While 67 per cent of the workers in the agriculture sector continue to stick to this occupation, 33 per cent move to other jobs in this ten-year period. Blue collar sector accounts for 14 per cent of the movement, white collar jobs 9 per cent, transportation sector 6 per cent and the remaining 4 per cent to service and recreation occupations. The least movement (about 9 per cent) is noticeable in the White Collar sector. It is surprising to note that most of the movement from this group is to the primary group. This may be due to combining "not stated" category with primary occupations. From the transport and communications group, the movement of 29 per cent is to blue-collar jobs. 14 per cent of the blue collar workers move to white collar jobs and 3 per cent to the transport and communications sector. The service and recreation occupations sector seems to force everybody to stick to itself.

Thus, as far as males are concerned, the 1951-1961 period noticed an average movement of 17.6 per cent of workers from their 1951 occupational categories. (If the service and recreation sector is ignored, the average movement works out as 22.0 per cent). If movements to white collar jobs and/or staying up there is taken to indicate "upward mobility", of the males in Canada 22.8 per cent experience upward mobility in this span.

7. Female Mobility Patterns

The mobility patterns of the Canadian females are more interesting. The movement out of the primary sector is indeed small (about 2 per cent). The movement out of service and recreation occupations is considerable (about 50 per cent) and most of this transition is to white collar jobs (32 per cent). The black to white collar change is 12 per cent (almost as high as that of males). But surprisingly enough, 20 per cent of the white collar job workers move out (down), mostly to service and recreation occupations (18 per cent). No female sticks to 'transport and communication' sector.

An overall average extent of mobility is 40.2 per cent in the ten-year span. Taking upward mobility as indicated by moving to white collar jobs and/or staying up there, on an average 37.8 per cent of the Canadian females were upwardly mobile in the ten-year span.

Thus, we are led to conclude that the females are generally more mobile and particularly more upwardly mobile in Canada in 1951-61.

8. Conclusions

This study has revealed that in 1951-61, the Canadian females have higher rates of mobility than the males. Since the earlier periods have not been studied, it is difficult to say whether the mobility has increased over time or not. But it could be said that the period 1961-71 and in the future too, female mobility is likely to increase in view of the women's rights movements and associated activities.

It has not been possible for us to study the mobility patterns in different provinces, between major urban concentrations, and between immigrants and non-immigrants. To gain such an understanding of the Canadian society, more research in this area is called for. This may help understand the society better and pave the way for major policy decisions.

References

Bartholomew, D.J. (1973). Stochastic Models for Social Processes (second edition) New York: John Wiley and Sons.

Blau, P.M., and O.D. Duncan (1967). The American Occupational Structure. New York: John Wiley and Sons.

Blumen, I., M. Kogan, and P. McCarthy (1955). Industrial Mobility of Labor as a Probability Process. Ithaca: Cornell University Press.

Ginsberg, R.B. (1972), "Critique of probablistic models: Application of the Semi-Markov Model to Migration", Journal of Mathematical Sociology, 2, 63-82.

Henry, N.W., R. McGinnis, and H.W. Tegtmeyer, (1971), "A finite model of mobility", Journal of Mathematical Sociology, 1, 107-118.

Hodge, R.W. (1966), "Occupational mobility as a probability process", Demography, 3, 19-34.

Joshi, D.D. (1967), "Stochastic models utilized in demography", World Population Conference 1965, Volume III. New York: United Nations.

Kenmey, J., and J.L. Snell (1960). Finite Markov Chains. Princeton: Van Nostrand.

Land, K.C. (1969), "Duration of residence and prospective migration", Demography, 6, 133-140.

Lipset, S.M., and R. Bendix (1962). Social Mobility in Industrial Society. Berkeley: University of California Press (third printing).

Lee, T.C., G.G. Judge, and A. Zellner (1970). Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data. Amsterdam: North-Holland Publishing Co.

McGinnis, R. (1968), "A Stochastic Model of Social Mobility", American Sociological Review 33, 712-722.

McGinnis, R., and N.H. Henry (1972), "Some properties of a stochastic attraction model",

in Population Dynamics (ed. T.N.E. Greville), New York: Academic Press.
Miller, G.A. (1952), "Finite Markov process in Psychology", Psychometrika, 17, 161-167.
Morrison, P.A. (1967), "Duration of residence approach and prospective migration: The evaluation of a stochastic model", Demography, 4, 553-561.
Myers, G.C., R. McGinnis, and G. Masnick (1967), "The duration of residence approach to a dynamic stochastic model of internal migration: A test of the axiom of cumulative inertia", Eugenics Quarterly, 14, 121-126.
Porter, J. (1965), The Vertical Mosaic. Toronto: University of Toronto Press.
Rogers, A. (1968), Matrix Analysis of Interregional Population Growth and Distribution. Berkeley: University of California Press.

TABLE 1

Transition Probability Matrix

1951 - 1961 Canada Occupations --- Males

	(1)	(2)	(3)	(4)	(5)
(1)	0.91255	0.00000	0.00572	0.03170	0.05003
(2)	0.14057	0.83446	0.02497	0.00000	0.00000
(3)	0.00000	0.29143	0.70857	0.00000	0.00000
(4)	0.00000	0.00000	0.00000	1.00000	0.00000
(5)	0.08797	0.13946	0.05850	0.04093	0.67313

Legend: (1) = White Collar Workers
(2) = Blue Collar Workers
(3) = Transportation and Communication Occupations
(4) = Service and Recreation Occupations
(5) = Primary Occupations

TABLE 2

Transition Probability Matrix

1951 - 1961 Canada Occupations --- Females

	(1)	(2)	(3)	(4)	(5)
(1)	0.79677	0.00000	0.02459	0.17864	0.00000
(2)	0.12344	0.71053	0.00372	0.09322	0.06909
(3)	0.53658	0.13029	0.00000	0.16060	0.17253
(4)	0.42930	0.00000	0.03261	0.50229	0.03580
(5)	0.00000	0.00000	0.02055	0.00000	0.97945

Legend: (1) = White Collar Workers
(2) = Blue Collar Workers
(3) = Transportation and Communication Occupations
(4) = Service and Recreation Occupations
(5) = Primary Occupations

APPENDIX

Table 1

Percentage Distribution of the Labour Force 15 Years of Age and

Over for Canada and Regions - Males

Occupation Division	Canada		New Foundland		Maritime Provinces		Quebec		Ontario		Prairie Provinces		British Columbia	
	1951	1961	1951	1961	1951	1961	1951	1961	1951	1961	1951	1961	1951	1961
(1) White Collar Workers	25.3	30.3	16.8	23.1	19.2	21.4	25.8	30.9	28.4	33.0	21.7	26.6	27.0	31.6
(2) Blue Collar Workers	35.1	35.0	31.3	36.2	30.8	32.6	38.0	37.7	40.3	37.6	22.2	25.8	37.9	37.3
(3) Transport and Communcation Occupations	7.2	7.5	8.2	9.6	8.3	8.2	7.5	8.2	7.1	7.2	6.0	6.5	8.4	8.1
(4) Service and Recreation Occupations	6.5	8.5	5.4	7.5	7.7	13.0	5.8	7.5	7.0	8.6	5.4	7.5	8.7	9.7
(5) Primary Occupations	24.6	16.0	37.2	20.0	32.5	19.8	21.0	12.7	16.1	11.1	44.0	31.3	16.5	10.4
(6) Occupation Not Stated	1.3	2.7	1.1	3.5	1.5	2.1	1.9	3.0	1.0	2.5	0.6	2.3	1.4	3.0

Source: 1961 Census of Canada, Labour Force, Table 7
Note: Category (5) in the text is found by combining (5) and (6) of these tables

APPENDIX

Table 2

Percentage Distribution of the Labour Force, 15 Years of Age and

Over For Canada and Regions - Females

Occupation Division	Canada		New Foundland		Maritime Provinces		Quebec		Ontario		Prairie Provinces		British Columbia	
	1951	1961	1951	1961	1951	1961	1951	1961	1951	1961	1951	1961	1951	1961
(1) White Collar Workers	54.1	55.9	56.4	61.9	55.4	57.9	47.4	51.1	56.4	58.0	55.9	54.2	61.0	61.8
(2) Blue Collar Workers	18.1	12.8	7.1	4.8	10.2	8.5	26.2	19.4	19.8	13.5	8.2	6.0	9.1	7.0
(3) Transport and Communication Occupations	2.8	2.2	2.1	2.1	2.6	2.4	2.5	2.0	3.0	2.1	2.5	2.4	4.1	2.3
(4) Service and Recreation Occupations	21.1	22.4	32.9	27.6	28.9	27.2	20.4	21.1	17.6	21.3	25.6	24.0	23.0	23.8
(5) Primary Occupations	2.8	4.3	0.8	0.4	1.4	1.6	2.2	3.2	2.1	3.3	7.0	10.9	1.7	2.0
(6) Occupation Not Stated	1.1	2.4	0.7	3.1	1.4	2.3	1.3	3.2	1.1	1.8	0.8	2.5	1.2	3.0

Source: 1961 Census of Canada, Labour Force, Table 7
Note: Category (5) in the text is found by combining (5) and (6) of these tables

THE RANDOMIZED RESPONSE TECHNIQUE, THE INTERVIEW, AND THE SELF-ADMINISTERED QUESTIONNAIRE: AN EMPIRICAL COMPARISON OF FERTILITY REPORTS*

Karol J. Krótki and Bonnie Fox, University of Alberta

Literature Review

In 1965, Warner devised the "randomized response" data-gathering technique as an attempt to increase the cooperation of respondents asked personal, confidential, or otherwise "sensitive" questions in an interview. Essentially, Warner felt that both refusal bias and response bias (i.e. due to untruthful answers) would be reduced if each respondent's privacy was protected by a method which randomized the appearance of the sensitive questions and concealed from the interviewer the exact question being answered. Thus, answers would "furnish information only on a probability basis" (Warner, 1965: 63).

The Warner technique consists of asking a respondent "one of two questions of the form: (1) I am a member of group A, and (2) I am not a member of group A" (Campbell & Joiner, 1973: 229). Since the probabilities associated with the selection of either question are known (i.e. built into the randomizing device), assuming truthful answers, an unbiased maximum likelihood estimate of the true proportion of the population in the stigmatized group can be made.

Let

π = the true proportion of A (the stigmatized group) in the population

P = the probability the randomized device chooses the sensitive question (1-P for the other question)

λ = the proportion of yes answers

Then

$$\lambda = P\pi + (1-P)\ (1-\pi)$$

Abul-Ela et al (1967) showed that this survey method of randomized response could be used to estimate t population proportions, where at least 1 and not more than (t-1) of them are stigmatizing. Of course one would need (t-1) simple random samples and (t-1) sets of different questions.

In an effort to further increase respondent cooperation, Simmons suggested a modification of the randomized response technique, whereby respondents select one of two unrelated rather than related questions (Horvitz et al., 1969). Two independent samples would be needed in order to estimate the population proportions in the two noncomplementary groups, unless the proportion of the population in the group mentioned in the unrelated question is known beforehand. In the latter case, which is the logical method to use, the estimating equation is:

$$\lambda = P\pi_A + (1-P)\pi_Y$$

where:

π_A = the true proportion of the population with sensitive attribute A

π_Y = the true proportion of the population with non-sensitive attribute Y (which is unrelated to A)

Horvitz et al. (1969) used this unrelated question randomized response technique, with π_Y known, to estimate the incidence of "illegitimate" births. Their estimate was very close to that obtained from information on the birth certificates from which the sample was chosen. Greenberg et al. (1969), after comparing the variances for estimates made using the Warner technique and those based on the unrelated question (with π_Y known or unknown) type of the randomized response technique, show that use of the latter will most likely entail greater statistical efficiency than the Warner technique. Moors (1971) also, in discussing the optimum model for the two-sample (i.e., π_Y unknown) unrelated question randomized response technique, shows that the unrelated question model is preferable to Warner's. Even when $\pi_Y = 1/2$ (i.e., the worst choice of π_Y), the variance of the estimates based on the optimized unrelated question technique is less than that of estimates derived using Warner's related question method.

Undertaking a massive field test of the unrelated question randomized response method, Greenberg et al. (1970) estimated one-year incidence of abortion, lifetime incidence of abortion, and use of oral contraceptives among North Carolina women. Their estimate of abortions in the previous year was similar (although obviously not comparable) to a 1961 Chilean survey estimate (Abernathy et al., 1970). This sample also produced an estimate of use of the birth control pill which was similar to one based on a national survey (Greenberg et al., 1970). Using another sample, two estimates of lifetime abortions were made - one assuming π_Y known and another assuming it had to be estimated. Estimates of π_A were found to be more accurate and less variable when π_Y is known than when it must be estimated. Finally, answers to several questions about the randomized response method itself showed that about 2/3 of the respondents believed their friends would not truthfully answer directly asked abortion questions. Moreover, 60 per cent of this North Carolina sample believed that their friends would not suspect a trick in the randomized procedure; 76 per cent said they themselves were convinced that the technique protected their privacy (Greenberg et al., 1970).

Another development in the history of randomized response data-gathering came when Greenberg et al. (1971) pointed out that quantitative information can be estimated as well. They proceeded to use the technique to derive reasonable estimates of the mean number of lifetime abortions as well as mean income. Perhaps the most important contribution of this method, and one reported by Greenberg et al. in this study, is the miniscule refusal rate accompanying use of the method.

Most recently Folsom et al. (1973) developed a new randomized response design which improves efficiency when two samples are required because π_Y is not known beforehand. The method consists of using two nonsensitive alternate questions in conjunction with the sensitive question. The authors also show the variance increases (around 50 per cent) when a weight = 1/2 is used instead of an optimal weight and conclude that their new design will never be more efficient than the simple alternate question model when π_Y is known. The practical significance of w = 1/2 lies in the fact that a coin can be used as the randomizing device, with all the advantages implied in interviewing response when compared with the mysterious box and differently coloured beads, required by the Warner technique.

Growth of Alberta Families Study

A stratified cluster sample of 1045 Edmonton women, between 18 and 54 years of age, was the basis of a comprehensive fertility study carried out between Nov., 1973 and Feb., 1974. Of the 2300 dwelling units originally selected, 794 either contained no eligible respondent or were vacant; the final sample of 1045 women represents 69.4% of the remaining sampled households. This sample was divided into three interpolated sub-samples in order to test the general data-gathering effectiveness of the randomized response technique. Respondents were either asked all fertility questions in the interview which included questions on abortion, were asked the "sensitive" questions by the randomized response method, or were given an anonymous mail-back questionnaire containing questions identical to those asked under the randomized response technique.

The specifics of the randomized response technique we used generally comprise a replication of those employed in the North Carolina studies. After an introduction to the randomized response method, instructions, and definitions (e.g., of abortion), 352 women were given a clear plastic box containing (35) blue and (15) red balls. The seven pairs of sensitive and unrelated statements (e.g. "I was born in the month of June") were printed on a card which was also handed to the respondent, blue marks appearing next to the former and red marks next to the latter. Respondents were told to shake the box, while the interviewer sat at a distance, and to answer the question marked with the color of the ball appearing in the window constructed in the box.

It was thus possible to compare the estimates obtained using the randomized method with those acquired by means of the anonymous questionnaire on the following seven variables: one-year incidence of abortions, lifetime incidence of abortions; incidence of premarital sexual intercourse, premarital pregnancy, "illegitimate" children, premarital use of contraceptives, and premarital abortions. In the interview, information on abortions within the previous year and lifetime abortions was acquired, and the estimates could be compared with those based on the randomized response technique and the questionnaire.[1] The purpose of our exercise was twofold: 1. to compare the estimates based on the three methods, to attempt to determine whether people directly asked "sensitive" questions on abortion were likely to lie (i.e., estimate response bias and its reduction using the randomized response technique) and to evaluate the accuracy of the randomized response-based estimates; 2. to compare the response rates for "sensitive" questions asked the three different ways. We hoped to be able to assess the usefulness of the randomized response technique.

Findings

In order to estimate the population proportions with the non-sensitive characteristics comprising the unrelated randomized response questions, we used either census data or data obtained in the GAFS survey. Four of the seven unrelated questions involved the statement that the respondent was born in a certain month. We averaged those proportions born in the specified months for the years 1925, 1935, 1945 (Canada, DBS). Another statement, "I was living in the same dwelling unit five years ago," was estimated using census figures which characterized five-year non-movers by age, sex, and province (Canada, 1961). For the quantity question, "How many children does your best friend have?" the distribution of children of the women in our sample was used (although we simply used the zero-non-zero proportions). Actually, there is no need to introduce the additional variance caused by these calculations into the estimating procedure. As Greenberg et al. (1969) point out, answers to the unrelated question can be built into randomizing device.

Before comparing the estimates based on the three data-gathering methods, we checked the similarity of our three samples. As table 1 shows, the samples are not significantly different in terms of their basic social profiles. Differences in the estimated population proportions should not, therefore, be due to gross differences in the composition of the three samples.

Our estimates using the randomized response technique are obviously similar to those based on responses to the self-administered questionnaire. In table 2 are reported the estimates.[2] Confidence intervals at the 95% level were placed around all of the point estimates although it is recognized that some of the estimated proportions were close to zero, and thus the normal approximation to the binomial distribution may not be appropriate.[3] A test for statistically significant differences between all pairs of estimates revealed no significant differences at the .05 level between any of the randomized response-based estimates and those obtained from answers to the anonymous questionnaire.[4] Only the estimate of lifetime abortions based on the randomized response method and that based on the interview were significantly different in the statistical sense.

A comparison of the estimated proportions themselves reveals a greater tendency for women to report "sensitive" events if the questions are randomized than if they are part of a self-administered questionnaire, and a greater tendency in both of these cases than in the interview situation. The estimate of the incidence of life-time abortions is probably most important, since it allows a comparison of all three methods, and the estimated proportions are not as close to zero as those for one-year abortions. The three estimates of lifetime abortions point toward the conclusion that women tend to give incorrect information when questioned on personal issues, and feel safer answering truthfully in a randomized response situation. One could even venture the conclusion, based on the lifetime abortion estimates, that people feel safer answering randomized questions than those on anonymous questionnaires. On the other hand, if the self-administered questionnaire is assumed to be as private as the randomized response technique, the absence of a statistically significant difference between the estimate based on the questionnaire and that obtained by the interview prohibits the conclusion that substantial response bias was introduced in the interview. On the whole, the other estimates, based on the randomized response technique and the questionnaire, were close enough to confirm the reliability of the former method.

While we have not conclusively proven a reduction in response bias with the randomized response method, the data-gathering technique definitely increases response rates. The response rates using the randomized response method were substantially higher (97 and 95 per cent) than those resulting from use of the mail-back questionnaire (73 per cent). Since the former allows the researcher to gather as much information as he might in the traditional interview, the randomized response technique no doubt is more useful than the questionnaire for gathering "sensitive" information. The question remains, however, whether the randomized response "game" is necessary: is the introduction of additional variance justified by the reduction of bias due to an avoidance of "sensitive" questions?

Our replication of the North Carolina questions asked about the randomized response technique reveals some interesting insights to the question of the need for a "game" in order to gather confidential information. In direct contrast to the North Carolina women, 68 per cent of the Edmonton sample thought their friends would truthfully answer a direct question on abortion. However, they were not asked if their friends who had had an abortion would answer a direct question about it. Only 63 per cent of the women thought their friends would find no trick to the randomizing device. A person's judgement of her friend's likely behavior may be grossly inaccurate. However, only 58 per cent of our respondents said they were sure of the privacy guaranteed by the randomized response technique, while 28 per cent were not sure the interviewer did not know which question was being answered. Our estimation of the importance of the randomized response method was reduced in the face of the high proportion of respondents saying friends would answer a direct abortion question, and the sizeable proportion personally doubting the privacy assured by the randomized response technique. Interviewers estimated that younger, better educated Canadian women had no qualms about answering the questions directly, especially in the context of a comprehensive fertility interview. One perceptive interviewer reported that only a few of her interviewees did not indicate the question they were answering in the RRT "game."

Table 1. Comparison of the three samples

	RRT	Ques.	Interview	X^2	sig.
Proportion ever married:	.844	.823	.818	.78	.68
Proportion over 29 years:	.524	.566	.494	3.47	.17
Proportion pregnant fewer than four times:	.884	.855	.836	2.79	.24
Proportion of British Isles ethnicity:	.384	.373	.332	2.04	.36
Education - fewer than 9 yrs:	.113	.089	.085		
9 - 12 years:	.571	.529	.523		
>13 years:	.316	.382	.392	4.84	.30
N =	346	269	327		

Table 2. Comparison of the 95% confidence intervals around sample proportions based on the three data-gathering methods

	RRT	Questionnaire	Interview
1. Abortion in the past 12 mos:	.032 (±.032)	.008 (±.001)	.003 (±.003)
2. Abortion during lifetime:	.090 (±.068)	.038 (±.023)	.015 (±.013)
3. Unmarried, sexual intercourse:	.623 (±.076)	.605 (±.059)	
4. Unmarried, became pregnant:	.190 (±.055)	.213 (±.049)	
5. Unmarried, gave birth:	.078 (±.063)	.075 (±.032)	
6. Unmarried, used contraceptives:	.326 (±.067)	.281 (±.054)	
7. Unmarried, had an abortion:	.021 (±.030)	.034 (±.022)	
N =	342	269	327

ENDNOTES

*We wish to thank Dr. John Fox for valuable computer and statistical assistance. Other colleagues in the Department of Sociology, University of Alberta, helped with various aspects of the main Edmonton survey and its RRT part, notably Dr. P. Krishnan and Mr. Roderic Beaujot. The Growth of Alberta Families Study (GAFS) has been supported as Family Planning Project 4470-8-1 of Health and Welfare Canada.

1. Between the final typing of the interview questionnaire and its original version, a change was made inadvertently which unfortunately omitted the month from the question on the date of marriage and thus precluded a rigorous comparison of all three techniques on all the questions. However, it should be possible to rescue some of the comparative information through approximations and the "no less than" approach.

2. These estimated proportions are not weighted and therefore should not be considered real rates in a strict sense. However, we compared the weighted and unweighted distributions on such variables as age, education, marital status, number of pregnancies, and ethnicity and found that they were not significantly different from each other (indices of dissimilarity were all quite small).

3. The variance of the RRT estimate, as reported in Greenberg et al. (1969) is: var $(\pi_A \pi_Y) = \frac{\lambda(1-\lambda)}{nP^2}$ Although we did not have a simple random sample, we used this formula.

4. The variance of the difference between two independent random variables is the sum of their variances. We computed a z-score.

REFERENCES

Abernathy, James, Bernard Greenberg, Daniel Horvitz
1970 Estimates of induced abortion in urban North Carolina, Demography, 7: 19-29.

Abul-Ela, Abdel Latif, Bernard Greenberg, Daniel Horvitz
1967 A multi-proportions randomized response model, Journal of the American Statistical Association, 62: 990-1008.

Canada, Dominion Bureau of Statistics
1925, 1935, 1945 Vital Statistics, Ottawa.

Campbell, Cathy and Brian Joiner
1973 How to get the answer without being sure you've asked the question, The American Statistician, 27: 229-31.

Folsom, Ralph E., Bernard G. Greenberg, Daniel G. Horvitz, and James R. Abernathy
1973 The two alternate questions randomized response model for human surveys, JASA 68: 525-530.

Greenberg, Bernard, Abdel-Latif, Abul-Ela, Walt Simmons, and Daniel Horvitz
1969 The unrelated question randomized response model: theoretical framework, JASA, 64: 520-39.

________________, James Abernathy, and Daniel Horvitz
1970 A new survey technique and its application in the field of public health, Milbank Memorial Fund Quarterly, 48: 39-55.

________________, Roy Kuebler, Jr., James Abernathy, and Daniel Horvitz
1971 Application of the randomized response technique in obtaining quantitative data, JASA, 66: 243-48.

Horvitz, Daniel, B.V. Shah, Walt Simmons
1969 The unrelated question randomized response model, Proceedings of the Social Statistics Section, American Statistical Association, 65-72.

Moors, J.A.
1971 Optimization of the unrelated question randomized response model, JASA, 66: 627-29.

Warner, Stanley
1965 Randomized response: a survey technique for eliminating evasive answer bias, JASA, 60: 63-69.

Warner, Stanley L.
1971 "The linear randomized response model", JASA, 66: 884-888.

RATIO BIAS AND VARIANCE IN DUAL SYSTEM ESTIMATION

Eli S. Marks, U.S. Bureau of the Census

Dual system estimation involves using one data collection system to estimate the completeness of reporting in the other data collection system. One then divides the estimates from System 1 by the estimated completeness rates provided by System 2.

Dual system estimation has been used to correct for "coverage" biases in demographic statistics (particularly in vital statistics). While it can be very valuable for this purpose, substantial ratio biases as well as increases in variance can result if the completeness rate used in the denominator of the estimate has a large coefficient of variation (.2 or more). While ratio bias and variance can be kept at low levels for estimates based on all the sample cases, estimates for subgroups can be badly biased and quite unstable. The current paper discusses techniques for dealing with the problem of ratio bias and variance in subgroup estimation.

Dual System Estimation

A typical example of dual system estimation is the use of a sample survey to estimate and correct for incompleteness of birth and death registration. The birth and death registration may be a 100% system or the registration may be confined to a sample of areas used to provide estimates (birth and death rates) for the entire country or region. The survey is conducted in a sample of the registration areas and birth and deaths reported in the survey are matched to the birth and death registers and classified as "registered" or "not registered".

A similar technique can be used to estimate and correct for incompleteness of coverage in a census or household survey. In census evaluation, a sample of small areas is selected and independently reenumerated. Persons and households enumerated in the survey are matched to the census schedules and classified as "enumerated" in the census or "not enumerated".

The two data collection systems may attempt to cover the same sample of the population or, as is the case with a census/post-enumeration survey, the areas covered by one system may be a subsample of the areas covered by the other. In either case, every attempt must be made to keep data collection for the areas covered by both systems as independent as possible. That is, it is desirable for the cases reported in one system to have the same (expected) proportion of cases covered by the other system as for the entire population. The situation may be diagrammed as follows:

In System 2	In System 1: Reported	In System 1: Not Reported	Total
Reported	M	U_2	N_2
Not Reported	U_1	Z	$N-N_2$
Total	N_1	$N-N_1$	N

Here there are N cases in the population of which N_1 are reported in System 1, N_2 are reported in System 2, and M are reported in both systems. N and Z (number of cases not reported in either system) are unknown but, if the two systems are independent,

$$(1) \quad W_1 = \frac{M}{N_2} \doteq \frac{N_1}{N} = P_1 \text{ and}$$

$$(2) \quad W_2 = \frac{M}{N_1} \doteq \frac{N_2}{N} = P_2 .$$

That is W_1 (the proportion of System 2 reports also reported in System 1) is an estimate of P_1, the completeness of System 1 reporting; and W_2 is an estimate of P_2, the completeness of System 2 reporting. W_1 and W_2 involve only the known quantities M, N_2 and N_1 and can, therefore be used to estimate N and Z by the formulae:

$$(3) \quad \hat{N} = \frac{N_1N_2}{M} = \frac{N_1}{W_1} \doteq \frac{N_1}{P_1} = N$$

$$(4) \quad \hat{Z} = \frac{U_1U_2}{M} = \frac{N_1N_2}{M} - N_1 - N_2 + M$$

$$\doteq N - U_1 - U_2 - M = Z$$

The values above may be for the whole population or for a class. For example, we might estimate the completeness of reporting in a census of all persons or of persons in a class such as "gainfully employed males, aged 15 to 19".

Ratio Bias and Variance

The quantities M, N_1, and N_2 and the estimates W_1, W_2, $\hat{N}$, and $\hat{Z}$ obtain when both systems try to cover the entire population (or to cover all members of a class of the population). If one or both systems involves a sample, we would use sample estimates m, n_1, n_2, w_1, w_2, $\hat{n}$ and $\hat{z}$.

The estimates w_1, w_2, $\hat{n}$ and $\hat{z}$ involve a ratio of a random variate or a product of random variates to another random variate. These ratios are consistent estimates of the corresponding population values W_1, W_2, $\hat{N}$ and $\hat{Z}$. The relvariances of w_1, w_2, $\hat{n}$ and $\hat{z}$ are given approximately by the formulae:

$$(5) \quad V_r^2 = \sum_i^t V_{x_i}^2 + \sum_i^u V_{y_i}^2 + 2\sum_{i<j}^t V_{x_ix_j} + 2\sum_{i<j}^u V_{y_iy_j} - 2\sum_{i,j}^{tu} V_{x_iy_j}$$

where

$$r = \frac{x_1x_2 \ldots x_t}{y_1y_2 \ldots y_u}, \qquad V_{x_i}^2 = \frac{\sigma_{x_i}^2}{X_i^2}$$

$$V_{y_i}^2 = \frac{\sigma_{y_i}^2}{Y_i^2}, \qquad V_{x_ix_j} = \frac{\sigma_{x_ix_j}}{X_iX_j},$$

$$V_{y_iy_j} = \frac{\sigma_{y_iy_j}}{Y_iY_j}, \qquad V_{x_iy_j} = \frac{\sigma_{x_iy_j}}{X_iY_j};$$

$\sigma^2_{x_i}$, $\sigma^2_{y_i}$, $\sigma_{x_i y_j}$ etc. are the variances and covariances of x_i, x_j, y_i and y_j; and X_i, Y_j are the expected values of x_i and y_j.

Thus, the relvariances of w_1, w_2 and $\hat{n}$ are:

(6) $V^2_{w_1} \doteq V^2_m + V^2_{n_2} - 2V_{mn_2}$

(7) $V^2_{w_2} \doteq V^2_m + V^2_{n_1} - 2V_{mn_1}$

(8) $V^2_{\hat{n}} \doteq V^2_{n_1} + V^2_{n_2} + V^2_m + 2V_{n_1 n_2} - 2V_{mn_1} - 2V_{mn_2}$

The dual system estimates from a sample are:

(9) $w_{1a} = \frac{m_a}{n_{2a}}$

(10) $w_{2a} = \frac{m_a}{n_{1a}}$

(11) $\hat{n}_a = \frac{n_{1a} n_{2a}}{m_a} = \frac{n_{1a}}{w_{1a}} = \frac{n_{2a}}{w_{2a}}$

where

w_{1a} = estimated completeness of Source 1 reporting for the a^{th} Class

w_{2a} = estimated completeness of Source 2 reporting for the a^{th} Class

$\hat{n}_a$ = dual system estimate of the total number of cases in the a^{th} Class

n_{1a} = estimated total number of cases in Class a reported by Source 1

n_{2a} = estimated total number of cases in Class a reported by Source 2

m_a = estimated total number of cases in Class a reported by both sources.

"Class a" may consist of all persons or all births or all deaths or of some subset, such as "Negro males age 20-24" or "births where father is an employed white collar or skilled worker" or "deaths due to cancer of White females age 40 to 54". The C.V. of each component of w_{1a} or w_{2a} or $\hat{n}_a$ depends, of course, on the sample design and upon the nature of the reporting in both sources.

While w_{1a}, w_{2a} and $\hat{n}_a$ are consistent estimates of W_{1a}, W_{2a} and $\hat{N}_a$, they are biased estimates and this bias can be considerable when the denominators of the ratios (n_{2a}, n_{1a} and m_a) have large coefficients of variation. Similarly, Equations (6), (7) and (8) are approximations to the correct relvariances and the approximation may be poor for a large coefficient of variation of the denominator of the ratio. Usually, there will be very little correlation between n_{1a} or n_{2a} and the completeness estimates w_{1a} and w_{2a}. In consequence, w_{1a} and w_{2a} will usually not be subject to appreciable ratio bias but $\hat{n}_a$ and its relvariance estimate will be subject to approximately the biases involved in using the reciprocal (1/x) of a random variate as an estimate of the reciprocal (1/X) of its expected value. Table 1 gives these biases for varying levels of the coefficient of variation of x for the case where x is a normal variate. Since 1/x becomes infinite for x=0, the distribution was truncated for values of $x \leq .01X$.

Table 1

Relative Biases of Using 1/x as an Estimate of 1/X and V_x as an Approximation to $V_{1/x}$ Where x is a Normal Variate with Expected Value X.

C.V. of x	Relative Bias of 1/x	C.V. of 1/x	Relative Bias of Taking $V_{1/x} = V_x$
.025	.001	.025	*
.05	.003	.050	*
.10	.010	.103	-.03
.15	.024	.162	-.07
.20	.046	.234	-.15
.25	.08	.38	-.34
.30	.14	.78	-.61
.35	.24	1.19	-.71
.40	.35	1.65	-.76

Note: It is assumed that when $x \leq .01(Ex)$, the sample is ignored and a new sample selected for determining 1/x.

* Less than .001

For a simple random sample, we usually have:

(12) $\text{C.V.}\ (m_a) \doteq \frac{1}{\sqrt{m_a}}$

While use of a simple random sample is rare in dual system estimation, Equation (12) provides a lower limit for the C.V. of the samples actually used and this C.V. will rarely exceed twice the values given by Equation (12). In practice, samples are usually designed to give fairly large values of n_{1a}, n_{2a}, and m_a for the sample as a whole and for major classes of the population. Thus, a sample to estimate completeness of birth registration and the true number of births in a country will ordinarily be designed to contain at least 1000 sample births for major population classes and this usually means $n_{1a} \geq n_{2a} > m_a > 600$ for such classes. For such cases, the C.V. of m_a will usually be less than 0.1 and the relative bias of $\hat{n}_a$ and of its variance estimate will be small (see Table 1).

For smaller classes, the values of n_{1a}, n_{2a} and m_a are often quite small. For very small classes, the value of making an estimate (biased or unbiased) of $\hat{n}_a$ is questionable. There are cases, however, where m_a is around 25 to 100, for which an estimate of $\hat{n}_a$ with a C.V. of 10 to 20 percent would be acceptable but not in combination with a relative bias of 3 or 4 percent. Also, even for smaller classes for which individual estimates may not be important, it is frequently desirable to have some idea of the general distribution. For example, a sample might give dual system estimates subject to only minor ratio biases for total number of male births or female births and even for the distribution of births by sex over 2 or 3 regions of equal size; but might not give satisfactory dual system estimates of births by region and age of mother, even if one uses 5-year age groups.

Various ideas have been suggested for avoiding the ratio biases of $\hat{n}_a$ estimates. One is simply to use the better of the single system estimates, n_{1a} or n_{2a}. However, one only takes on the cost and problems of dual system estimates in situations where there is very serious non-reporting bias. For very small classes, the ratio bias may be greater than the response bias but there is a middle range (say $10 < m_a < 50$) in which the ratio bias is large enough to be troublesome but is still smaller than the response bias of the single system estimate.

Another solution to the problem of ratio bias is to apply to n_{1a} or n_{2a}, completeness rates estimated for some larger class -- for example, apply to the reported birth registrations by age of mother an estimate of the completeness of all birth registrations. This would mean estimates of the form:

$$(13)\quad \check{n}_{ab} = \frac{n_{1ab}}{w_{1a}} = \frac{n_{1ab}n_{2a}}{m_a}$$

where w_{1a}, n_{2a}, m_a are as defined above for the a^{th} class and

n_{1ab} = estimated total number of cases reported by Source 1 for the b^{th} subclass of Class a.

For example, n_{1ab} might be the sample estimate of number of male births registered for mothers age 15 to 19; n_{2a}, the number of male births estimated from a sample survey matched to the registrations; and m_a, the estimated number of sample survey births found (by the matching) to be registered. From Equation (5), the relvariance of $\check{n}_{ab}$ is:

$$(14)\quad V^2_{\check{n}_{ab}} = V^2_{n_{1ab}} + V^2_{n_{2a}} + V^2_{m_a} + 2V_{n_{1ab}n_{2a}} - 2V_{n_{1ab}m_a} - 2V_{n_{2a}m_a}$$

While estimates of the form $\check{n}_{ab}$ are not subject to the ratio bias of the regular dual system estimates they may suffer from the fact that the completeness of reporting for some subclasses of Class a may be much lower than the overall (average) completeness of Class a reporting and may be much higher for other subclasses. We may, in fact, have reasonably good evidence of such variation in completeness of reporting between subclasses within the same general class. For example, completeness of birth registration for both very young and very old mothers might be considerably worse than completeness of reporting for mothers in the middle range, as evidenced by matching with the registration only 40% of the sample survey births for mothers under 20 and over 45 while about 70 percent of the other reported sample survey births were found to be registered. To deal with this problem, Francis C. Madigan and Alejandro Herrin of the Mindanao Center for Population Studies have suggested using for each subclass, all cases reported in either source plus a proportional distribution of the dual system estimate of the cases missed by both sources. This means estimates of the form:

$$(15)\quad \hat{n}_{ab} = \frac{x_{ab}\hat{n}_a}{x_a} = \frac{x_{ab}n_{1a}n_{2a}}{x_a m_a}$$

where $\hat{n}_a = \frac{n_{1a}n_{2a}}{m_a} = x_a + \frac{u_{1a}u_{2a}}{m_a}$ = usual dual system estimate for Class a

$u_{1a} = n_{1a} - m_a$ = estimated number of Source 1 cases of Class a not matched in Source 2

$u_{2a} = n_{2a} - m_a$ = estimated number of Source 2 cases of Class a not matched in Source 1

$x_a = m_a + u_{1a} + u_{2a} = n_{1a} + n_{2a} - m_a$

= estimated number of cases of Class a reported in either Source 1 or Source 2

n_{1ab} = estimated number of cases of Subclass ab reported in Source 1

n_{2ab} = estimated number of cases of Subclass ab reported in Source 2

m_{ab} = estimated number of cases of Subclass ab reported in both Source 1 and Source 2

$u_{1ab} = n_{1ab} - m_{ab}$ = estimated number of Subclass ab cases reported in Source 1 but not in Source 2

$u_{2ab} = n_{2ab} - m_{ab}$ = estimated number of Subclass ab cases reported in Source 2 but not in Source 1

$$x_{ab} = m_{ab} + u_{1ab} + u_{2ab}$$

$$= n_{1ab} + n_{2ab} - m_{ab}$$

= estimated number of Subclass ab cases reported in either Source 1 or Source 2.

From Equation (5), the relvariance of $\hat{n}_{ab}$ is given by:

$$(16)\quad V^2_{\hat{n}_{ab}} = V^2_{x_{ab}} + V^2_{n_{1a}} + V^2_{n_{2a}} + V^2_{x_a} + V^2_{m_a}$$
$$+ 2V_{x_{ab}n_{1a}} + 2V_{x_{ab}n_{2a}}$$
$$+ 2V_{n_{1a}n_{2a}} + 2V_{x_a m_a}$$
$$- 2V_{x_{ab}x_a} - 2V_{n_{1a}x_a}$$
$$- 2V_{n_{2a}x_a} - 2V_{x_{ab}m_a}$$
$$- 2V_{n_{1a}m_a} - 2V_{n_{2a}m_a}$$

While Equation (16) looks fairly complex, the individual terms can be estimated fairly easily. Formulae for estimating the variances and covariances of n_{1a}, n_{2a} and m_a are given by Marks, Seltzer and Krotki (2). These formulae can be extended fairly easily to estimating the variances of x_a and x_{ab} and their covariances with each other and with n_{1a}, n_{2a}, and m_a. As noted by Tepping (3), the variance of $\hat{n}_{ab}$ can also be estimated by "calculating a linear combination of sample totals for each primary sampling unit and then estimating the variance of the sum of those combinations". That is, one would compute for each primary sampling unit, the variable:

$$(17)\quad y_{abi} = \hat{n}_{ab}\left\{\frac{x_{abi}}{x_{ab}} + \frac{n_{1ai}}{n_{1a}} + \frac{n_{2ai}}{n_{2a}} - \frac{x_{ai}}{x_a} - \frac{m_{ai}}{m_a}\right\}$$

where x_{abi}, n_{1ai}, n_{2ai}, x_{ai}, m_{ai} are estimates of X_{ab}, N_{1a}, N_{2a}, X_a and M_a derived from the i^{th} primary sampling unit only (i.e., are PSU totals weighted up to allow for the sampling probabilities). For a sample of k PSU's, we would have

$$(18)\quad \bar{y}_{ab} = \frac{\sum_i^k y_{abi}}{k} = \hat{n}_{ab}$$

and the estimated variance would be:

$$(19)\quad s^2_{\hat{n}_{ab}} \doteq s^2_{\bar{y}_{ab}} = \frac{\sum_i^k (y_{abi} - \bar{y}_{ab})^2}{k(k-1)}$$

Other methods of variance estimation are discussed by Tepping (3) and by Frankel (1).

1. Martin R. Frankel, Inference from Survey Samples, the University of Michigan, 1971, 173 pp.

2. E. S. Marks, W. Seltzer, K. Krotki, Population Growth Estimation, The Population Council, 1974.

3. Benjamin J. Tepping, "Variance Estimation in Complex Surveys", Proceedings of the Social Statistics Section 1968, American Statistical Association, Washington, pp. 11-15.

AN APPRAISAL OF THE MODEL TABLES OF VITAL RATES

S. Mitra, Emory University

1. The Models

The search for parametric models that can adequately describe a particular series of age-specific vital rates began in the Population Branch of the United Nations in the early 50's which resulted in the publication of a set of model life tables suitable particularly for underdeveloped countries (U.N., 1955). From 158 life tables of various countries and periods, the UN analysts noted that a second degree polynomial in ${}_nq_x$ provides excellent approximation of ${}_nq_{x+n}$ where the parameters defining the polynomial can be obtained by the method of least squares. This provided a tool to build up the entire ${}_nq_x$ series from an initial value of ${}_1q_0$ and the successive polynomial approximations. The estimated values of ${}_nq_x$ were then used to generate values of all life table functions.

Coale and Demeny created four categories of life tables (1966), on the basis of greater homogeneity in the patterns of mortality within each category that produced high correlations between successive ${}_nq_x$ values. For these categories labelled North, South, East and West, linear regressions of ${}_nq_x$ and log ${}_nq_x$ on $\overset{0}{e}_{10}$ were obtained by the method of least squares and both of these approximations were used to develop estimates of ${}_nq_x$ values, and thus the engire life table was obtained from an initial value of $\overset{0}{e}_{10}$.

Similar experimentation with age-specific fertility rates (Mitra, 1965), resulted in a set of model fertility tables. Here also, high correlations were observed among successive age-specific rates, between an age-specific fertility rate and the general fertility rate and also between an age-specific fertility rate and the sex age adjusted birth rate. Later, another series of model fertility tables was developed by using Pearson's Type I curve to describe the age-specific rates (Mitra, 1967). The latter model, unlike the former, depends upon more than one parameter, and it was shown that the number of independent parameters can be reduced from four to two or even to one, under certain reasonable assumptions.

2. The Problems With the Use of Model Tables

Although, the model tables were designed to provide estimates of age-specific vital rates, they were far from being flawless. As Gabriel and Ronen (1958) pointed out that estimates of $\overset{0}{e}_0$ from ${}_1q_0$ by graphic interpolation among the life tables used by the UN show that the model life tables tend to overestimate life expectancy (by an average of 2.133 years), or underestimate mortality. Coale and Demeny mentioned the logical problem of using the same ${}_nq_x$ as both the dependent and the independent variable in the regression equations. They also mentioned that the use of any other ${}_nq_x$ as the initial value would have been equally justified but would have produced different estimates of the same life table functions.

Gabriel and Ronen also examined the joint variation of ${}_1q_0$ and $\overset{0}{e}_0$ and experimented with a second degree polynomial in ${}_1q_0$ as an estimator of $\overset{0}{e}_0$ obtained by the method of least squares. A comparison with the model values revealed that the model values were higher as well as lower than their quadratic estimator at higher and lower levels of the continuum of $\overset{0}{e}_0$. However, they did not carry out further their investigation of the discrepancies between model values and their estimates by alternative methods.

If they did, they would have seen that since the model life expectancies increase non-linearly (almost like a quadiatic), with decline in ${}_1q_0$, a property also shared by their quadratic approximations, such two curves may not intersect at more than one point. In that case, the difference between the two estimates will change its sign only once as ${}_1q_0$ moves from one to the other end of the continuum. Their first observation about consistent upward bias in the model values, also seem to have a logical explanation that may follow from a simple illustration. Consider l_5 and l_{10} of the life table expressed as (for $l_0 = 1$)

$$l_5 = (1-{}_1q_0)(1-{}_4q_1)$$

$$l_{10} = (1-{}_1q_0)(1-{}_4q_1)(1-{}_5q_5) \qquad (1)$$

For simplicity, assume that ${}_4q_1$ is estimated from ${}_1q_0$ by a linear regression, and ${}_5q_5$ from that of ${}_4q_1$. Using the superscript ~ to denote estimates,

$${}_5\tilde{L}_5 = \tfrac{5}{2}(\tilde{l}_5+\tilde{l}_{10}) =$$

$$\tfrac{5}{2}(1-{}_1q_0)[(1-{}_4\tilde{q}_1) + (1-{}_4\tilde{q}_1)(1-{}_5\tilde{q}_5)] \qquad (2)$$

where ${}_1q_0$ is the initial value and therefore not an estimate, so that the expected value

$$E({}_5\tilde{L}_5) = \frac{5}{2}(1-{}_1q_0)[E(1-{}_4\tilde{q}_1) + E\{(1-{}_4\tilde{q}_1)(1-{}_5\tilde{q}_5)\}] \quad (3)$$

Note that ${}_5q_5$ is dependent upon ${}_4q_1$ and therefore the expected value

$$E\{(1-{}_4\tilde{q}_1)(1-{}_5\tilde{q}_5)\} = E(1-{}_4\tilde{q}_1)E(1-{}_5\tilde{q}_5) + r\,\sigma_1\sigma_5 \quad (4)$$

Where σ_1, σ_5 and r stand respectively for the standard deviations of ${}_4q_1$ and ${}_5q_5$ and the correlation coefficient between them. Thus (4) can be written as

$$(1-{}_4q_1)(1-{}_5q_5) + r\,\sigma_1\sigma_5,$$

and after substitution in (3),

$$E({}_5\tilde{L}_5) = \frac{5}{2}(l_5 + l_{10} + r\sigma_1\sigma_5) \quad (5)$$

which for r>0 has an upward bias. Thus, $\overset{0}{e}_0$ which is obtained by summing the L_x values will have a tendency to overestimate the actual life expectancy. While such was not the case when the model life expectancies were compared with those derived from the life tables developed by Gabriel and Ronen, the explanation lies in the fact that their life tables were different from the UN model life tables in that the former were derived by assuming a linear relationship between consecutive ${}_nq_x$ values, whereas the estimation formula was assumed to be quadratic in the latter.

There is at least one other fact that has so far remained in the dark and this is concerned with the efficiency, or lack of it, of the methods used in the construction of UN model life tables as well as Mitra's model fertility tables. While it may be true that the correlation between ${}_nq_{x+n}$ and ${}_nq_x$ is higher than that between ${}_nq_{x+n}$ and other q values preceding ${}_nq_x$, the efficiency of the estimate of ${}_nq_{x+n}$ from ${}_nq_x$ is questionable when the latter, in turn, is estimated from the q value preceding it and so on. As an example, consider the following where, for reasons of simplicity, the relationship between the q values, has been assumed as linear. Thus,

$${}_4q_1 = a_1 + b_1({}_1\tilde{q}_0) \quad (6)$$

$${}_5q_5 = a_5 + b_5({}_4\tilde{q}_1) \quad (7)$$

where the variables on the left hand side of the equation are the expected values and those with ~ as superscript indicate the observed values. If ${}_5q_5$ is estimated from ${}_4q_1$ which, in turn, is estimated from the initial value of ${}_1q_0$, then (7) can be written as

$${}_5\hat{q}_5 = a_5 + b_5a_1 + b_5b_1({}_1\tilde{q}_0) \quad (8)$$

It is easy to see that (7) and (8) will produce identical results for all values of ${}_4q_1$ only when the ${}_1q_0$ and ${}_4q_1$ are perfectly correlated. Further, (8) expresses ${}_5q_5$ as a linear function of ${}_1q_0$ which will generate best estimate of the former only when the expression coincides with the linear regression of ${}_5q_5$ on ${}_1q_0$ obtained by the method of least squares. The residual or the error sum of squares in the latter case is proportional to

$$1-r_{05}^2 \quad (9)$$

and for (8), which of course is an unbiased estimator, the residual sum of squares is

$$(1-r_{05}^2) + (r_{05}-r_{01}r_{15})^2 \quad (10)$$

where

r_{01} = corr (${}_1q_0$ and ${}_4q_1$)

r_{05} = corr (${}_1q_0$ and ${}_5q_5$)

r_{15} = corr (${}_4q_1$ and ${}_5q_5$)

Thus, unless $r_{05} = r_{01}r_{15}$, (8) will be an inefficient estimator of ${}_5q_5$ than the one provided by the straightforward regression of ${}_5q_5$ on ${}_1q_0$. From these results, it is quite apparent, that the consistent use of the initial parameter to estimate the remaining parameters of the life table is still the best, no matter how low the correlation gets with increase in the age difference and no matter how high the correlation is between two consecutive parameters. This provides the statistical justification of the life tables constructed by Gabriel and Ronen as well as those by Coale and Demeny.

However, Coale and Demeny's choice of $\overset{0}{e}_{10}$ as the initial parameter has more than one oddity as they themselves have acknowledged. But they do not seem to realize that the treatment of ${}_nq_x$ or log ${}_nq_x$ as dependent upon $\overset{0}{e}_{10}$, is a statistical manipulation that cannot be justified on logical grounds. In fact, the use of ${}_1q_0$, but not any ${}_nq_x$, to estimate the rest, seems to be perfectly logical, at least in a generational life table, for which ${}_1q_0$ precedes in time, and therefore, can be regarded as a determiner of the remaining ${}_nq_x$ values. From that point of view, any ${}_nq_x$ can be regarded as dependent upon any, a few, or all of the preceding ${}_nq_x$'s and not vice versa.

It can similarly be argued that since $\mathring{e}_{10}$, even in a cross-sectional life table, can be obtained only when all ${}_nq_x$ values, beginning at x equal to 10 are available, any definition of the latter as dependent upon the former suffers from a logical absurdity. Further, Coale and Demeny's use of the regression of ${}_nq_x$ on $\mathring{e}_{10}$ at one end, that of log ${}_nq_x$ on $\mathring{e}_{10}$ at the other, and the average of the two regressions in the middle of the $\mathring{e}_0$ continuum does not seem to have been based on sound logical or technical principles. Such a solution to the problem of achieving better fit is neither unique, nor does it follow from a theoretical model. One could as easily divide the range of $\mathring{e}_{10}$ into three or four sections and fit a linear regression for each of these, as for these sub-sections, the correlations would definitely turn out to be even higher to justify such a procedure. There is one other point that should also be mentioned in this context. The authors do not seem to be concerned when the ${}_nq_x$ values so obtained could not reproduce the initial $\mathring{e}_{10}$ that was the principal, if not the sole pivot of the entire structure. On the other hand, they are satisfied with the observation that the linear regression of calculated $\mathring{e}_{10}$ on ${}_nq_x$ fits the observed ${}_nq_x$'s even better.

These authors have omitted from their study, an analysis of the intercorrelations of ${}_nq_x$'s for the four regional categories, although they have done so for the log $({}_nq_x)$'s. The reason for such omission is not clear, but it seems that those intercorrelations, if obtained, would have been higher than the corresponding values of the regions combined into one group, and would have justified the development of regional tables following the method outlined by Gabriel and Ronen.

3. Mathematical Models

The controversies about the dependence independence relationship, or about the selection of the appropriate regression model are somewhat unavoidable as long as attempts to settle the issues revolve only around statistical relationships. The soluntion of this problem seems to rest, therefore, on the development of mathematical models that can describe the vital rates or functions thereof with reasonable accuracy, and when the formulation of such models is based on the observable characteristics of these variables. Such models were proposed in the past by Gompertz (1825), modified later by Makeham (1860) and Perks (1932) in the area of mortality, as well as by Wicksell (1931) and UN (1963) for the graduation of age-specific fertility rates. Needless to say that these models were proven to be somewhat less than adequate and others were proposed for the graduation of life table functions (Mitra, 1965) and of age-specific fertility rates (Mitra, 1967). The latter model, tried specifically on the Canadian data was found to be quite satisfactory (Mitra and Romaniuk, 1973). All of these models are expressions of the functions of vital rates as functions of the most important variable in the area of vital rates, namely, age.

For the fertility rates, it was observed that their pattern of variation with age resembles the mathematical form of that of Pearson's Type I, and under certain plausible assumptions, the number of unknown parameters of this function can be reduced to two or even to one which renders the model very useful in cases where such data are lacking. Mortality rates or other life table functions derived therefrom are somewhat difficult to graduate, however, a polynomial approximation of the logarithm of life expectancy was found to be quite adequate. In fact, a quadratic function of age, determined by an initial value of $\mathring{e}_0$, and an independent estimate of the maximum value of the life expectancy and the corresponding age (usually less than 5) were found to be sufficient for developing the series of $\mathring{e}_x$ values (Mitra, 1971), and these can then be used to generate the entire life table. Experiments with a few male model life tables (UN, ibid) demonstrated the usefulness of this method. Presumably, the goodness of fit will improve even further, if the life tables are classified into homogeneous groups, as Coale and Demeny did for developing their series of model tables.

Graduation formulas for other life table functions like force of mortality (μ_x) or the number of survivors (l_x) at age x are yet to be discovered. Because of the difficulty of integrating a function that has a polynomial as an exponent, such graduation formulas are not deducible from the polynomial function found suitable for the logarithm of life expectancies. This does not imply that the mathematical relationships among life table functions should be used as criteria for testing the validity of the models, even in cases where such mathematical operations are difficult to perform. Like any other model, the justifications for these models should also be derived primarily from the logical steps that were used for their theoretical formulations, and finally, from the extent to which they demonstrate their ability to reproduce reality.

4. Tests for Critical Values

More important than finding consistencies among the mathematical models of different life table functions are several tests that can be used to determine the validity of a life table. Preliminary tests can be made by comparing the values of a life table function with the general pattern that is quite well known. Not so well known are some properties possessed by the life table functions, e.g., the expectation of life attains a maximum value at an age (usually before 5) where its reciprocal is also equal to the force of mortality (Mitra, 1971). In general, this age is less than the age at which the force of mortality assumes its minimum value which in turn is less than the age at which the life table survivorship function has a point of inflection, or where its first derivative assumes a maximum or a negatively minimum value (Mitra, 1973). The significance of these three optimum values of the life table functions and their interrelationships need hardly be stressed in judging the adequacy of model life tables.

As is customary in such cases, estimates of life table functions obtained either by correlations or from mathematical models will be subject to sampling errors. For the mathematical models, the parameters will have to be estimated from an available set of life tables, usually by the method of least squares, and as a result, the sampling error of the estimate of any life table function will be determined by the sampling errors of these parametric estimates. Derivations of these estimates will, however, be approximate in most cases, as demonstrated in the following example where the mathematical model is given by

$$\log_e \left(\overset{0}{e}_x\right) = a+bx+cx^2 \quad (11)$$

When a, b and c are determined by the method of least squares, the variances of $\log_e\left(\overset{0}{e}_x\right)$ can be obtained for each x, and from the approximate relation that

$$V\{f(x)\} \approx V(x)\{f'(x)_{x=E(x)}\}^2 \quad (12)$$

$$V\left(\overset{0}{e}_x\right) \approx \left(\overset{0}{e}_x\right)^2 V\left(\log_e \overset{0}{e}_x\right) \quad (13)$$

substituting observed $\overset{0}{e}_x$ for the expected value. Similarly,

$$V\left(\frac{1}{\overset{0}{e}_x}\right) = V\left(\overset{0}{e}_x\right) \Big/ \left(\overset{0}{e}_x\right)^4 \quad (14)$$

Again, $\log_e T_x = -\int \frac{d_x}{\overset{0}{e}_x}$ (15)

Therefore,

$$V(\log_e T_x) = V(\overset{0}{e}_x)\left[\frac{d}{d\overset{0}{e}_x}\int \frac{-dx}{\overset{0}{e}_x}\right]^2 \quad (16)$$

Now

$$\frac{d}{d\overset{0}{e}_x}\int \frac{-dx}{\overset{0}{e}_x} = \frac{d}{dx}\int \frac{-dx}{\overset{0}{e}_x} \cdot \frac{dx}{d\overset{0}{e}_x}$$

$$= -\frac{1}{\overset{0}{e}_x}\Big/\frac{d\overset{0}{e}_x}{dx} = -\frac{1}{\overset{0}{e}_x(\overset{0}{e}_x\mu_x-1)} \quad (17)$$

Where $\mu_x = -\frac{1}{l_x}\frac{dl_x}{dx}$ is the force of mortality.

Substituting in (16)

$$V(\log_e T_x) \approx \frac{V(\overset{0}{e}_x)}{(\overset{0}{e}_x)^2(\overset{0}{e}_x\mu_x-1)^2} \quad (18)$$

and

$$V(T_x) \approx T_x^2 V(\log_e T_x) \quad (19)$$

$V(l_x)$ can be obtained from the relationship

$$V(T_x) = l_x^2 V(\overset{0}{e}_x) + (\overset{0}{e}_x)^2 V(l_x) \quad (20)$$

by noting that in the difference equation

$$\Delta T_x = \overset{0}{e}_x \Delta l_x + l_x \Delta\overset{0}{e}_x \quad (21)$$

l_x and $\overset{0}{e}_x$ can be regarded as independent. These and other results similarly obtained can be used to test the efficiency of the estimates of life table functions.

5. Summary

An analysis of the methods used for the development of model life tables by the United Nations, regional model life tables by Coale and Demeny and model fertility tables by Mitra revealed certain shortcomings and scopes for improvement. The UN model tables were found to be biased and lacking in efficiency by Gabriel and Ronen who had suggested certain modifications. The reasons for some of their deficiencies (also present in the model fertility tables), not elaborated by Gabriel and Ronen have been developed in this paper. Some of the problems associated with the construction of the regional model life tables have also been outlined, and the advisability of searching for justifiable mathematical models have been proposed. Such models, developed earlier by Mitra in the areas of fertility and mortality were found to describe the general pattern of these series reasonably well. Since the parameters of the model have to be estimated by following statistical procedures, the efficiency of any estimate or a function thereof depends upon the standard errors of the estimates of those parameters. Some of these formulas have been obtained for life table functions derived from a simple mathematical model proposed earlier for the construction of model tables.

References

Coale, A.J. and Demeny, Paul, Regional Model Life Tables and Stable Populations, Princeton University Press, Princeton, N.J., 1966.

Gabriel, K.R. and I. Ronen, "Estimates of Mortality from Infant Mortality Rates", Population Studies, Vol. 12, No. 2. 1958.

United Nations, "Age and Sex Patterns of Mortality, Model Life Tables for Under-Developed Countries", ST/SOA/Series A/22,1955.

Makeham, W.M., "On the Law of Mortality and Construction of Annuity Tables", Journal of the Institute of Actuaries, Vol. 8, 1860

Perks, W., "On Some Experiments in the Graduation of Mortality Statistics", Journal of the Institute of Actuaries, Vol. 63, 1932.

Mitra, S., "Model Fertility Tables", Sankhyā, The Indian Journal of Statistics, Vol. 27, Series B, 1965.

Mitra, S., "Graduation of Life Table Functions", Proceedings of the 1971 Conference of the American Statistical Association. Also, see "On Some Properties of the Expectation of Life $\overset{0}{e}_x$" in Contributions by Indian Authors to Second World Population Conference, Government of India Publications, 1965.

Mitra, S., "The Pattern of Age-Specific Fertility Rates", Demography, Vol. 4, No. 2, 1968.

Mitra, S., and Romaniuk, A., "Pearsonian Type I Curve and Its Fertility Projection Potentials", Demography, Vol. 10, No. 3, 1973.

SOME DYNAMIC INDICATORS OF NEIGHBORHOOD CHANGE: A CASE STUDY

Eric G. Moore, Queen's University
Stephen Gale, University of Pennsylvania

1. INTRODUCTION

Social indicators serve a number of different functions in the analysis of social policy, the description of social change and social reporting for administrative purposes (Land, 1972). Although much of the current literature concerns broad regional and national interests, there is also a need to develop indicators at the local or neighborhood level. Many planning and policy decisions concerned with actions such as zoning or the location of public facilities need both sounder background evaluation and subsequent monitoring, procedures which would benefit from the development of appropriate indicators.

Data inputs to planning and policy activities serve three main functions: (1) to provide basepoint information such as the need for specific services; (2) to generate insights regarding processes thereby providing a greater understanding of the way in which a given problem arises or changes; and (3) to provide a basis for evaluation of the impact of decisions. Our particular concern in this paper is with data relating to population composition and housing. All too often, the decennial census is the only source of such data available to the planner. Unfortunately, the Census can only serve the first of these three functions adequately; both the time between successive censuses and the inability to link individual records longitudinally greatly restricts its value in explanation and evaluation.

In this paper we demonstrate the potential value of comprehensive micro-level longitudinal files for neighborhood analysis. Such files permit the identification of the sets of transactions which underly the changes in population and housing characteristics and thereby provide planners with a basis for assessing the extent to which changes in such characteristics are responsive to public actions. Essential to the development of this argument is a conceptualization of change processes at the local level. Such a framework was established in an earlier paper (Gale and Moore, 1972) and we reproduce its basic elements in Figure 1. The most important points are (i) the concept of an occupancy pattern which is the specific assignment of households with given characteristics to dwellings with given characteristics (the A matrices), and (ii) the differentiation between changes in occupancy which arise from residential mobility (the M matrices) and the changes due to demographic processes and structural modification which take place in units which do not experience a household relocation (the S matrices). Our purpose here is to apply this conceptualization to a specific problem, namely the analysis of changes in the degree of overcrowding in one neighborhood in Wichita, Kansas.

2. ANALYSIS

2.1 The Definition of Overcrowding

A very simple definition of overcrowding is adopted to keep the amount of computing to a minimum. Although Federal guidelines contain far more complex definitions (taking into account sex, age, and relationships of all members of the household), this simplified procedure is sufficient in exploring the <u>relative</u> contribution of different <u>transactions</u> to change in overcrowded units. In the subsequent analysis, the following situations are defined as being overcrowded:

i) a one-bedroom dwelling containing three or more persons
ii) a two-bedroom dwelling containing four or more persons
iii) a three-bedroom unit containing five or more persons.

Dwellings with four or more bedrooms were excluded from the analysis.

2.2 Data

The necessary data were obtained from the annual enumeration of population and housing of Wichita-Sedgwick County Metropolitan Area, Kansas in 1971, 1972 and 1973 (for a full description of these files, see Gschwind (1973)). The enumeration includes all dwelling units in the county and records a wide range of data for both the dwelling unit and the occupant households; these data are recorded in such a way that specific dwellings and households can be linked in successive years thus permitting the detailed documentation of change.

This illustrative study focuses on one small area comprising four census tracts (15, 27, 28, 29) just to the west of the central business district. Its population is white lower middle class living mainly in single family homes. During the first year (1971-2) the degree of overcrowding remained stable but dropped markedly in the second year (1972-3). However, even in the first year, different components of the system affected overall overcrowding in different ways and it is important to realize the ways in which these components are in balance.

In each year it is possible to assign each dwelling unit to the categories

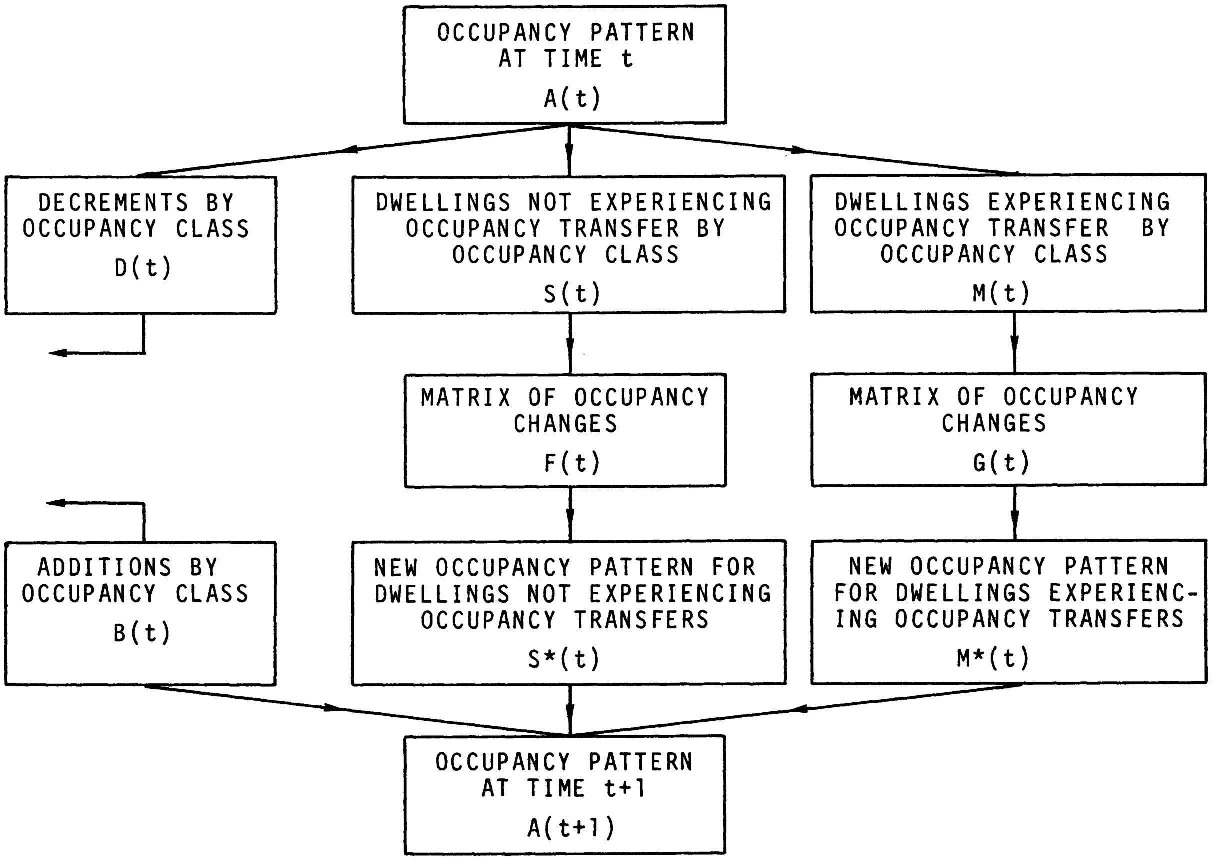


FIGURE 1: SOURCES OF CHANGE IN THE OCCUPANCY PATTERN OF A NEIGHBORHOOD

"overcrowded" (O) or "not overcrowded" (NO) according to the definition in 2.1; from this assignment the occupancy matrices can be constructed in the form shown in Figure 2. The set of occupancy matrices for the years 1971-2 and 1972-3 corresponding to the conceptual framework in Figure 1 are set out in Figure 3. The change matrices F and G are not reproduced here for lack of space but are available on request.

		O	NO
Owned Units	1 bedroom	7	158
	2 bedroom	182	1171
	3 bedroom	151	498
Rented Units	1 bedroom	62	591
	2 bedroom	170	528
	3 bedroom	88	113

FIGURE 2: INITIAL OCCUPANCY MATRIX FOR 1971

A(1971)

		O	NO
Owned Units	1 bedroom	7	158
	2 bedroom	182	1171
	3 bedroom	151	498
Rented Units	1 bedroom	62	591
	2 bedroom	170	528
	3 bedroom	88	113

D(1971)

0	2
2	12
0	8
1	19
3	7
0	0

S(1971)

6	143
164	1090
142	470
23	309
76	299
57	67

M(1971)

1	13
16	69
9	20
38	263
91	222
31	46

B(1971)

0	4
8	28
3	12
14	52
5	28
8	8

S*(1971)

6	144
166	1074
138	474
31	314
97	278
55	69

M*(1971)

5	21
10	72
5	35
53	281
62	213
27	53

A(1972)

11	169
184	1174
146	521
98	647
164	519
90	130

D(1972)

1	0
2	7
1	2
4	12
3	5
1	3

S(1972)

5	145
164	1059
133	468
26	297
100	285
50	77

M(1972)

5	24
18	108
12	51
68	338
61	229
39	50

B(1972)

1	1
0	8
0	7
2	49
3	21
0	2

S*(1972)

4	138
144	1096
125	494
16	296
110	281
35	75

M*(1972)

3	28
16	112
12	40
35	377
55	241
25	57

A(1973)

8	165
160	1216
137	541
53	722
168	543
60	134

FIGURE 3: SOURCES OF CHANGE IN OCCUPANCY PATTERN FOR OVERCROWDED UNITS 1971-3

2.3 Components of Change in Overcrowding

Shifts in the overall pattern of overcrowding arise from the four elements identified in Figure 1; from demolitions, from new construction, from residential mobility and from changes in the stable population. From the planning viewpoint, it is desirable to develop measures of the impact of these four processes. At the simplest level we can identify the proportion of overcrowded units in each component of change as is done in Figure 4. From this representation it is seen that each process acts in different ways. Both demolition and construction seem to possess little consistency from one year to the next reflecting the locational specificity of decisions in this sector. The relative behavior of the S and M matrices is more revealing; in both years, the dwellings experiencing occupancy transfers exhibited noticeably greater reductions in overcrowding than those which did not. This might lead to the inference that it is mobility which is the most important component of change but such a conclusion requires more critical analysis.

In the previous paper (Gale and Moore, 1972, p. 258), we defined rates of change for each occupancy class due to each of the four elements. Using the notation that a_{jk} represents the number of observations in cell jk of the matrix A, (in this study, for example, a_{11} (1971) has the value 7 representing the number of overcrowded units in 1 bedroom owned units at the beginning of the period, 1971-2) we define the following proportionate rates:

i) due to demolitions: $d_{jk}/a_{jk} = \gamma_{jk}$

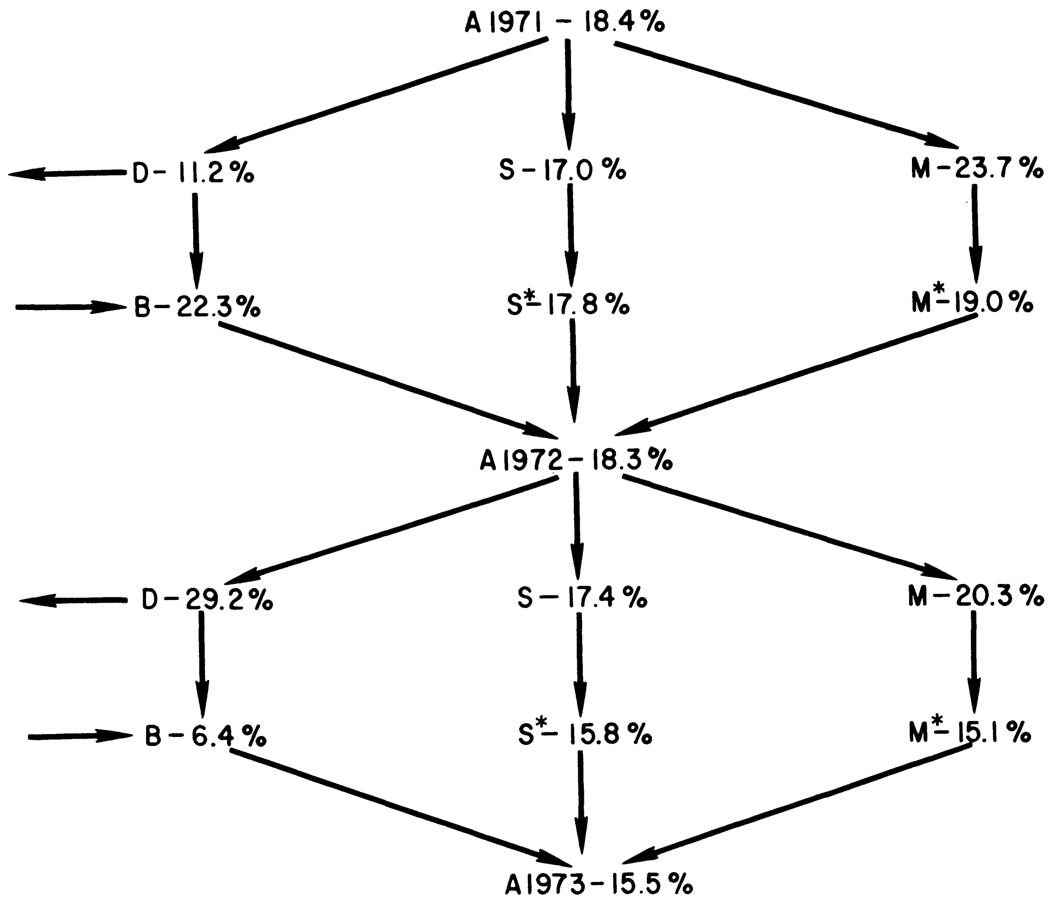
ii) due to additions: $b_{jk}/a_{jk} = \delta_{jk}$

iii) due to occupancy transfers: $(m^*_{jk}-m_{jk})/m_{jk} = \alpha_{jk}$

iv) in the remaining stock: $(s^*_{jk}-s_{jk})/s_{jk} = \beta_{jk}$

The composite rate for iii) and iv), λ_{jk} is the rate of change in that part of the stock unaffected by construction activity and is defined by

$$\lambda_{jk} = \theta_{jk}\cdot\alpha_{jk} + (1-\theta_{jk})\cdot\beta_{jk}$$



% OVERCROWDED UNITS FOR EACH COMPONENT—
CENSUS TRACTS 15,27,28,29
Figure 4

Parameter		γ		δ		α		β		θ		λ	
Year		1	2	1	2	1	2	1	2	1	2	1	2
Structure Type													
	1	*	-090†	*	090	*	*	*	*	*	*	*	*
Owned	2	-011	-011	044	*	-375	-111	012	-122	088	099	-022	-121
	3	*	-007	020	*	*	*	-028	-060	060	083	-052	-055
	1	-016	-041	226	020	395	-485	350	-385	613	723	377	-457
Rented	2	-018	-018	030	018	-320	-098	278	100	535	379	-047	024
	3	*	-011	091	*	-129	-359	-035	-300	352	438	-068	-326
All Owned		-006	-011	032	001	-231	-114	-006	-096	076	103	-023	-098
All Rented		-012	-022	084	014	-113	-315	173	-085	500	488	030	-197
All Units		-009	-017	043	009	-129	-232	053	-092	282	298	002	-134

*insufficient observations †decimal point omitted

TABLE 1: PARAMETER VALUES FOR COMPONENTS OF CHANGE, 1971-2 AND 1972-3

		Mobility Rate		Fraction Becoming Overcrowded	
		O	NO	O	NO
Structure Type					
	1	*	082†	*	*
Owned	2	088	059	308	100
	3	060	040	*	187
	1	613	444	281	102
Rented	2	535	424	152	180
	3	352	408	364	235
All Owned		076	056	286	133
All Rented		500	378	225	147
All Units		282	210	234	145

*insufficient observations

†decimal point omitted

TABLE 2: DIFFERENTIAL IMPACTS OF MOBILITY FOR OVERCROWDED (O) AND NOT OVERCROWDED (NO) UNITS, 1971-2

where θ_{jk} is the mobility rate of occupancy class jk. The values of γ, δ, α, β, θ, λ for overcrowded units by structure type for 1971-2 and 1972-3 are given in Table 1.

The parameter values in Table 1 provide a more critical perspective on the nature of change. It shows that although the total number of events involved in the demolition and construction processes are relatively small, they are critical if the changes generated by mover and nonmover households are in balance; such was the case in the first year with the building activity resulting in an overall 3.4% increase in overcrowded units and the other two elements essentially cancelled each other out. In the second year, however, there was both a reversal of the impact of nonmover households resulting in a reduction in overcrowded units and an increased rate of reduction in overcrowding due to mobility. These two effects result in 13.4% decrease in overcrowded units as against a 0.8% reduction from the building sector. From a planning perspective, perhaps the most important point is that the major effects are found in the rented sector. In particular, 1 bedroom rented units possess large changes for all parameters in the direction of reduced overcrowding. Such an observation would suggest that this particular sector of the market should be examined in greater detail to determine if a single apartment block or localized set of units are the ones most responsible for the change and, if so, the specific reasons for these effects.

The relationship between population mobility and overcrowding is of particular importance given the large changes which are effected in the study area through mobility. Two further questions are of interest here. Are overcrowded units more likely to experience a household relocation than those which are not overcrowded? Once a relocation takes place, is an overcrowded unit more likely to be overcrowded with its new occupants than one which was originally not overcrowded? Table 2 provides a comparison of mobility rates and of the proportion of units overcrowded after relocation for each occupancy class for 1971-2. From this we can see that overcrowded units, in general, are subject to higher levels of mobility, but that there is a much greater likelihood of an overcrowded unit remaining overcrowded after a new household moves in for almost every structure type. Given these counteracting tendencies, the precise role of mobility is still indeterminate. Further analysis is required, the most promising direction being to distinguish dwelling units on the basis of condition, the inference being that it is the overcrowded poorer units which are more likely to remain overcrowded.

3. CONCLUSION

The material presented in this paper is illustrative and the specific parameters only refer to one area for a short period of time. Perhaps of greater interest is the degree to which these parameters vary over different neighborhoods within a city. For example, the study area is characterized by an older population which appears to be in the process of "thinning out". A systematic examination of these parameters for rooming house areas or for upper income suburban neighborhoods would undoubtedly tell us more about the dynamics of small area change within a metropolitan area. However, even within the limited context of this study, we have identified important differences in the contribution of various processes to changes in overcrowding. The types of files available in Wichita permit identification of specific sectors in which change is occurring most rapidly and which deserve more detailed examination from the planner. The development of such files and the search for dynamic indicators in a wide variety of problem areas offers much promise for improvement of the base for planning decisions.

REFERENCES

Gale, S. and E.G. Moore, "Some Dynamic Indicators of Neighborhood Change", Proceedings of the Social Statistics Section, American Statistical Association, 1972, 255-259.

Gschwind, R.A. (1973), "The Intergovernmental Enumeration, Wichita-Sedgwick County, Kansas: 1971-1973", Working Paper No. 2, Research on Metropolitan Change and Conflict Resolution, Philadelphia, Peace Science Unit, University of Pennsylvania.

Land, K.C. (1972), "Social Indicator Models: An Overview", paper presented at the Annual Meetings of the American Association for the Advancement of Science, Washington, D.C.

ACKNOWLEDGEMENT

The support of the National Science Foundation Grant GS-39387 and the assistance of the Wichita-Sedgwick County Metropolitan Area Planning Department are gratefully acknowledged.

AN EVALUATION OF THE INFLUENCE OF RESIDENCE HALL LIFE STYLE ON A CHANGE OF COLLEGE STUDENT'S ATTITUDES

Marilyn Nouri, Oneonta, New York
Ted Walbourn, School Psychologist, Baldwinsville, N.Y.

1. Introduction

Residence hall living on college campuses has changed considerably in the last decade. In order to keep residence halls filled, it has been necessary to adapt to rules and regulations of the modern college student. At the same time, colleges have been concerned with the effect of the relaxing of controls on student attitude and behavior.

Little research has been done on the influences of residence hall living on the student, particularly in the area of comparison of various styles of residence hall living. Much of the literature indicates the importance of the environment on the student. The interaction of the student with his physical, social and psychological environment does much to influence his perception of self and the world around him. (French,1963;Baker,1966;Mirande,1968;Pace and Stern,1958;Pervin,1967)

This study will investigate the influence of style of residence hall living on change in student's attitudes in three specific areas: premarital sexual behavior, concept of self, and male-female interaction. These three topics were chosen because of much of the criticism, both negative and positive, of residence hall living revolves around these areas. Much of the controversy is over coeducational living. On the positive side it is claimed that it improves the quality of male-female interaction, that many brother-sister relationships develop, that it is a more natural environment and that the student is better prepared to face the "real world" when he graduates. On the negative side it is stated that coeducational living causes an increase in promiscuity, little studying is done and a general breakdown of moral values.

Previous research gives some insight into which of the above views might be more accurate. Changes of attitudes regarding premarital sexual behavior has been most exhaustedly researched but not as it relates to coeducational living. Studies indicate that individual's attitudes liberalize regarding sex as they go through the courtship process until they marry and have children (Reiss,1967; Mirande,1968). There has been some liberalization of sex codes in the last few years.

Male-female interaction in adolescence has been studied mainly as it relates to dating. Lowrie (1951) found that students date for four reasons: (1) mate selection, (2) recreation, (3) anticipatory socialization, and (4) adult role clarification. Coeducational living might bring similar results. The changing of one's self concept is also highly related to one's reference groups. Residence hall living is only a small part of the total college environment that one experiences when he goes away to school.

It is reasonable to assume that attitudes of students do change while they are in college. Individuals are constantly going through the socialization process in what is called continuing socialization. When a student comes to college, he is removed from his home environment and put into quite a different situation. He is given new knowledge by professors; he is constantly surrounded by peers whose ideas and behavior patterns may be very different from those he has experienced in the past. The college environment usually provides a new freedom away from parental supervision that the individual has not experienced before. Due to the trend away from close supervision by college officials, the student is much more free to set his own regulations regarding personal behavior. Whether or not living in one residential life style or another will add to the changes is difficult to hypothesize. Attitude change is a very complex process resulting from many factors and whether this single factor is significant enough in itself is not known. The following three general hypotheses were delineated:

1. There is a significant change in attitudes from freshman to the senior year in college.
2. There is a significant difference in attitudes among students who choose one life style over another.
3. The degree of change in attitudes in one semester will be significantly related to the life style in which one is living.

Attitudes here has been broken down into three specific areas - premarital sexual behavior, concept of self, and male-female interaction.

2. Method

The sampled population consists of the resident students at one of the two year colleges of the State University of New York. A stratified random sample of 300 students was selected from the college housing listing (a sampling fraction of 20%). The criteria for stratification consisted of class standing, sex, and residence hall life style. The three residence hall life styles that were compared in this study were:

1. one-sex dorms: all of the students in the bui-

lding were of one sex. There were no female one-sex dorms.
2. coed by wing dorms: one wing of the building contained males, the other wing, females. Visitation between the sexes was not allowed after midnight.
3. coed by suite dorms: one suite of a residence hall was occupied by males and down the hall a neighboring suite was occupied by females. There were no restrictions on visitation. The response return to the questionaire first administered in September 1971 was 75%. The follow-up study was administered at the end of the fall semester with response return of 50% of the original sample.

Operational measures of specific concepts have been developed. It is assumed that if a relationship can be established between operational measures, then this is evidence that a relationship exists between the concepts they are measuring. This can only be assumed if the operational measures do measure the concepts being studied. Evidence of a relationship, then, is support for the general and empirical hypotheses. Paired t-test and analysis of variance were used to test the relationships.

The questionaire used in the study is a combination of Pervin's Instrument for Transactional Analysis of Personality and Environment and Riess's Scale of Premarital Sexual Permissiveness. Nine characteristics related to social behavior were chosen from Pervin's scale to give a composite self concept and were used to measure a degree of change in that concept. Reiss's scale is composed mainly of questions with six alternative answers to measure strength of attitudes regarding premarital sexual permissiveness. The authors added questions used for background information about each subject and questions to determine attitudes regarding male-female interaction. Degree of certainty scaling was used to measure this area.

3. Results

Tables one to five show the results of the statistical analysis. The results indicate little change in attitudes. Change from the freshman to senior year in college occurred in only two areas. The freshmen women had significantly less permissive attitudes regarding premarital sexual behavior than did senior women. However, by the time of the follow-up study, the significant difference had disappeared. The other difference was found between freshmen and senior men. Male freshmen consider themselves to be significantly less comfortable with the opposite sex than do senior males. This difference also disappeared by the time of the follow-up study.

When attitudes were compared by life style among freshmen women, senior women, freshman men and senior men, only one significant difference was found.(See Table II) Freshmen women living in the coed by wing dorm were found to be significantly less comfortable with the opposite sex than were freshmen women living in the coed by suite dorm.

The comparison of change in attitudes after having lived in the residence halls indicated different results.(See Table III) Somewhat unexpected results came out of the change of atti-

Table I

CHANGES IN ATTITUDES FROM FRESHMAN TO SENIOR YEAR IN COLLEGE. (t-test)

	Before	After
Regarding self		
All students	0.3405	
Regarding premarital sexual behavior		
All students	1.7847	0.5824
Females	2.5518*	1.0925
Males	0.1913	0.2867
Regarding male-female interaction		
All students	1.3205	
Females	0.6824	
Males	1.9739*	0.7980

*Significant beyond the 5% level of confidence.

Table II

DIFFERENCES IN ATTITUDES AMONG STUDENTS LIVING IN VARIOUS LIFE STYLES. (mean vales and analysis of variance)

Regarding self[1]	Mean	F Value
All students		1.3075
Coed by suite	32.3333	
Coed by wing	29.6938	
All males	32.1290	
Freshman females		0.0002
Coed by suite	28.6875	
Coed by wing	28.7368	
Freshman males		0.9872
Coed by suite	36.0000	
Coed by wing	31.2000	
All males	33.6000	
Senior females		0.0567
Coed by suite	31.7777	
Coed by wing	30.9090	
Senior males		0.5220
Coed by suite	32.7333	
Coed by wing	28.5555	
All males	30.7500	
Regarding premarital sexual behavior[2]		
All students		2.1121
Coed by suite	8.2923	
Coed by wing	7.3469	
All male	8.4687	
Freshman females		.0605
Coed by suite	6.2352	
Coed by wing	6.0000	
Freshman males		1.6722
Coed by suite	9.6250	
Coed by wing	9.1000	
All males	8.1250	
Senior females		.3243
Coed by suite	8.1764	
Coed by wing	7.5454	
Senior males		.7849
Coed by suite	9.3333	
Coed by wing	8.0000	
All males	8.8125	(con'd)

Table II, continued.

	Mean	F Value
Regarding male-female interaction[3]		
All students		1.6657
Coed by suite	17.0000	
Coed by wing	18.1836	
All males	20.1875	
Freshman females		6.2243*
Coed by suite	12.7058	
Coed by wing	18.6842	
Freshman males		0.0136
Coed by suite	21.7647	
Coed by wing	22.2000	
All males	21.7500	
Senior females		0.0078
Coed by suite	14.3333	
Coed by wing	14.6363	
Senior males		0.3381
Coed by suite	19.6666	
Coed by wing	17.0000	
All males	18.6250	

[1]The higher the mean value the more sociable they consider themselves to be.
[2]The higher the mean value the more permissive the attitudes.
[3]The higher the mean value the less comfortable they are with the opposite sex.
*Significant beyond the 5% level of confidence.

tudes regarding self. Senior women living in the coed by wing dorm became significantly less positive in their concepts of themselves, as did freshman and senior men living in the all male dorm. It is possible that the measure of self concept is inaccurate. The measure of self concept was developed from the discrepancy scores and it may not be valid to use it to measure change in self concept.

In attitudes regarding premarital sexual behavior the only change that took place was among freshman females in the coed by wing dorm. That group changed to become significantly more liberal in their attitudes regarding premarital sexual behavior than before the experience. However, when they were compared to the freshman females living coed by suite after the experiment, no differences were found.

In attitudes regarding male-female interaction, again the only group that changed significantly were the freshman females in the coed by wing dorm. They changed in the direction of being more comfortable with the opposite sex. However, even with the change in attitudes, they were still significantly less comfortable with the opposite sex than were freshman girls living in the coed by suite dorm.

It appears that life style has little relationship to attitudes and attitude change. There are very few differences betweeen the groupings when they move into the life style and it causes very few changes. The only significant explanable difference is the freshman females in regard to male-female interaction. If girls are less comfortable with men they tend to choose to live in coed by wing dormitories. Coed by wing living helps to make them more comfortable with men but not to the level of the freshman girls living coed by suite.

It was stated in the theoretical orientation that coming to college may put the students into a new environment which will cause attitude change.

Table III

DIFFERENCES IN DEGREE OF CHANGE IN ATTITUDES AMONG VARIOUS LIFE STYLES. (Paired t-test)

Degree of change among life styles--Self	
Coed by suite	
Freshman females	0.0666
Senior females	0.8834
Freshman males	0.9472
Senior males	1.8944
Coed by wing	
Freshman females	1.7145
Senior females	3.0596*
Freshman males	1.4384
Senior males	1.4743
All male	
Freshman males	1.8047*
Senior males	2.4861*
Degree of change among life styles--Premarital sexual attitudes	
Coed by suite	
Freshman females	1.6898
Senior females	1.0895
Freshman males	1.1423
Senior males	1.3193
Coed by wing	
Freshman females	2.2641*
Senior females	0.0000
Freshman males	0.0000
Senior males	0.0000
All male	
Freshman males	0.0804
Senior males	0.7992
Degree of change among life styles--Male-female interaction	
Coed by suite	
Freshman females	0.0652
Senior females	0.0719
Freshman males	0.0272
Senior males	0.6629
Coed by wing	
Freshman females	2.0909*
Senior females	1.7634
Freshman males	0.8310
Senior males	0.5143
All male	
Freshman males	1.2043
Senior males	0.1833

*Significant beyong the 5% level of confidence.

Whether or not this in fact happens depends upon the groups of students and the attitude being analyzed. The male sub-culture has apparently already given males a liberal outlook on premarital sexual behavior so that the college experience causes no further significant liberalization of values. On the other hand, the college experience does change the attitudes of the female regarding premarital sexual behavior. She becomes significantly more liberal in her views within one semester.

The college experience also helps both sexes to feel more comfortable with the opposite sex. Freshman men differed significantly from senior men in this respect but the difference had disappeared by the follow-up study. Life style appeared to have no effect on this change. Only freshman girls living coed by wing felt significantly less comfortable with men before the study; this situation imporved by the follow-up study, but not to the level of the coed by suite freshman girls.

No meaningful analysis of the college experience's effect on self can be made since the results follow no particular pattern.

4. Summary

Change in attitude did take place. The college experience did cause men to become more comfortable with the opposite sex and women to develop more liberal sexual attitudes. Freshman men already had such liberal sexual attitudes that there was little room for change. The place of residence played little or no significant role in these changes. Freshman girls who chose to live coed by wing dorm were significantly less comfortable with the opposite sex than freshman girls who chose to live coed by suite and although they changed to being more comfortable, they were still signfcantly less comfortable than the suite girls at the end of the study. Feeling comfortable with the opposite sex appears to be a selective factor in choice of life style. It is possible that if they had been forced to live by another life style they would have been much less satisfied.

This research and another part of this project not covered in this paper lend support to the idea that various kinds of life styles should be provided in college residence halls so that the student can choose to live in the style most appropriate for his needs and interests. None of the life styles investigated contributed to changes in student attitudes.

Footnotes

1. Baker, S.R. "The Relationship Between Student Residence and perception of Environmental Press," The Journal of College Student Personnel, pp. 222-224, July, 1966.

2. French, John. "The Social Environment and Mental Health." Presidential Address before the Society for the Psychological Study of Social Issues, The American Psychological Association, Philadelphia, September 2, 1963.

3. Lowrie, Robert H. "Dating Theories and Student Responses." American Sociological Review. 16 (1951) pp. 334-340.

4. McDaniel, Clyde O. "Dating Roles and Reasons for Dating." Journal of Marriage and the Family, 31 (1959), pp, 97-106.

5. Mirande, Alfred M. "Reference Group Theory and Adolescent Sexual Behavior." Journal of Marriage and the Family, November 1968, pp. 572-577.

6. Pace, Robert C. and Stern, GeorgeG. An Approach to the Measurement of Psychological Characteristics of College Environments. 1958.

7. Pervin, Lawrence. A Twenty-College Student X College Interaction Using TAPE (Transactional Analysis of Personality and Environment).

8. Pervin, Lawrence A. and Rubin, Donald B. "Student Dissatisfaction with College and the College Dropout: A Transactional Approach." The Journal of Social Psychology, 1967, 72, pp. 285-295.

9. Reiss, Ira L. The Social Context of Premarital Sexual Permissiveness. Holt, Rinehart and Winston. New York, 1967.

10. Udry, J. Richard. The Social Context of Marriage. J.B. Lippincott Co., Philadelphia, 1966.

THE ADJUSTMENT OF SELECTED FOREIGN STUDENTS ATTENDING IOWA STATE UNIVERSITY[1]

Mostafa M. Noury, Westfield State College

Foreign students' perception and attitudes toward selected aspects of the host society are major problems affecting their adjustment processes. This paper reports an empirical study done on two groups of foreign students - Indian and European students. One purpose of this study has been to examine foreign students' perception and attitudes toward selected aspects of the host society, and to consider the possible consequences of these perceptions and attitudes on their adjustment processes. To determine if differences exist between the two groups with respect to perception, attitudes and adjustment is the second objective.

The findings of this study indicate that differences do exist between the two groups and that European students have shown more favorable perception, attitudes and adjustment to the host society than Indian students. Whereas perception has been found to be significantly related to adjustment, attitudes fall short of statistical significance. Reasons and explanation for the unexpected relationships are given.

INTRODUCTION

Today foreign students are not only a common sight on campuses and universities in the United States, but also due to the apparent increase in their numbers in recent years, the subject of foreign students in the United States has been met with concern of both educators and social scientists. As a result, efforts have been directed toward understanding the problems of adjustment resulting from their cross-cultural experiences (Smith, 1956; Barry, 1966).

The foreign students attending universities in the United States, come from very diverse social, economic and academic backgrounds. They arrive at a campus where they encounter new values and associations that are different from those in their past experience. In order to participate in the American society foreign students are required not only to accomodate but also to adjust to the host society. Because of the very diverse cultural backgrounds and systems, the foreign students find it difficult in the beginning to adapt, adjust or conform to the norms of the American society. The degree of adjustment among individuals varies according to their cultural backgrounds. Where differences are greatest between home and host culture, problems of adjustment may be expected to be more acute than when differences are slight. (Sewell, 1961).

The present study addresses itself to the adjustment of two groups of foreign students attending Iowa State University - mainly students from India[2], and students from Western Europe and Scandinavian countries[3]. More specifically, the main purposes of this study are:

1. To investigate the relationship between selected perceptual and attitudinal variables and adjustment among the aggregate (Indian and European students combined) as well as similar relationships within the individual groups (Indian and European groups separately).
2. To determine if differences exist between the two groups with respect to perception, attitudes and adjustment.

DEPENDENT AND INDEPENDENT VARIABLES

Adjustment

Adjustment - the dependent variable was defined as the response in behavioral patterns toward some selected aspects of the host society. In order to operationalize this variable, an index of 24 quantifiable statements of several segments[4] of adjustment to the host society was developed. Some statements or items were given a five-point continuum, others three-points, while others were on a two-point continuum. The adjustment of the Indian group as well as the European group was calculated by adding the scores of the 30 individuals in each group over the 24 items included in the adjustment scale. These 24 items that constituted the adjustment scale had a possible maximum score of 48 points and a possible minimum of 6 points.

Perception

Perception as well as attitudes are usually inferred from the behavior of the individual and are not measured directly. The underlying assumption is that the individual's response to a statement with a positive or negative judgment is a type of behavior which can be taken as a measure of the perception and the attitudes the individual has in regard to the dimension of which the statement is a measure.

In this study perception was defined as the means by which people (students) form impressions of and, hopefully, understand one another (Tannenbaum et al., 1961). Four dimensions of the foreign student's perceptions of some aspects of the host society were developed. These were perception of work in the United States, education in the United States, American people, and perception of social life in America. Seven[5] items made up the original measure of perception. While responses to six items were based on a five-point continuum, responses to one item were based on a three-point continuum. The seven items that formed the perception scale had a possible maximum score of 14 points and a possible minimum of zero.

Attitudes

Using Allport (1935) definition, an attitude was defined as a mental and neural state of readiness, organized through experience, exerting a direction or dynamic influence upon the individual's response to all objects and situations with which it is related. Four dimensions of the foreign students attitudes toward the host society were undertaken. These were attitudes toward: work in the United States, education in the United States, American people, and attitudes toward social life in America. The sixteen items that made up the foreign students' attitudes toward the United States in the four areas cited above were ranked on a five-point continuum. These sixteen items had a possible maximum score of 32 points and a possible minimum of zero.

PROCEDURE

To achieve the study objectives, a simple correlation analysis was applied in order to determine the degree of association or relationship. The t-test for differences between the two groups was used to test the hypothesized differences between the groups of foreign students along the variables investigated.

SAMPLE

This study was based on a total sample of size 60 male foreign graduate students attending Iowa State University. Students were selected from two cultural areas; i.e., India and West Europe. The sample was drawn from a list of the Indian students and the West European students. This list was selected from a list comprising the entire population of foreign students attending Iowa State University in the Spring Quarter of 1970. The population list was compiled from two sources. The Foreign Student Advisor's list provided the basic listing of foreign students at Iowa State. This list was supplemented from the records of the Admission's Office. The population included in the sample was drawn randomly as follows: 30 Indian students from a total group of 63, representing the developing countries, and 30 West European students out of a total of 51, representing the industrialized nations.

FINDINGS

Tables 1, 2, and 3 present the correlation coefficients among the variables considered on the aggregate as well as the individual group level. Table 4 presents results of the t-test for differences between the two groups along the variables investigated.

As can be seen from Table 1, the correlation coefficients between favorable perception and adjustment among the aggregate group of foreign students were significant at the five per cent level of probability. Along the four dimensions of perception, only the foreign student's favorable perception of work in the United States was not related to adjustment (Table 2). On the individual group level, whereas the four dimensions of perception were not related to adjustment among Indian students, perception of social life in America was found to be related to adjustment among European students (Table 2).

With regard to the expected relationship between attitudes and adjustment among the aggregate, the computed correlation coefficients showed that no significant positive relationships existed between these two variables (Table 1). Along the four attitudinal dimensions, only favorable attitudes toward education and adjustment were significantly related (Table 3). On the individual group level, the correlation coefficients between the four attitudinal dimensions and adjustment fell short of statistical significance (Table 3).

Results of the t-test for the difference in adjustment between the means of the two groups (Table 4) revealed that significant differences existed between the two groups, indicating that European students displayed patterns of adjustment more favorable to the host society than Indian students. Significant differences also existed between the two groups along the four perceptual dimensions (Table 4). With regard to attitudes, significant differences existed between the two groups along three of the attitudinal dimensions, namely, attitudes toward work, Americans, and toward social life in America (Table 4).

DISCUSSION AND CONCLUSIONS

The first objective of this study was to examine the relationships between selected perceptual and attitudinal variables and adjustment among the aggregate as well as similar relationships within the individual groups. As can be readily seen from the data in Table 2, three of the four dimensions of perception were found to be statistically related to adjustment for the aggregate of the two groups. However, the data did not support the expectation that the student's perception of work in the United States was related to adjustment. This might be due to the smallness of the sample size or to the trivial relationship between perception and adjustment in the case of Indian students, which might have depressed the correlation coefficient for the aggregate of the two groups.

Although three dimensions of perception of the United States were found to be statistically related to adjustment for the aggregate of the two groups, the individual group scores fell short of statistical significance, except for the Europeans perception of social life in America which was found to be significantly related to adjustment at the one per cent level of probability. This latter relationship was expected since social life in Western Europe and Scandinavian countries is to some degree similar to that of the United States. As to the unexpected relationship between perception and adjustment with regard to work, education, and Americans, an examination of the responses of the two groups revealed that European students have scored higher than the Indian students on all of the perceptual items (Table 5). This unexpected relationship might also be due to the small sample size as in the case of perception of work, a clustering at one end of the value continuum as in the case of perception of education, or the fact that the relationship is in the same direction and is cumulative in both groups as in the case of perception of Americans.

With regard to the hypothesized relationship between attitudes and adjustment among the aggregate of the two groups, only attitudes toward education in the United States were found to be statistically supported. This might be due to the fact that since both groups were deeply involved in the American educational system, this exception might not be decisive. The fact that both groups saw it rewarding to attend Iowa State University might indicate that some degree of commonality in attitudes toward education could be expected.

In examining the relationship between attitudes and adjustment for each group separately, the four attitudinal dimensions were not statistically supported. However, the European students' attitudes toward Americans and social life in America were found to be close to the significance level (p = .36). This might be due to the cultural similarity factors, such as European students may find in the United States a niche for themselves socially and academically, and thus tend to express more favorable attitudes to the host society, Indian students find the United States culturally different socially and academically and thus tend to express patterns of attitudes toward the United States less favorable than that of the European students.

Three reasons could be stated as possible explanations for the lack of support of the hypothesized relationship between attitudes and adjustment. Firstly, the sample size was relatively small, secondly, some deficiencies in the measuring instrument might be involved such as relatively few items in certain attitudinal dimensions, and thirdly was the limitation of the study to a narrow spectrum of the student population.

The second objective of this paper was to investigate the differences in perception, attitudes, and adjustment between Indian and European students. The aggregated measure of the four dimensions of perception showed significant differences in perception between the two groups. In examining the differences between the two groups along the four attitudinal dimensions, the findings revealed that significant differences existed between the two groups in the dimensions of work, Americans, and attitudes toward social life in America. Whereas attitudes toward education were significantly related to adjustment, no significant difference in attitudes toward education between the two groups could be detected (Table 4). This apparent inconsistency could be explained methodologically. If attitudes toward education and adjustment were found to be significantly related for the aggregate of the two groups, while the individual group scores fell short of statistical significance, this was indicative of accumulative condition of this factor. In other words, attitudes toward education for both groups might be in the same direction. This explanation gained support from the data, where a clustering was found at one end of the value continuum.

With regard to differences in adjustment between the two groups, the measure of this variable showed significant differences between the means of the two groups. This was expected since Western European culture is to some degree similar to that of the United States, and hence, European students displayed patterns of adjustment more favorable to the United States than the Indian students.

On the basis of the data presented in Tables 1, 2, 3, and 4, it was concluded that differences in perception, attitudes, and adjustment existed between the two groups. The foreign students' perceptions of selected aspects of the United States were found to be not only related to their adjustment to the host society, but also European students have expressed patterns of perception more favorable, and more closely related to their adjustment than the Indian students. These conclusions were in line with several research studies conducted by Becker (1966), Morris (1960), Sewell and Davidsen (1961), and Cora du

Bois (1962).

SUMMARY

This study had two objectives. The first one was to examine the relationship between selected perceptual and attitudinal variables and adjustment among the aggregate as well as similar relationship within the individual groups. The second objective was to determine the differences in perception, attitudes, and adjustment between Indian and European students. Four dimensions of perception as well as attitudes were investigated in the study. These were the foreign students' perceptions and attitudes in the areas of work and education in the United States, of Americans, and of social life in America. Some specific segments of the foreign students' adjustment to the host society were also investigated.

A random sample of 60 foreign students, 30 Indians and 30 West European and Scandinavian students, attending Iowa State University on the Spring Quarter of 1970, were selected for the study. The hypothesized relationships between the variables were developed and tested by means of the simple correlation analysis and the t-test for the differences between the two groups.

The findings of this study revealed that while perception was found to be significantly related to adjustment, attitudes fell short of statistical significance, both on the aggregate as well as the individual group level. Apparent differences in perception, attitudes, and adjustment existed between the two groups. Along the four perceptual and attitudinal dimensions significant differences existed between the two groups, except for attitudes toward education in the United States. Reasons as well as explanations for the unexpected relationships were given in the Section on Discussion.

BIBLIOGRAPHY

Allport, G.W. Attitudes. In Murchison, Carl, ed. A handbook in social psychology. Pp. 798-844. Clark University Press, Worcester, Massachusetts. 1935.

Barry, J. The Thai students in the United States. A study in attitude change. Unpublished Ph.D. dissertation. Library, Columbia University, New York, New York. 1966.

Becker, T.S. Perceptions and attitudinal changes among foreign students on the U.C.L.A. campus. Unpublished Ph.D. dissertation. Library, University of California, Los Angeles. 1966.

Du Bois, C. Foreign students and higher education in the United States. Harvard University Press, Cambridge, Massachusetts. 1962.

Morris, R.T. The two-way mirror. National status in foreign students adjustment. The University of Minnesota Press, Minneapolis, Minnesota. 1960.

Selltiz, C., A.L. Hoposon, and S.W. Cook. The effects of situational factors on personal interaction between foreign students and Americans. Jour. of Social Issues 12: 33-44. 1956.

Sewell, W. and O. Davidsen. Scandinavian students on an American campus. University of Minnesota Press, Minneapolis, Minnesota. 1961.

Smith, M. Brewster. Cross-cultural education as a research area. Jour. of Social Issues 12: 2. 1956.

Tannenbaum, R.W., R. Irving, and F. Massarik. The process of understanding people. In Bennis, W.G., E.H. Schein, D.E. Berlew, and F.I. Steele, eds. Interpersonal dynamics. Pp. 725. The Dorsey Press, Homewood, Illinois. 1961.

---

[1]The research reported in this paper is part of an unpublished M.S. Thesis, "The Adjustment of Selected Foreign Students at Iowa State University," Iowa State University Library, 1970.

[2]The students from India will hereafter be indicated as Indian students.

[3]Students from Europe and Scandinavian countries will hereafter be referred to as European students. The reason for grouping West European and Scandinavian students under one single nationality group is that several previous investigators have considered the Europeans as a single group vis-à-vis the non-Europeans, with regard to extent of associations with Americans and other variables (Morris, 1960, Selltiz et al., 1963).

[4]The segments of adjustment to the host society were adjustment to: religion, food, dress, parties, language, picnics, sports, travel, music, use of nicknames, reading a book about America, visiting American families, association with Americans, participation in campus organizations, and marrying an American girl. The degree of the foreign student involvement with his home country

was also taken as a measure of adjustment such as the number of letters written and received from home, reading news papers and magazines of native country, and celebrating native festivals in the U.S.

[2]The seven items developed to measure the foreign students' perception of: work, education, Americans, and American social life were three, one two, and one item, respectively.

Table 1. Correlation coefficient between perception, attitudes, and adjustment "all dimensions combined"[1]

Factors	Adjustment		
	Indians	European	Total
Percep.	.097	.301	.317*
Attitudes	.271	.306	.194

[1]i.e. perception and attitudes toward work, education, Americans, and American social life.
*Significant at the .05 level of confidence.

Table 2. Correlation coefficient between perception factors and adjustment.

Percep. factors	Adjustment		
	Indian students	European students	All students (combined sample)
Percep. of work	.008	.206	.201
Percep. of educ.	-.042	.230	.415**
Percep. of Amer.	.086	.130	.285**
Percep. of Amer. social life.	-.039	.470**	.342**

*Significant at the .05 level of confidence.
**Significant at the .01 level of confidence.

Table 3. Correlation coefficient between attitude factors and adjustment

Attitude factors	Adjustment		
	Indian students	European students	All students (combined sample)
Toward work	-.170	.204	.203
Toward education	.180	-.024	.303*
Toward Americans	.318	.331	.165
Toward Amer. social life	.109	.334	.129

*Significant at the .05 level of confidence.

Table 4. Differences between the two groups.

Factor	t-value
Adjustment	9.678**
Att. toward work	2.562*
Att. toward education	.334
Att. toward Americans	3.671*
Att. toward Amer. social life	3.036*
Percep. of work	2.793*
Percep. of education	2.547*
Percep. of Americans	3.636*
Percep. of Amer. social life	2.986*

*Significant at the .05 level of confidence ($t_{.05} = 1.672$).

**Significant at the .01 level of confidence ($t_{.01} = 2.30$).

Table 5. Items of perception in relation to selected categories.

Item number	Max. range of scores for each group.	Sum scores of Indians n = 30	Sum scores of Europeans n = 30
Work 1	0 - 60	17	24
2	0 - 60	12	8
3	0 - 60	14	28
		43	60
Education			
4	0 - 60	21	44
Americans			
5	0 - 60	16	25
6	0 - 60	19	30
		35	55
Amer. social life			
7	0 - 60	16	25
Total		112	184

PROCEDURES FOR ESTIMATING M.C.D. POPULATIONS FOR STATE REVENUE SHARING

C. D. Palit, H. Reinhardt, J. Sweet, and H. Winsborough
University of Wisconsin

H. Krebs, F. Backus, G. Ferwerda, D. Giovannini, and L. Golicz
State of Wisconsin

In November, 1971, the Wisconsin Legislature changed the basis for the sharing of tax revenues with its minor civil divisions to one which is based on estimates of the populations of the minor civil divisions made annually. There being 1800-plus minor civil divisions in 72 counties in Wisconsin, this presented a formidable task. The State Department of Administration was charged with making the estimates, and early in 1972 contracted with the University of Wisconsin for methodological development. At that time an informal "seminar" was organized which included some academic demographers, statisticians, and people from state government, who had been making county estimates for Wisconsin, and were familiar with some of the data sources that are useful in making population estimates.

Rather than rehearse the entire history of the enterprise, I'd like to discuss three aspects of our experience which may be of general interest:

1. What we came to view as the dimensions of the problem;
2. How we came to the solution we did;
3. How the method was tested.

One of the most important aspects of the problem, as we quickly realized, was the complementary problem of the large number and small size of the units to be estimated. In 1973 there were 1,872 cities, villages, and towns in Wisconsin, and what was remarkable about them is that 88 percent had populations of less than 2,500. Table 1 shows the distribution by type and size.

If that is contrasted with the estimation problem at the county level, where only 12.5 percent of the counties of Wisconsin have populations less than 10,000, and none less than 2,500, it is clear first, that a good-sized data set was necessarily involved, and that a set of that size is difficult to monitor year to year.

Second, it was imperative to use data that had a uniform collection and management system. Building permit data, each municipality applying its own rules; or utility meters, each of several hundred companies using its own standards; or school census data, several hundred districts responding with no State surveillance; all proved fatally flawed. Third, in many of our communities, the smallness of the numbers was going to lead to considerable random variability and therefore instability of estimates.

The second limitation on our enterprise was the legislative requirement that the estimates be current. This had been interpreted by the court to mean that the estimate effective date and year of production must be the same. Thus, we had to produce January 1, 1973 estimates by August 1, 1973.

Table 1

MUNICIPALITIES BY TYPE BY SIZE
Wisconsin, 1970

Municipality	Town		Village		City		Total		
Size - 1970	No.	%	No.	%	No.	%	No.	%	Cum. %
Less than 500	385	30.3	184	46.1	10	4.9	579	30.9	30.9
500 - 1,000	518	40.8	121	30.3	10	4.9	649	34.7	65.6
1,000 - 2,500	289	22.8	69	17.3	62	30.5	420	22.4	88.0
2,500 - 5,000	61	4.8	13	3.3	43	21.2	117	6.3	94.3
5,000 - 10,000	12	0.9	7	1.8	32	15.8	51	2.7	97.0
Over 10,000	5	0.4	5	1.2	46	22.7	56	3.0	100.0
Total	1,270	100.0	399	100.0	203	100.0	1,872	100.0	100.0

The implications of this limitation were that any data sources that could not be made available in time for August 1 estimates were in effect useless to us.

The third influence or dimension of this problem was clearly the fact that a considerable sum of money was riding on the outcome of these estimates. The revenue-sharing amounted to $35 per capita and that, for almost all communities, is a sizable part of their budget. This therefore subjected the whole process to close scrutiny and made for some very unacademic political pressures.

However, the greatest limitation on the methods that could be chosen was the availability of data. We set out to assess the problem by reviewing the conventional methods for small area population estimation and the data required for use of these methods. What we found was discouraging. Vital events, for instance, which are indispensable for the Vital Rates method, Component II and Composite, and are sometimes part of the Ratio Correlation method, were simply not available at the M.C.D. level at all, and they were usually not available by the August 1 deadline.

School enrollment data, which are essential for the Component II and the Composite methods and are frequently used in the Ratio Correlation method could not be used at the M.C.D. level because the geographic allocation is by school districts which respect no political boundaries.

A third standard data set that would have been useful in both Component and Composite is Medicare enrollments, but this too we found was unavailable at the M.C.D. level, and in any case was not available in time to be useful for our estimates.

So, what we were reduced to, was finding administrative data that could be used symptomatically. I'll spare you a listing of the dozens of sets that we investigated and rejected. What we came down to as reliable administrative data sets associated with population were motor vehicle registration data and data from state income tax returns.

Given these data limitations, the choice of the method was clearly a restricted one. What I'd like to examine now is the process by which we decided on the method chosen.

These results of our data search reduced our choice of method to a Ratio Correlation procedure, which has the appeal of being in wide use and being familiar in the literature, or a Censal Ratio type procedure. Unfortunately the only data series long enough for us to construct a ratio correlation regression equation was the motor vehicle data. The tax data had a serious discontinuity in the early 60's as a result of passage of a withholding provision into the revenue statute. The other choice, you will recall, was the Censal Ratio method, where a symptom population ratio in the census year is updated to the estimate year, and that estimated ratio, with the current symptom count, is used to produce a population estimate. The Censal Ratio procedure allowed the use of both the motor vehicle data and income tax data to produce estimates.

One would prefer more than one symptomatic data set and therefore one would prefer to use the Censal Ratio method, we nevertheless performed a test to determine the direction in which we should go.

The test consisted of making estimates of the 1970 populations for 71 of Wisconsin's counties from the 1960 base, and then comparing the errors of the estimates. This approach permitted us not only to compare our alternatives to each other, but also to evaluate our alternatives against the results of other county estimating procedures tested in the Federal-State Cooperative Program. For our test, we used three estimation methods: a Ratio Correlation procedure based on the automobile series, a conventional Censal Ratio procedure patterned after the Vital Rates procedure, but using motor vehicle data.

We referred to the third method as the Ratio Difference Estimator. This estimator was also a Censal Ratio procedure, but unlike conventional Censal Ratio procedures, uses a difference estimator to update the Symptom-population ratio of the census year to the time for which the estimate is desired.

Briefly the procedure is as follows: for each area for which the population is to be estimated, the ratio of symptom to population for the base year is calculated, viz.,

$$r_o(u) = \frac{\left\{\begin{array}{l}\text{Symptom at time zero, or}\\ \text{base year for the } u^{th} \text{ area}\end{array}\right\}}{\left\{\begin{array}{l}\text{Population count at time}\\ \text{zero for the } u^{th} \text{ area}\end{array}\right\}}$$

A similar ratio $R_o(S)$ = is computed for the area S; where S is a larger area containing all the "u" areas, i.e.,

$$R_o(S) = \frac{\text{Total symptom at time zero for S}}{\text{Total population in S at time zero}}$$

An independent estimate for the population at time t is obtained and used to estimate the ratio,

$$R_t(S) = \frac{\text{Total symptom for S at time t}}{\text{Total population for S at time t}}$$

We call this estimate $\hat{R}_t(S)$ and use it to estimate the ratio,

$$r_t(u) = \frac{\text{Symptom at time t for } u^{th} \text{ area}}{\text{Population at time t for } u^{th} \text{ area}}$$

with difference estimator,

$$\hat{r}_t(u) = r_o(u) + \hat{R}_t(S) - R_o(S).$$

The estimate $\hat{r}_t(u)$ is then used to generate an estimate of the population at time t in the u^{th} area by dividing the symptom for the u^{th} area at time t with the estimating ratio, i.e., the estimated population for the u^{th} area at time t is,

$$\hat{P}_t(u) = \frac{\text{Symptom at time t for } u^{th} \text{ area}}{\hat{r}_t(u)}$$

The test comparisons were made by reference to a series of five measures of accuracy. The first was the mean percentage absolute deviation from the census counts. This is conventional and gives a good overall measure of average performance. Its disadvantages are that it is:

1) unweighted, i.e., a 5% error on 50 weighs the same as a 5% error on 700,000; and that 2) since it deals in absolute values, it gives no indication as to the presence or absence of bias.

The second measure of accuracy used was the simple percent of deviations that were positive. This enables one to evaluate the estimates for the presence of bias.

The third was the mean square error.

And finally, two measures were used that are not well-represented in the literature of population estimation, the number and percent of misallocation. Misallocation is calculated as the sum of the absolute values of the deviations, that sum divided by two. This has the advantage of being a weighted measure, of being intuitively appealing inasmuch as it actually measures the number of people who wound up in the wrong jurisdiction as a result of estimation errors. For purposes of tax sharing in particular, this was the most direct measure of performance we could devise, for by multiplying the misallocation number by $35 you have an exact amount of money misallocated among the jurisdictions estimated. The percent misallocation is the percent of the total population which was misallocated.

Table 2 represents the various test measures as produced in the Federal-State Cooperative Program in the test for Wisconsin counties, and then those of the ratio correlation, conventional censal ratio and censal ratio with a difference estimator. While the ratio correlation method and the ratio difference method have similar percent errors, the mean square error seems to indicate that the distribution of errors is more favorable for the ratio difference method. This is seemingly confirmed by the percent misallocation figures.

While these results were gratifying at the county level, it told us little about how this method would perform at the M.C.D. level. For an M.C.D. test, we used 1960 data and 1970 data and compared the estimate with the 1970 census counts. This could only be done with the motor vehicle registration data, which one could expect would be less accurate alone than in combination with other data series. The mean deviation of the 1,800-plus minor civil division estimates was about 10% and the misallocation about 3.2% after 10 years. Graph 1 shows the distribution of errors. Table 3 shows the error measures of this test.

Table 2
ERROR MEASURES OF VARIOUS ESTIMATION METHODS
Wisconsin Counties
1960 Based 1970 Estimates

Method	% Error	M.S.E.	Mis-allocated	% Mis-allocated
Vital Rates	5.49	19836576	90320	2.05
Regression*	4.52	14896030	78312	1.77
Component*	6.43	79084064	118312	2.68
Composite*	3.85	7980251	59422	1.35
Ratio Difference**	1.96	5258575	35902	0.81
Censal Ratio**	2.25	15965140	47834	1.08
Ratio Correlation**	2.20	17552576	48929	1.11

* These are the results from the Federal State Co-op Program adjusted so as to sum to the 1970 Census total for the 71 counties used. (Wisconsin now has 72 counties.)

** Only one Symptomatic data set, viz., number of passenger automobiles registered to each county was used to make these estimates.

Graph 1
DISTRIBUTION OF PERCENT ERROR OF 1960 BASED ESTIMATES FOR 1970 MCD POPULATIONS (AUTOS ONLY)

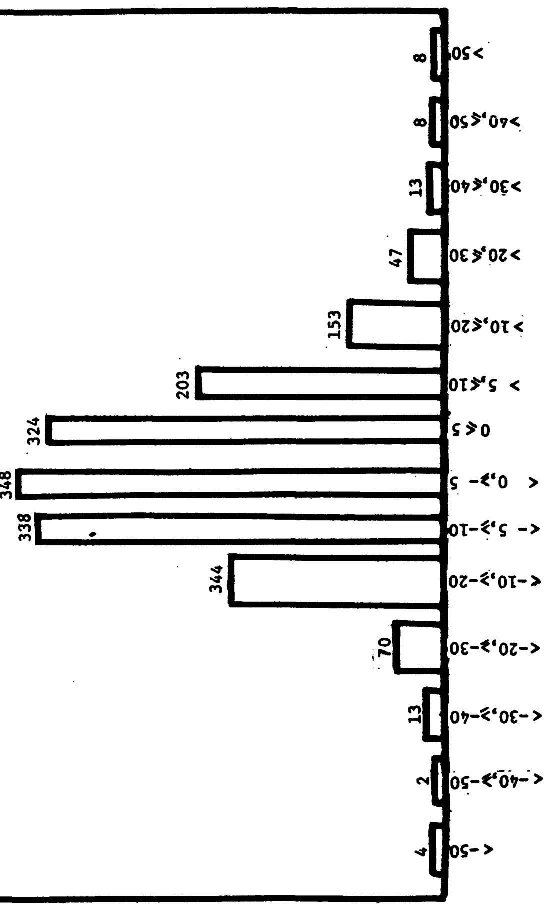


Table 3

Summary Statistics on the Estimation Errors of 1960 Based Ratio-difference Estimates for 1970 Wisconsin M.C.D. Populations

Average Percent Error	10.03
Mean Square Error	440,262
Misallocation	140,753
Percent Misallocation	3.19

The only hard testing available for the income tax data was a test of Dane County, and its 60 M.C.D.'s. This was fortuitous in that Dane County had been one of the sites of a dress rehearsal census in 1968. While the tax data by itself did not perform any better than the motor vehicle data, what we found was that the averaging of the tax estimates with those made from motor vehicle data reduced the observed errors. Mean percent absolute error between 8 and 9%, misallocation percent 2.4. The percent misallocation of the various tests are summarized in Table 4.

Based on our detailed evaluation of all these tests, we made the following choices in constructing the estimation model for minor civil divisions:

First, that it should be a two-step process that proceeds from the state control total to county estimates; these county estimates then serve as county control totals to the M.C.D. estimates within each county.

Our empirical results confirmed the suggestion made earlier to us by Peter Morrison that this would improve the estimates at the M.C.D. estimates level.

The second decision was that the Censal Ratio method, with the difference estimator as the updating mechanism, be used on three separate series of data and the unweighted average of these be the final estimate for the M.C.D. The three series were passenger automobiles, income tax filers, and dollar value of exemptions claimed.

The third decision concerning the actual estimation model was to apply the Censal Ratio method estimating model to the uncountable population only, that is, removing the institutional population from the base population and getting an independent count of that institutional population at the estimate year to add to the estimate generated by the Censal Ratio method.

Table 4

MISALLOCATION RESULTS FOR SOME ESTIMATES OF WISCONSIN POPULATIONS USING THE RATIO-DIFFERENCE ESTIMATOR

		% of Total Population Misallocated
1	1960 Based 1970 estimates for Wisconsin Counties (autos only)	0.81
2	1960 based 1970 estimates for Wisconsin Municipalities (autos only)	3.19
3	1968 based 1970 estimates for Dane County's municipalities (autos only)	1.6
4	1968 based 1970 estimates for Dane County's municipalities (Average of auto, files, and $ exemption estimates)	1.5
5	1973 Estimates for Wisconsin municipalities with special census in 1973	2.0

The difficulties were surmounted and the estimates were made on time, to the surprise of some of us.

As a result of the preliminary estimates, some 80 challenging cities submitted claims of underestimation. Of these, we found about 20 to be substantial, generally meaning some data error had occurred. The data in these cases were adjusted before the final estimates were made. Final estimates did not satisfy everybody; some 14 municipalities brought suit for judicial review of the estimation method. After an evidentiary hearing in the circuit court, which consumed some fifteen days, the judge strongly upheld the reasonableness of this method.

There have been some 50 special censuses conducted in Wisconsin M.C.D.'s around the time of the estimates, and these generally confirm our expectations as to the probable precision of the estimates. The average error is about 6% and misallocation is about 2%.

Reference: U. S. Bureau of the Census, Current Population Reports, Series P-26, No. 21, "Federal-State Cooperative Program for Local Population Estimates: Test Results--April 1, 1970."

LINEAR COMPRESSIONS USING THE IMAGE FACTOR ANALYSIS CRITERION

Michael D. Pore, University of New Orleans

Image analysis, as developed by Guttman [1953], is the partitioning of an observable random vector, X, into an image vector, WX, and an anti-image vector, (I-W)X.

(1) $$X = WX + (I-W)X$$

where WX is the minimum squared error linear estimate of X such that

$$\text{diag } W = \phi.$$

Diag W (implies W is a square matrix) is a diagonal matrix with elements on the main diagonal identical to the main diagonal of W.

Let R be the sample correlation matrix associated with a random sample of $p \geq N$ observations (each $N \times 1$) on X.

Theorem 1: The coefficient regression matrix, W (defined above), can be shown to by any solution to the system

(2) $$\text{Diag } W = \phi$$

(3) $$WR = R - D$$

for some diagonal matrix D.

A literature search has not revealed a general solution to the system (2) and (3), but only a solution in the case that R is non-singular. This solution follows by post-multiplying (3) by R^{-1} to yield

(4) $$W = I - DR^{-1},$$

and then impose the restriction (2) to show that

(5) $$D = (\text{diag } R^{-1})^{-1}.$$

Current computer programs in image covariance factor analysis, for example in the Statistical Programs in the Social Sciences (SPSS), use (4) and (5) but are not operable for singular R. The property that each element of (5) is non-negative, and several other properties shown by Guttman [1956] have made (5) a nearly universal estimate of the covariance of the "unique" variables in factor analysis. In order that image analysis might be more generally used and reasonable estimates for factor analysis developed, a general solution to Theorem 1 is now developed. The following preliminary results are used to develop Theorem 2. R^+ represents the pseudo-inverse of R, as defined in Boullion and Odell [1971].

Lemma 1: There are no zeros on the main diagonal of R^+.

Notation: Let $A = \text{diag}(I - R^+R) = (I - \text{diag } R^+R)$.

Theorem 2: A solution to the system of equations

$$W R = R - D$$

$$D \text{ is diagonal}$$

$$\text{diag } W = \phi$$

is

(6) $$W_0 = I - (I - AA^+)(\text{diag } R^+)^{-1} R^+ - A^+(I - R R^+$$

and

(7) $$D_0 = (I - AA^+)(\text{diag } R^+)^{-1}.$$

In general, all solutions to Theorem 2 are of the form

$$W = I - H_1 (I - A A^+) R^+ + H_2 (I - R R^+)$$

for some H_1 and H_2 where H_1 is diagonal. Theorem 2 presents the solution when H_2 is also diagonal.

Factor Analysis.

Factor analysis hypothesizes the existence of an integer M, $0 < M < N$, and random vectors Y and U such that

(8) $$\underset{N\times 1}{X} = \underset{N\times M}{\Lambda} \underset{N\times M}{Y} + \underset{N\times 1}{U}$$

where Λ is an $N \times M$ matrix of constant coefficients and

(9) $$E(Y) = \phi, \; E(U) = \phi,$$

(10) $$\text{Cov}(Y,Y) = I$$

(11) $$\text{Cov}(Y,U) = \phi$$

(12) $$\text{Cov}(U,U) \text{ is diagonal.}$$

This model implies the following partitioning of the correlation matrix:

(13) $$R = \Lambda\Lambda^T + \text{Cov}(U,U)$$

Two general methods exist for fitting data to this model. The first (1) is to initially estimate $\Lambda\Lambda^T$ with $R-D_0$, then use one of three criteria to reduce the rank of $R-D_0$. The three criteria are principal factor analysis (PFA), cannonical factor analysis (CFA), and alpha factor analysis (AFA). These three criteria are discussed later. The second method (2) is to initially estimate $\Lambda\Lambda^T$ with the image covariance matrix, $G = \text{Cov}(W_0X, W_0X) = W_0 R W_0^T = R + D_0 R^+ D_0 - 2D_0$, then use a "rank reducing" method on G.

Neither method has been possible for singular R, although iterative approximation schemes have been developed for the first (1) method. Tests of significance for "goodness of fit" exist for only one method, CFA, and then only under assumptions of normality. Hence, the present model-fitting methods are unsatisfactory, often giving rough approximations with gross errors.

Y is interpreted as the common factors. This means that conceptually

ΛY and WX are identical, yet Y is not a linear combination of X. (A proof is in Pore [1973]). Also, image analysis makes a strong argument against the factor analysis model including equation (12): since the anti-image covariance matrix Cov(I-W)X, (I-W)X, is not diagonal.

These criticisms, coupled with the experience of researchers declaring that equation (12) was not a significant assumption in their analyses, has led the author to drop this assumption and continue the analysis. This "relaxed" factor analysis model can not only be precisely fit with any set of data (for all M, $0 < M < N$) but many such solutions exist. Hence, to restrict it for more meaningful interpretations, the common factors, Y, are restricted to being a linear compression of X. This is

(14) $$Y = BX$$

where B is a full rank $N \times M$ matrix of unknown constant coefficients. The "modified" factor analysis model (MFA) is given by

$$X = \Lambda Y + U,$$

where $$E(Y) = \phi, \quad E(U) = \phi,$$

$$\mathrm{Cov}(Y,Y) = I,$$

$$\mathrm{Cov}(Y,U) = \phi$$

and

(15) $$Y = BX.$$

It can easily be shown that

(16) $$\Lambda = RB^T$$

and

(17) $$U = (I - RB^TB)X,$$

with the only restriction on B being that it is $M \times N$ and satisfies

(18) $$BRB^T = I.$$

The researcher need no longer estimate common factor scores for particular individuals. They are specified in the model (15) as a linear compression of the observation. This also makes interpretation of Y a precise linear compression of X.

In being so general the MFA model allows data to be fitted to the model in ways that are meaningless to the researcher. That is, thare exist matrices, B, such that the MFA model is satisfied, yet y is uninterpretable. The problem is that

$$BRB^T = I$$

does not sufficiently restrict B to meaningfull solutions. Classical partitioning procedures may be imposed to optimize interpretability. One application of this model, MPC, is to partition B as

$$B = PA,$$

so that AX are the principal components of X. P then reduces the rank to the M largest principal components and scales B so that (18) is satisfied. This is one meaningful way to select B, but it is not in keeping with the concepts of factor analysis: that is, it does nothing toward separating error and underlying factors, etc. The following method is designed to do just that.

Application: B is partitioned so that

(19) $$B = PW$$

where WX is the image of X and P, an $M \times N$ matrix, reduces the rank of B and scales it so the (18) is satisfied. (See (1) and Theorem 2 for an explicit definition of W).

The superior properties of the image vector in factor analysis have been discussed. It is, conceptually, the type of common factor "filter" for which the researcher uses factor analysis. Y, then, represents a linear compression of the image of X. The precise type of linear compression will depend on the researcher's criteria for optimization. Classical criteria include principal components, canonical correlation, and generalizability. Methods using each of these criterion are now presented.

MFA Applied to PFA Criterion (MPF).

PFA is the method where Cov(U,U) is estimated with D_0 of Theorem 2, and then the largest M principal components of $R-D_0$ are extracted. MFA follows the same principle. The main difference is that the classical factor analysis model is not being approximated, hence estimates of Cov(U,U) are no longer diagonal. Image theory justifies using the anti-image covariance matrix as an initial extimate of D. The procedure is, then, to extract the largest M principal components from the image covariance matrix.

The image covariance matrix can be shown to be

(20) $$G = R + D_0R^+D_0 - 2D_0.$$

Application: Let E be the $N \times N$ p.s.d. diagonal matrix with the eigenvalues of G on the main diagonal in descending order and H the respective $N \times N$ eigenvector matrix such that

(21) $$G = HEH^T = [H_1 \quad H_2] \begin{bmatrix} E_1 & \phi \\ \phi & E_2 \end{bmatrix} \begin{bmatrix} H_1^T \\ H_2^T \end{bmatrix} = H_1E_1H_1^T + H_2E_2H_2^T$$

for partitionings of E and H such that E_1 is $M \times M$ p.d., and H_1 is $N \times M$.

Now P in (19) is

(22) $$P = E_1^{-\frac{1}{2}} H_1.$$

Notice that

$$BRB^T = E_1^{-\frac{1}{2}} H_1 WRW^T H^T H_1^T E_1^{-\frac{1}{2}}$$

$$= E_1^{-\frac{1}{2}} H_1 G H_1^T E_1^{-\frac{1}{2}}$$

$$= E_1^{-\frac{1}{2}} E_1 E_1^{-\frac{1}{2}}$$

(23) $$= I.$$

BX is the scaled linear compression of the image variables that retains the M largest principal components of the image variables. Those researchers partial to PFA will find the same favorable principles in MPF, but applied to the image variables, rather than a hypothetical set of variables with covariance R-D.

MFA Applied to CFA Criteria (MCF).

CFA is the method of constructing factors, Y, that have maximum canonical correlation with the observations X. There are at least two possible ways to apply this principle to MFA. These are: (1) maximizing the correlation of Y and X, as above, with the restriction Y = PWX (2) maximizing the canonical correlation of Y and WX, with the restriction Y = PWX. The first method is what will be used to develop MCF. The second method remains undeveloped.

Proceeding in the fashion developed for CFA:

(24) $$Cov(Y,Y) = D$$

(25) $$Cov(X,X) = R$$

(26) $$Cov(X,Y) = RW^T P^T = \Lambda.$$

Hence, $\frac{Y}{X}$ has the super covariance matrix

$$\Sigma = \begin{bmatrix} I & \Lambda^T \\ \Lambda & R \end{bmatrix}$$

Following the same procedure as Anderson [1957] (Chapter 12),

$$|\Lambda\Lambda^T - \lambda R| = 0$$

but since our estimate of $\Lambda\Lambda^T$ is G, the image covariance matrix, then

(28) $$|G - \lambda R| = 0.$$

Now since R can be written

$$R = Q[e_i]Q^T$$

where $[e_i]$ is an $r \times r$ p.d. diagonal matrix, r is the rank or R, and Q is an $N \times r$ matrix of eigenvectors of R respective to $[e_i]$ then (28) implies

$$[e_i]^{-\frac{1}{2}} Q^T G Q [e_i]^{-\frac{1}{2}} - \lambda I = 0.$$

where the λ is the canonical correlation involved.

Application: Let $[u_i]$ be the p.d. diagonal matrix of eigenvalues of $[e_i]^{-\frac{1}{2}} Q^T G Q [e_i]^{-\frac{1}{2}}$ in descending order, and S the respective matrix of eigenvectors.

It can be shown that

(29) $$P = [u_i]_1^{-\frac{1}{2}} S_1^T [e_i]^{-\frac{1}{2}} Q^T.$$

Notice that (18) is assured by

$$WRW^T = G$$

and it follows that

$$PGP^T = I.$$

The researcher preferring CFA will find MCF more theoretically defensible, yet based on the same optimization procedure. MCF also incorporates the properties of image analysis, hence optimization is bi-dimensional, rather that simply in the one aspect, correlation.

MFA Applied to AFA Criteria (MAF).

The AFA criterion is to define the common factors in such a way that their reliability coefficient is maximized. MAF will use the same reliability coefficient, α, that AFA used. (Tryon [1957] has shown that practically all reliability coefficients are equivalent to α.) The α coefficient is

$$\alpha = \frac{N}{N-1}\left[1 - \frac{\omega^T(\text{Trace } \Lambda\Lambda^T)\omega}{\omega^T \Lambda\Lambda^T \omega}\right]$$

where ω is the coefficient vector in defining the common factors

(30) $$Y = \omega^T WX.$$

By substituting the initial estimate of $\Lambda\Lambda^T$ with G (defined by (20)) yields

(31) $$\alpha = \frac{N}{N-1}\left[1 - \frac{\omega^T(I - D)\omega}{\omega^T G \omega}\right]$$

which, in turn, yields an identical eigenvalue equation as in AFA (see Kaiser and Caffrey [1965]) except R - D is replaced by G. Hence,

(32) $$(H^{-1}GH^{-1} - U^2 I)e = \phi$$

where $H^2 = I - D$.

Application: Let $[e_i]$ be the $N \times N$ diagonal matrix of eigenvalues of $H^{-1}GH^{-1}$, in descending order. Also let Q be the respective eigenvector matrix. It can be shown that

$$P = [e_i]_1^{-\frac{1}{2}} Q_1^T H^{-1}.$$

Notice that

$$PGP^T = I,$$

hence the restrictions for MFA are satisfied. Since P is the objective of the procedure above, it differs slightly from AFA; but the principle, general method of analysis, and interpretability are all identical to AFA. Although AFA is not as widely used as the other forms of factor analysis, researchers may find an increasing need for psychometric sampling (or Q-analysis, as it is sometimes called: [Rummel, 1970]). As researchers do, they may find AFA claims the type of analysis they are looking for; but MAF will do likewise, and it will fit a model precisely, as opposed to the approximation technique of classical methods.

REFERENCES

Anderson, T.W. An Introduction To Multivariate Statistical Analysis. New York: John Wiley & Sons, Inc,, 1957.

Boullion, Thomas L., and Odell, Patrick L. Generalized Inverse Matrices. New York: John Wiley & Sons, Inc., 1971.

Guttman, L. Image theory for the structure of quantitative variates. Psychometrika, 1953, 18, 277-296.

__________. Some necessary conditions for common factor analysis. Psychometrika, 1954, 19, 149-161.

__________. "Best Possible" systematic estimates of communalities. Psychometrika, 1956, 21, 273-285.

Kaiser, H.R., and Caffrey, J. Alpha factor analysis. Psychometrika, 1965, 30, 1-14.

Pore, Michael D. Factor analysis as a data compression technique. Unpublished Ph.D. Dissertation, Texas Tech University, 1973.

Rummel, Rudolph J. Applied Factor Analysis. Evanston, Illinois: Northwestern University Press, 1970.

Tryon, R.C. Reliability and behavior domain validity: reformulation and historical critique. Psychological Bulletin, 1957, 54, 229-249.

WHAT PROPORTION OF ITEM NON-RESPONSE SHOULD LEAD TO DELETING THE CASE?

Charles H. Proctor, North Carolina State University

Introduction

In processing survey questionnaire data on attitudes there sometimes arises the dilemma of having either to impute many item responses because a respondent has not answered every item or to delete him from the study. Because this problem may arise at the stage of cleaning the data even before it is decided how to analyze the attitude scores, we will consider only the use of the grand average item score imputed in place of an item non-response. Other alternative ways of fitting item scores based on further knowledge of the respondent and on what parameters are to be estimated or tested and how they are tied to the distribution of scale scores can, of course, be proposed. However, it has been reported that using the mean to replace a missing value [the "zero order" method in Afifi and Elashoff, 1967], yields a relatively acceptable mean square error of the estimated regression coefficient so long as x and y are not too closely correlated. Such a result would likely hold true in many of the present cases.

In addition to dealing only with this one method of fitting missing responses, there are some other rather narrow assumptions to be made concerning: (1) the problem to be solved (namely, detecting a shift in a sub-group mean of attitude scores from the overall population mean), (2) the model (namely, independence, homogeneity and near normality of measurement errors) and (3) the pattern of non-response (namely, one respondent missing a few item responses). I would contend that the recommendations for a wider variety of estimation and testing problems on the association between other variables and the attitude score, under a more realistic model, with more complex patterns of non-response will turn out to be similar to those discovered below for this particular case. At any rate this paper is not intended to delimit exactly the scope of the recommendations, but mainly to point toward a way of posing and solving such a problem.

Distributional Assumptions

The quantity of interest, a scale score, is taken to be the unweighted average of item scores and the items are assumed to be about equally intercorrelated. Each item might have possible answers of the form {Strongly Disagree, Disagree, Undecided, Agree, Strongly Agree} with corresponding item scores of {1, 2, 3, 4, 5}. Of the, let us say, K items in the scale a respondent may not answer PK of them, where P is thereby defined as the proportion of non-response. Q = 1-P becomes the proportion of items responded to.

The model equations for these item and scale scores will be written in the spirit of psychological scale theory [Gulliksen, 1950] as if for continuous random variables with the pious hope that the actual discreteness of the distributions will not upset the conclusions too badly. The first model equation for an item score then becomes:

$$(1) \qquad y_{ik} = a_i + e_{ik} ,$$

where y_{ik} is the i^{th} respondent's score on the k^{th} item. The quantity a_i is his true score, while e_{ik} represents measurement error in i's response to the k^{th} item. Both of these terms are taken to be random and all such quantities are assumed to be mutually independent with homogeneous variance. In actuality there would also be included in model equation (1) a fixed effect, a λ_k say. A reasonable restriction is $\Sigma \lambda_k = 0$ and upon taking the average item score the λ_k's disappear so they can be ignored without losing generality.

The second model equation is for the i^{th} respondent's scale score

$$(2) \qquad y_i = a_i + \bar{e}_i^{(K)} ,$$

in which

$$\bar{e}_i^{(K)} = \sum_{k=1}^{K} e_{ik}/K .$$

The respondents of the sub-group of interest will be numbered so that the last one, the n^{th} one, is the non-response case. The imputed item scores for the n^{th} respondent are, as mentioned already, made equal to the grand average of the population surveyed, namely:

$$(3) \quad y_{nj} = \mu_o \quad \text{for} \quad j = 1, 2, \ldots, PK .$$

The n^{th} respondent's scale score is then expressible as:

$$(4) \qquad y_n = P\mu_o + Qa_n + \bar{e}_i^{(QK)} ,$$

where $\bar{e}_i^{(QK)}$ is the mean of the QK answered-item measurement errors.

Testing Problem and Test Statistics

It is supposed that the n respondents belong to a special, so called, treatment group wherein their average true score is μ_1, perhaps differing from μ_o. That is, $E(a_i) = \mu_1$ when i is randomly drawn from the treatment group, but $E(a_i) = \mu_o$ if i is randomly drawn from the study population. The objective of the study is taken to be to decide whether $\mu_1 = \mu_o$. Again to simplify matters, we assume that μ_o is known so much more precisely than is μ_1 that it can be supposed that μ_o is known without any uncertainty. The two procedures: (a) delete the n^{th} case or (b) impute the PK item responses, then lead to the following two test statistics having the following variances.

If the n^{th} case is deleted then one computes:

$$(5)\quad \bar{y}_{(n-1)} = \sum_{i=1}^{n-1} y_i/(n-1) \equiv \bar{a}_{(n-1)} + \bar{\bar{e}}^{(K)}_{(n-1)} \quad .$$

From the model equation and supposing the variance of the a_i's to be σ_a^2, the mean and variance of $\bar{y}_{(n-1)}$ are found as:

$$(6)\qquad E(\bar{y}_{(n-1)}) = \mu_1 \ ,$$

and

$$(7)\quad V(\bar{y}_{(n-1)}) = (\sigma_a^2 + \sigma_e^2/K)/(n-1) = V_a \quad \text{say},$$

where σ_e^2 is equal to $V(e_{ik})$ for all values of i and of k.

Incidentally, since the quantities σ_e^2 and σ_a^2 have now been introduced, the reliability of the scale score, namely the square of the correlation between a_i and y_i, may be computed as:

$$(8)\qquad \rho = \sigma_a^2/(\sigma_a^2 + \sigma_e^2/K) \ .$$

When the n^{th} case is assigned item scores as in (3) one computes:

$$(9)\qquad \bar{y}'_{(n)} = n^{-1}\sum_{i=1}^{n} y_i = \frac{n-1}{n}\ \bar{a}_{(n-1)} + \frac{Qa_n}{n} + \frac{P\mu_o}{n} + \frac{n-1}{n}\ \bar{\bar{e}}^{(K)}_{(n-1)} + \frac{\bar{e}^{(QK)}}{n} \quad .$$

The mean and variance of this statistic then are seen to be:

$$(10)\qquad E(\bar{y}'_{(n)}) = \frac{n-1+Q}{n}\ \mu_1 + \frac{P}{n}\ \mu_o = [(n-P)\mu_1 + P\mu_o]/n$$

and

$$(11)\quad V(\bar{y}'_{(n)}) = [(n-1+Q^2)\sigma_a^2 + (n-1+Q^{-1})\sigma_e^2/K]/n^2 = V_b, \quad \text{say} \ .$$

In order to test for a deviation of μ_1 from μ_o one could choose between the following ratios. If the n^{th} case is deleted the test statistic is:

$$(12)\qquad Z_a = (\bar{y}_{(n-1)} - \mu_o)/\sqrt{V_a} \ ,$$

or if item scores are imputed the test statistic is:

$$(13)\qquad Z_b = (\bar{y}'_{(n)} - \mu_o)/\sqrt{V_b} \ .$$

Comparison of the Two Test Statistics

Notice the differing non-centralities of these statistics in:

$$(14)\qquad E(Z_a) = (\mu_1 - \mu_o)/\sqrt{V_a} \ ,$$

and in:

$$(15)\qquad E(Z_b) = \frac{n-1+Q}{n}(\mu_1 - \mu_o)/\sqrt{V_b} \ .$$

The probability of detecting a given difference between μ_1 and μ_o will be larger with Z_a than Z_b if the quantity:

$$(16)\qquad CV_a = V_a/(\mu_1 - \mu_o)^2$$

is less than:

$$(17)\qquad CV_b = V_b/(\tfrac{n-1+Q}{n})^2(\mu_1 - \mu_o)^2 \ .$$

Thus, the two methods will be equally precise if $CV_a = CV_b$, or if $CV_b - CV_a = 0$. When CV_b exceeds CV_a then method (a), deleting the case, is more attractive.

Using equation (8) so as to introduce the reliability ρ into the expression and solving $CV_b - CV_a = 0$ for the break-even values of Q leads to:

$$(18)\quad [(n-1)\rho - 1]Q^3 - 2(n-1)Q^2 + (n-1)(1-\rho) = 0$$

Notice that the first derivative with respect to Q of this expression is

$$(19)\quad 3[(n-1)\rho - 1]Q^2 - 4(n-1)Q = -Q[3(n-1)(1-Q\rho) + Q + (n-1)] ,$$

and, since ρ and Q are always positive and less than 1, this quantity will be negative. Thus the attractiveness of method (a) will always decrease with increasing Q. That is to say, as the rate of responding increases, the second method, namely imputing item scores, becomes more attractive, while if a case presents too high a rate of non-response it may be best to delete it.

Dividing equation (18) by n and then allowing n to increase leads to:

$$(20)\qquad \rho Q^3 - 2Q^2 + 1-\rho = 0 .$$

Equation (20) does not depend on the sub-group size and so is available for use in a variety of estimation and testing situations.

Solving equation (20) for Q at reliabilities (values of ρ) that one meets in common practice (that is, between .6 and .8) the proportion required is found to range from 30 to 50 percent of the items to be answered. With a scale reliability of .75, only persons answering 38 percent or less of the items would be deleted. The accompanying Figure 1 and Table 1 show in more detail how Q depends on ρ and n. Comparison of the columns $n = \infty$ and $n = 10$ in Table 1 shows them to be sufficiently close so that the curve for $n = 10$ was omitted from Figure 1.

Recommending Deletion or Imputation when Cleaning the Data

Using Figure 1 plus knowledge of ρ, one can thus decide what is best to do for any respondent, depending on his rate of item non-response, Q. It is of interest to examine the relative precision of imputation versus deletion, as measured by the ratio CV_a/CV_b when Q varies. This can indicate how serious it would be to ignore the recommendation based on Q and blindly impute or delete.

The ratio, RP, is expressible in terms of ρ as:

$$(21)\quad RP = \frac{CV_a}{CV_b} = \frac{[1 + Q/(n-1)]^2}{1 + [\rho Q^2 + (1-\rho)/Q](n-1)} .$$

As n grows larger the precision of imputing relative to deleting becomes quite insensitive to the proportion of item non-response so let's look first at the case n=2. That is, replace n-1 by 1 in formula (21). For concreteness, let $\rho = 3/4$ and $K = 6$. Then the possible values of Q are zero, 1/6, 1/3, 1/2, 2/3, 5/6 and 1 with the corresponding values of relative precision of zero, .540, .970, 1.333, 1.626, 1.846 and 2.000. The break-even value seems to be near $Q = 2/6$ as can also be verified from Table 1 which gives it as $Z = .346$. The drop in relative precision to 54% and the rise to 133% for non-response rates adjacent to $Q = 2/6$, suggest some serious losses in information if the recommendation is ignored. However, taking n=2, as was done in this calculation, is tantamount to supposing that the non-response is occurring for one of just two cases in the only sub-group of interest within the study population. In actual applications, the size of the sub-group of interest would average nearer a half of the total sample size for the whole survey.

In the case that n=2 the relative precision of .540 at $Q = 1/6$ expresses the fact that if one of the two respondents in a special sub-group answers only one of six items and we impute his item scores on the other five and then lump the two scale scores together, this processing decision results in only 54% of the precision obtainable by deleting him and using only his partner in testing for the sub-group's distinctiveness. If, for example, each case costs $10 (to select, to measure and to tabulate) so that the pair costs $20 then a relative precision of .540 means roughly obtaining a return of only $10.80 or $5.40 per case.

When the proportion of item non-response is near the break-even point, it does not matter which method is used, so no loss occurs. When the proportion is very near zero or near one there is no question about what to do and no loss occurs. It is only when item response is near .20 or near .70 and one may elect the wrong procedure, say imputing or deleting respectively, that losses arise. Even in such cases it is seen that it will only be an extreme situation that could cause a loss of half the precision in the case. The moral is that one would do well to guess at values for ρ and n and use the resulting break-even value of Q in processing the data and expect only a minor penalty for not using the correct or true values of ρ and n. In short, the rule: Impute if item response is above 50% and delete if less, should cover many cases of actual survey practice.

Discussion

A brief word concerning the distributional assumptions behind the condition $CV_b - CV_a = 0$ may be in order. The form of all distributions is taken to be normal and the various variances, at least the value of ρ, are supposed known. As usual, one does not expect a real situation to conform to such suppositions and only expects, as we do in the present case, that non-normality and the need to estimate variances will not upset the conclusions concerning the break-even value of Q. The condition itself insures that no matter what difference exists between μ_1 and μ_0 the test statistics for the two cases will exhibit the same power.

A further refinement in the result may be to introduce some cost considerations. There is probably a greater processing expense involved in imputing than in deleting. However, I personally feel that it is an additional loss to delete any information furnished by a respondent who expects his answers to be used. Also, deleting cases runs a greater risk of introducing the selection bias associated with refusals. Thus I am content with equation (18). At any rate the method would be to express both CV_a and CV_b in terms of the total size of the survey, say n_T, whereupon the size of the treatment sub-group would be set at, say, $n = \pi\, n_T$. Also the costs, say C_a and C_b, would need to be shown as functions of n_T. The break-even level of Q would be derived by equating the change in C_a/CV_a per unit increase in n_T to the change in C_b/CV_b per unit increase in n_T.

References

Afifi, A. A. and Elashoff, R. M. Missing Observations in Multivariate Statistics, II. Point estimation in simple linear regression. Journal of the American Statistical Association, 1967, 62, pp. 10-29.

Gulliksen, Harold. Theory of Mental Tests. John Wiley, 1950, Chapter 16.

TABLE 1

Proportion of Items Answered, Q, at Break-even Point between Deleting Case and Imputing Item Scores for Levels of Scale Reliability ρ and Treatment Sub-group Size n

Reliability	Treatment Sub-group Size			Reliability	Treatment Sub-group Size		
ρ =	n = ∞	n = 10	n = 2	ρ =	n = ∞	n = 10	n = 2
.05	.695	.682	.607	.55	.512	.503	.452
.10	.683	.670	.596	.60	.484	.476	.429
.15	.670	.656	.584	.65	.453	.446	.404
.20	.654	.642	.571	.70	.419	.413	.377
.25	.638	.626	.557	.75	.382	.377	.346
.30	.621	.609	.542	.80	.340	.336	.311
.35	.603	.591	.527	.85	.293	.290	.271
.40	.583	.572	.510	.90	.237	.235	.222
.45	.561	.551	.492	.95	.165	.164	.158
.50	.537	.528	.473				

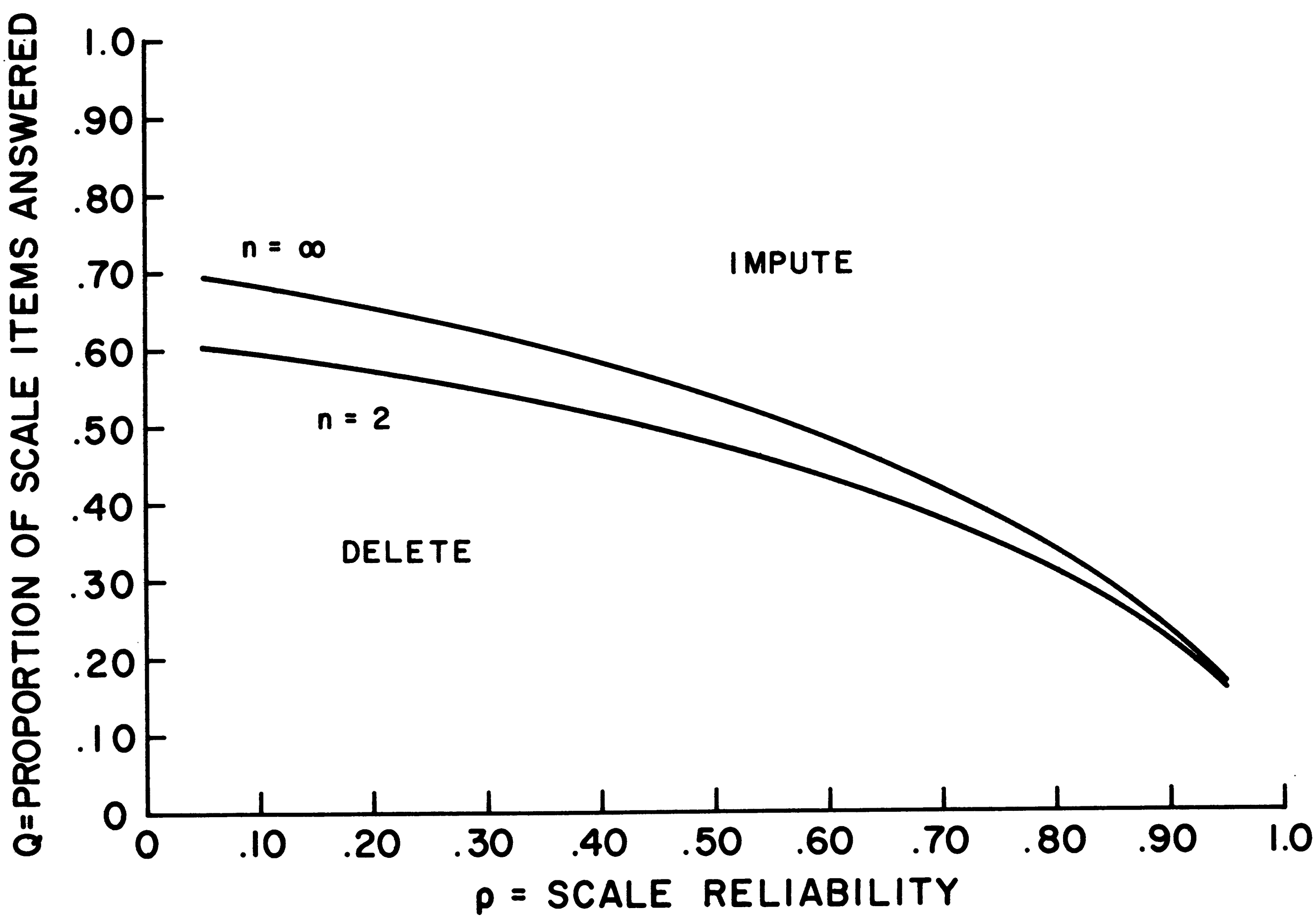


FIGURE 1

Break-even Proportion of Item Non-response as a Function of Scale Reliability

A FACTOR ANALYTIC APPROACH TO DISCARDING VARIABLES

A. C. Rencher, Brigham Young University

SUMMARY

Often, when a large number of variables are observed, little information is lost if some of the variables are discarded. A method based on factor analysis is proposed for selecting the variables to be rejected. An example is given wherein the method is used to reduce the number of questions on a questionnaire.

1. INTRODUCTION

In some multivariate situations it is desirable or even necessary to reduce the number of variables and work with a "good" subset. A method such as principal component analysis could be used to obtain a set of new variables defined as linear combinations of the original variables. Often a relatively small set of such new variables will preserve most of the information. However, in some cases the new variables are not useful. For example, suppose it is required to reduce the size of a questionnaire or test. It is usually not practicable to use new questions defined as linear combinations of the original ones. Even if such new questions could be identified and phrased meaningfully a problem arises in their use. If the subjects are asked to respond to the new questions it will be found that their answers are correlated. Much information is thereby lost since the new variables would be independent if obtained as orthogonal functions of the original variables.

In most common multivariate situations, the intended analysis is of greater interest than a prior reduction in the number of variables. However, in many cases the results will be only slightly affected if certain variables are discarded before performing the analysis. The ensuing savings in the number of required measurements would be especially helpful in situations which recur routinely.

2. ALTERNATIVE SELECTION METHODS

Several methods have been suggested for selection of variables. The following three approaches are discussed extensively by Jolliffe (1972, 1973).

(1) A stepwise multiple correlation method which discards successively the variable having the largest multiple correlation with the remaining variables. An empirically determined stopping point is suggested.

(2) Principal component methods which associate a variable with each of the principal components and discard those variables associated with the principal components having the smallest corresponding eigenvalues. An empirically determined cut off point is proposed.

(3) Clustering methods which segregate the variables into groups and select one variable from each of the resulting groups. The clustering stops when an empirically determined termination point is reached.

These three methods together with variations are compared by Jolliffe using both real and artifical data. None of the methods was found to be uniformly superior to the others.

3. A METHOD BASED ON FACTOR ANALYSIS

Intuitively, it seems one should be able to examine the correlation matrix and from the patterns therein discover which variables can be most readily sacrificed as providing the least additional information above that already available in their fellows. The three methods listed in Section 2 are in fact based on extracting information from the correlation matrix. Other simpler but less efficient approaches readily suggest themselves. For example, the sum of squares of each row of the correlation matrix could be examined and the variable with the largest sum of squares deleted. This process could be continued in a stepwise fashion after deletion of the appropriate row and column at each step. It appears however, that this procedure would be inferior to the first method reported in Section 2. It might be considered only if time or computational limitations prescribed its use. A far more satisfactory procedure for examining the correlation matrix is now described.

Consider the usual (orthogonal) factor analysis model for p variables expressed in terms of m factors.

$$
\begin{aligned}
x_1 &= \lambda_{11}f + \lambda_{12}f_2 + \dots + \lambda_{1m}f_m + e_1 \\
x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2m}f_m + e_2 \\
&\vdots \\
x_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pm}f_m + e_p
\end{aligned} \tag{1}
$$

In matrix form (1) can be expressed as $\underset{\sim}{x} = \underset{\sim}{\Lambda}\underset{\sim}{f} + \underset{\sim}{e}$. The following will be assumed: $E(\underset{\sim}{x})=\underset{\sim}{0}$, $\text{cov}(\underset{\sim}{x}) = \underset{\sim}{\Sigma}, E(\underset{\sim}{f}) = \underset{\sim}{0}$, $\text{cov}(\underset{\sim}{f}) = \underset{\sim}{I}$, $E(\underset{\sim}{e}) = \underset{\sim}{0}$, $\text{cov}(\underset{\sim}{e}) = \underset{\sim}{\Psi} = \text{diag}(\Psi_1,\dots,\Psi_p)$, and $\underset{\sim}{e}$ and $\underset{\sim}{f}$ are independent. From these assumptions it follows that

$$\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}. \tag{2}$$

In practice the correlation matrix is often used in place of $\underset{\sim}{\Sigma}$.

If two variables x_i and x_j have nearly identical loadings on the m factors, they contribute similar information and one of the two may be deleted. The following weighted distance function is suggested as a measure of the degree of closeness or similarity of x_i and x_j:

$$d_{ij}=w_1(\lambda_{i1} - \lambda_{j1})^2+w_2(\lambda_{i2}-\lambda_{j2})^2+\dots+ w_m(\lambda_{im}-\lambda_{jm})^2 \quad . \tag{3}$$

If the λ_{ik} are estimated by factoring $\underset{\sim}{\Sigma}$ (or $\underset{\sim}{\Sigma}-\underset{\sim}{\Psi}$) using eigenvalues and eigenvectors, appropriate weights are provided by the eigenvalues. Thus, w_1 is the largest eigenvalue, w_2 is the next largest, etc. If another method has been used to estimate the loadings or if they have been rotated, appropriate weights can be found from

$$w_k = \sum_{i=1}^{p} \lambda_{ik}^2 \qquad k = 1, 2, \ldots, m.$$

The following procedure is suggested to determine which variables can best be discarded. Find d_{ij} for each pair x_i and x_j, $i \neq j$. Arrange these in ascending order. Then from the pair of x's with smallest distance function value, retain the one with smaller Ψ value and delete the other. Similarly, one of the two variables with next smallest d_{ij} can be deleted. This process can continue until as many variables as desired are discarded. In each case when choosing between two variables it seems preferable to retain the one with smaller value of Ψ. If two variables are found to be similar and later each is found to be close to a third variable, two of the three may be deleted. This suggests the possible use of a clustering procedure. However, cluster analysis is not recommended due to the problem to be discussed next which can best be handled in the pairwise framework.

Negative correlations have an effect on the location of points in the space of factor loadings. If the response on a question is given on a multipoint ordered scale between poles, a reversal of the poles will change the sign of the correlation of the given question with all others. From (2) it is clear that

$$\sigma_{ij} = \sum_k \lambda_{ik}\lambda_{jk}, \; i \neq j.$$

A change in sign produces

$$-\sigma_{ij} = \sum_k (-\lambda_{ik})\lambda_{jk} = \sum_k \lambda_{ik}(-\lambda_{jk}).$$

Thus, either $(\lambda_{i1},\lambda_{i2},\ldots,\lambda_{im})$ or $(\lambda_{j1},\lambda_{j2},\ldots,\lambda_{jm})$ has all signs reversed and appears on the opposite side of the origin from its original position. If two points are close together in the factor loading space, a reversal in sign of the correlation between the two variables they represent would place the two points very far apart. The distance function d_{ij} would then fail to detect the close similarity between the two variables. Allowance can be made for this possibility when making each comparison by considering $\Sigma\, w_k(\lambda_{ik} + \lambda_{jk})^2$ as well as $\Sigma\, w_k(\lambda_{ik}-\lambda_{jk})^2$ in each case and choosing the smaller as the distance between the two variables in question. Thus (3) becomes

$$d_{ij} = \min \{\sum_k w_k(\lambda_{ik}-\lambda_{jk})^2, \sum_k w_k(\lambda_{ik}+\lambda_{jk})^2\}. \quad (4)$$

If the number of variables to be discarded has not been predetermined, some visual assistance in arriving at an appropriate value can be obtained by ordering the distances and plotting them. The plot can then be examined for a turning point where a noticeably steeper ascent begins.

4. EXAMPLE

A teacher evaluation survey was administered periodically to determine students' rating of their university teachers. The major part of the questionnaire consisted of 22 questions about various aspects of teacher effectiveness. When a revision was contemplated, information was sought as to whether some of the questions could be deleted with little sacrifice in information. A shorter questionnaire would meet with less student resistance.

A factor analysis was performed using a correlation matrix obtained from past survey results. It was found that 17 factors were required to explain 95% of the variation. The distance function (4) was calculated for each of the 231 pairs of variables. The smallest 20 of these are given in Table 1. The remaining distances ranged up to a high of 2.875.

TABLE 1

Pairs of questions with smallest distance functions

Question Pair	Distance	Question Pair	Distance
19,20	.023	14,19	.657
21,22	.041	2,19	.797
8,20	.151	2,20	.877
8,19	.170	5,19	1.016
1,16	.172	8,14	1.072
2, 7	.174	7,19	1.082
8,10	.352	5,20	1.112
10,19	.453	2,14	1.113
10,20	.473	2, 8	1.132
14,20	.603	13,19	1.163

Note that the pair of variables that are closest, 19 and 20, are also both very close to 8. Thus, only one of these three variables need be retained.

As an aid to determining how many questions might appropriately be deleted, the first 15 distances are plotted in Figure 1. There is a definite change in pattern from the sixth to seventh value. It seems clear that a variable can be deleted from each of the first two pairs with almost no loss of information. The next four distances also appear small enough to warrant deletion of one variable from each pair if desired. Beyond that point, however, further deletion may not be justified.

(FIGURE 1 HERE)

5. DISCUSSION

The methods discussed by Jolliffe (see Section 2) appear to be well suited to situations where a rather small subset of the original variables is adequate. The present method is recommended where relatively few (up to one-third, say) of the variables are to be discarded. This situation can be readily identified when the factor analysis indicates that a rather large number of factors are required to "explain" the data.

The method of this paper has the added advantage of simplicity of operation. It is easily programmed and can be included as an option in a standard factor analysis routine. The resulting ordered distance values can be examined meaningfully by someone with little statistical expertise.

The number of factors to be used is somewhat arbitrary. Good results will be obtained by retaining at least enough to "account for" about 95% of the variance in the system.

REFERENCES

JOLLIFFE, I. T. (1972). Discarding variables in a principal component analysis. I: Artificial data. Appl. Statist., 21, 160-173.

JOLLIFFE, I. T. (1973). Discarding variables in a principal component analysis. II: Real data. Appl. Statist., 22, 21-31.

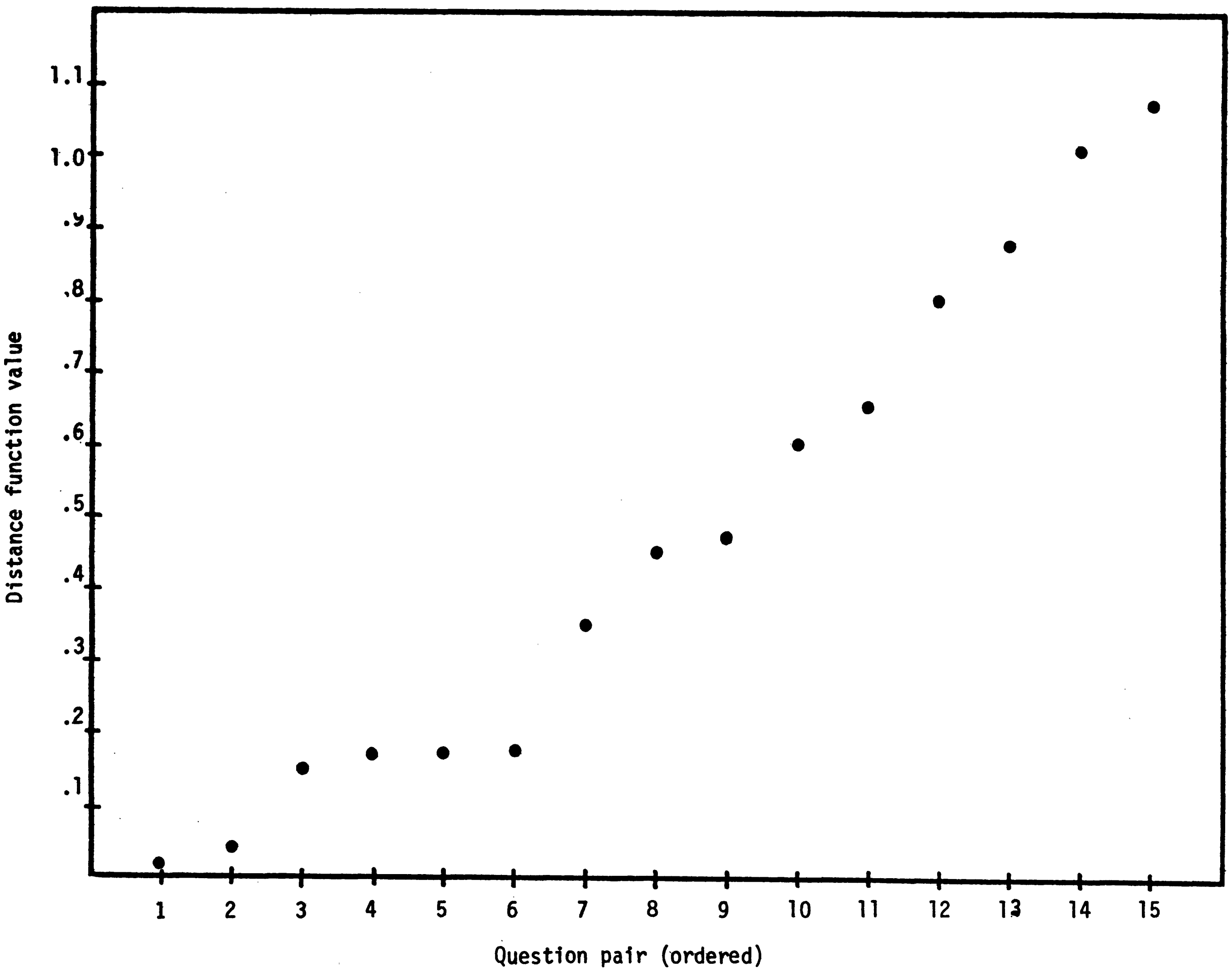


Figure 1. The 15 smallest distance values.

A MODEL OF BIRTH PROBABILITIES SPECIFIC FOR AGE AND PARITY OF WOMEN

Harry M. Rosenberg, Carolina Population Center, University of North Carolina
Ralph E. Thomas, Battelle Memorial Institute, Columbus Laboratories

Formal models of aggregate human fertility have focussed, in the main, on the variable, age of woman, and its relation to the probability of having a birth. The literature includes early work such as Lotka's graduation of the maternity function[1] by the normal or Gaussian distribution, Wicksell's fit by a Pearson Type III curve, and Hadwiger's exponential fit.[2] Recent work by Mitra and Romaniuk has used the Pearson Type I curve for graduation of fertility rates by age of woman; Brass has used polynomials; and Murphy and Nagnur have fit Gompertz functions to fertility rates, cumulated by age of woman.[3] Considerably less attention has been given to the variable parity and its relation to the probability of birth, even though some, such as Ryder, have come to emphasize the importance of this variable. Ryder has stated, "The most important fertility variable is birth parity, not merely because it influences fertility, but because it is fertility."[4] Accordingly, in some of Ryder's recent work, birth analyses are carried out entirely in terms of parity indices rather than birth rates specific for age.

Conceptually, we need not adopt the view that parity of women is a more useful perspective, or a more powerful correlate of birth probability than age of women. Instead, parity and age can be treated as complementary perspectives on fertility, combined in a more basic index of aggregate fertility, the birth rate specific for both age and parity, hereinafter called the "birth probability."

Pioneering work with birth probabilities done by Whelpton was reinforced by Karmel's discussion of Whelpton's work.[5] More recently Murphy assessed the stable population implications of birth probability assumptions.[6] The present paper attempts to redirect attention to rates specific for both age and parity of women by constructing a general structural model that captures the major features of a birth probability time series for American women during the years 1917 to 1968.

THE MODEL The attempt has been made in developing the birth probability model to construct one which, while showing good correspondence with the data, also embodies relations and parameters that can be related to certain social and biological aspects of human fertility. Such a model can have value for both estimation and explanation.

Research upon which this paper is based was supported by Grant No. IR01 HD05981-01 from the National Institute for Child Health and Human Development (Center for Population Research). Valuable assistance in this study has been provided by Norman Ryder, Office of Population Research, Princton University; John Patterson, Division of Vital Statistics, National Center for Health Statistics; Amy J. Kuntz, Department of Mathematics, University of North Carolina, Chapel Hill; and Roger W. Cote, Battelle Memorial Institute, Columbus Laboratories.

Evaluation of arrays of birth probabilities for individual birth cohorts of women indicates that the likelihood of having a birth increases monotonically with increasing parity n of women, controlling for age x; and decreases monotonically with increasing age of women after age 22 years, controlling for parity.[7] A model that can represent these relations is as follows:

$$b(x,n)=\begin{cases} 1, & x-\lambda n<m_1+e^{-1/A}(m_1-m_2) \\ -A\ln\left(\dfrac{x-m_1-\lambda n}{m_2-m_1}\right), & m_1+e^{-1/A}(m_2-m_1)\le x-\lambda n\le m_2 \\ 0, & x-\lambda n>m_2 \end{cases} \quad (1)$$

where λ represents the number of years 'consumed' by each additional birth; m_1 represents the age at menarche; m_2, the age at menopause, and the parameter A denotes a scale factor. The end conditions restrict birth probabilities from exceeding 1.0 at the onset of menarche, and the lower restriction prevents birth probabilities from becoming negative after menopause.

Conceptually, parameter m_1 is interpreted as representing the chronological onset of female fertility; m_2 as representing the termination of the fertile period; $x-m_1$, then, represents the number of years since the menarche. The parameter λ is the reproductive 'cost' in years per additional child, and the entire expression in the numerator of equation (1), namely $(x-m_1-\lambda n)$ represents nonreproductive years at age x. The denominator m_2-m_1 can be viewed as a representation of the fertile period.

Thus the ratio $\frac{x-m_1-\lambda n}{m_2-m_1}$, varies between zero and one when $m_1\le x-\lambda n\le m_2$. Because the numerator represents the complete span of her reproductive years, it is seen that the ratio is that proportion of her reproductive life that to that point has not been used for reproduction. A low value for this ratio, which implies a high potential for future childbearing, can result under either of two conditions: (1) when the woman is near age m_1, or (2) when n is large so that the residual $(x-m_1)-\lambda n$ is small. In the first case, the ratio will be small because of the woman's chronological youth; in the second, because her reproductive youth has been maintained through repeated childbearing, reflecting a large value for n. Small values of the ratio, then, are associated with probabilities close to the value one; large values of the ratio are associated with probabilities close to zero.

It is a significant feature of the model that a reproductive history of high fertility is associated with a high probability of having a next child. Another way of stating this is to say "the more children a woman has had, the more she is likely to have, except as she ages."

Another way of picturing the process qualitatively is as a series of monotonically decreasing probability schedules by age of woman, one for each parity class of woman. The schedules increase in height by $-A\ln\left(\frac{x-m_1-\lambda(n+1)}{x-m_1-\lambda n}\right)$

with increasing parity as illustrated in Figure 1, such a woman remaining childless would be associated with the continuous downward trajectory indicated as "I", while a woman having a first birth at age x, shown as path "II", would, at that age, experience a discontinuity as she moved to a higher, but decreasing, trajectory of women in parity class one. Although Figure 1 shows a series of birth probability schedules, they are, in fact, only a single curve, as represented in equation 1.

The parameters m_1, m_2, and λ are assumed to be biological processes that may change slowly over time. In contrast, the parameter A is assumed to reflect mainly aggregate fertility responses of a birth cohort to exogenous social and economic factors. For this reason, the behavior of A is of particular analytic interest and is pursued in another paper. Previous work along these lines has been done by Campbell and others.[8]

PARAMETERIZING THE MODEL Parameters of the model were estimated using data from the Division of Vital Statistics, National Center for Health Statistics, U.S. Department of Health, Education, and Welfare.[9] Birth probabilities are actually derived measures, in which the numerator represents births of a specific birth order; that is, first births, second, third, fourth births, etc., by age of woman.[10] The denominator represents an independent estimate of women at risk, by age and parity, to have a birth of that order. Thus, between ages m_1 and m_2, all parity zero women are assumed to be at risk to have first births; parity one women, at risk to have second births, etc. The numerator and the denominator are each subject to the errors and biases characteristic of their respective sources, namely, the vital registration system and the decennial censuses. These include problems of under-reporting, misclassification of age and parity of women, and misreporting of birth order of child. Moreover, it is likely that historically the extent and type of biases in the data arising in both sources has changed, so that some perturbations over time in data series actually reflect changes in the quality of data rather than changes in fertility.

Furthermore, biases and distortions in the rates are likely to be concentrated among certain age and parity combinations, where the probability of having a birth is very high. Such a group would be young, high parity women among whom the risk of birth is high. For such numerically small groups, rates tend to be less stable and reliable than those of other age-parity combinations.

Parameter estimation, for all cohorts combined (1877 through 1954), involved a two-stage procedure in which, first, λ was estimated, and second, the parameters m_1, m_2-m_1, and A were derived by a linear least squares program,[11] using the estimated value of λ and the following transformation of equation (1): $(x-\lambda n)=m_1+(m_2-m_1)e^{-b/A}$ (2)

Estimates of parameters for all cohorts combined were as follows: λ= 1.83 years; m_1= 8.58 years; m_2-m_1= 30.83 years; and A= 0.209. Our goal is to develop a model whose parameters can be interpreted as social or biological. However, the initial estimates of the parameters do not entirely satisfy the objective. For example, the estimated value for m_1, interpreted as the age of menarche, is about nine years, well below the recorded experience for this country.[12] The length of the fertility period, to which the difference m_2-m_1 refers, is also less than the span indicated by the data set.

In validating the model, subsequent to establishing single values for each parameter based on the combined cohort estimates, we found that a single estimate for the parameter A resulted in consistent underestimates of birth probabilities for parity zero women. As a consequence, separate estimates were made of parameter A for each parity.

In general, the fit between observed and model-generated estimates of cohort birth probabilities is good; it reveals, interestingly, that positive or negative deviations persist over successive ages rather than showing random movement around predicted values between successive ages. Autocorrelation of this kind might represent actual aggregate fertility behavior in which lower parity specific rates are subsequently compensated by higher rates and the reverse; or, it could reflect either data or model bias. For all the parities combined in the 1908 cohort, the relative difference (root mean square expressed as a percent) between the actual and the fitted data was 1.6 percent. This was the best fit achieved for cohorts 1877-1954; the poorest fit was for the 1897 cohort for which the model values deviated from the actual values by about six percent.

MARKOV PROCESS REPRESENTATION As Keyfitz indicated in summarizing the work of E. M. Murphy, birth rates specific for age and parity lend themselves quite naturally to representation as Markov processes.[13] In the Markovian model, the birth probabilities are the likelihood that a randomly-selected woman in a parity class will move, or make a transition, from parity state n to parity state n+1. The birth probability model here presents a particular kind of Markov process that can be characterized as nonhomogeneous and irreversible. Nonhomogeneity refers to the characteristic that the state transition probabilities are conditional on age of the woman, while irreversibility means that women can only move to higher parity classes, and never return to lower ones.

With these observations, the transition matrix M(x), as a function of calendar age (x), may be written as follows:

$$
M(x) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7+ \end{array}
\begin{array}{c}
\begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7+ \end{array} \\
\left[\begin{array}{cccccccc}
q_0 & p_0 & \cdots & & & & & 0 \\
 & q_1 & p_1 & & & & & \cdots \\
 & & q_2 & p_2 & & & & \cdots \\
 & & & q_3 & p_3 & & & \\
 & & & & q_4 & p_4 & & \\
\cdots & & & & & q_5 & p_5 & \\
\cdots & & & & & & q_6 & p_6 \\
0 & \cdots & & & & & & 1
\end{array}\right]
\end{array},
$$

where the p_n, q_n denote transition probabilities, and the parity states are shown by the labels bordering the matrix. Only the diagonal and the super-diagonal have non-zero entries; all other entries are zero. The symbol p_n denotes the probability that a randomly selected woman of age x and parity n will make a transition to parity state n+1 in the coming year; the probability that no such transition is made is denoted by $q_n = 1-p_n$, $n=0, \ldots, 6$.

Next, at the beginning of a given year, let $V(x) = (v(0), \ldots, v(7))$ denote a parity distribution of 1000 randomly selected women of age x among the parity states 0 through 7. The expected parity distribution at the beginning of the next year is then given by

$$V(x+1) = V(x)M(x).$$

Similarly, by incrementing age by one year we have $V(x+2) = V(x+1)M(x+1)$, and the replacement of V(x+1) yields

$$V(x+2) = V(x)M(x)M(x+1).$$

Proceeding recursively, it follows that

$$V(x_m) = V(x_n) \prod_{x=x_n}^{x_m} m(x),$$

where x_n and x_m denote the ages of menarche and menopause, respectively. As an example, if 1000 women of age 14 are all taken to be in parity state 0, then V(14) = (1000, 0, 0, 0, 0, 0, 0, 0) and the expected parity distribution at age 49 is given by

$$V(49) = V(14) \prod_{x=14}^{49} M(x).$$

This result shows how the parity distributions can be computed from the matrices of transition probabilities. These transition probabilities may be taken equal to the reported birth probabilities. Alternatively, the published birth rates may first be 'smoothed' by fitting the birth probability model developed in this paper.

THE MODEL FOR YOUNGER WOMEN The model postulates decreasing birth probabilities with increasing age of women. And, in general, the fit between recorded and fitted values is quite good for women in the age range 22 to 47 years. Before age 22 years, however, observed values are considerably below those expected on the basis of the model. For example, the model yields 0.328 for the birth probability of a 15-year-old woman of zero parity, compared with a reported value for this birth birth probability of 0.0087. However, the reported figures do show increasing birth probabilities with increasing parity, controlling for age of women.

The large discrepancies between the observed and the model birth probabilities at lower ages are believed to be attributable, in large measure, to the relatively small number of women who are married at young ages. Proportions of women ever married increase from about one out of 100 at age 14 years to over seven out of ten at age 22, stabilizing for recent cohorts at about 90 percent ever married at age 30 years.[14] Marriage, we know, is neither a necessary nor sufficient condition for childbearing, given the incidence of premarital conception, which may or may not be followed by wedlock, together with the prevalence of marital infertility; however, marriage can serve as a proxy of changes in exposure to the risk of having children at younger ages.

For simplicity, consideration of marriage as an explicit element was deliberately omitted in developing the model in its present, preliminary form. There are, nevertheless, ways in which it could be incorporated into a somewhat more complex model. Keyfitz, for example, suggests that the matrix formulation include parity states specific for nuptiality.[15] But the major restriction on such an approach are data limitations with respect to the legitimacy status of newborns and the marital status of women at risk to have children.

A somewhat less formidable way of approaching the problem, but one also constrained by data limitations, is normalizing reported birth probabilities by estimated proportions of women-ever-married prior to estimating model parameters. The effect of normalizing birth rates by marital status can be illustrated using 1960 data on births and women ever married by age, as shown below:

Age Groups of Women	Birth Rate 1960 [16]	Proportion of Women Ever Married, 1960 [17]	Birth Rate for Ever Married, Normalized Women 1960
15-19 years	0.089	0.118	0.754
20-24	0.258	0.720	0.358
25-29	0.197	0.869	0.227
30-34	0.113	0.891	0.127
35-39	0.056	0.881	0.064
40-44	0.016	0.860	0.019

Normalization by marital status, this illustrates, changes the standard maternity function to one compatible, in its general magnitude, with our birth probability model.

DISCUSSION The availability of fifty years of birth rates specific for the age and parity of American women provides a rich data base for structural and historical analysis. The general structural model

$$b(n,x) = -A(n)\ln\left(\frac{x - m_1 - \lambda n}{m_2 - m_1}\right)$$

approximates the major features of the truncated fertility function cohorts of women born during 1896 to 1945. The basic features of the model are the following: (1) two parameters that relate conceptually, though not quantitatively in our initial work, to the biological states of menarche and menopause; (2) a parameter for the socio-biological process of childspacing; and (3) a scale factor that appears to reflect a cohort's responses to socio-economic forces, both short-term and long-term.

The model serves to organize the age and parity specific birth probabilities in a manner that is intuitively appealing and is useful as an investigative tool. However, certain shortcomings of the model should be noted: (1) some modifications should be made to extend the range of the model downward to age 14 years by procedures such as those discussed here; (2) further adjustments to the parameters identified as predominantly biological may be desirable to bring them more closely into alignment with independently-derived values; and (3) further studies should be made to minimize biases in the model, and to ensure the validity of reported rates, through adjustment if necessary.

Future work will involve detailed historical analysis of the behavior of each of these parameters and examination of the covariation between these parameters and related biological and social variables. In addition, initial work indicates that the model may be helpful in extending Ryder's work on the relation between cohort and period fertility indices.[18] Also, it may be useful in projecting age-specific period fertility rates on a more secure cohort fertility base.

General Model Equation:

$$b(x,n) = -A\ln\left(\frac{x-m_1-\lambda n}{m_2-m_1}\right)$$

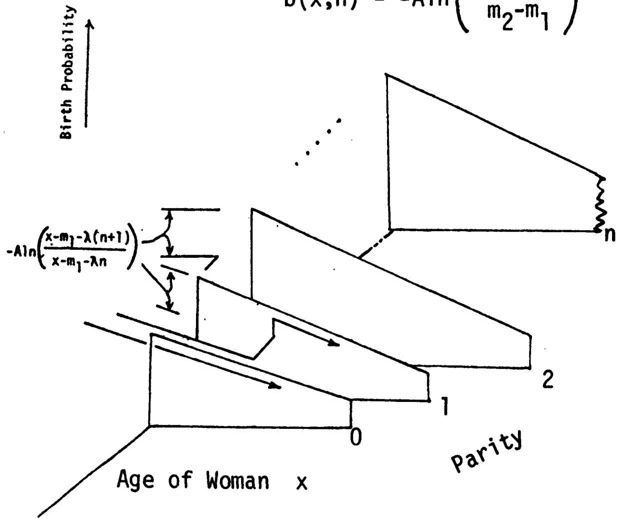


Figure 1. Hypothesized Relation Between Birth Probability and Age of Woman, II.

Footnotes

1/ In developing the birth probability model, we do not explicitly incorporate the force of mortality, ordinarily expressed as p(x) in mathematical treatments of population renewal. We use the expressions maternity function and fertility function interchangeably.

2/ J. Alfred Lotka, *Theories analytique des associations biologiques.* Part II. Analyse demographique avec application particuliere a l'espece humaine, Actualities scientifiques et Industrielles, No. 780. Paris: Hermann and Cie, 1939; S. D. Wicksell, "Nuptiality, Fertility, and Reproductivity," *Skandinaviske Aktuarietidskrift,* 1931, pp. 125-127; and H. Hadwiger, "Eine analytische Reproductionsfunktion fur biologische Gesamtheiten," *Skandinaviske Aktuarietidskrift,* 1940, pp. 101-113. For a concise summary of the earlier work, see Nathan Keyfitz, *Introduction to the Mathematics of Population,* Reading, Massachusetts: Addison-Wesley Publishing Company, 1968. pp. 140-169.

3/ S. Mitra, "The Pattern of Age-Specific Fertility Rates," *Demography*, Volume 4, 1967, pp. 894-906; A. Romaniuk, "A Three Parameter Model for Birth Projections," *Population Studies* Volume 27, Number 3, 1973, pp. 467-478; William Brass, "Graduation of Fertility Distributions by Polynomial Functions," *Population Studies,* Volume 14, Number 2, 1960, pp. 148-162; Edmund Murphy and Dhruva N. Nagnur, "A Gompertz Curve that Fits: Applications to Canadian Fertility Patterns," *Demography*, Volume 9, 1972, pp. 35-50.

4/ Norman B. Ryder, "The Measurement of Fertility Patterns," in Mindel C. Sheps and Jeanne C. Ridley, editors, *Public Health and Population Change,* Pittsburgh, Pa.: University of Pittsburgh Press, 1965, p. 293. Parity refers to the number of births a woman has had; and birth order to whether the child is a first, second, third, etc child. Thus, women who have had no children are referred to as zero parity women; those who have had a first child (birth order one child), as parity one women, etc.

5/ P. K. Whelpton, "Reproductive Rates Adjusted for Age, Parity, Fecundity, and Marriage," *Journal of the American Statistical Association,* Volume 45, 1950, pp. 119-124.

6/ Edmund M. Murphy, "A Generalization of Stable Population Techniques," unpublished Ph.D. dissertation, Department of Sociology, University of Chicago, 1965.

7/ Development and parameterization of the model is restricted to women aged 22-47 years. See the section below on problems of fitting the model to younger women.

8/ See, for example, Campbell's discussion of the short-term covariation between birth probabilities and indices of economic activity in Clyde V. Kiser, Wilson H. Grabill, and Arthur A. Campbell, *Trends and Variations in Fertility in the United States,* Cambridge, Massachusetts: Harvard University Press, 1968, Chapter 12.

9/ For 1959-1968, the data are from unpublished worksheets; for 1917-1958, the data are from P. K. Whelpton and A. A. Campbell, "Fertility Tables for Birth Cohorts of American Women," *Vital Statistics--Special Reports,* Volume 51, Number 1, January 29, 1960.

10/ See footnote 4 above.

11/ P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences,* New York: McGraw-Hill, Inc., 1969, pp. 208-212.

12/ The estimated age at menarche in the United States is between 12 and 13 years. See J. M. Tanner, *Growth at Adolescence,* Springfield, Massachusetts; Charles C. Thomas, 1962.

13/ Keyfitz, *op. cit.*, pp. 328-331, and footnote 6 above. For a general treatment of Markov processes, see J. G. Kemeny, and J. L. Snell, *Finite Markov Chains,* Princeton, New Jersey: D. Van Nostrand Company, 1960.

14/ U. S. Department of Commerce, Bureau of the Census, *1960 Census of Population* PC(2)-4E, "Marital Status," Table 1.

15/ Keyfitz, *ibid.*, p. 328

16/ U. S. Department of Health, Education and Welfare, Division of Vital Statistics, *Vital Statistics of the United States, 1960,* Volume I, "Natality," Table 1E. Birth rates by age of women are equal to the summation of birth rates by age and parity of women weighted by the parity distribution of women at each age.

17/ See footnote 14. Proportions ever married at midpoint of age group, *e.g.*, age 17 years for age group 15-19 years, etc.

18/ Norman B. Ryder, "The Process of Demographic Translation," *Demography*, Volume 1, Number 1, 1964 pp. 74-92; "The Structure and Tempo of Current Fertility," in *Demographic and Economic Change in Developed Countries,* for the National Bureau of Economic Research, Princeton, New Jersey: Princeton University Press, 1960, pp. 117-136.

DISCOURAGED WORKERS AND UNEMPLOYMENT: A FURTHER ANALYSIS BY COLOR

Marc Rosenblum, John Jay College, City University of New York

Summary and Conclusions

Not only have nonwhites failed to advance in recent years when compared with whites in the frequency of their labor force activity, but an analysis of Current Population Survey (CPS) and related data suggests their situation is actually deteriorating. Moreover, whatever gains have been recorded are all in the female group and attributable primarily to increased labor force attachment by sex rather than by race.

And, while comprehensive demographic analysis is beyond the scope of this paper, the unfavorable trends in several measures of labor force behavior seem to closely follow proportional nonwhite increases in working age population. Particularly during periods of tight and deteriorating labor markets, like the present, this would suggest a near zero-sum situation in the aggregate distribution of employment gains by race, or certainly a less flexible one than needed to accommodate all job-seekers under present institutional and labor market arrangements.

Furthermore, several labor force measures under evaluation all indicate a growing difference in worker discouragement along racial lines. Thus, the apparent rise in withdrawal from the labor force by blacks -- indicated by reduced participation rates -- is in part a shift to discouraged worker status rather than lessened attachment to work.[1]

What the analysis will specifically show is (1) an upward trend in the proportion of job-market discouragement for nonwhites while that for white workers is declining.

(2) a growing difference by race in the ratios of discouraged unemployment to regularly calculated unemployment.

(3) a growing difference by race in labor force participation rates calculated by both regular and discouraged methods, with the greater black difference indicative of proportionally higher job-related discouragement.

(4) a larger proportion of nonwhites surveyed as not-in-the-labor force but wanting to work relative to whites.

(5) greater proportions of blacks who are discouraged workers -- for both job-market and all reasons -- in relation to those persons in each racial category not-in-the-labor force and wanting employment.

(6) a higher proportion of blacks who are job-market discouraged relative to population, with a difference that varies both with economic conditions and demographic change.

Background and Introduction

Attempts to explain differences in employment and unemployment by race have in the past focused on education, occupational levels and several other variables. At the same time, researchers over the years have developed more understanding of labor force behavior in general.

Part of this overall effort, particularly during the past two years, has extended to persons counted as not-in-the-labor force by CES interviews.(Only those overtly seeking work are enumerated as unemployed, hence part of the labor force). A portion of those not-in-the-labor force, known as discouraged workers, are persons expressing a desire to work but not looking because they believe it impossible to find a job.

Worker discouragement has interested several investigators, from the standpoint of subemployment or insufficient earnings.[2] This author's focus is on people discouraged by insufficient labor demand and quantitative aspects of their response at different unemployment levels.

At the root of problem is an anomaly: persons must indicate some specific job search activity to be officially considered unemployed, even where their knowledge of labor market conditions indicates such activity to economically irrational. Given these conditions it is perfectly reasonable for an individual to avoid a job search that has a low probability of success. But clearly such behavior results in understated aggregate joblessness under official definitions of unemployment.

Discouraged unemployment is interpreted to cover those persons currently available for (and realistically able to) work. Regular (as defined by BLS) unemployment coincides with this measure except for the job-market discouraged. Thus, workers discouraged for economic reasons increase during slack periods, and decline when labor demand rises.

Although the number of persons so affected is not large relative to the overall labor force, redefinition of worker discouragement as a component of unemployment would have a clear impact on the government's most widely-followed economic statistic. That impact was determined to have a mean of 15% over the past seven years, with variation dispersed procyclically.[3]

These findings suggest that some discouraged workers are quite sensitive to levels of economic activity, and have an attachment pattern not unlike the majority of workers -- except that in the absence of jobs they are more passive in their search behavior. Our knowledge of job hunting patterns indicates that costs are incurred in relation to expected return. As the prospective wage level and probability of finding work diminish (as in the case of many nonwhites) it is possible that passive -- rather than active -- search effort may be entirely rational.

To follow the number of persons whose lack of work can be alternately regarded as unemployment or nonparticipation the author derived a "discouraged unemployment rate."[4] If people discouraged for job-market reasons (but not those discouraged for personal reasons) are assigned to the unemployed category, we obtain a more comprehensive measure of manpower underutilization.

A distinction between people available for work and the officially unemployed is thus made. If the unemployment rate is viewed (aside from its literal definition) as an index of inadequate manpower demand, use of the "discouraged" version permits that important marginal element -- the job-market discouraged -- to be analyzed.

Initially the entire labor force, and its divisions by age and sex, were reviewed. In this paper the element of race is introduced to determine the extent of variation in discouraged worker behavior on those dimensions.

Trends in Job-Market Discouragement and Participation

The total number of persons discouraged for job-market reasons fluctuates procyclically, but in less elastic terms than the overall participation to employment relationship. By race, the nonwhite proportion of this total has risen almost continuously from 22.4% in 1967 to 28.1% in 1973; or by 25.6%. The uptrend, particularly since 1970, is attributable primarily to more discouragement among females.

By comparison, the proportion of blacks in the civilian labor force went up 1.1% during the like time period. This was a net figure, since the proportion of males declined 2.1% while females increased their activity by 5.3%. Table indicates these proportions.

Table 1

DISTRIBUTION OF JOB-MARKET DISCOURAGEMENT AND CIVILIAN LABOR FORCE BY COLOR, 1967-73

Sex &	Discouragement		C.L.F.	
Race	1967	1973	1967	1973
W Male	.252	.268	.569	.548
W Female	.524	.451	.319	.339
Total W	.776	.719	.888	.887
NW Male	.063	.077	.064	.063
NW Female	.161	.204	.048	.050
Total NW	.224	.281	.112	.113
Sum Total	1.000	1.000	1.000	1.000

Thus, similar patterns are emerging in both labor force participation and job related discouragement. Since 1967, practically all gains in the labor force have been concentrated among white females. The magnitude of this is made clearer by Table 2. In that table, differences are shown between the 1973 total labor force and a 1973 labor force derived by using the component of change method.[5]

Table 2

DIFFERENCE BETWEEN ACTUAL 1973 LABOR FORCE AND DERIVED 1973 LABOR FORCE BY USING 1967 PARTICIPATION RATES (Thousands)

Sex	NW	W	Both
Male	-346	-979	-1,325
Female	8	2,977	2,985
Total	-338	1,998	1,660

Since 1967, participation rate change has accounted for a net increase of 1.66 million persons over the number who would otherwise have been in the labor force. 2.99 million of them are women, -1.33 million are men.

Similarly, 2.0 million are white, -.34 million are black. These changes are examined further below in connection with related demographic effects.

Discouraged Unemployment by Color

The employment pattern for nonwhite persons is more sensitive to cyclical fluctuations in economic activity. In the traditional sequence, blacks are the last to be hired and the first laid off. This well-known phenomenon is reflected in unemployment rates that are approximately twice the rate for whites, and around 1.8 times the overall national rate. A better understanding is obtained when discouraged unemployment, as defined above, is measured. Figure 1 (following complete text) shows this for nonwhite unemployment relative to the white labor force.

For the period under study, a U-shaped sequence is evident. Two facts stand out:

(1) Discouraged unemployment for blacks is proportionately greater, with that line almost continuously above the path of regular unemployment. For 1973, the difference in ratios was at a maximum for the series, suggesting a worsening of nonwhite employment conditions.

The exception, 1968, is explained by high levels of economic activity (albeit war-related) and declining unemployment. With the overall jobless rate below 4%, employers were hiring nonwhite workers at a rate that brought the black unemployment rate down faster than the white -- so the ratio increased. Discouraged workers coming into the labor force as work was available, on the other hand, were more evenly distributed in terms of the ratio, so the $(u_1/u)^*$ measure kept falling.

(2) The ratio appears to move with a lag. Its low for 1970 and 1971, came at a time when unemployment rose and a recession was under way. Given that black unemployment moves in something less than a 2:1 pattern to declining economic conditions, the ratios do not turn upward until deterioration of employment is clear-cut.

And, as pointed out above, blacks are more likely to suffer discouragement in the labor market. The ratio of discouraged unemployment thus remains higher than the ratio for unemployment calculated by standard methods.

Further analysis of unemployment rate ratios by color builds on previous findings. In the author's prior study, it was noted that a ratio of discouraged to regular unemployment rates, $(U/L)^*(U/L)$ is inversely related to the unemployment rate, while the number of persons discouraged for job-market reasons moved positively with the unemployment rate.[6] The inverse relationship results from a decline in the ratio during periods of high unemployment. Then, the number of unemployed persons increases rapidly, lessening the relative difference between both types of unemployment -- discouraged and regular.

The same technique is then applied to the data by race. Taking the difference in means for 29 quarterly observations (I-67 through I-74), nonwhite discouraged unemployment averaged 16% higher while the corresponding white measure averaged 14% above regular unemployment. This difference is statistically significant.[7]

On the basis of color, it was further hypothesized that nonwhites would show a more sensitive relationship between the number of job-market discouraged and unemployment rates. This would be in keeping with the job-search behavior literature, given existing information networks, transportation patterns and discrimination.[8] The results confirm this hypothesis, as shown below.

Table 3
RELATIONSHIP OF PERSONS DISCOURAGED FOR JOB-MARKET REASONS AND UNEMPLOYMENT RATES, BY COLOR, I-67 to I-74

	W			NW		
SEX	r	R^2	t	r	R^2	t
MALE	.45	.20	2.61*	.62	.38	4.10***
FEMALE	.75	.56	5.81***	.77	.59	6.20***
BOTH SEXES[a]	.67	.45	4.66	.73	.53	5.56

a Significance not tested for both sexes. See footnote 7.
* Significant at .05 level.
*** Significant at .001 level.

While the correlations vary for different subgroups by age, they are for each group by race greater for non-whites. In the face of cumulatively greater impediments to obtaining employment, black persons are more likely to withdraw from the officially-defined labor force.

The effect of discouraged unemployment can be further adduced by comparing labor force participation rate differences by race. Customarily, reduced participation rates during times of economic slack are interpreted as withdrawal from the labor force. When job-market discouragement is counted as unemployment (as this paper does for analytical purposes), a portion of this cyclical variation is removed -- since only those persons unable or unwilling to participate for various other reasons are considered not in the labor force

The difference between regular and discouraged participation rates is much higher for nonwhites than for whites. While the extent of difference is not uniform across subgroups, its direction is similar in every case by age and sex, their totals and the total of both sexes.*

Discouragement for Other Than Job-Related Reasons

Workers are considered discouraged for personal reasons in CPS surveys -- as opposed to job-market reasons -- if their response indicates:

(a) The lack of necessary schooling, training, skills or experience.
(b) Employers think they are too young or too old.
(c) Other personal handicaps in finding a job.

Elsewhere, the author argues that (b) -- the age-mismatch group -- can be considered a job-market variable.[9] Leaving (a) and (c) as personal elements of discouragement, it may then be surmised why the association above is almost three times as strong for whites. Other personal handicap (c) includes responses reflect ing racial discrimination, while item (a) covers areas where blacks have historically lagged behind the general population.

Another way of viewing the same phenomenon is to regress the total of all discouraged workers on unemployment -- using the discouraged unemployment rate (U/L)*. On this measure, the number of nonwhites is more sensitive to changes in economic conditions. The relationship is:

	r	R^2	t
White	.56	.31	3.48**
Nonwhite	.76	.57	6.01***

** Significant at the .01 level.
*** Significant beyond the .001 level.

The pool of nonwhite workers discouraged for personal reasons may include at least some whose perceived inability to find work is linked to economic conditions The number of persons in this category is not large, however, since the mean of total black discouragement for the period under study is only 160,000, with 24,000 of these categorized as for personal reasons.

The Effect of Population Growth on Labor Force Participation

Between 1967 and 1973 the nonwhite population of working age (16 years and above) increased at an annual rate of 3.52%. During the same period, comparable figures for whites showed a 1.69% rise.

If racial characteristics were fully disregarded in the labor market one could assume that trends in participation by color would likewise be indistinguishable. This is not the case. Black participation is declining while white participation is not. Using race as a proxy for the variables on which employment decisions are made, the white and nonwhite experience can be compared by a component-of-change analysis.

This is seen in table 4. Expansion of the black labor force is a result of rapid population growth, despite a decline in male participation rates. For whites, on the other hand, a modest ongoing drop in male participation is more than offset by continued growth in the female sector, exceeding the population rise.

Since labor force increase is tied to employment growth, a sustained period of subemployment places limits -- not measured here -- on its size during any period. Population growth in excess of labor force expansion thus results in declining participation rates unless a more than proportional absorption of persons from specific age groups is effected, i.e. a changing dependency ratio.

It is not the focus of this paper to follow all the paths opened up by the above analysis, other than the relationship between worker discouragement and declining participation. As developed below, however, the growing divergence between discouraged-to-regular unemployment ratios in racial terms cannot be fully understood without inclusion of demographic factors.

Table 4

LABOR FORCE GROWTH BY COMPONENT OF CHANGE
1967-73
(percent per decade)

Race& Sex	Labor Force Growth	Attributable to Population Growth	Attributable to Participation Rate Change
NW			
Male	19.5	30.5	-11.0
Female	34.7	34.3	0.4
Total	25.8	32.1	-6.3
W			
Male	12.2	15.7	-3.5
Female	36.4	16.3	20.1
Total	20.5	15.9	4.6

Assessing the Evidence

The economic state of nonwhites has been amply documented in the poverty literature.[10] Thus, among the things this paper has done is present a series of labor force measures that amplify and confirm findings derived by other methods of analysis. In that sense, no surprises emerge.[11]

Yet the questions remain. What causes the differentials described above? And perhaps more importantly, why are they getting worse? Obviously, no single answer is possible. Is there nevertheless some theoretical framework within which these findings may be reconciled?

A number of possible factors might be considered individually and in combination. They are:

1. Some degree of labor market inequality continues to exist. In part this is due to racial group differences by occupation, by geographical distribution, by skill and training, expectation and aspiration levels, and by education as it bears on the above. That is, part of the inequality is institutionally based, including the effect of market segmentation. Some degree continues to be subjective or discriminatory.
2. Changes in participation rates by sex are interacting with demographic changes by race, leaving relatively more minority

group persons without employment under existing patterns. Recognizing the existence of the labor market barriers referred to above, a conjectural sequence would see additional nonwhites absorbed into jobs in some relation to their existing proportion of the total labor force.

This absorption could increase (a) under favorable, full employment conditions, (b) as the population proportion dropped or (c) or if market barriers were lowered. It could diminish, or not increase, (a) when employment was scarce and competition for jobs keener, (b) if the population proportion rose, or (c) if market barriers remained unchanged.

Where the rate of population growth exceeds the absorption rate the result would be, holding all other labor market factors constant, (a) reduced participation rates, and (b) increased discouragement. If, on the other hand, population rose less rapidly than the absorption figure, (a) participation would increase, and (b) discouragement would drop proportionally.

The ratio of job-market discouragement to population (D_i/P_i), shown in table 6, bears out this formulation.

Table 5

JOB MARKET DISCOURAGEMENT AS A PERCENTAGE OF NONINSTITUTIONAL POPULATION - BY COLOR

Year	Color W	Color NW	Difference (NW - W)
1967	0.39	0.95	0.56
1968	.35	.81	.46
1969	.28	.69	.41
1970	.33	.82	.49
1971	.39	1.00	.61
1972	.38	.97	.58
1973	.32	.99	.67

The difference declined to a low of 0.41 in 1969, at the peak of Vietnam-related full employment. The worsening labor market situation of nonwhites is indicated by the increase to 0.67 in 1973.

This explanation is tentative,and a frankly post-hoc attempt to account for part of the worsening trend in black-white labor force patterns.[12]

The basic existence of racial differences in discouraged worker behavior stems from historical fact. Labor market practices and institutions have tended to preserve past inequalities, or to accept them in a broader social setting not limited to purely economic spheres.

That these differences are becoming more pronounced despite efforts over the past decade to offset them, however, appears related to the interaction of demographic and labor force variables.[13]

In that sense, the problem becomes more intractible. For the trend to be overcome or offset, an increasing number of nonwhite persons would have to gain jobs.

Given the present outlook of slow growth and rising overall joblessness this would involve some redistribution of employment rather than, as under ideal conditions, a bigger slice of an expanding pie. Continuation of recent trends in the distribution of discouraged workers by race thus is more likely, at least in the period ahead.

Footnotes

1. The terms nonwhite and black are used interchangibly, although only 90% of nonwhites are,strictly defined, black.
2. Thomas Vietorisz, Jean-Ellen Giblin and Robert Meir. Some Indicators of Subemployment. New York: Research Center for Economic Planning, 1974; Sar A. Levitan and Robert Taggart. The Employment and Earnings Inadequacy Index: A New Social Indicator. Baltimore: John Hopkins University Press, 1974.
3. Marc Rosenblum, Discouraged Workers and Unemployment: A Review of Recent Evidence. A paper read to the New York State Economics Association, March 30, 1974, Buffalo, New York (forthcoming in that Association's 1974 Papers and Proceedings).
4. Ibid.
5. This is done by taking base year (1967) participation rates multiplied by actual 1973 population to obtain the labor force that would have existed had all change been demographically based, compared with actual change. This comparison, relative to real change, indicates the influence of each factor.
6. Rosenblum, op. cit.
7. t=-2.50, significant at .02 level for the total of both sexes by race. In this paper some test results are presented comparing the combined sums of both sexes by color, others comparing the totals of each sex and color. The comparison best illustrating that aspect of labor force behavior being discussed was used as both sets of tests cannot be run on the same data without violating the assumptions of independence necessary for employing t-tests.

8. A thorough review of the job search literature is contained in Robert J. Gordon, "The Welfare Cost of Higher Unemployment." Brookings Papers on Economic Activity, 1973,1, 133-195.

9. Marc Rosenblum, "Discouraged Workers and Unemployment," Monthly Labor Review, 97, 1974, 28-30.

10. A recent example is Bradley Schiller, The Economics of Poverty and Discrimination. Englewood Cliffs: Prentice-Hall, 1973.

11. Some surprises in related areas could almost be anticipated from a reading of this data. For example, this past July it was announced that income differentials by race had continued to increase in the United States, further reversing the trend toward a narrower gap prevalent from 1965 to 1970. The higher proportion of job-related discouraged workers among blacks would suggest such an outcome. See U.S. Department of Commerce, Bureau of the Census. The Social and Economic Status of the Black Population in the United States, 1973. Current Population Reports, P-23, No. 45, 1974.

12. It is possible, if not very likely, to attribute the increased differences to a rise in overt discrimination and/or labor market inequalities. It is also possible to try a neoclassical approach, such as "relative wage" theory.

Workers, striving to maintain desired living standards, alter their participation in line with a ratio of current real wages to desired ones. In the face of inflation eroding real wages recently (lowering this ratio), participation and unemployment would rise. And, by implication, so would discouragement. This rise in labor force would take the form of an influx of secondary workers since participation rates of primary workers are almost inelastic, and near a maxim.

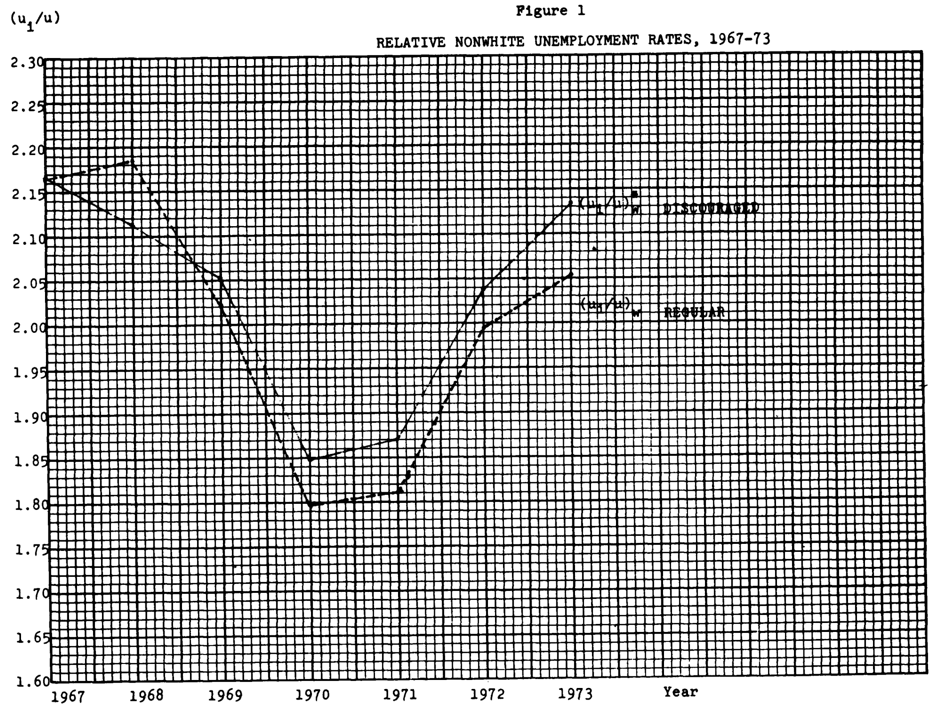
But, given declining relative wages for a greater proportion of nonwhite workers (see footnote 14) an increase in black participation rates would be expected. As the data indicates, this has not occurred, on a regular or discouraged basis, for primary or secondary workers, or relative to whites. Thus, neither of these alternatives is plausible. For elaboration of the "relative wage" theory, see Michael L. Wachter, "A Labor Supply Model for Secondary Workers." Review of Economics and Statistics, 54, 1972, 141-51.

13. A reader is inclined to dismiss this formulation as a job-fund variant of the wages-fund theory should bear in mind that the sequence seems to explain some of what did occur during the period of investigation. Institutions, customs and practices are all subject to change at future times, although with greater difficulty when the desired direction of change runs counter to tangible impacts of demographic changes.

* Several tables illustrating these rates, ratios and differences are available as an appendix to this paper from the author, Department of Economics, John Jay College, 445 West 59th Street, New York, New York, 10019

Figure 1

RELATIVE NONWHITE UNEMPLOYMENT RATES, 1967-73



INITIATIVE, INNOVATION, AND AGING IN MASTER CHESS

Ernest Rubin, American University

Prefatory Remarks

Performance ratings of chess masters in tournaments generally decrease after they have reached their prime. The time interval of optimal play of most masters appears to occur between the ages of 25 and 35 years. The general phenomenon of declining creativity and efficiency with age has, of course, been widely observed and commented upon throughout history.

There is, however, some basis to challenge the axiomatic character of this generalization. The purpose of this paper is to examine certain aspects of aging in master chess, particularly in relation to initiative and innovation during competitive play.

Aging in a Competitive Situation

A certain amount of individual impairment normally (or otherwise) occurs over time for all living things. In a competitive situation, however, two factors should be noted, the age of the individual and of his opposition. In sports that depend on individual rather than team performance, there is the further dichotomy, namely, how players fare in tournament and match play.

This analysis reviews the experience of chess grandmasters in tournament play. Few chess masters have engaged in serious match play throughout their entire competitive careers; tournament experience provides, therefore, a time sequence that more adequately provides data regarding age and performance.

Illustrative of the dual age process is the information presented in Table 1, Median Age of Lasker's Opponents in Selected Chess Tournaments: 1895-1936. Lasker's career spanned forty-seven years (1889-1936). For thirty-two years (1904-1936) or almost 70 percent of his active career, Lasker exceeded the median age of his opposition. The age differential widened with time after 1904. During the last decade of his chess activity (1927-36) he was the oldest competitor in his final four tournaments (1934-36).

Another example (Table 2) provides the age distribution data of the qualifying world champion interzonal tournaments from 1948-73. S. Gligoric (1923-), the Yugoslav grandmaster, has played in eight of the nine interzonal tournaments between 1948 and 1973. For these tournaments the mean age has varied between 30 and 37 and the median age between 29.5 and 38.3 years. Since 1955, Gligoric has exceeded the mean and median age of his competitors. His performance in the interzonals has fluctuated considerably since 1948 and has trended downward since 1958.

That the performance results of the chess master are influenced or affected by his age and that of his opposition has been commented on by various chess experts. Edward Lasker (1885-), an international chess master of broad experience, has stated:

> "...Chess history, like the history of competitive physical sports, had shown that youth was bound to triumph.
>
> "Physical endurance often made up for the greater experience and the finer positional understanding of an older master. Besides, the young players were catching up with the knowledge of their teachers with remarkable ease."

And psychologist Dr. Reuben Fine, a chess grandmaster, provides the complementary aspect:

> "...while older people (past 50) can retain skills they once had with little loss of ability, they find it difficult to learn new ones. It is this weakness which shows up in chess masters, because chess is always progressing, and those who stand still soon fall behind."

Innovation in Competition

A measure of innovation, though not necessarily an exclusive one, may be ascertained by examining "brilliancy" in chess play. To be sure, individuals regard such play in a very subjective manner--i.e., there are no standard and uniform rules or criteria to determine what constitutes brilliant play. Certain attempts have been made, however, to provide guidelines as to the elements of chess brilliancies.

For the present purpose, I accept the verdict of chess judges (or juries) that have determined which games deserve the accolade and the prize money for brilliancy or best-played games. Such determinations have been made since 1876.

In Table 3, Age Distribution of Brilliancy Prize Winners and Their Opponents in Master Chess: 1876 - 1955, data are provided, based on 77 games. Of interest, regarding the age factor, is that there is little discernible statistical difference, i.e., 0.4 years between the mean ages of the winners and losers. The breakdowns, by color of winner and loser, indicated only a slight difference of 0.5 years in the means between winner (white) and loser (black). Somewhat surprising, however, is the result between winner (black) and loser (white), showing a difference of 2.5 years, in favor of the older player with the black pieces. Possibly the small number of cases (20) affect the latter result. (Applying a t test indicates no statistically significant difference in these means).

The results suggest that age and innovation in master chess, as indicated by brilliancy prize awards, are substantively independent. Or to put these results in another way, older players contrive brilliancies on par with younger players.

Combinatorial skill, a primary indicator of chess talent, does not appear to be adversely affected over the active career of chess masters.

The Age Distribution of a Brilliancy Prize Winner

Also of interest is the relationship of aging to brilliancy prize awards for individual chess masters. Le Lionnais discusses 239 brilliancy prize games covering the period 1876-1949. Alekhine (1892-1946) obtained 19 brilliancy prize awards, exceeding by far, the performances in this category of about 200 chess masters.

Alekhine's active chess career of almost four decades covered the years 1908-45. With the exception of two years, 1935-37, Alekhine was world chess champion from 1927 until his death in 1946. He obtained his first brilliancy prize in 1916 and his last in 1942.

Data in Table 4 show Alekhine's age, brilliancy prizes awarded and tournament participation. His most frequent brilliancies occurred in the ten-year period 1922-31 when he was in his thirties. It was in this period that he won the world championship from Capablanca (1927) and achieved highest rankings in major tournaments.

For the present purpose it is worth noting that 40 percent of his brilliancies occurred after age 40. Further study is needed to obtain a sample providing comparable analysis for other masters who have been awarded at least 10 or more brilliancy prizes. Such data could provide an input toward a defensible model of the general relationship of aging and chess performance under standardized conditions of competition.

Individual Tournament Performance and Age

The tournament careers of six grandmasters were examined in some detail. These competitors were W. Steinitz (1836-1900), J. H. Blackburne (1841-1924), S. Tarrasch (1862-1934), E. Lasker (1868-1941, S. Tartakower (1887-1956), and J. R. Capablanca (1888-1942).

The period of chess competition represented by this group is almost a century; Steinitz and Blackburne began their chess careers in the early 1860's while the last tournament in which Tartakower participated was 1954. Contrasting styles of master play underwent many changes in this sequence of competing chess generations. The six players selected are representative of three generations, two grandmasters from each time period. On the average these generations are separated by about twenty-five years. In each pairing there is a world champion and another grandmaster. Their tournament records reveal individual performance patterns of considerable interest.

Some Comparative Aspects of the Six Grandmasters

A summary (Table 5) provides comparative relevant information for each player over his active chess career. Perhaps the first point to be noted is that the profile for each player is unique although there are certain similarities. The records of Steinitz, Capablanca and Tartakower provide performance profiles that suggest more general models or types.

While in every case a peak score is achieved after the first decade of play, the Capablanca profile provides the smoothest transition over time of the decrease in wins, the increase in draws, and relative stability in losses. The Steinitz profile, after his peak period shows the decrease in wins coupled with the increase in draws and losses over time. Tartakower reaches a peak and remains virtually level in the proportion of wins but, in his last decade of play (60-69), decreased in draws and increased in losses.

Additional information regarding the results for other grandmasters, playing the white and black sides, are needed. These tentative results indicate that the complex phenomenon called aging is of great variability as regards the initiative in chess. It is apparent that some masters, as revealed by their records, decline much faster than others. In the case of Tartakower, unlike Tarrasch, the decline is hardly apparent. For Blackburne his results after 70 were better, percentagewise, than in his sixties; but he only played 43 games in his last decade of play compared with 180 games a decade earlier, when he was 60-69.

Some variation regarding the optimal quinquennium age in tournament performance for eleven selected grandmasters is provided in Table 6. Two players, Tarrasch and Reti, did best in the age group 20-24 years. Four grandmasters, Blackburne, Tartakower, Alekhine and Keres were in the 25-29 year set. World champions, Steinitz, Lasker and Capablanca are in the 30-34 year age category. Gligoric in the 35-39 years and Nimzovich in the 40-44 years groups complete the table. This quinquennium breakdown, while of interest, contains an important limitation, namely, the number of games played. Thus, less than 75 games were played by 7 of the 11 grandmasters.

Older vs. Younger Competitors

As illustrative, let us consider the tournament record of Lasker's 36 losses between 1889-1936, taking into account his age and that of his victorious opponents. These data, including a breakdown by white and black, are detailed in Table 7.

The mean age of Lasker's losses is 44 years with a standard deviation of 17 years; the mean age of his opponents is around 34 years with a standard deviation of 11 years. The age distribution of Lasker's losses is U-shaped (comparable or analogous to a mortality curve). By contrast, the age distribution of the successful opponents is a reverse J-shaped curve. Additional data may reveal that these results are the rule with possibly few exceptions.

Of interest is the final performance in a major tournament of sixteen outstanding masters. These data, covering the 75 year period 1878 to 1953, provide detail as to age, score, rank and number of competitors in these tournaments. The variation of results is considerable, as is the age spread of 32 to 80 years.

Only three players in this group (Anderssen, Lasker and Maroczy) scored .500 or better. Maroczy's record is perhaps the most unusual in this group; he drew every game. Tarrasch at Bad Kissingen in 1928 lost 3 and drew 8 games. Some of the grandmasters, even at the terminal point in their careers, were very hard to beat. The lowest score was that of Gunsberg in the great St. Petersburg tournament of 1914; he lost 8, and drew 2 games.

To whom do the oldest players lose in these tournaments? Part of the answer was suggested in the discussion on Lasker's losses. The results for some of the players may provide clues to enlarge or refine the solution to the question posed.

At Nottingham in 1936, when Lasker was 68, he lost three games, one each to R. Fine (1914-), S. Reshevsky (1911-) and S. Flohr (1908-); their respective ages then were 22, 25 and 28 years. Capablanca at the A.V.R.O. tournament of 1938 was 50. He lost four games; one each to P. Keres (1916-), Botwinnik (1912-), Alekhine (1892-1946) and M. Euwe (1901-). The respective ages of these players were 22, 26, 46 and 37 years. (The last three noted held the world chess championship during their careers). When Tarrasch made his farewell appearance at Bad Kissingen in 1928, at age 66, he lost three games, drew eight games and won none. His losses were to E. Bogolyuboff (1889-1952), A. Nimzovich (1886-1935) and R. Reti (1889-1929); they were respectively 39, 42 and 39 years old in 1928. Blackburne was 73 when he played in the St. Petersburg tournament of 1914. He lost five games, one each to Capablanca, Lasker, Tarrasch, Marshall and Janowski. These opponents in 1914 were respectively 26, 46, 52, 37 and 46 years of age. Burn was 64 in 1912 when he played at Breslau, losing seven games in a field of 18 competitors. He was beaten by Duras, Rubinstein, Teichmann, Schlechter, Tarrasch, Spielmann and Treybal. The respective ages of the victors were 30, 30, 44, 38, 50, 29 and 30.

In the foregoing none of the selected masters (Lasker, Tarrasch, etc.) lost to an older player. There may be a few exceptions to this result; for example, Gunsberg, age 60, lost to Blackburne, age 73, in the St. Petersburg Tournament of 1914. Perhaps examples of this type would suggest that when both players are beyond a certain age, e.g., 55-59 years, age differences are not of consequence. By and large, however, the older player (50 years and over) usually lost to players at least one and a half generations younger and less frequently to competitors that were closer to him in the time continuum.

The significance of this last point is related to the ideas suggested by Reuben Fine and Edward Lasker. They maintain that older players do not keep up with newer developments in the field of chess and that younger players become familiar with the chess style, strengths and weaknesses of the older players.

The phenomenon of aging in a competitive situation is quite complex. It involves physiological and psychological factors as well as cultural and/or technological factors relating to the state of the art.

The discussion throughout this paper relates only to men. If sufficient data existed for women chess players, both as to number of players, tournaments, and for a substantial time period, a comparable analysis would, I believe, be quite valuable. Data for women players is relatively scarce, although tournament competition among women has expanded in recent years. There is very little data for tournaments consisting of men and women competitors at master strength.

Summary of Findings

The foregoing discussion has attempted to focus on individual performance, over time, of certain identifiable aspects in master chess competition, namely, innovation and initiative. An attempt was also made to indicate the nature of the competitive milieu; how, with the increasing age of the master, there also develops, after he has reached the age of 35 years, an increasing difference between his age and the average age of his competitors. Or to put it another way, in tournament play, the master not only ages after 35, but his collective opposition grows younger.

Innovation in tournament competition, as measured by the winning of brilliancy prizes, is independent of the age factor. Thus the average age and standard deviations for seventy-seven winners and losers revealed no statistically significant differences. Further, the age distribution of the winners was approximately that of all competitors in a sample of many tournaments. This point may be stated in another way, i.e., the proportion of participation by age, in tournaments, was consistent with the age groups obtaining brilliancy prizes.

Another factor of interest, namely, the initiative, revealed generally expected results, but with certain exceptions. For a selection of six grandmasters, drawn from three successive generations, I obtained profiles, by age, of their overall scores, wins, draws, and losses. Additional detail on the playing with white and black pieces was developed for Emanuel Lasker and J. R. Capablanca.

The profile of each player is unique. Three distinct profiles appear as models for characterizing broader groups of chess masters. The Tartakower model is, perhaps, of special interest; it appears to reflect virtually no decline in performance with age and an amazing stability in the quality of play with regard to the three categories of wins, draws and losses. The Capablanca model is quite distinct, to wit, after peak performance, a continual decrease in wins, increase in draws and only a marginal slight increase in losses. The Steinitz model is characterized, after the optimal performance period, by a steady decrease in wins coupled with an increase in losses and in draws.

Lasker's profile is comparable to that of Capablanca. Of further interest is the data of games in which Lasker had the white and black pieces. It will be recalled that white has the initiative by rule of the first move.

In his last three decades of play (40-69) Lasker won 44 games with white and 37 games with black; whereas in his first two decades of competition (20-39) he won 67 games with white and 52 games with black. Thus there is only a slight diminution in the initiative since the corresponding percentages of wins with white are 56 percent (early period) against 54 percent (later period).

Lasker's experience with draws perhaps best demonstrates the decline (though small for Lasker) in the initiative. Between 20 years and 40 years Lasker drew 40 games, of which 22 were with white and 18 with black. After 40, Lasker drew 62 games, 19 with white and 33 with black. Somewhat unusual is the lifetime result in that Lasker's total draws are evenly divided between white and black.

Throughout his tournament career Lasker lost 36 games, 11 with white and 25 with black. The proportion of losses with white is considerably greater after than prior to 40.

While the data on Lasker is based on total recorded tournament games, comparable data for Capablanca were developed on a (non-random) sample covering 25 percent of his total tournament games. Capablanca, over his career, won twice as often with white as with black and drew twice as often with black as with white. Of the 28 games he lost in thirty years of tournament play, about 90 percent were lost with the black pieces. Further detailed data are required to pinpoint the effect of age on the initiative in the case of Capablanca.

It appears, however, that declines in the advantage conferred by the initiative will vary considerably. Tartakower's record would imply no decline in the initiative. For most masters there is a decrease in performance after their peak period associated with white (the initiative) as well as with black. This result, however, cannot be attributed to a decline in mental ability or chess talent.

Concluding Remarks

On the basis of this preliminary exploration some conjectures appear plausible.

First, there are many patterns of "aging" in master chess. Not only are the rates of change, but the quality of the changes in performance over time, an individual matter. Additional study is required to define age profiles in master chess players.

A second point relates to the matter of chess ability; the power to reason in the strategic sense, is maintained almost unimpaired throughout the career of most masters. The effect of the younger opposition appears to be felt, first in physical and then in psychological terms.

The stress and strain of tournament play has several sequential aspects which differentially affect the older players. Sustained concentration, under time constraint rules, for a tournament period of 10 to 30 days probably favors the younger players. Recuperative power generally occurs sooner in younger than in older persons.

The psychological aspect is probably related to the inflexibility of the older player in adapting to new developments (e.g., in the openings) and to the adaptability of the younger player to the older player's style. The value of experience supersedes ambition and youth only to a point.

The second conjecture implies that it would be useful to study a group of masters (if possible grandmasters) born within, say, a five year age period. I think that the birth period of players born between 1880 and 1884, or 1885 and 1889, or in a consecutive five year period between 1880-1889 could provide a suitable cohort. Such a study would analyze how members of the cohort played against each other, and how they fared against competitors outside the cohort.

Another component of the aging problem in chess relates to stability of performance. There is a suggestion in some of the data that the master's performance, after the optimal period, undergoes a transition in the sense of becoming erratic. Younger players will, from time to time, be off form, that is, play quite badly in a particular tournament. This phenomenon appears to occur more frequently after a master has passed his peak period; it deserves systematic statistical investigation.

Age is recognized at the lower bound; there is a world junior championship and competition is restricted to players under 20. It would be useful and informative if tournaments could be based on suitable age ranges.

It is hoped that some of the conjectures that have been advanced will be analyzed. The records of master chess provide a data source of considerable value, provided that they are used with an understanding of their limitations. The implications of the present work for domains other than chess require separate explication.

NOTES AND REFERENCES

1. Because of space limitations additional information regarding this paper may be obtained from the author.

2. The age of the player at the time of the final tournament is shown in parentheses.

Anderssen (60); Bird (69); Blackburne (73); Burn (64); Capablanca (50); Gunsberg (60); Lasker (68); Maroczy (59); Marshall (54); Mieses (80); Pillsbury (32); Steinitz (63); Tarrasch (66); Tartakower (66); Tchigorin (57); Thomas (67).

Table 1

Median Age of Lasker's Opponents in Selected Tournaments[1/]: 1895-1936

Date	Tournament	Number of Opponents	Median Age of Opponents[2/]	Age of Lasker	Difference in Years (Lasker Younger -) (Lasker Older+)
1895	Hastings	21	42	27	-15
1896	Nurnberg	18	35	28	- 7
1899	London	14	38	31	- 7
1904	Cambridge Spring	15	34	36	+ 2
1909	St. Petersburg	18	28	41	+13
1914	St. Petersburg	10	37	46	+ 9
1923	Mahrisch-Ost.	13	35	55	+20
1924	New York	10	38	56	+18
1925	Moscow	20	34	57	+23
1934	Zurich	15	39	66	+27
1935	Moscow	19	30	67	+37
1936	Nottingham	14	37	68	+31

1/ The tournaments selected consist of 10 or more of Lasker's opponents.
2/ Median age rounded to nearest year. The median age of the 187 opponents in the foregoing selected tournaments was 35.5 years.

Source: Emanuel Lasker (two volumes of tournament and match games), edited by J. Gilchrist (The Chess Player, Nottingham, England, 1967-68).

Table 2

Age Distribution of Chess Masters in World Champion Interzonal Tournaments: 1948-73

Age Group	1948	1952	1955	1958	1962	1964	1967	1970	1973[a/]	Total	Percent Distribution
	TOURNAMENT YEAR										
15-19 years	0	0	1	1	1	0	1	1	0	5	2.3
20-24 years	3	2	2	5	0	1	3	1	8	25	12.0
25-29 years	2	5	4	5	6	6	1	3	6	38	18.0
30-34 years	1	2	6	4	9	8	8	7	4	49	23.1
35-39 years	6	6	4	3	3	1	5	7	7	42	19.8
40-44 years	7	4	2	3	2	6	2	2	2	30	14.2
45-49 years	0	2	2	0	1	0	0	2	4	11	5.2
50-54 years	0	0	0	0	1	1	1	0	2	5	2.3
55-59 years	0	0	0	0	0	1	1	1	2	5	2.3
60-64 years	1	0	0	0	0	0	0	0	1	2	1.0
Total......	20	21	21	21	23	24	22	24	36	212	100.0
Mean.......	37.0	35.1	33.2	30.5	33.6	25.4	34.3	35.4	36.0	34.6	---
Standard Deviation..	9.2	7.7	7.8	7.4	7.4	8.4	9.2	7.3	11.5	9.0	---

a/ Two interzonal tournaments were held for 36 qualifying chess masters in 1973. Eighteen competitors made up each tournament; the age data were consolidated.

Sources: Anne Sunnucks, The Encyclopedia of Chess, (St. Martin's Press, N.Y., 1970), pp. 380-391 for the years 1948-67. Chess Life and Review, January 1971, Vol.26, No. 1, p. 8 and September 1973, Vol. 28, No. 9, pp. 499-500 for the years 1970 and 1973 respectively.

Table 3. Age Distribution of Brilliancy Prize Winners and Their Opponents in Master Chess: 1876-1955

Age Group	Total: Winner	Total: Loser	Winner Played: White	Winner Played: Black	Loser Played: White	Loser Played: Black
10-19.........	0	1	0	0	1	0
20-29.........	21	23	18	3	7	16
30-39.........	38	33	26	12	6	27
40-49.........	10	11	8	2	3	8
50-59.........	5	6	3	2	1	5
60-69.........	2	3	1	1	2	0
70-79.........	1	0	1	0	0	1
Total.........	77	77	57	20	20	57
Mean..........	36.2	36.6	35.5	38.0	34.5	36.0
Stand.Deviation	10.4	11.3	11.0	10.0	14.3	10.2
Median........	34.6	34.4	34.0	35.8	33.3	34.6

Source: Fred Reinfeld, Great Brilliancy Prize Games of the Chess Masters (Collier Books, Crowell-Collier Publishing Co., N. Y. 1961); I. A. Horowitz, The Golden Treasury of Chess, (Galahad Books, N. Y., 1969).

Table 4

Brilliancy Prizes Awarded A. Alekhine, by Age and by Tournament Participation: 1908-45

Alekhine's Age	Number of Tournament Participations[1/]	Number of Brilliancy Prizes Awarded[2/]	Ratio of Prizes to Tournament
15-19	5	0	.000
20-24	13	1	.077
25-29	8	1	.125
30-34	8	6	.750
35-39	7	4	.571
40-44	16	2	.125
45-49	12	3	.250
50-54	11	2	.182
Total	80	19	.238

1/ The number of tournaments shown is incomplete; the actual number is probably between 85 and 90.

2/ It is not known whether brilliancy prizes were awarded in every tournament.

Source: François Le Lionnais, Les Prix de Beauté aux Echècs. (Rev. Ed., Payot, Paris 1973).

Table 5

Percentage Tournament Scores of Grandmasters by Age Group

Player	20-29	30-39	40-49	50-59	60-69	70-79	Lifetime
Steinitz	.710	.820	.715	.615	.580	---	.685
Blackburne	.460	.750	.695	.560	.655	.590	.640
Tarrasch	.675	.805	.620	.540	.480	---	.585
Lasker	.765	.815	.795	.765	.595	---	.740
Capablanca	.825	.870	.725	.665	---	---	.765
Tartakower	.575	.610	.650	.640	.625	---	.630

Table 6

Optimal Tournament Performance in a Five-Year Age Period of Eleven Selected Grandmasters

Player	Age Period	Games Played	Percent Distribution: Wins	Draws	Losses	Score
Tarrasch..........	20-24	29	.655	.069	.276	.690
Reti..............	20-24	190	.500	.205	.295	.603
Blackburne........	25-29	33	.515	.182	.303	.606
Tartakower........	25-29	71	.479	.380	.141	.669
Alekhine..........	25-29	43	.651	.349	0	.826
Steinitz..........	30-34	58	.672	.104	.224	.724
Lasker............	30-34	43	.767	.186	.047	.860
Capablanca........	30-34	38	.789	.211	0	,895
Gligoric1/........	35-39	197	.467	.459	.074	.697
Nimzovich.........	40-44	188	.543	.340	.117	.713
Keres2/...........	25-29	110	.573	.336	.091	.741

1/ S. Gligoric (1923-) of Yugoslavia is an International Grandmaster.

2/ P. Keres (1916-) of the U.S.S.R. is an International Grandmaster.

Table 7
Age Distribution of Victorious Tournament Opponents
of Emanuel Lasker, by Color of Pieces: 1889-1936

Lasker's Age Group	Number of Games Won by Successful Opponents, Age Group: 20-29	30-39	40-49	50-59	60-69	Games lost By Lasker Total
20-29..........	7	3	1	2	1	14
30-39..........	1	2	0	1	0	4
40-49..........	2	1	0	0	0	3
50-59..........	1	2	0	0	0	3
60-69..........	8	0	4	0	0	12
Total..........	19	8	5	3	1	36

Opponent (White Pieces)

Lasker's (Black Pieces)	20-29	30-39	40-49	50-59	60-69	Total
20-29..........	6	3	0	1	1	11
30-39..........	2	1	0	0	0	3
40-49..........	1	2	0	0	0	3
50-59..........	1	1	0	0	0	2
60-69..........	5	0	1	0	0	6
Total..........	15	7	1	1	1	25

Opponent (Black Pieces)

Lasker's (White Pieces)	20-29	30-39	40-49	50-59	60-69	Total
20-29..........	1	0	1	1	0	3
30-39..........	0	0	0	1	0	1
40-49..........	0	0	0	0	0	0
50-59..........	0	1	0	0	0	1
60-69..........	3	0	3	0	0	6
Total..........	4	1	4	2	0	11

SOCIAL INDICATOR CONSTRUCTION USING ORTHOGONAL CONTRASTS AND CLUSTER ANALYSIS

James M. Sakoda and James P. Karon, Brown University

What we have to say is not particularly novel or earthshaking, but may be of interest and use to those engaged in enterprises involving the construction of social indicators. The use of orthogonal contrasts makes it possible to construct indices which are independent of one another and free of spurious correlations. It also permits the creation of more than one index from a variable. We have applied this approach in part to an analysis of 212 census tracts in Rhode Island, using Count 4 census tract data from the 1970 census of population and housing. We constructed a variety of indices and subjected the correlation coefficients among them to a cluster analysis procedure. On the basis of the composition of the various clusters we have concluded that many of the social indicators we have constructed are meaningful and that the use of the orthogonal contrast approach to index construction helps to create a variety of indices from a single variable.

There are two alternatives to use of orthogonal contrasts. The first is to calculate an overall index such as median education, median income, median age as a single index for a continuous variable. The use of orthogonal contrast applies to variables consisting of three of more categories, but continous variables can still be broken up into several categories, and hence one can elect either approach. When several categories of a variable are set up or are available in the original data source, there is a variety of combinations of proportions or ratios which can be constructed. A standard approach is to construct _k_ dummy variables for a _k_ category variable. Dummy variables are traditionally calculated by scoring one category as 1 and the remaining ones as zero, which is essentially the same as calculating a proportion in each category based on the total number in all categories. With three categories, such as single, divorced or separated and married we can construct three proportions:

proportion single, proportion divorced or separated and proportion married.

The difficulty with the dummy variable approach is that there is a spurious negative correlation among the proportions, since they must add up to 1.0. In multiple regression this lack of independence is taken care of by dropping one of the dummy variables and letting the multiple regression procedure adjust the effects of negative correlations. Such variables, however, are often not suitable as social indicators, since each variable reflects the combined effects of the remaining variables. In factor analysis the spurious negative correlations can result in bipolar factors, which are not meaningful.

Setting up orthogonal contrasts involves starting with or creating several categories, such as 5 or 10 year age groupings, but it is desirable to using groupings which are meaningfully different. For example, one can break up age into life cycle categories. We selected four categories, 0-5, 6-24, 25-64 and 65 and over. These groupings correspond to young children, youths, active adults, old people. Orthogonal contrasts involve setting up a set of k-1 independent comparisons, and can be checked most easily for orthogonality by setting up weights so that they add up to zero and also so that the sum of crossproducts of weights add up to zero. One standard approach to orthogonal contrast is to pick one variable and compare it with the remaining categories, drop that category and pick another category to compare with the remaining ones, and continue in this way until only a single proportion is justified.

Table 1.
Orthogonal Comparisons for Three Categories

a	b	c	Proportions
+2	-1	-1	a / (a + b + c)
0	+1	-1	b / (b + c)

In Table 1 is shown the weights and proportions for three categories, and in Table 2, for four categories. For four categories _a_ is compared with a combination of _b_, _c_, and _d_; _b_ with a combination of _c_ and _d_; and _c_ with _d_.

Table 2.
Orthogonal Comparisons for Four Categories: I

a	b	c	d	Proportions
+3	-1	-1	-1	a / (a + b + c + d)
0	+2	-1	-1	b / (b + c + d)
0	0	+1	-1	c / (c + d)

Each proportion has a different base denominator term. There is a discussion of orthogonal contrasts in both Snedecor and Blalock. The order of picking and dropping a category is the choice of the user, and he should pick the order that he considers to be most meaningful.

Table 3.
Orthogonal Comparisons for Four Categories: II

a	b	c	d	Proportions
+1	+1	-1	-1	(a + b) / (a + b + c + d)
+1	-1	0	0	a / (a + b)
0	0	+1	-1	c / (c + d)

For four categories or more there is another possible approach, shown in Table 3. That is to divide the categories into two groups and then take each half separately and proceed to pick orthogonal contrasts. This approach we used for our AGE variable and is shown in Table 4. With four categories there is another possible combination which involves the notion of two main effects and their interactions. The weights and proportions are shown in Table 5. In using orthogonal contrasts it is necessary to settle for a coherent set of indices, and it is the set that is most useful that has to be constructed. From a practical point of view one may choose to

Table 4.
Orthogonal Comparisons Applied to AGE

+1	+1	-1	-1	Variable 1. Young (0-5, 6-24) / Total population
+1	-1	0	0	Variable 2. Very young (0-5) / 0-24
0	0	-1	+1	Variable 3. Aged (65+) / 25 and over

Table 5.
Orthogonal Comparisons for Four Categories: III

a	b	c	d	Proportions
+1	+1	-1	-1	(a + b) / (a + b + c + d)
+1	-1	+1	-1	(a + c) / (a + b + c + d)
+1	-1	-1	+1	(a + d) / (a + b + c + d)

use nonorthogonal comparisons, but care should be taken to avoid spurious correlations.

We constructed 42 indices, not all of which were based on orthogonal contrasts. A few were medians, some were based on only two categories, and some did not follow the orthogonal approach, but were based on hunches as to the significance of a variable. These variables were subjected to a cluster analysis procedure developed by the senior author. The procedure can be described briefly as involving the following steps:

1. Pick a key variable on the basis of the highest variance of rows of absolute r's in the 42 x 42 correlation matrix from among variables not already included in clusters or tested previously as a key variable. The variance does a good job of picking good key variables for distinctive clusters.

2. Add variables to the cluster on the basis of correlation with the key variable corrected for uniqueness, generally using .70 as a cutoff point. This is a loose criterion which helps to prevent formation of clusters between two independent dimensions.

3. Accept the cluster if a minimum number of variables (usually three or four) is included in the cluster. A variable ending up in more than one cluster is allotted to the cluster with which it has the highest relationship.

4. Continue in this way until all variables are included in clusters or have been tested as key variables.

The cluster analysis is followed by a multiple group factor analysis and a procedure to improve the solution. This procedure has produced six factors, all of which were interpretable. We feel that this is due to the sensitivity of the clustering procedure. In many social area studies only three interpretable factors are found. Our six factors were: socio-economic status, group quarters, migration, residential or suburban status, ethnic status and instability. The three uncommon factors are group quarters, migration, and ethnic status. The function of the cluster or factor analysis is to examine the relationships among the indices and thereby to seek meaning for them. We were particularly interested in knowing whether the indices were meaningful and whether by use of the orthogonal contrast procedure we could increase the use of a single variable to measure different dimensions. We were gratified to find that this happened in the case of AGE, FAMILY TYPE, MARITAL STATUS, NATIVITY, MIGRATION and LABOR FORCE. The three indices based on age categories, for example, ended up in three different clusters. The proportion young went into the migration cluster, the proportion very young into the group quarters factor, and the proportion aged into the residential one. This happened in spite of the fact that the correlation between proportion young and very young was -.683.

Another one of our successes was marital status. Proportion ever-married went into the group quarters factor and proportion divorced or separated into the instability factor. Similarly, nativity produced a foreign stock index which identified with the instability factor. Not all orthogonal contrasts showed this diversity of meaning. Male labor force resulted in two variables, proportion professional-managerial and proportion blue collar, both of which represented socioeconomic status. The housing indices were not organized systematically according to orthogonal contrast principles, but at least three different denominator terms were used. Housing items are represented in three clusters--residential, socioeconomic status and group quarters.

Conclusion: Our conclusion is that where we applied orthogonal contrast principles there was considerable success in creating two or more indices which represented different meaningful factors. We might have benefited from still additional orthogonal indices in some instances. We also feel that factor analysis has generally not solved the problems of determining the number of dimensions and rotation to a meaningful position, and there remains a valid need for a good cluster analysis program. Those interested in a more detailed paper should write to us. The cluster analysis program is written in DYSTAL, a set of FORTRAN subroutines allowing for dynamic storage allocation, and is one of the programs in our DYSTAL II tape. If you are interested you can write to the senior author about obtaining a copy of the tape.

Table 6.
Description of 42 Indices

Factors: S = Socio-Economic Status; G = Group Quarters; M = Migration; R = Residential (Suburban); E = Ethnic (Nonethnic); I = Instability (Stability)

Source	Factor	No.	Name	Index
Age	M	1	Young	Population, age 0-24 / Total population (r_{12} = -.137, r_{13} = -.683, r_{23} = -.011)
	G	-2	Very young	Population, age 0-5 / Population, age 0-24
	R	-3	Aged	Population, age 65 and over / Population, age 25 and over
	G	4	Group quarters	Population living in group quarters / Total population
Family type	I	-5	H-W families	Husband-wife families / All families
	M	20	Couples with children	Married couples with children under 6 (should have used H-W families with children under 6 to be consistent)
		6	Children	Population, age 0-16 / All families
Marital status	G	-7	Ever-married	Ever-married persons, age 14 and over / Population, age 14 and over (r_{78} = -.342)
	I	+8	Div-sep	Divorced or separated persons, age 14 and over / Ever-married persons, age 14 and over ($r_{9,10}$ = -.210)
Nativity	E	+9	Foreign stock	Foreign born or native of foreign or mixed parentage / Total population
	I	+10	Black	Black population / Population native of native parentage (both parents)
	M	11	Recent migrant	Foreign born persons who immigrated from 1945-1970 / Total foreign born population
	M	12	Out of staters	Persons born out of state / Total population
	M	-13	Stayers	Persons, age 5 or more, living in same house five years ago / Population, age 5 or more (excluding Armed Forces)
Migration	E	+14	Old residents	Persons who moved into present unit 1949 or earlier (excluding persons born into unit) / All persons who have ever changed housing unit ($r_{14,15}$ = -.218)
	M	15	Recent movers	Persons who moved into present unit from 1968-1970 (excluding persons born into unit) / Persons who moved into present unit from 1950-1970 (excluding persons born into unit)
		16	H.S. graduates	Persons, age 18-24, who have completed high school / Population, age 18-24
		17	College graduates	Persons, age 25 or more, who have completed college / Population, age 25 or more
		18	Median education	Median educational level of population, age 25 or more
		19	Fertility	Number of children ever born to females, age 15-44 / Ever-married females, age 15-44

	E	+21	Female labor force	Females, age 16-64, not inmates and not attending school but in the labor force / Females, age 16-64, not inmates and not attending school
		-22	Male unemployment	Civilian males, age 16 and over, unemployed (but in labor force) / Total civilian male labor force ($r_{22,23}$ = .340, $r_{22,24}$ = .139, $r_{23,24}$ = .664)
Male and female labor force	S	23	Prof.-managerial	Civilian males, age 14 and over, employed in professional, technical, managerial or administrative work / Total civilian male employed population, age 14 and over
	S	-24	Blue collar	Civilian males, age 14 and over, employed in blue collar occupations (not farmers or farm managers) / All civilian employed males, age 14 and over, other than professional, technical, managerial or administrative
	G	28	Males	Male population / Total population
	S	25	Family income	Median family income
	I	+26	Welfare	Families or unrelated individuals, age 14 and over, receiving public assistance or welfare payments / All families and unrelated individuals, age 14 and over, with income
	I	+27	Poverty	Families in government determined poverty status / All families
	R	-29	Old housing	Occupied or vacant year-round housing units built 1939 or earlier / All occupied and vacant year-round housing units
		-40	Vacancy	Vacant housing units for rent or sale 6 months or more / All occupied and vacant year-round housing units
	R	30	One-unit housing	One unit, detached, occupied housing / All occupied housing units
	S	31	1 1/2 baths	Occupied housing units with one and a half or more baths / All occupied housing units
		-32	Air conditioning	Occupied housing units with no air conditioning / All occupied housing units
	S	33	Rent	Median gross rent of renter occupied units
		-34	Four or more	Occupied housing units in structures of four or more stories / All occupied housing units
	S	35	Dishwasher	Occupied housing units with a dishwasher / All occupied housing units
		36	Owner-occupied	Housing units which are owner occupied or being bought / All occupied housing units
	R	41	Young owner	Husband-wife families with head below age 45 in one unit, owner occupied housing (not mobile homes or trailers) / All owner occupied housing units
	S	42	New homes	One unit, owner occupied housing of value $25000 or more, built from 1950-1970 / All one unit, owner occupied housing units
	R	-39	Commercial use	One family, occupied housing units which are used for commercial purposes / All one family, occupied housing units

G	37	Persons per room	Total population / Total number of rooms not in group quarters
S	38	Value of housing	Median value of owner occupied, one family houses which are on a place of ten acres or less and have no business or medical office on the property. (Mobile homes, trailers, cooperatives and condominiums are not included)

Bibliography

Blalock, Hubert M., Jr., Social Statistics, 2nd edition, McGraw-Hill, 1972.

Dent, Franklin Owen, Aspects of Change in Urban Social Spatial Structure, Doctoral Dissertation, Brown University, Department of Sociology, 1972.

Rummel, R.J., Applied Factor Analysis, Evanston: Northwestern University Press, 1970.

Shevky, Eshref and Bell, Wendell, Social Area Analysis, rev. ed., Chicago: University of Chicago Press, 1955.

Shevky, Eshref and Williams, Marilyn, The Social Areas of Los Angeles, Berkeley: University of California Press, 1949.

Snedecor, George W., "Nonorthogonal Sets of Comparisons," Statistical Methods, 4th edition, Ames, Iowa: The Collegiate Press, Inc., 1946, pp. 406-408.

Thomson, Godfrey H., The Factorial Analysis of Human Ability, New York: Houghton Mifflin, 1951.

Thurstone, L. L., Multiple Factor Analysis, Chicago: University of Chicago Press, 1947.

Timms, Duncan W.G., The Urban Mosaic, London: Cambridge University Press, 1971.

Tryon, R.C., "Cumulative Communality Cluster Analysis," Educational and Psychological Measurement, 18 (1958), pp. 3-35.

Tryon, R.C.,and Bailey, D.E., Cluster Analysis, New York: McGraw-Hill, 1970.

NONSAMPLING ERRORS

Eugene Sauls, University of Arkansas
and
James A. Millar, University of Arkansas

Introduction

There are two classes of nonsampling errors which occur in a survey. These classes are herein designated as nonresponsive errors and improper response errors. Nonresponsive errors arise when an inquiry is not answered. Improper response errors arise when a respondent offers an incorrect answer to an inquiry.

This paper reports on a study which investigated the existence of each of these classes of errors. The study was conducted with the use of deposit accounts in a credit union and loan accounts in a bank. Confirmations of the correctness of account balances were requested from the depositors and debtors after the account balances were adjusted.

The study shows that nonresponse and improper errors do exist and must be recognized by surveyors.

Nonresponses

Nonresponse errors occur because the subject is unable to respond or because the subject refuses to answer. The subject may be unable to respond because of a lack of knowledge or because of an inability to understand the inquiry. The refusal may result because the subject views the inquiry as an unjustifiable invasion of privacy or because the subject simply will not take the time to respond.

The subject may be asked questions for which he lacks the knowledge to answer. In some instances the lack of knowledge may be a function of background characteristics. For example the nonresponsive rate to a question concerning educational philosophies may be higher among less educated subjects. If the lack of knowledge is a function of the characteristics of the subject, a bias is injected into the results. This bias may be compensated for by stratifying the population by its relevant characteristics. In some cases the subjects favoring a particular position may become more knowledgeable because they seek information supporting their position. This form of bias cannot be compensated for; therefore, conclusions apply only to the population to which the subjects belong. For example, one might expect a higher nonresponse rate from athiest on a question concerning support for the divinity of Jesus Christ than would be expected from Christians. The degree of devoutness would presumably affect the nonresponse rate. If the population cannot be stratified according to its characteristics or the ratio of each strata to the population is unknown, an immeasurable bias is present.

The interviewer may pose the question in such a way that the subject is not certain of the questions. Rather than attempt to interpret the question, the subject may choose to not respond. If the question is worded so that any reservation as to the question would not be a function of the characteristics of the subject, no bias is present.

The subject may refuse to respond on the grounds that the question is an unjustifiable invasion of privacy. There would generally be no reason to believe that those who refused to respond for this reason would be any different from those who responded.

A subject's view of his time constraints may result in nonresponse errors. A subject who felt that he was busy or that his time is too valuable to be spent answering questionnaires may refuse to respond. Perhaps a bias is frequently encountered because of the subject's perceived time constraint. For example, public officials and business executives would probably be much more inclined to not respond because of the time involved than would a trademan or a laborer.

Improper Responses

Improper response errors occur because the respondent (1) does not know the correct answer but offers one nevertheless, (2) does not understand the question and therefore misinterprets the question, (3) is influenced by the interviewer the offer the answer perceived by the subject as the answer most suitable to the interviewer or (4) deliberately misstates his position in order to influence others.

Opinion surveys have a relatively high probability of being made of subjects who do not know the answer. For example, a subject may respond that blue is his favorite color and he may actually believe he is telling the truth whereas a thorough analysis of his preference through choice situations may reveal that red is his favorite color. In most cases, however, the lack of knowledge is probably not a function of some characteristic of the population; therefore, there would generally be no significant bias in the overall results of a large sample. Technical questions may result in a serious bias because some groups of the population do not have sufficient knowledge to answer but answer anyway.

Misinterpretation of a question poses a serious problem. There might be many instances where the interpretation of the question is a function of some characteristic of the population, e.g., education.

Prior Studies

Robert Ferber reported the results of a study in which interviews were conducted with 188 urban subjects and 33 farm subjects.[1] The subjects were queried about their financial holdings. The balances of certain deposit accounts held by the subjects were known. The nonresponse rate was seventy-nine percent of subjects contacted. More importantly, nonresponders had a significantly higher average balance than responders. Population estimates would be significantly below the true balance in cases where this condition existed.

Research Design

In the financial audits of organizations it is standard practice for the auditors to circularize the accounts, i.e., confirmation requests are sent to creditors and debtors asking them to confirm the balance or to report a difference. The evidence gathered through this procedure assists the auditor in arriving at a conclusion as to the reasonableness of the account balance.

Confirmation requests were sent under routine audit conditions to samples of deposit accounts of the MSU Employees Credit Union and loan accounts of the Continental Illinois National Bank and Trust Company of Chicago.

Deposit Accounts

The experiment was conducted on the 456 time deposit accounts as of February 29, 1968. Twenty-two accounts were deleted for various reasons. Certificates are issued to the depositors, thus there was tangible evidence on the account balance. The balance of the account represented principal only and averaged approximately $3,000.

Three samples were drawn: K_1 - a sample of fifty accounts which were circularized without adjusting the balances of the accounts. This sample was the control group.

K_2 - a sample of thirty accounts, the balances of which were adjusted by a positive adjustment of ten percent - (rounded to the nearest fifty dollars) with a maximum adjustment of $500. One recipient in this sample learned that the confirmation request was part of a study; therefore, that account was deleted from the sample, leaving twenty-nine accounts in the sample.

K_3 - a sample of thirty accounts, the balances of which were adjusted by negative ten percent (rounded to the nearest fifty dollars) with maximum adjustment of $500.

Second requests were mailed fourteen days after the first request to those who did not respond to the first requests.

Loan Accounts

Direct personal loan accounts were used in this part of the study. The average balance of the personal loan accounts in the sample was approximately $1,200 and the average balance of the automobile loan accounts was approximately $1,800. The experiment was conducted during March and April 1968. Two samples were drawn:

- b_1- a sample of one hundred accounts which were circularized without adjusting the balances of the accounts. This sample represented the control group.
- b_2- a sample of thirty accounts, the balances of which were adjusted by a positive adjustment of approximately ten percent.

Second requests were mailed to sample b_1 about two weeks after the first requests were mailed. Second requests were not sent to sample b_2 because of the adverse reaction to the first requests.

Hypotheses

Disregarding the effects on statistical models, the effects of nonresponse and improper response errors on conclusions drawn from questionnaires are a function of the extent of errors and the degree of errors. The extent and degree of errors varies with the circumstances. There can be no universal criteria for acceptable error rates. In order to provide test criteria, arbitrary error rates were set.

The test criterion for proper responses was 0.70 when there was no effort to follow up on nonresponses. When second requests were sent to nonresponsers to the first requests, the test criterion was 0.90 (approximately = 0.70 + (0.70) x (0.30). Improper response errors were tested against a 0.05 test criterion. Nonresponse errors were tested against the nonresponse error rate in the control group.

The hypotheses are designated as "D" for deposits for "L" for loans. The second identifier is: "P" for improper response or "N" for nonresponse. The third identifier is the sequential number of the hypothesis.

H_o:DP1 - The proportion proper responses to overstated amounts is 0.90.

H_a: The proportion of proper responses to overstated amounts is less than 0.90.

H_o:DP2 - The proportion of proper responses to understated amounts is 0.90.

H_a: The proportion of proper responses to understated amounts is less than 0.90.

H_oD1 - The proportion of improper responses to overstated amounts is 0.05.

H_a: The proportion of improper responses to overstated amounts is more than 0.05.

H_o:D14 - The proportion of improper responses to understated amounts is 0.05.

H_a: The proportion of improper responses to overstated amounts is more than 0.50.

H_oD15 - The proportion of improper responses to overstated amounts is equal to that of understated amounts.

H_a: The proportion of improper responses to overstated amounts is not equal to that of understated amounts.

H_oDN6 - The proportion of nonresponses to incorrect amounts is equal to that of correct amounts.

H_a: The proportion of nonresponses to incorrect amounts is not equal to that of correct amounts.

H_oDN7 - The proportion of nonresponses to overstated amounts is equal to that of understated amounts.

H_a: The proportion of nonresponse to overstated amounts is not equal to that of understated amounts.

H_oDP2 - The proportion of proper responses to understated amounts is 0.90.

H_a: The proportion of proper responses to understated amounts is less than 0.90.

H_o:D13 - The proportion of improper responses to overstated amounts is 0.05.

H_a: The proportion of improper responses to overstated amounts is more than 0.05.

H_o:D14 - The proportion of improper responses to understated amounts is 0.05.

H_a: The proportion of improper responses to overstated amounts is more than 0.50.

H_o:D15 - The proportion of improper responses to overstated amounts is equal to that of understated amounts.

H_a: The proportion of improper responses to overstated amounts is not equal to that of understated amounts.

H_o:DN6 - The proportion of nonresponses to incorrect amounts is equal to that of correct amounts.

H_a: The proportion of nonresponses to incorrect amounts is not equal to that of correct amounts.

H_oDN7 - The proportion of nonresponses to overstated amounts is equal to that of understated amounts.

H_a: The proportion of nonresponse to overstated amounts is not equal to that of understated amounts.

H_oDP8 - The proportion of proper responses to first request on incorrect amounts is equal to that of second requests.

H_a: The proportion of proper response to first requests on incorrect amounts is not equal to that of second request.

H_oDP9 - The proportion of proper responses is independent of the age of the recipient.

H_a: The proportion of proper responses is not independent of the age of the recipient.

H_oDP10 - The proportion of proper responses is independent of the size of amount.

H_a: The proportion of proper responses is not independent of the size of the amount.

H_oLP11 - The proportion of proper responses to overstated amount is 0.70.

H_a: The proportion of proper responses to overstated amounts is less than 0.70.

H_oLI12 - The proportion of improper responses to overstated amounts is 0.05.

H_a: The proportion of improper responses to overstated amounts is greater than 0.05.

H_oLN13 - The proportion of nonresponses to overstated amounts is equal to that of correct amounts.

H_a: The proportion of nonresponses to overstated amounts is not equal to that of correct amounts.

H_oDLP14 - The proportion of proper responses to incorrect deposit accounts is equal to that of loan accounts.

[1]Robert Ferber, "The Reliability of Consumer Surveys on Financial Holdings: Time Deposits," Journal of the American Statistical Association (March 1965), 148-163.

Test Results

Hypothesis	Test Statistic	Decision
1	*Z=.00106	reject
2	*Z=.17361	Not reject
3	*Z=.00184	reject
4	*Z=.18630	Not reject
5	**t= 1.06	Not reject
6	**t= .74	Not reject
7	**t= .05	Not reject
8	**t= .10	Not reject
9	***F= .200	Not reject
10	***F= 1.06	Not reject
11	****Z= 3.19	reject
12	****Z= 1.25	Not reject
13	χ^2 = .17	Not reject
14	***F= 7.16	reject
15	***F= 1.57	Not reject
16	***F= 4.64	reject

*Hypergeometric

**Bennett and Franklin Test

	.05	.01
t	2.77	3.64

***Bartletts Transformation

	.05
F	4.46

****Binomial

THE IMPACT OF SCHOOL DESEGREGATION PLAN FEATURES ON WHITE REJECTION[1]

Wen-Fu P. Shih, Everett F. Cataldo, Micheal W. Giles,
and Douglas S. Gatlin
Florida Atlantic University

INTRODUCTION

The present study examines selected factors associated with the withdrawal of whites from desegregated public schools to segregated private schools. The analysis is based upon a sample survey of parents of elementary and secondary school children in seven desegregated school districts in Florida, conducted in the winter and spring of 1973. Several official school district sources were consulted to identify white rejecters--parents with a child attending public school in a district in 1971-72, but enrolled in a local private school in 1972-73. A total of 1,386 white rejecters were interviewed. For comparison purposes, interviews were obtained from a sample of 2,112 white compliers--parents with a child attending public school in the same district in both 1971-72 and 1972-73. The data have been weighted to reflect the estimated "true" ratio of compliers to rejecters within the districts.[2]

Several factors have been associated with rejection of desegregation.[3] Each of the factors was an element in the implementation of desegregation in the seven school districts under examination. The independent variables and their operationalizations are as follows:

1. Busing to achieve racial balance.
   a) No Bus - Bus: children were not bused in 1971-72, but were scheduled to be bused in 1972-73.
   b) No Bus - No Bus: children were not bused in either 1971-72 or 1972-73.
2. School reassignment to achieve racial balance.
   a) Transferred: children were assigned to different schools between 1971-72 and 1972-73.
   b) Not transferred: children remained in the same school between 1971-72 and 1972-73.
3. Percent black enrollment.
   a) Black ratio increased: the proportion of black students in assigned public schools increased between 1971-72 and 1972-73.
   b) Black ratio not increased: the proportion of black students in assigned public schools stayed the same or decreased between years.
4. Distance from school.
   a) Distance increased: the mileage from home to school increased between 1971-72 and 1972-73.
   b) Distance not increased: the mileage from home to school remained the same or decreased between years.

The racial balances in public schools attended by respondents' children in 1971-72 were taken from official school records. Racial balance data for the 1972-73 school year and information on transfers were recorded for the public schools to which rejecters' children were assigned, and for the public schools attended by compliers' children. Information on the busing and distance variables was obtained from the survey instrument. The contribution of each of these factors to rejection is examined through the use of conditional probabilities.

FINDINGS

If each of the factors contributes to white withdrawal, we would expect the probability of rejection to increase with the onset of busing, school transfer, black ratio increase and distance increase. As Table 1 shows, each of these factors appears to be significantly related to rejection. The probability of rejection is greater under the presence of these conditions, and lower under their absence.

TABLE 1

CONDITIONAL PROBABILITY OF REJECTION UNDER EACH CONDITION

Condition	Probability	z†
Busing:		
No Bus - Bus	.053	8.674*
No Bus - No Bus	.029	
Reassignment:		
Transferred	.047	9.840*
Not transferred	.023	
Percent black enrollment:		
Black ratio increased	.038	6.780*
Black ratio not increased:	.022	
Distance from school:		
Distance increased	.043	5.783*
Distance not increased	.028	

†Standard normal deviate[4]

*Significant at $\alpha = .01$

These four factors, however, are confounded by intercorrelation. For example, children who are transferred may, simultaneously, experience the onset of busing to a school a greater distance from home. Table 2(a) indicates that for those parents whose children are transferred to different schools, neither busing nor increased distances, alone, influence the decision to reject. Parents whose children were bused were no more likely to reject than those whose children were not bused. Similarly, parents whose children traveled increased distances to school were no more likely to reject than those whose children did not.

TABLE 2

PROBABILITY OF REJECTION UNDER THE CONDITIONS OF TRANSFER, BUSING AND DISTANCE INCREASE

a)

Condition	Probability	z
Transferred:		
Bused	.044	.652$^{n.s.}$
Not - bused	.040	
Transferred:		
Distance increased	.042	1.697$^{n.s.}$
Distance not increased	.064	

b)

Condition	Probability	z
Transferred:		
Bused and distance increased	.042	.196$^{n.s.}$
Not - bused and distance not increased	.043	

n.s.: not significant at $\alpha = .05$

Perhaps the onset of busing and increased distances, conjointly, influence rejection decisions for the transfer group. If this is the case, those who experience busing and increased distances should be more likely to reject than those who experience neither. As Table 2 (b) shows, this is not the case. Among those transferred, parents whose children experienced both conditions were no more likely to reject than parents whose children experienced neither condition. Thus, we may conclude that the reassignment variable, and not the busing and distance variables, appears to be the main determinant of rejection.

Whether or not students are transferred, they may experience an increase in the black ratio in their assigned schools. When white students are reassigned to achieve racial balances, the transfers are often from predominantly white schools to schools that were predominantly black, particularly at the outset of desegregation. Students not reassigned may experience an increase in the black ratio as a result of black students from other neighborhoods being transferred into their schools to replace white students reassigned elsewhere. Table 3(a) shows that an increase in the black ratio contributes to rejection irrespective of reassignment. For both the transferred and not transferred groups, the probability of rejection increases with an increase in the black ratio. Does this mean that black ratio increase and not school transfer is the determining factor in rejection? Table 3(b) shows that when the reassignment variable is controlled by the black ratio variable, the effect of being transferred is not entirely eliminated. Reassignment to a different school remains related to rejection when the black ratio is increased, but not when the black ratio remains the same or decreases. For those who experienced an increase in black ratio, parents whose children are transferred are more likely to reject than parents whose children are not. Thus, there is a tendency for rejection to increase when the black ratio increases, particularly when this condition is combined with reassignment to a different school.

TABLE 3

PROBABILITY OF REJECTION UNDER THE CONDITIONS OF TRANSFER AND BLACK RATIO INCREASE

a)

Condition	Probability	z
Transferred:		
Black ratio increased	.049	5.951^*
Black ratio not increased	.021	
Not transferred:		
Black ratio increased	.026	2.362^{**}
Black ratio not increased	.020	

TABLE 3 (continued)

b)

Condition	Probability	z
Black ratio increased:		
Transferred	.049	6.735^*
Not transferred	.026	
Black ratio not increased:		
Transferred	.021	$.309^{n.s.}$
Not transferred	.020	

**Significant at α = .05

Let us now reexamine the busing variable to determine its potential effect on rejection in combination with the school reassignment and black ratio variables. We observed above that for the transferred group the onset of busing did not contribute to rejection. Is this the case even in conjunction with an increase in the black ratio? Table 4(a) shows that irrespective of busing, parents of transferred children who experience an increase in the black ratio are more likely to reject than those who do not. Among the former, Table 4(b) further indicates that those who are bused are no more likely to reject than those who are not bused. Thus, the onset of busing appears to make no appreciable contribution to rejection.

TABLE 4

PROBABILITY OF REJECTION UNDER THE CONDITIONS OF TRANSFER, BUSING AND BLACK RATIO INCREASE

a)

Condition	Probability	z
Transferred and bused:		
Black ratio increased	.050	4.269^*
Black ratio not increased	.021	
Transferred and not bused:		
Black ratio increased	.043	3.187^*
Black ratio not increased	.021	

TABLE 4 (continued)

b)

Condition	Probability	z
Transferred and black ratio increased:		
Bused	.050	$.996^{n.s.}$
Not - bused	.043	
Transferred and black ratio not increased:		
Bused	.021	$0.0^{n.s.}$
Not - bused	.021	

DISCUSSION

This paper has examined the effects of four desegregation plan features on the decision of parents to withdraw their children from desegregated schools. Examined bivariately, each of the four factors appeared to be related to rejection. Under more detailed controls, however, the effects of busing and distance from school were washed out. The factor most consistently associated with rejection was black ratio increase. Thus, it would appear that some measure of white withdrawal will take place whenever desegregation is implemented, since by definition desegregation means increasing the proportion of blacks attending schools with whites.

Beyond this, however, our findings suggest that busing can be employed by school decision makers to achieve satisfactory racial balances without fear of incurring additional rejection.

The tendency for rejection to increase with black ratio increases held independent of whether school reassignment occurred. However, the probability of rejection under the condition of increase in the black ratio was greater when transfer took place than when it did not. When the condition of black ratio increase was not present, parents of transferred children were no more likely to reject than parents of children who were not transferred. This suggests that school reassignment can be employed as a useful desegregation plan feature, provided that transfers do not take place simultaneously with increases in the black ratio. Of course, this would not be possible in the initial year of desegregation. It would be possible thereafter, but only if black ratios in all of a district's schools were approximately the same. Thus, a policy of equalizing racial balances throughout a school district receives empirical support from our study.

FOOTNOTES

[1]The research reported herein was conducted under a grant from the National Science Foundation, Division of Social Systems and Human Resources, GI-34955. The opinions expressed are those of the authors and should not be construed as representing the opinions or policy of any agency of the United States Government.

[2]The sample weights are based on the ratio of new rejecters to compliers for 1973-74 among the original group of compliers, all of whom were screened in January, 1974, to determine their most recent compliance/rejection status. The 1973-74 ratios were used to estimate the actual distributions of compliers to rejecters for 1972-73. Since rejecters were over-sampled, the N for compliers was weighted upwards, accordingly.

[3]Micheal W. Giles, Douglas S. Gatlin, and Everett F. Cataldo, "The Impact of Busing on White Flight," Social Science Quarterly, forthcoming; Micheal W. Giles, Everett F. Cataldo, and Douglas S. Gatlin, "White Flight and Percent Black: The Tipping Point Reexamined," Florida Atlantic University, mimeo; Carl Hansen, Danger in Washington (West Nyack, N.Y.: Parker Publishing Company, 1968); Arthur L. Stinchcombe, Mary McDill, Dollie Walker, "Is There a Racial Tipping Point in Changing Schools?" Journal of Social Issues (25,2) 1969, pp. 127-34; Nicolaus Mills, ed., The Great School Bus Controversy (New York: Teachers College Press, 1973); U.S. Commission on Civil Rights, Racial Isolation in the Public Schools (Washington, D.C.: U.S. Government Printing Office, 1967).

[4]Testing criterion z is presented in John H. Mueller, Karl F. Schuessier, and Herbert L. Costner, Statistical Reasoning in Sociology 2nd ed. (Boston: Houghton Mifflin, 1970) pp. 429-431.

Where $z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

COLLEGE STUDENT MOBILITY IN THE UNITED STATES IN FALL 1968: GRAPH-THEORETIC STRUCTURING OF TRANSACTION FLOWS

Paul B. Slater, West Virginia University

1. Introduction

In 1960 I. Richard Savage, a statistician, and Karl W. Deutsch, a political scientist, published a study of the 1928 trade flows among North Atlantic nations (Savage and Deutsch, 1960). They employed an origin-destination independence model to control for the substantial differences among countries in the sizes of their total imports and exports. Deviations -- measured by relative acceptance coefficients -- from this indifference or null model were interpreted as indicators of the magnitudes of trade linkages. It was suggested that these values be employed as dependent variables in additional analyses which would seek to explain variation in trading ties on the basis of distance, political relations, etc. Subsequently, Leo A. Goodman, a statistician and sociologist, proposed that the model should be estimated in such a manner that zero values would be assigned to flows among nations with no trading relations (Goodman, 1963, 1964). Due to the difficulties of obtaining data on trade within countries and developing a theory to explain such commerce, Savage and Deutsch had themselves nullified the diagonals of their flow table and indifference model.

A wide variety of transaction flow data -- for example, diplomatic and tourist exchanges, international mail deliveries, and newspaper and periodical circulations -- have been analyzed by these methods. Most of the applications and discussions of the null model have been reported in the political science literature (Lijphart, 1964; Puchala, 1970; Clark and Welch, 1972; Hughes, 1972; Chadwick and Deutsch, 1973; Clark, 1973; Mann and Chadwick, 1973; Seligson, 1973), while a few have appeared in geographical sources (Soja, 1968; Smith, 1970a, b; Williams and Zelinsky, 1970; Britton, 1971; Bradford, 1973).

The major innovation and methodological development in this area since the work of Savage, Deutsch, and Goodman has been accomplished by a political scientist, Steven J. Brams (1966a, 1966b, 1968a, 1968b, 1969). He used graph theoretic algorithms to cluster and hierarchically structure political units on the basis of salient linkages; that is, relative acceptance coefficients greater than some specified threshold level. [Nystuen and Dacey (1961), Filani (1972), Rouget (1972), Campbell (1972, 1974), and Taliaferro and Remmers (1973) also employed graph theory to study transaction flows. They did not, however, focus on deviations from an origin-destination independence model as Slater (1974) did in studying interindustrial flow tables. Fisher (1969) employed a "lockstep progressive merger procedure", while Ghosh (1970) and Roy (1971) minimized a contiguity index in structuring interindustry tables.]

A 51 x 51 matrix classifying the entire Fall 1968 American college student population by state (and District of Columbia) of residence and enrollment is the subject of the transaction flow analysis reported here. These data were compiled through the intensive efforts of the National Center for Educational Statistics (Wade, 1970, 1971). 2,445 of the 2,495 institutions of higher learning to which questionnaires were sent responded. The residence-enrollment patterns of the 50 non-responding institutions were imputed on the basis of a 1963 survey. Information was collected and published on a more disaggregated basis (sex, type of college, and degree program) than that studied here. Data for individual institutions have not been released. Post-1968 surveys have not been conducted and are not, at present, planned due to both the expense involved and recent legal decisions which have allowed many students to become in-state residents and thus avoid discriminatory fees levied against nonresident students.

The approach of Brams is paralleled to a considerable extent in this investigation of these flow data. However, in addition to an origin-destination independence model, models of symmetry and quasi-symmetry are also estimated. (Non-diagonal cell estimates in these two models are independent of the omission or inclusion of the diagonal entries.) Departures from these two models reflect imbalances of flows between pairs of states. The quasi-symmetry analysis controls for the fact that total outflows from states are unequal to total inflows, while the symmetry study does not. Residuals -- not relative acceptance coefficients -- are utilized in all three analyses, as measures of deviation, due to their greater resistance to fluctuations in the ratios of small observed to expected values. Questions concerning the distribution of the sizes of "consignments" do not arise here, as they can in trade flow studies (Savage and Deutsch, 1960; Goodman, 1963). Unlike each monetary unit expended in trade, each individual student is an essentially independent unit.

2. A Quasi-Independence Model

The multiplicative model

$$x_{ij} = r_i c_j, \tag{1}$$

where x_{ij} is the number of Fall 1968 American students resident in state i enrolled in all institutions of higher learning in state j was estimated. The diagonal entries, x_{ii}, are typically omitted from transaction flow analyses. Their values are often unavailable or, if available, of dominant magnitudes that strongly perturb estimation procedures. Since a line (loop) from a point to itself is not an admitted concept in the theory of directed graphs, the diagonal entries are, in any case, irrelevant in directed graph (digraph) studies. Iterative techniques which are necessary for the estimation of (1), if any of the ij-cells are disregarded (Goodman, 1968), were employed in this study due to the omission of the diagonal cells. (The extents to which students attend colleges in their states of residence are not, therefore, revealed in the results presented here.) (1) is, in such cases, referred to as a quasi-independence model.

One of several asymptotically equivalent formulas for the residuals from this model

$$(x_{ij} - r_i c_j)/(r_i c_j)^{\frac{1}{2}} \tag{2}$$

is employed as a measure of deviation, not relative acceptance coefficients

$$(x_{ij} - r_i c_j)/r_i c_j \tag{3}$$

nor absolute deviations

$$x_{ij} - r_i c_j \tag{4}$$

[The solution of a set of 102 (51 + 51) normal equations could have been used to adjust the standardized residuals (2) for their asymptotic variances (Haberman, 1973). However, considerable cost would have been involved in computing the adjustments. In addition, modifications would have been small, since no state had more than thirteen percent of all the Fall 1968 out-of-state students resident or enrolled in it. For exploratory purposes, therefore, the standardized residuals appeared adequate.]

Residuals incorporate a square root scale. They are, consequently, in a sense intermediate between relative acceptance coefficients (logarithmic scale) and absolute deviations (original scale). Brams could conceivably have employed residuals in his studies. He expressed dissatisfaction with the exclusive usage of either relative acceptance coefficients or absolute deviations, since the former were biased against nations with large trade flows, while the latter, on the other hand, gave undue importance to relatively small changes in large trade flows. Therefore, his standards for classifying linkages as salient required that both these measures had to be greater than certain levels (Brams, 1966a).

There are 2,550 (51 x 51 - 51) residuals from the quasi-independence model (1). A salient linkage criterion which sets all residuals greater than some chosen threshold to 1 and all other residuals to 0 can, of course, be established in many ways. It would be of interest to vary thresholds and observe how interrelationships revealed by the application of graph-theoretic procedures are altered (Gotlieb and Kumar, 1968; Doreian, 1969; Auguston and Minker, 1970; Osteen, 1974). The stability of relationships could be evaluated by doing so, as in clustering by fuzzy sets (Zadeh, 1965).

After preliminary investigation of the matrix of residuals, it was decided to set the 150 largest residuals to 1 and the remaining 2,400 to 0. The 151st largest residual, 46.3, was found with the use of a partial sorting algorithm (Chambers, 1971). If only a few order statistics are required, this method yields the desired information at less expense than if complete sorts are performed. This choice was made to strike a balance between richness of detail, significance of conclusions, and ease of computation and presentation. The results obtained are displayed in Figure 1. (The states of Alaska and Hawaii are represented by circles. Readers unfamiliar with the names and locations of the states should refer to a map which yields such information.) If the ij-entry of the Boolean adjacency matrix obtained by use of 46.3 as a threshold is 1, an arrow is directed from state i to state j. This indicates that state i sent substantially more students to state j than would be expected if the null model held. There is, of course, extensive departure from (1) because factors such as distance, tuition differentials, and quality of educational facilities are not accounted for in it. A summary measure of the degree to which it fails to hold is given by the percentage discrepancy statistic

$$(\sum_{ij}\sum 50|x_{ij} - r_i c_j|)/\sum\sum x_{ij} \text{ for all } x_{ij},\ i \neq j \tag{5}$$

The numerical value of this statistic is the percentage of the 1,104,622 students attending out-of-state colleges in the Fall of 1968 that would have had to have been redistributed in order to achieve quasi-independence (Brams, 1966a). It was computed to be 40.7. Since the maximum possible value is 50, state of residence and state of enrollment are strongly associated. Most of the 150 salient linkages are, in fact, between contiguous pairs of states, while the remainder occur between relatively proximate pairs. (It should be realized that there are certainly substantial within-state effects present in the data. Residents of one section of a state might have different preferences for out-of-state colleges than those of another.)

By exponentiating a 51 x 51 Boolean matrix with 201 non-zero entries (150 for the saliences plus 51 on the main diagonal), the reachability matrix of the digraph formed by the 150 links can be obtained (Harary et al., 1965). [The power to which it is necessary to raise the matrix is termed the diameter of the graph. Algorithms for the multiplication of sparse Boolean matrices were employed for this purpose (IBM, 1970). These were much more efficient than conventional matrix multiplication routines would have been. Graph algorithmic programming languages that are also of value in computing reachability matrices have been developed (Rheinboldt et al., 1972).] One state (A) is reachable from another (B) if a sequence of directed links exists from B to A. Strong components -- which may be interpreted as regions -- of the graph can be easily determined from the reachability matrix. These are maximal sets, any member of which is reachable from any other.

The graph formed by the 150 salient linkages was found to possess five strong components. Two of them were sets consisting of single states -- Alaska and Montana. The other three were non-trivial. One was a western region composed of Arizona, California, Hawaii, Idaho, Nevada, Oregon, Utah, and Washington. Another was a northeastern region comprised of Connecticut, Delaware, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The fifth region was a large diffuse one comprised of the remaining thirty (midwestern and southeastern) states plus the District of Columbia. Virginia, a member of this region, was reachable from another member, Colorado, in a minimal path of nine links. (The diameter of the graph was thirteen.)

If the threshold for saliences were raised, more regions (strong components) with fewer states would result. By lowering the threshold, on the other hand, coarser regional distinctions would appear. When, during preliminary investigations, the largest 330 residuals were deemed salient, every state was reachable from every other one. Therefore, there was only one strong component or region, the entire nation.

One of the most interesting aspects of regionalizing in this fashion is isolating the links, if any, between the strong components. Since an acyclic digraph is obtained by condensing a digraph with respect to its strong components, no two strong components can be mutually salient. For example, Montana and Alaska -- both strong components -- are each salient to the same member of the western region -- Washington. However, no member of the western region is salient to either of these two states, thus preventing their incorporation into it. Wyoming, a constituent of the largest region, is salient to its neighboring states of Montana and Utah, but neither Montana nor any western state is salient to Wyoming. Pennsylvania, which is in the northeastern region, is salient to two states that it borders, Ohio and West Virginia. However, the only states salient to Pennsylvania are other northeastern ones -- Delaware, New Jersey, and New York.

Strong components may contain cliques; that is, maximal complete subgraphs. Every member of a clique is salient to every other member. There are five three-state cliques in the 150-link digraph. These are {Idaho, Oregon, Washington}, {California, Oregon, Washington}, {Connecticut, Massachusetts, New York}, {Alabama, Mississippi, Tennessee}, and {Alabama, Georgia, Tennessee}. Clearly, cliques need not be mutually exclusive, in the manner of strong components. [Gotlieb and Kumar (1968) combine overlapping cliques to obtain "diffuse" classes.] Therefore, the latter are more appropriate for regionalization purposes. [Taliaferro (1973) regionalized the counties of Kansas on the basis of weak components; that is, maximal sets of counties all mutually reachable by sequences of undirected links. The weak components of a digraph are the strong components of the digraph's symmetric closure (Harary et al., 1965; Harary and Miller, 1970).]

3. A Symmetry Model

A quasi-independence model can be estimated for any rectangular table of counts. The study of square tables of counts -- where row and column classifications coincide -- is richer, however, in that models of symmetry and quasi-symmetry can also be estimated (Ireland et al., 1969). Under an hypothesis of symmetry, the estimate of the ij or ji cell of a square table is simply

$$(x_{ij} + x_{ji}) / 2 \tag{6}$$

These values were computed for the Fall 1968 student flow table. The associated 2,550 residuals were then calculated, and the 150 largest were again deemed salient. (The 151st largest residual was 12.1. 21.8 percent of the students would have had to have been redistributed for symmetry to hold.) These saliences are exhibited in Figure 2. An arrow directed from state i to state j indicates that state i sent substantially more students to state j than it received from state j. For several reasons, however, Figure 2 would be unacceptably complex if arrows were used to represent all 150 saliences. One reason is that contiguity is not as prominent a factor as in Figure 1, while another is that there can be no two-headed arrows in Figure 2. Also, a very high density of emitted saliences was found in the neighboring states of Connecticut, New Jersey, New York, and Pennsylvania. Therefore, symbols are used to represent most of the saliences directed from these four states.

In addition to transmitting to 32 states plus the District of Columbia, the matrix of residuals revealed that all of the other fifty units received more students from New Jersey than it sent there. Even to New York, which itself transmitted to 29 units (including Pennsylvania, which transmitted to 19), New Jersey sent 24,534 students, while receiving only 6,689 in return. Overall, 16,831 out-of-state students were enrolled in New Jersey's institutions of higher education, while 116,536 New Jerseyites were attending colleges outside their home state.

Another interesting aspect of Figure 2 is the large number of saliences incident to Utah. This phenomenon is probably in large part attributable to Mormon students enrolling in Brigham Young University. The catalog for this school, like many others, lists the states and foreign areas from which the students come. As of August 18, 1969, 9,513 of 27,277 students were from Utah. 16,348 were from the other 49 states plus the District of Columbia. In the eastern portion of the nation, Tennessee has a dominant role similar to Utah's in the west. Tennessee and Utah -- in strong contrast to New Jersey -- have large net inflows of students. (Tennessee received 31,763 and sent 12,973, while the comparable data for Utah were 19,108 and 3,047.) The quasi-symmetry model controls for such differences between inflows and outflows.

Several of the one-way linkages under the quasi-independence model (Figure 1) are shown in Figure 2 to also be strongly asymmetrical. (For example, Wyoming → Utah; Montana → Washington; Florida → Tennessee.) Also, some two-way linkages exhibit the same property. (Illinois → Indiana; Florida → Georgia; Idaho → Utah, for instance.) As an illustration of the magnitude of the 150 asymmetries, a few -- not selected on any systematic basis -- are listed.

Ill.	sent	12,052	to	Wisc.	and	recd.	2,428.
Iowa	"	2,428	"	Neb.	"	"	725.
Maryland	"	1,903	"	N. C.	"	"	193.
Pa.	"	6,080	"	W. Va.	"	"	428.
New York	"	1,083	"	Arizona	"	"	109.
Florida	"	997	"	Miss.	"	"	154.

There are no cycles (intransitivities) in Figure 2. If one unit can be reached by a series of directed links from another unit, there is, therefore, no return path. Hence, there are no non-trivial strong components. The points in such an acyclic digraph can be hierarchically ordered so that a point is assigned to a kth level if, and only if, the longest path to the point is of length k (Brams, 1968a, b, 1969; Rouget, 1972). A point at any level of the hierarchy is either salient to a point at the next higher level or has a salience from a point at the immediately preceding level. Table 1 presents the hierarchical arrangement -- based on seven levels -- for the digraph of Figure 2. (Since Arkansas and North Dakota are isolated, that is, have no saliences associated with them, they are omitted from the table.) No state is salient to New Jersey (6), while Maine (0), for instance, is not salient to any other state. Maryland (3) is salient to Ohio (2), which in turn is salient to Kentucky (1). Two states, not at adjacent levels, can be salient. However, the salience must be directed from the higher level to the lower level.

4. A Quasi-Symmetry Model

The marginal (row and column) sums of the cell estimates in the quasi-independence model (1) and the sums of symmetrically located pairs of cell estimates in the symmetry model (6) are identical to the corresponding sums in the flow table. By iteratively constraining a 51 x 51 table, the entries of which are all initially equal, to possess first one and then the other of these two properties, convergence to estimates under an hypothesis of quasi-symmetry can be obtained. Cell estimates in such a model are the closest possible -- in the sense of minimum discrimination information -- to being symmetric, given that the marginal sums are inhomogeneous; that is, the sums of rows and columns corresponding to the same classifications are unequal (Ireland et al., 1969). The 150 largest residuals from the quasi-symmetry model are presented in Figure 3. The 151st largest residual was 6.9. 11.5 percent of the students would have had to have been redistributed for quasi-symmetry to hold. Since symmetry is equivalent to quasi-symmetry plus marginal homogeneity, the difference, 10.3, of the two percentage discrepancy statistics, 21.8 and 11.5, is the percentage of departure from (6) attributable to marginal inhomogeneities.

New Jersey is salient to nine states in Figure 3, a considerable decline from the 33 saliences in Figure 2. By controlling for the large disparity between the number of students leaving New Jersey and the number entering, most of the asymmetries under (6) were diminished. Similarly, the saliences associated with Connecticut were reduced from 11 to 4. (20,884 out-of-state residents were enrolled in Connecticut's colleges, while 42,965 Connecticut residents attended college outside the state.) Interestingly, Virginia, also a substantial net exporter of students, has a salience from Connecticut under the hypothesis of quasi-symmetry, but not symmetry.

On the other hand, the number of saliences associated with Pennsylvania increased from 19 to 22. (Pennsylvania was a relatively weak net exporter of students, receiving 62,302 and sending 76,420.) New York's saliences decreased in number from 29 to 27. (New York received 69,424 students and sent 135,981.) The number of saliences from Ohio increased from 6 to 11. (Ohio was a net importer, receiving 57,428 and sending 50,215.) Florida is shown in Figure 3 to be a strong attracter of students from Northeastern states, possibly for climatic reasons. (Climatic differences may also explain Minnesota's salience to California.) California and Hawaii have exceptional appeal for Florida residents. The number of saliences to Utah decreased, while those to Tennessee and Texas increased.

Perhaps the most surprising salience revealed in Figure 3 is that of Oregon to Pennsylvania. Its presence results in the digraph associated with Figure 3 being cyclic, in contrast to the acyclic digraph corresponding to Figure 2. (For instance, Pennsylvania → California, California → Oregon, Oregon → Pennsylvania. This is not, however, the maximal cycle containing the Oregon → Pennsylvania link.) In the Fall of 1968, Pennsylvania sent 82 students to Oregon, while Oregon sent 225 to Pennsylvania. Oregon received roughly a quarter more students than it sent, while Pennsylvania sent approximately a quarter more than it accepted from out-of-state.

5. Conclusions

Substantial departures from quasi-independence, symmetry, and quasi-symmetry models of the Fall 1968 state-to-state table of college students occur. By concentrating on the largest residuals from these models, with the use of graph theoretic procedures, dominant patterns of deviations can be perceived. The procedures employed are of general value in analyzing the rich body of transaction flow data.

ACKNOWLEDGEMENT

The data analyzed were brought to the author's attention by William H. Miernyk.

REFERENCES

Auguston, J. G. and Minker, J., 1970, "An Analysis of Some Graph Theoretical Cluster Techniques," Journal of the Association for Computing Machinery, 17, 571-588.

Bradford, B. C., 1973, "On the Structure of Soviet Urban Migration," Geography Department, University of Wisconsin, Madison, Wisconsin, U.S.A.

Brams, S. J., 1966a, "Transaction Flows in the International System," American Political Science Review, 60,880-898.

Brams, S. J., 1966b, "Trade in the North Atlantic Area: An Approach to the Analysis of Transformations in a System," Peace Research Society Papers, 6, 143-164.

Brams, S. J., 1968a, "Measuring the Concentration of Power in Political Systems," American Political Science Review, 62, 461-475.

Brams, S. J., 1968b, "DECOMP: A Computer Program for the Condensation of a Directed Graph and the Hierarchical Ordering of its Strong Components," Behavioral Science, 13, 344-345.

Britton, J. N. H., 1971, "Methodology in Flow Analysis," East Lakes Geographer, 7, 22-36.

Campbell, J., 1972, "Growth Pole Theory, Digraph Analysis and Interindustry Relationships," Tijdschrift voor Economische en Sociale Geographie, 63, 79-87.

Campbell, J., 1974, "Selected Aspects of the Interindustry Structure of the State of Washington, 1967," Economic Geography, 50, 35-46.

Chadwick, R. W. and Deutsch, K. W., 1973, "International Trade and Economic Integration: Further Developments in Trade Matrix Analysis," Comparative Political Studies, 6, 84-110.

Chambers, J. M., 1971, "Algorithm 410 Partial Sorting (ML)," Communications of the Association for Computing Machinery, 14, 357-358.

Clark, C., 1973, "The Impact of Size on the Savage-Deutsch RA Statistic: A Research Note," Comparative Political Studies, 6, 109-122.

Doreian, P., 1969, "A Note on the Detection of Cliques in Valued Graphs," Sociometry, 32, 237-242.

Filani, M. O., 1972, "Changing Patterns of Central Places and Functional Regions: Temporal and Spatial Dynamics of Air Traffic Flows in the United States," Ph. D. Thesis, Pennsylvania State University, State College, Pennsylvania.

Fisher, W. D., 1969, Clustering and Aggregation in Economics (Baltimore: Johns Hopkins Press).

Ghosh, A. and Sakar, H., 1970, "An Input-Output Matrix as a Spatial Configuration," Economics of Planning, 10, 133-142.

Goodman, L. A., 1963, "Statistical Methods for the Preliminary Analysis of Transaction Flows," Econometrica, 31, 197-208.

Goodman, L. A., 1964, "A Short Computer Program for the Analysis of Transaction Flows," Behavioral Science, 9, 176-186.

Goodman, L. A., 1968, "The Analysis of Cross-Classified Data: Independence, Quasi-Independence and Interactions in Contingency Tables With or Without Missing Values," Journal of the American Statistical Association, 63, 1091-1131.

Gotlieb, C. C. and Kumar, S., 1968, "Semantic Clustering of Index Terms," Journal of the Association for Computing Machinery, 15, 493-513.

Haberman, S. J., 1973, "The Analysis of Residuals in Cross-Classified Tables," Biometrics, 29, 205-220.

Harary, F., Norman, R. Z., and Cartwright, D., 1965, Structural Models: An Introduction to the Theory of Directed Graphs (John Wiley: New York).

Harary, F. and Miller, H., 1970, "A Graph-Theoretic Approach to the Analysis of International Relations," Journal of Conflict Resolution, 14, 57-63.

Hughes, B. B., 1972, "Transaction Data and Analysis: In Search of Concepts," International Organization, 26, 659-680.

IBM Corporation, 1968, System 360 Matrix Language (MATLAN), Manual H 20-0479-1.

Ireland, C. T., Ku, H. H., and Kullback, S., 1969, "Symmetry and Marginal Homogeneity of an r x r Contingency Table," Journal of the American Statistical Association, 64, 1323-1341.

Lijphart, A., 1964, "Tourist Traffic and Integration Potential," Journal of Common Market Studies, 2, 251-262.

Mann, C. A. and Chadwick, R. W., 1973, "Ranull-Transaction Flow Analysis," Behavioral Science, 18, 72-73.

Nystuen, J. D. and Dacey, M. F., 1961, "A Graph Theory Interpretation of Nodal Regions," Papers and Proceedings of the Regional Science Association, 7, 29-42.

Osteen, R. E., 1974, "Clique Detection Algorithms Based on Line Addition and Line Removal," SIAM Journal of Applied Mathematics, 26, 126-135.

Puchala, D., 1970, "International Transactions and Regional Integration," International Organization, 24, 732-763.

Rheinboldt, W., Basili, V., and Meszteniji, C., 1972, "On A Programming Language for Graph Algorithms," BIT, 12, 220-241.

Rouget, B., 1972, "Graph Theory and Hierarchisation Models," Regional and Urban Economics, 2, 263-296.

Roy, M. K., 1971, "A Note on the Computation of an Optimal Ordering for an Input-Output Matrix," Economics of Planning, 11, 95-97.

Savage, I. R. and Deutsch, K. W., 1960, "A Statistical Model of the Gross Analysis of Transaction Flows," Econometrica, 28, 551-572.

Seligson, M. A., 1973, "Transactions and Community Formation: Fifteen Years of Growth and Stagnation in Central America," Journal of Common Market Studies, 11, 173-190.

Slater, P. B., 1974, "Exploratory Analyses of Interindustrial Transaction Tables," Environment and Planning, 6A (1974), 207-214.

Smith, R. H. T., 1970a, "Concepts and Methods in Commodity Flow Analysis," Economic Geography, 46, 404-416.

Smith, R. H. T., 1970b, "Inter-regional Trade as a Component of National Unity," in Selected Papers, 21st International Geographical Congress, Vol. 2, Ed. S. P. Chattenjee (National Committee for Geography, Calcutta), pp. 285-293.

Soja, E. W., 1968, "Communications and Territorial Integration in East Africa: An Introduction to Transaction Flow Analysis," East Lakes Geographer, 4, 39-57.

Taliaferro, J. D. and Remmers, W. W., 1973, "Identifying Integrated Regions for Health Care Delivery," Health Services, 18, 337-343.

Wade, G. H., 1970, Residence and Migration of College Students, Fall 1968: Basic State-to-State Matrix Tables (National Center for Educational Statistics, Washington).

Wade, G. H., 1971, Residence and Migration of College Students (National Center for Educational Statistics, Washington).

Williams, A. V. and Zelinsky, W., 1970, "On Some Patterns in International Tourist Flows," Economic Geography, 46, 549-567.

Zadeh, L. A., 1965, "Fuzzy Sets," Information and Control, 8, 338-353.

TABLE 1

Hierarchical ordering of states based on 150 largest residuals from symmetry model

Level	States
6	New Jersey
5	Connecticut, New York
4	Pennsylvania
3	Florida, Maryland
2	Alaska, Hawaii, Illinois, Montana, Ohio, South Carolina
1	Arizona, California, Delaware, Georgia, Idaho, Iowa, Kentucky, Nevada, New Mexico, Virginia, Washington, Wyoming
0	Alabama, Colorado, District of Columbia, Indiana, Kansas, Louisiana, Maine, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Nebraska, New Hampshire, North Carolina, Oklahoma, Oregon, Rhode Island, South Dakota, Tennessee, Texas, Utah, Vermont, West Virginia, Wisconsin

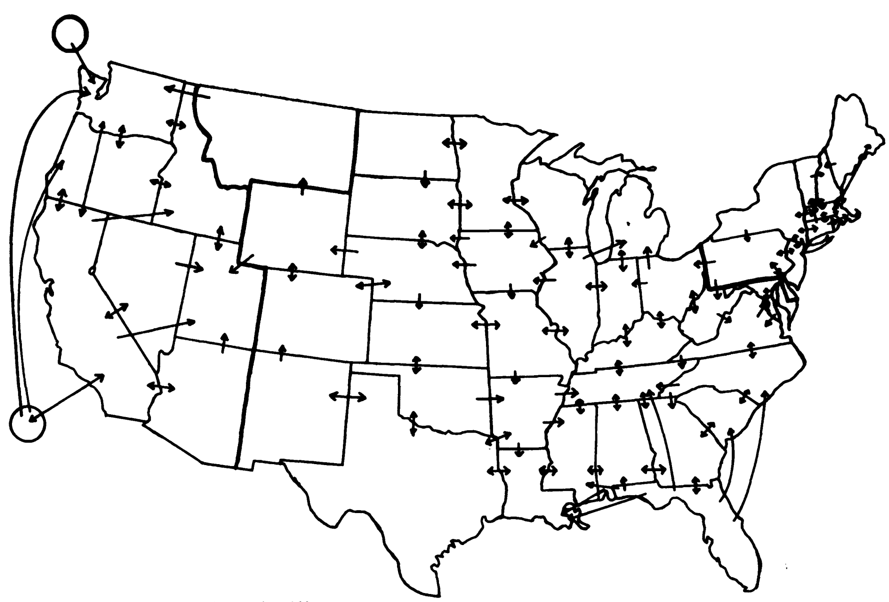


Figure 1. 150 largest residuals from quasi-independence model, and regionalization based upon strong components.

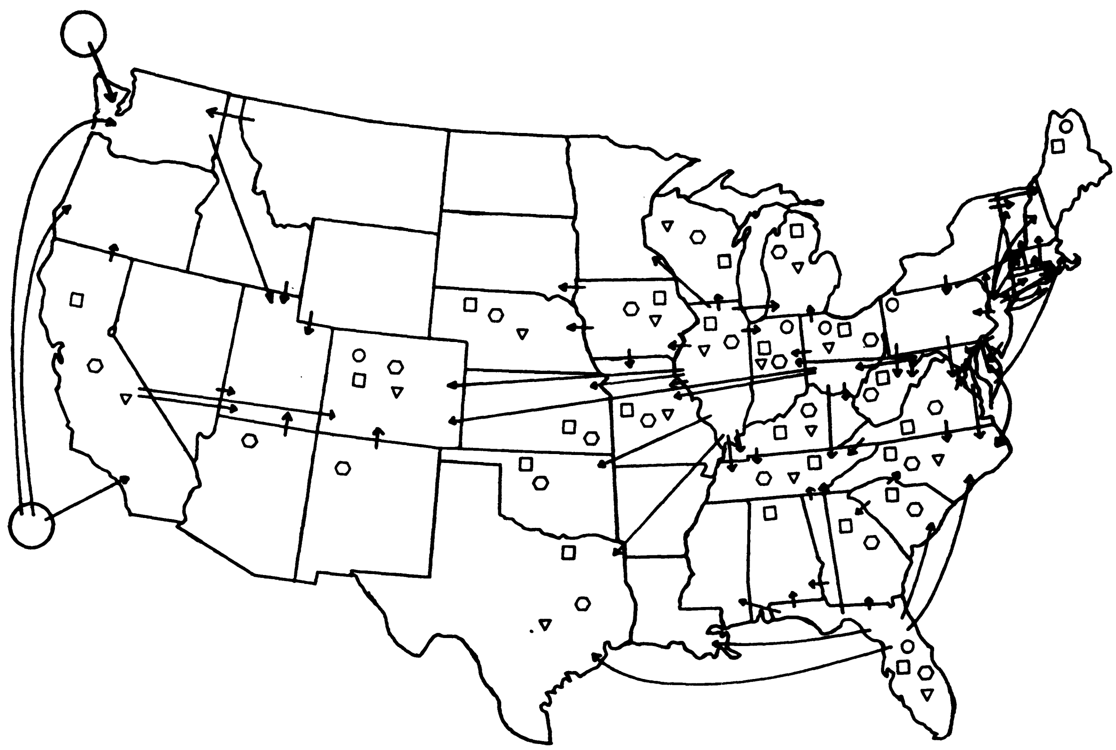


Figure 2. 150 largest residuals from symmetry model.

O : ← Connecticut
□ : ← New Jersey
⬡ : ← New York
△ : ← Pennsylvania
District of Columbia ← {Connecticut, New Jersey, New York, Pennsylvania}

Figure 3. 150 largest residuals from quasi-symmetry model.

O : ← Connecticut
□ : ← New Jersey
⬡ : ← New York
△ : ← Pennsylvania
▷ : ← Ohio
District of Columbia → {Illinois, North Carolina, Ohio}

A COMPARISON OF THREE TECHNIQUES FOR ELICITING TRUTHFUL ANSWERS TO SENSITIVE QUESTIONS

Linda L. Smith, Walter T. Federer and Damaraju Raghavarao[1], Cornell University

1. Introduction

The problem of attempting to obtain truthful answers to sensitive, incriminating, or embarrassing questions has received increased attention since the appearance of Warner's [7] 1965 publication describing the randomized response (RR) technique. Some of this work may be found in the literature [1,2,3,4,8] cited at the end of this paper. An alternative to the RR method, utilizing balanced incomplete block design concepts, was presented by Raghavarao and Federer [5]; their method is designated as the block total response (BTR) procedure. A randomized form of the block total response (RBTR) technique is described by Smith [6]. She compared the three techniques in a designed experiment with v=7 questions in b=7 incomplete blocks of size k=3 questions each on a set of 7n=84 individuals enrolled in a statistics class. This paper includes the results obtained by Smith [6].

Briefly, the RR technique utilized here consisted of presenting the respondent with two questions, A a sensitive one or one of interest and B a nonsensitive one whose answer was known. The randomizing device used was such that 70% of the time the respondent, anonymously, answered question A and 30% of the time he answered question B; an affirmative answer to question B should occur 2/3 of the time. Each respondent was presented k=3 questions with a different randomizing device and a different nonsensitive question being used for each of the three questions presented.

The BTR method consists of using a balanced incomplete block design with parameters v=b, k=r, and λ. The respondent gives a total score for the particular set of k questions presented to him; responses to individual questions are not required and thus, under proper scoring of questions, the answers to individual questions are unknown to anyone but the respondent. From the set of b blocks it is possible to estimate average response to each question but not to obtain an individual's response. For this study a balanced incomplete block design with v=7=b, k=3=r, and λ=1 was utilized.

The RBTR procedure consists of randomly and anonymously allotting a block of k questions to the respondent. The respondent draws a cork from a jar containing corks numbered 1 to b, such that each number occurs an equal number of times. He keeps the cork and writes down a total for the block number on his cork. He then folds the questionnaire and places it in a sealed (or locked) box containing other respondents' answers. For the first respondent a few blank sheets could be inserted in the box to give the respondent a feeling of security regarding his anonymity. For this particular study, only seven corks were included in a jar at one time. This was done because the interviewing was done with groups of six to 14 students. Thus, in the RBTR procedure the interviewer does not know which block of questions was answered by the respondent and in addition the respondent's answers to individual questions are unknown. This double degree of security may be necessary in certain cases. The idea of using a locked box in which respondents placed their answers was described by Arnold J. King to one of the authors (WTF) and was used by King in a survey on income of Iowa farmers. The farmers had to be assured that the information they gave would not end up in the hands of the Internal Revenue Service.

In general survey practice with the RBTR technique, a plastic jar with n corks numbered 1, n corks numbered 2, n corks numbered 3, ..., n corks numbered 7, could be used. Then when one gets down to the last few corks, a check could be made to determine which numbers remained. If only one or two numbers remain, several corks with the same number could be added to give a sense of anonymity to the respondee. It was found that students felt much more secure about their anonymity with procedure III than with the other two procedures. This appears to show up in their answers to questions 2 and 4, where the percentage obtained by procedure III was much higher than for procedures I and II.

2. Procedure

It was decided to use 84 individuals in the study, which, as determined by the results, was too small a sample. The individual respondents were from an introductory statistics class and were interviewed in eight groups, ranging in size from six to 14 students. Since a comparison of the three techniques was being made, the question of representativeness of the sample is not involved. A copy of a four-page questionnaire presented to one student follows:

-page 1-

Introduction: This survey is designed for the purpose of comparing three methods of obtaining truthful answers to questions that an individual might feel are embarrassing, incriminating, or both. All three methods are aimed at keeping the individual's answers completely anonymous. We are not interested in how particular individuals answer specific questions, but rather in estimating the number of individuals responding in a certain way (either yes or no).

Results of this survey will be summarized and analyzed and then presented to you at a later date.

After finishing this survey, some of you may want to construct a similar survey using one or more of these three methods and your own questions for your term project.

When you have completed all three sections, please write down any comments you have concerning the three procedures. Did you feel as though your answers were really anonymous? Why or why not? Write comments on back of this page.

-page 2-

I. Randomized Response Procedure

In this method you are to answer one of two questions according to the procedure indicated. You give only one answer in each pair of questions, and because of the way this method is designed the interviewer cannot tell which question <u>you</u> have answered. The interviewer knows only the proportion of times the question is answered.

1. Answer question A if the last digit of your social security number is 0,1,2,3,4,5 or 6. Answer question B if the last digit is 7,8 or 9.

Question A: Does your parent earn more than $25,000 a year?

Question B: Is your number 7 or 8 (as opposed to 9)?

Answer: Yes_____ No_____

2. Answer question A if your birthday falls between January 1 and September 12. Answer B if your birthday falls between September 13 and December 31.

Question A: Have you smoked any marijuana during the past two weeks?

Question B: Is your birthday between September 13 and November 25?

Answer: Yes_____ No_____

3. Draw a cork and note its color, and if it is white, also note its number. Answer question A if it is red and question B if it is white.

Question A: Are you under 21 years of age?

Question B: Was the number on your cork 1?

Answer: Yes_____ No_____

-page 3-

II. Block Total Response Procedure

For this procedure you are asked to give only a total of your answers to the three questions. Please do not, unless you wish to, indicate what you answer for individual questions, but rather a total for all three. Because you give just the total, the interviewer cannot tell how you have answered individual questions. For example if your total was 4 and if you had been asked questions 1, 2, and 3, you could have answered in any of the following ways:

1.	0	1	1.	0	1	1.	0	1
2.	2	3	2.	2	3	2.	2	3
3.	1	0	3.	1	0	3.	1	0
	=	4		=	4		=	4

	Possible Answers	
3. In general, are you happy with your decision to come to Cornell?	Yes (1)	No (0)
4. While at Cornell, have you ever stolen money or any other article worth more than $5.00 from a roommate, friend, employer, or anyone else?	Yes (3)	No (2)
6. Have you smoked any marijuana during the past two weeks?	Yes (2)	No (3)

Total for the three questions: _____

-page 4-

III. Randomized form of Block Total Response

Without looking, select a cork from the jar. Observe the number on the cork. This number will correspond to a certain set of questions that you will answer. The seven questions are:

1. Are you under 21 years of age?
   Yes (0) No (1)
2. Did you cheat in any way on the Stat 200 prelim that you took last week?
   Yes (2) No (3)
3. In general, are you happy with your decision to come to Cornell?
   Yes (1) No (0)
4. While at Cornell, have you ever stolen money or any other article worth over $5.00 from a roommate, friend, employer, or anyone else?
   Yes (3) No (2)
5. Does your parent earn more than $25,000 a year?
   Yes (0) No (1)
6. Have you smoked any marijuana during the past two weeks?
   Yes (2) No (3)
7. Are you enrolled in the College of Agriculture and Life Sciences?
   Yes (1) No (0)

Do not give answers to individual questions but rather a total of your three answers.

If the number on your cork was 1, give a total for questions 1,2,4 Answer_____

If number was 2, give total for 2,3,5 Answer_____

If number was 3, give total for 3,4,6 Answer_____

If number was 4, give total for 4,5,7 Answer_____

If number was 5, give total for 5,6,1 Answer_____

If number was 6, give total for 6,7,2 Answer_____

If number was 7, give total for 7,1,3 Answer_____

Now fold this page in half and place in the box provided. This is to insure that your answers remain anonymous.

From the above seven questions listed under III, it should be noted that questions 2 and 4 are highly sensitive ones; question 6 may or may not be sensitive; questions 3 and 5 may be sensitive for some individuals. Under the conditions of the study questions 1 and 7 should not be sensitive for any student. Note that all the scores are not zero and one. The scoring system used was such as to allow several ways in which at least some scores could be obtained. The problem of a good scoring procedure is a difficult and unsolved problem. Some block totals are obtainable in only one way which allows the interviewer to ascertain the respondent's answers to individual questions. Although the scoring procedure used is not perfect, it is a much better one than simply coding all yes answers zero and all no answers one. To improve the scoring procedure one could include a quantitative variable, for example age to nearest birthday, as one of the nonsensitive questions. Alternatively, one could include a nonsensitive question, whose answer is known, in every block.

The seven questions were grouped into v=7 subsets of k=3 questions each in a balanced incomplete block arrangement as follows:

$$\begin{array}{llll} T_1 = Q_5, Q_6, Q_1 & & T_5 = Q_6, Q_7, Q_2 \\ T_2 = Q_4, Q_5, Q_7 & & T_6 = Q_1, Q_2, Q_4 \\ T_3 = Q_3, Q_4, Q_6 & & T_7 = Q_2, Q_3, Q_5 \\ T_4 = Q_7, Q_1, Q_3 & & \end{array} \qquad (1)$$

The three procedures were coded as follows:

I = Randomized Response Procedure (RR)

II = Block Total Response Procedure (BTR)

III = Randomized Form of Block Total Response Procedure (RBTR)

Each person answered a set of questions using each of the three methods. Since the order in which the techniques were presented might somehow affect the result, the following set of six sequences of the techniques was used to obtain a balanced arrangement among the groups and order of presentation:

$$S_1 = \text{I, II, III} \qquad S_4 = \text{II, III, I}$$
$$S_2 = \text{I, III, II} \qquad S_5 = \text{III, I, II}$$
$$S_3 = \text{II, I, III} \qquad S_6 = \text{III, II, I}$$

Then for 42 respondees the following setup was used where seven students did S_1, another seven did S_2, etc.

Sets of Seven People in Each S_i

	1	2	3	4	5	6	7
S_1	T_7T_6	T_1T_7	T_2T_1	T_3T_2	T_4T_3	T_5T_4	T_6T_5
S_2	T_2T_3	T_3T_4	T_4T_5	T_5T_6	T_6T_7	T_7T_1	T_1T_2
S_3	T_3T_5	T_4T_6	T_5T_7	T_6T_1	T_7T_2	T_1T_3	T_2T_4
S_4	T_4T_7	T_5T_1	T_6T_2	T_7T_3	T_1T_4	T_2T_5	T_3T_6
S_5	T_5T_2	T_6T_3	T_7T_4	T_1T_5	T_2T_6	T_3T_7	T_4T_1
S_6	T_6T_4	T_7T_5	T_1T_6	T_2T_7	T_3T_1	T_4T_2	T_5T_3

where the first T in each pair corresponds to procedure I and the second T corresponds to procedure II. Since it was not known which T_i an individual would select in procedure III, it was not possible to balance procedure III with the other two. For the RR and BTR procedures the same set of T_h questions for both groups was excluded. A pair of orthogonal latin squares of order seven was used to construct the above design by simply excluding the row of the pair of latin squares which contained the T_hT_h sets of questions. Note that under procedure III, a student may have answered one of the T_i in the pair T_hT_j for procedures I and II.

After the first six blocks of seven had been completed by the 42 respondees, another 42 students followed the same setup to obtain the 84 responses. The experiment was conducted during a one week period in March 1974.

3. Summarized Data and Calculations

The total number of "yes" responses for each question and for each S_j (j=1,2,...,6) for the randomized response technique is:

Question	Number "Yes" I.1	I.2	Sequence	Number "Yes" I.1	I.2
1	12	13	1	10	10
2	3	5	2	8	13
3	15	13	3	9	7
4	3	4	4	8	13
5	4	6	5	11	12
6	6	7	6	9	7
7	12	15			

The total of the responses for each T_h from all 12 S_j for both sets of 42 students for the BTR (II) and the RBTR (III) procedures are:

Total	II	III
$Y_{1.}$	51	42
$Y_{2.}$	43	44
$Y_{3.}$	65	65
$Y_{4.}$	22	21
$Y_{5.}$	77	75
$Y_{6.}$	67	65
$Y_{7.}$	55	57

The calculations for the randomized response technique are:

P(yes answer) = P(Yes on 1st question) + P(Yes on 2nd question)

= P(1st question chosen) · P(Yes/1st question) + P(2nd question chosen) · P(Yes/2nd question)

Rearranging and substituting in the known values, we have

$$P(\text{Yes/1st question}) = \frac{P(\text{Yes answer}) - (.3)(\frac{2}{3})}{0.7} = \hat{X}_{h.},$$

where P(Yes answer) is the proportion of "Yes" responses for a particular question and P(Yes/2nd question) = 2/3 for all three nonsensitive questions used in the RR method.

The estimate for question one is obtained as $\hat{X}_1 = [(12+13)/3(12) - .20]/0.7 = 0.706$. Note that there are 3(12) = 36 individuals who answered each question using the RR method. The proportions for the remaining questions are given in Table 1.

For a given question h asked of a population of individuals it is assumed that there is a true mean $\bar{X}_h$, that there is an individual response X_{hi} for the ith individual, and that $X_{hi} - \bar{X}_h = e_{hi}$ represents a deviation of individual i from the population mean. Then $E[e_{hi}|h] = 0$ and if the ith and i'th individuals' responses are independent then we can say that the e_{hi} are identically and independently distributed with zero mean and common variance. The responses to the seven questions may be represented as:

$$\text{(2)} \qquad \begin{aligned} &X_{1i} = \bar{X}_1 + e_{1i}, \quad X_{2i} = \bar{X}_2 + e_{2i}, \quad X_{3i} = \bar{X}_3 + e_{3i} \\ &X_{4i} = \bar{X}_4 + e_{4i}, \quad X_{5i} = \bar{X}_5 + e_{5i}, \quad X_{6i} = \bar{X}_6 + e_{6i} \\ &X_{7i} = \bar{X}_7 + e_{7i} \end{aligned}$$

where $X_{1i}, X_{2i}, X_{3i}, \ldots, X_{7i}$ are answers to 1,2,3,...,7 respectively, $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_7$ are population means for questions 1,2,3,...,7, and e_{hi} is a deviation of an answer X_{hi} from the population mean $\bar{X}_h$. Let Y_{hj} be the total of the answers for the jth respondent answering the hth set of questions; then

$$\text{(3)} \qquad \begin{aligned} &Y_{1j} = X_{5j} + X_{6j} + X_{1j} \qquad Y_{5j} = X_{6j} + X_{7j} + X_{2j} \\ &Y_{2j} = X_{4j} + X_{5j} + X_{7j} \qquad Y_{6j} = X_{1j} + X_{2j} + X_{4j} \\ &Y_{3j} = X_{3j} + X_{4j} + X_{6j} \qquad Y_{7j} = X_{2j} + X_{3j} + X_{5j} \\ &Y_{4j} = X_{7j} + X_{1j} + X_{3j} \end{aligned}$$

Using (2) in (3), and omitting the e_{hi} terms we obtain estimates for individual questions as follows:

$$\hat{X}_1 = [Y_{1.} + Y_{4.} + Y_{6.} - (Y_{2.} + Y_{3.} + Y_{5.} + Y_{7.})/2]/3n$$
$$\hat{X}_2 = [Y_{5.} + Y_{6.} + Y_{7.} - (Y_{1.} + Y_{2.} + Y_{3.} + Y_{4.})/2]/3n$$
$$\hat{X}_3 = [Y_{3.} + Y_{4.} + Y_{7.} - (Y_{1.} + Y_{2.} + Y_{5.} + Y_{6.})/2]/3n$$
$$\hat{X}_4 = [Y_{2.} + Y_{3.} + Y_{6.} - (Y_{1.} + Y_{4.} + Y_{5.} + Y_{7.})/2]/3n$$
$$\hat{X}_5 = [Y_{1.} + Y_{2.} + Y_{7.} - (Y_{3.} + Y_{4.} + Y_{5.} + Y_{6.})/2]/3n$$
$$\hat{X}_6 = [Y_{1.} + Y_{3.} + Y_{5.} - (Y_{2.} + Y_{4.} + Y_{6.} + Y_{7.})/2]/3n$$
$$\hat{X}_7 = [Y_{2.} + Y_{4.} + Y_{5.} - (Y_{1.} + Y_{3.} + Y_{6.} + Y_{7.})/2]/3n$$

where n is the number of people answering a given set of questions. Substitution of numerical values for $Y_{h.}$ in the above equations results in a set of values, some of which exceed unity. Subtracting the number to the left of the decimal results in the proportions given in column three of Table 1.

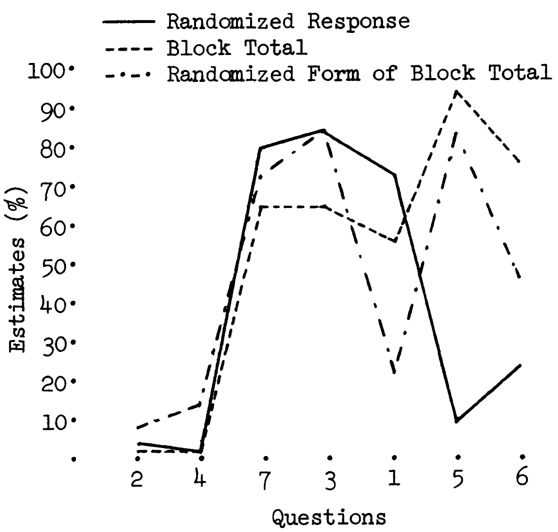
Table 1. Estimated proportions, $\hat{X}_h$, for each of seven questions by three different procedures

Estimate	I - RR	II - BTR	III - RBTR
$\hat{X}_1$	0.71	0.56	0.21
$\hat{X}_2$	0.03	0.01	0.08
$\hat{X}_3$	0.83	0.64	0.83
$\hat{X}_4$	0.01	0.01	0.12
$\hat{X}_5$	0.11	0.93	0.83
$\hat{X}_6$	0.23	0.76	0.46
$\hat{X}_7$	0.79	0.64	0.71

The method for obtaining the estimated proportions using the randomized form of the block total is the same as given above for the block total response. The only difference with this technique is in the manner in which an individual receives the set of three questions to be answered. With this method he chooses a number at random (without replacement) and answers a set of questions corresponding to that number. Thus, the interviewer not only does not know what the respondee has answered for a particular set of questions, but he does not even know which set of questions the respondee has answered.

From the class roster, we could check to find out the number of students who were actually enrolled in the College of Agriculture and Life Sciences (question 7 from the survey). Out of the 97 students enrolled in the course, 71, or 73%, of them were in the College. This value of 0.73 corresponds fairly closely to the estimates obtained by using our three techniques. The discrepancies are 0.06, -0.09, and -0.02 for methods I, II, and III respectively. The graphical representation of the above results is given in Figure 1.

Figure 1. Responses for seven questions from three methods



Considerable discrepancies were obtained for question 5 by the randomized response technique and the other two procedures. From discussions with students it would appear that a figure of 11% of parents with income over $25,000 is an unusually low figure, but that a figure of 93% by the block total response method is somewhat high. Likewise, it is doubtful if 76% of the students smoked marijuana during the past week (question 6 and the block total response technique).

The difference between block totals for question set 1,5, and 6 for the BTR and RBTR methods was relatively large, 51 versus 42. This difference resulted in quite different estimates for these questions by the two different methods. This could be explained as sampling variation and hence the conclusion that the sample size of n=12 was too small for comparing procedures. It was large enough to gain experience in conducting the three procedures in an experiment, and the total of 84 individuals allows a contrast of the three methods.

4. Estimated Variances for the Three Methods

The variance of an estimated percentage $\hat{X}$ from a binomial distribution would be $\hat{X}(1-\hat{X})/kn$. For the randomized response procedure, this variance is divided by π^2, where π is the fraction answering the sensitive question; in our case $\pi = 0.7$. The variance for each question was computed and is given in the second column of Table 3.

Two different methods have been devised for estimating the variances for the block total response procedure. The same methods may be used for the RBTR technique. The Y_{hi}, h=1,2,...,v=7 and i=1,2,...,n, are statistically independent since a simple random sample of n individuals was selected to answer each set T_h of questions and the different sets T_h were used on a different set of n individuals. Hence we may compute the variance of a given $\bar{Y}_{g.} = \Sigma_{i=1}^{n} Y_{gi}/n$ as:

$$(5) \qquad V(\bar{Y}_{g.}) = \left[\sum_{i=1}^{n} Y_{gi}^2 - \left(\sum_{i=1}^{n} Y_{gi}\right)^2 \Big/ n\right] \Big/ n(n-1) .$$

Using formula (5), the various variances were obtained and are presented in Table 2

Table 2. Estimated variances for each $Y_{g.}$ for the BTR and the RBTR procedures

Question Set	II - BTR	III - RBTR
T_1	0.0170	0.0833
T_2	0.0827	0.0657
T_3	0.0372	0.0221
T_4	0.0884	0.0777
T_5	0.0221	0.0625
T_6	0.0221	0.0372
T_7	0.0221	0.0322
Sum	0.2916	0.3807

The first method of estimating the variance of the $\hat{X}_h$ in Table 1 for the BTR method is

(6) $V(\hat{X}_h)$ = [Sum of the variances of $\bar{Y}_{g.}$ for the three blocks where question h occurred plus one-fourth of the variances of the $\bar{Y}_{g.}$ in the blocks where question h did not occur]/9 = [Sum of all 7 variances plus 3(sum of variances of $\bar{Y}_{g.}$ where question h occurred)]/36 .

For example, question 1 occurred in sets T_1, T_4, and T_6 with corresponding block means $\bar{Y}_{1.}$, $\bar{Y}_{4.}$, and $\bar{Y}_{6.}$. The estimated variance of $\hat{X}_1$ is computed from (6) as:

$$V(\hat{X}_1) = [V(\bar{Y}_{1.}) + V(\bar{Y}_{4.}) + V(\bar{Y}_{6.}) + \{V(\bar{Y}_{2.}) + V(\bar{Y}_{3.}) + V(\bar{Y}_{5.}) + V(\bar{Y}_{7.})\}/4]/9$$
$$= [0.2916 + 3(0.0170 + 0.0884 + 0.0221)]/36$$
$$= 0.0187 .$$

The remaining variances are computed from (6) and are given in the third column of Table 3. Likewise, the variances for the RBTR method are computed from (6) and are presented in the fourth column of Table 3.

The second method of computing the variances of the $\hat{X}_h$ is somewhat like a variance component procedure in that negative estimates of the variance can occur. The method is to use variances instead of block totals $Y_{s.}$ in equations (4) and solving. For example for $h = 1$,

$$V(\hat{X}_1) = [V(\bar{Y}_{1.}) + V(\bar{Y}_{4.}) + V(\bar{Y}_{6.}) - (V(\bar{Y}_{2.}) + V(\bar{Y}_{3.}) + V(\bar{Y}_{5.}) + V(\bar{Y}_{7.}))/2]/3$$
$$= [3\{V(\bar{Y}_{1.}) + V(\bar{Y}_{4.}) + V(\bar{Y}_{6.})\} - \sum_{s=1}^{7} V(\bar{Y}_{s.})]/6$$
$$= [3\{0.0170 + 0.0884 + 0.0221\} - 0.2916]/6$$
$$= 0.0152 .$$

The remaining estimated variances for both the BTR and RBTR procedures may be computed in a similar manner and are presented in Table 3.

Table 3. Estimated variances of $\hat{X}_h$ times 10^4 for the RR, BTR, and RBTR procedures

Variance	RR	First method BTR	First method RBTR	Second method BTR	Second method RBTR
$V(\hat{X}_1)$	117	187	271	152	356
$V(\hat{X}_2)$	16	136	216	-154	25
$V(\hat{X}_3)$	80	204	216	252	26
$V(\hat{X}_4)$	6	199	210	224	-10
$V(\hat{X}_5)$	55	182	257	123	272
$V(\hat{X}_6)$	100	145	246	-104	205
$V(\hat{X}_7)$	94	235	277	480	395
Average	67	184	241	139	181

Owing to relatively small sample size, n=12, comparisons of variances for individual questions are of little value. Instead, consider the average variances over all questions for the 7n=84 individuals. By the first method of computing the estimated variances, the average of the variances for the BTR method, 0.0184, and the RBTR method, 0.0241, is roughly three times that for the RR procedure, 0.0067. For the second method of estimating variances for the BTR and RBTR methods, the average variance is roughly twice that for the RR method. Since the estimated $\hat{X}_h$ (Table 1) for the RR method are generally higher (or lower) than the corresponding ones for the BTR and RBTR methods, the average variance for the RR method is smaller than if the $\hat{X}_h$ from the other two procedures had been obtained. Hence, the average variance for the RR procedure is underestimated for comparisons with the other two procedures.

In comparing the estimated variances for the RR and BTR procedures, consider the simplified situation wherein the variances of all $\hat{X}_h$ are equal to σ^2, and further suppose that responses to all k questions in a block are independent. Then, for k=3, $V(\hat{X}_h) = 4\sigma^2/3n$ for the first method. Since the RR estimate is obtained on kn=3n individuals, the variance is computed as $\sigma^2/3n\pi^2 = \sigma^2/3n(.7)^2 = \sigma^2/1.47n$. The ratio of the variances is $\sigma^2/3n(.7)^2/4\sigma^2/3n$ which is approximately equal to two. Thus, from variance considerations only the RR procedure is more efficient than the BTR method used here.

If cost is also considered, it will take k=3 or more times as long to administer the RR procedure as either of the other two procedures. Increasing the length of the interview and the length of time an interviewer spends with the respondent may be factors completely offsetting any gain in variance efficiency.

5. Discussion

After explaining the three techniques to the students and after attempting to convince them that their answers would truly be anonymous, several students remained skeptical. Some stated that no matter what an interviewer told them, they still would not answer a "sensitive" question truthfully, if a truthful answer could incriminate or embarrass them. They believed that if someone was ingenious enough to think of these techniques, they probably were ingenious enough to determine what an individual's answers had been. The problem of convincing the respondent of his anonymity appears to be the biggest problem associated with obtaining truthful answers. Interviewer training can help to some extent.

On questions 2 and 4, the most sensitive and incriminating ones, the responses for the RR and BTR procedures were much lower than for the RBTR method. Perhaps the reason for this is the increased anonymity obtained in the RBTR procedure, as students felt most anonymous with this method. This result brings out the fact that although the RR procedure may convince a proportion of the sample that their responses remain anonymous, a fraction remains unconvinced. The same is true for the BTR method. This would indicate that the RBTR method should be used in preference to the BTR method and that the technique of using a sealed box as in the RBTR method, with the RR procedure may be useful in practice to increase the truthfulness of responses.

It should be noted that birth dates are probably not uniformly distributed over a year. Hence, the assumption that 30% of the people have birth dates between September 13 and December 31 may not be correct. Since the RR procedure is affected by discrepancies in the proportion π, it might be better to have people whose birth dates fall on the 11, 12, ..., 19th day of a month, answer question B. Then, $1-\pi$ would be 108/365 ignoring leap years or [108/366 + 3(108)/365]/4 including leap years. People may have birth dates uniformly distributed over the days of a month during the year but probably not throughout the year.

Lengthy interviews may be considerably shortened using the BTR or RBTR methods, especially if questions are of a sensitive nature and if the RR technique is used on each question. Hence, situations may arise in which the BTR method is more efficient when both cost and variance are considered. Regardless of efficiency, the surveyor should use whatever method produces the most truthful answers. In this connection the BTR and RBTR procedures can be considered as alternatives to the RR method. There may be, however, procedures yet to be devised, which will replace all these procedures.

References

[1] Abul-Ela, A. A., Greenberg, B. G., and Horvitz, D. G., "A Multi-Proportions Randomized Response Model," Journal of the American Statistical Association, 62 (September 1967), 990-1008.

[2] Campbell, C. and Joiner, B. L., "How to Get the Answer Without Being Sure You've Asked the Question," The American Statistician, 27, No. 5 (December 1973), 229-31.

[3] Greenberg, B. G., Abernathy, J. R., and Horvitz, D. G., "Application of the Randomized Response Technique in Obtaining Quantitative Data," Proceedings of Social Sciences Section, American Statistical Association (August 1969), 40-3.

[4] Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R., and Horvitz, D. G., "The Unrelated Question Randomized Response Model Theoretical Framework," Journal of the American Statistical Association, 64 (June 1969), 520-39.

[5] Raghavarao, D. and Federer, W. T., "Application of BIB Designs as an Alternative to the Randomized Response Method in Survey Sampling," Number BU-490-M in the Mimeo Series of the Biometrics Unit, Cornell University (December 1973).

[6] Smith, L. L., "A Comparison of Three Techniques for Eliciting Answers to Sensitive Questions," B.S. Thesis in Statistics and Biometry, Cornell University (June 1974).

[7] Warner, S. L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60 (March 1965), 63-9.

[8] Warner, S. L., "The Linear Randomized Response Model," Journal of the American Statistical Association, 66 (December 1971), 884-8.

[1]Located at Cornell University in Summer, 1973, at the University of Guelph in 1973-74, and at Temple University for 1974-75.

LACK OF FIT AND THE MULTIPLE REGRESSION OF ANOVA MODEL*

Dwight E. Spence, University of Pennsylvania

ABSTRACT

Suggested in this paper is a test for lack of fit and its implications for model change applicable to the general regression analysis model, especially when used to handle analysis of variance with qualitative independent variables. Draper and Smith (1967) have developed a procedure to test for lack of fit of the general regression analysis model in the case where the data includes repeat measurements of Y at the given X values. However, when the independent variable(s), X, is qualitative, as is more usually the case in multiple regression-ANOVA (MR/AV), Draper's estimate of "pure error" will be exactly that value given as the residual mean square estimate. Thus, any possibility for such a test of lack of fit, i.e., for finding "pure error" to be less than the residual mean square, will be obviated in this MR/AV case. An alternative is to obtain repeated treatment levels from which an estimate of pure error can be obtained. Use of this suggested alternative pure error term in a test for lack of fit is used as an initial indicator as to whether attempts should be made to (a) find dependent variable transformations and (b) try various possible independent variable product-variables.

FITTING THE REGRESSION ANALYSIS MODEL IN EXPERIMENTATION

INTRODUCTION

The following is not intended as an exhaustive presentation of the use of the regression analysis model vis-a-vis experimental design data. In part, it is a review of the major aspects of this subject; however, the emphasis is on testing the adequacy of a specific model against real data. It is with respect to this latter point that we have ventured beyond what is already available in the literature on this subject. The first section will examine the data coding techniques to be used in applying the regression model to experimental design data; the special concern here (as throughout this paper) is for the case of continuous dependent variables and categorical independent variables. Section II considers methods to be used in testing for lack of fit of a specific model to a given set of observations. Finally, the focus of Section III centers on the large spectrum of alternative regression models available for application in data analysis.

SECTION I

ONE-DIMENSIONAL ANOVA

If we are to use multiple regression analysis to do analysis of variance, the ANOVA data must be recoded prior to applying the multiple regression model (MR/AV). Thus, in this section we will consider the problem of recoding with regard to simple ANOVA, factorial designs with equal cell sizes, and factorial designs with unequal cell sizes.

First the case of simple ANOVA. The most direct approach to recoding in MR/AV is referred to as dummy coding, and has as its objective a vector representation of each data observation such that observations in different cells are distinctly represented in terms of a binary system of 1's and 0's (Kerlinger and Pedhazur, 1973). Recall that the general linear regression model

$$Y = \beta_0 Z_0 + \beta_1 Z_1 + . . . + \beta_n Z_n + e \qquad (1)$$

where $Z_i = f(x_1, x_2, \ldots, x_m)$; $Z_0 = 1$
can also be expressed in the form

$$Y = \mu_Y + \beta_1(Z_1 - \mu_Z) + . . . + \beta_n(Z_n - \mu_Z) + e \qquad (2)$$

where $\mu_Y - \beta_1\mu_Z - . . . . - \beta_n\mu_Z = \beta_0 Z_0$

which, when compared with the analysis of variance model

$$Y = \mu_Y + \alpha_i + . . . + \delta_j + e \qquad (3)$$

reveals the equality (Hays, 1973)

$$\beta_1(Z_1 - \mu_Z) = \alpha_i \qquad (4)$$

However, in the present case of categorical independent variables where the variables are being recoded each level of each factor-variable is represented in the model. Consider a factor-variable with three levels, then observations at the first level are represented as the row vector

[1 1 0]

while observations at the second and third levels would have vector representations of

[1 0 1] and
[1 0 0] respectively. The first column is a unit vector; the second column has a "1" if an observation is a member of the first level and a "0" otherwise; the third column has a "1" where the observation belongs to the second level and a "0" otherwise. In contrast, the third level is indicated by the fact that its members belong to neither level one nor two, i.e., zeros in both columns two and three. Therefore, if we have three observations at each of the three levels, the result of recoding would be the following matrix:

$$\begin{matrix} X_0 & X_1 & X_2 \end{matrix}$$
$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Thus the number of predictor variables "increases" by one (and alas the number of Df decreases!); and the general linear regression model is

$$Y = \beta_0 Z_0 + \beta_1 Z_1 + \beta_2 Z_2 + e \qquad (1)$$

*Presented at the National Convention of the American Statistical Association, 1974

even though we are still concerned with only one factor with three levels. Computation of the square of the multiple correlation coefficient leads to an F test as follows:

$$F = \frac{R^2/P}{(1-R^2)/n-P-1}$$

where P = number of predictors;
n = number of observations.
The obtained F value with its respective number of degrees of freedom has the same level of probability as that which we could have computed using the typical ANOVA F-test

$$F = \frac{MS_{Bet.}}{MS_{With.}}$$

In place of recoding through the use of dummy coding, we can use either "effect coding" or "orthogonal coding." Besides doing the same thing as the first technique, these latter two recoding techniques provide a convenient approach to the multiple comparison of means. Effect coding differs from dummy coding in that it uses "-1's" instead of "0's" to represent that group which was identified by the fact of non-membership in any other group. Thus, rather than just 1's and 0's, the coding entails 1's, 0's and -1's. Using this coding approach post hoc multiple comparisons of means can be carried out by what in effect is a statistical test of the significance of the difference between any two or more relevant regression coefficient estimates, $\hat{\beta}$. In terms of effect coding, the above matrix would take the form

$$\begin{array}{ccc} X_0 & X_1 & X_2 \end{array}$$
$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

In contrast, orthogonal coding leads to the opportunity to perform planned orthogonal comparisons among group means. To do this the column vectors must consist of terms such that each vector is orthogonal to all other vectors, with the exception of the initial unit vector which is omitted in many computer routines. Each column vector in turn expresses a separate orthogonal contrast. Again we can construct the matrix based on the above hypothetical ANOVA problem using orthogonal coding thus:

$$\begin{array}{ccc} X_0 & X_1 & X_2 \end{array}$$
$$\begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 1 & 1 & \frac{1}{2} \\ 1 & 1 & \frac{1}{2} \\ 1 & -1 & \frac{1}{2} \\ 1 & -1 & \frac{1}{2} \\ 1 & -1 & \frac{1}{2} \\ 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

Column two gives a contrast between the means for group one and group two, while column three compares the average of group one and two with group three--orthogonal to the first comparison. A test of significance applied to a particular regression weight ($\hat{\beta}$) indicates whether the comparison specified by the corresponding column vector is significant or not, i.e., if $\hat{\beta}$ is significantly different from zero, the comparison of column vector X_1 is also significant with respect to the difference between group one and group two means.

Factorial Design:

For the case of multiple independent variables, each expressed in terms of the categorical level of measurement, the coding systems dealt with above can be shown to have direct applicability through generalizations of the principles used in the simple ANOVA situation. Suppose our analysis is of a 2 x 3 factorial design with two observations per cell. The resulting orthogonally coded matrix would take the form

$$\begin{array}{cccccc} (X_0) & (X_1) & (X_2) & (X_3) & (X_{12}) & (X_{13}) \end{array}$$
$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 0 & -2 & 0 & 2 \\ 1 & -1 & 0 & -2 & 0 & 2 \\ 1 & -1 & 0 & -2 & 0 & 2 \\ 1 & -1 & 0 & -2 & 0 & 2 \\ 1 & -1 & 0 & -2 & 0 & 2 \\ 1 & -1 & 0 & -2 & 0 & 2 \end{bmatrix}$$

The X_0 column vector is again a vector of 1's and refers simply to the overall mean level of our regression model. Vector X_1 refers to factor A of two levels and uses a 1 to indicate membership in the first level and -1 for membership in the second level, contrasting these two levels. Columns X_2 and X_3 of course refer to independent variable B which has three levels. Thus X_2 provides a comparison between levels one and two of variable B; whereas, levels one and two are compared with level three in column X_3. Finally, X_{12} and X_{13} account for the possible interaction between the two independent variables for any given dependent variable. These last to vector are formed by simply taking the cross-products of the relevant main effects column, e.g., X_{12} consist of cross-products between the elements of vectors X_1 and X_2. The several F-ratios for main and interaction effects can be computed through the computation of R^2 values based on subsets of the orthogonal column vectors. For example, if vectors X_{12} and X_{13} are separately analyzed we get a R^2 value indicating the amount of variance accounted for by the interaction of the two (or more!) factors.

An added complication enters the picture if the cell sizes contain unequal numbers of observations. This stems from the fact that the main and interaction effects are no longer orthogonal to each other; and therefore, different results can be obtained depending on the order in which the total variance is partitioned into the variance of the separate components (Overall and Klett, 1973). Briefly, these are the three possible partitions. Simply consider the problem of calculating the variance due to the main effect of variable A in the hypothetical 2 x 3 factorial design mentioned above.

a) R^2 can be computed for the column vector(s) of main effect A, and the variance thus accounted for can be determined by the product

Total Variance $\cdot R^2_{x_1}$ {Variance due to main effect A}

b) If R^2 for main effect B is subtracted from R^2 for both main effects A and B, then the variance accounted for by main effect A is given by the product

Total Variance $\cdot [R^2_{x_1,x_2,x_3} - R^2_{x_2,x_3}]$

{Variance due to main effect A}

c) Yet a third approach might entail taking the difference between R^2 computed for effects A, B and interaction and R^2 computed for effects B and interaction; in this instance, the variance due to main effect A would be given by the product

Total Variance $\cdot [R^2_{x_1,x_2,x_3,x_{12},x_{13}} - R^2_{x_2,x_3,x_{12},x_{13}}]$

{Variance due to main effect A}

In each of these three efforts to accounting for the variance due to main effect A different answers are obtainable given the situation of unequal cell sizes. It is the second strategy, (b), that corresponds to the ANOVA technique as used in experimentation (see Overall and Klett, 1972).

Aside: Wolf and Cartwright have recently (1974) proposed a quite different approach to this problem of coding in MR/AV. Their approach requires that an experimenter solve for the coded matrix (x) using the matrix solution to the normal equations $b = (x'x)^{-1}x'y$. Of course y is the N×1 vector of dependent variable observations. The J×1 (where J = the number of comparisons) b vector is obtained by the matrix formula $b = c(V'V)^{-1}V'y$ where V is an N×K (K = the number of categories) matrix of dummy coded 1's and 0's, indicating membership (1) and non-membership (0) in the several categories; and C is a J×K matrix of weights for the comparison of category means. An example of C provided by Wolf and Cartwright is

$$C = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -\frac{1}{4} & -\frac{1}{2} & 1 & 0 \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

which applies in one-dimensional ANOVA when the cell sizes are equal. Thus computing the vector b allows for the generation of a coded matrix (x) to be used in the multiple regression analysis of ANOVA (by solving for x in $b = (x'x)^{-1} x'y$)). The end product is a coded matrix which in appearance is considerably different from the matrices obtainable using either of the three coding techniques that we have considered up to this point. Likewise, Wolf and Cartwright state that the computed b, error and F values all differ from those gotten using a coding system consisting of 1's, 0's and -1's. However, they do reveal that the inferences made using these two differing approaches do not, themselves, differ. The statistical conclusions remained the same.

SECTION II

TESTING FOR LACK OF FIT

If we make the very important shift in reference away from simply testing the relationship between the independent and dependent variables vis-a-vis a specific test statistic to testing the adequacy of the test statistic model with respect to the data at hand, we will at least become aware of the presen-e of diverse models; and at most we will find independent/dependent variable relationships which would have otherwise been overlooked given the limitations of any one specific model. This point is singularly fundamental throughout this and the next section: whenever we test for a relationship within a set of observations, we do so in the context of a very specific model imposed on the data; and thus, must consider the possibility that an alternative model might prove more revealing (without simply capitalizing on chance characteristics of the data).

Draper and Smith (1966) published a technique to test for the lack of fit of a regression model. The key component is the computation of a "pure error mean squares" term (S_e^2) used to evaluate a lack of fit mean squares term (MS_L). To compute this pure error value the data must contain repeat observations on the dependent variable (Y) at several levels of the independent variable (X). Thus, pure error refers to the variability of Y within levels of X:

$$\text{Pure Error } (SS_e) \quad \sum_{j=1}^{k} \sum_{i=1}^{n_i} (Y_{ij} - \bar{Y})^2 \quad .$$

$$\text{Lack of Fit } (SS_L) \quad \text{Residual (SS)} - \sum\sum(Y-\bar{Y})^2$$

This test for lack of fit follows the F distribution with

$$F = \frac{SS_L/Df_L}{SS_e/Df_e} = \frac{S_L^2}{S_e^2} = \frac{\text{Lack of Fit (MS)}}{\text{Pure Error (MS)}}$$

A significantly large F value indicates that the residual error term consists of more than simple error variability, i.e., the model being considered fails to account for some non-error variance.

Other questions might be raised at this point; however, the crucial one here is: What happens if the independent variables (X) are categorical rather than continuous? After all, this is more typically the case in social science data analysis using ANOVA. If we attempt to apply the above technique to MR/AV given categorical independent variables, the finding is that this approach invariably gives us a value for S_e^2 exactly the same as the residual MS value. And thus we have no basis for ever finding a model inadequate. It is for this case that we would like to suggest an alternative approach.

Perhaps the pure error term (S_e^2) should be computed by splitting the data in half for each level of theindependent variable, computing a mean for each half cell of data and looking at the variability of half cell means within the several independent variable levels.

$$\text{Pure Error (SS)}' \quad \{\sum\sum(\bar{Y}-\bar{\bar{Y}}_i)^2 \frac{n_i}{2}\}$$

where $\bar{Y}$ is half cell mean;

$\bar{\bar{Y}}$ is a mean for the entire cell.

This presumably would be comparable to a treatment (SS) value under the condition of no systematic relationships in a set of data. Thus, the F distribution would take the form

$$F = \frac{\text{Pure Error) (MS)}'}{\text{Residual (MS)}'} = \frac{\sum\sum \frac{n_i}{2} (\bar{Y}-\bar{\bar{Y}})^2/Df_e}{\text{Residual (SS)}'/Df_{res}}$$

and a significantly small F value would indicate that the residual term is inflated for the given regression model. Of course this requires reasonably large sample sizes.

SECTION III

ALTERNATIVE REGRESSION ANALYSIS MODELS

A test for lack of fit is here intended to imply both an entire system of candidate models and a willingness to explore the possible usefulness of any one of these models. Again the context to be assumed is that of continuous dependent variables and categorical independent variables. Two major types of alternative models are to be dealt with. The maneuver distinguishing the first type will be the gamut of variable transformations which may be applied to the dependent variable, whereas the strategy of the second class of alternative models revolves around the generation of "new" independent variables through taking the cross-products of the original independent variables producing what will be called "interaction variables." Of course there exists no compelling reason why both of the above manipulations cannot be applied simultaneously.

There is quite a good deal of literature devoted to the subject of data variable transformations (Bartlett, 1947; Box, 1962; Tukey, 1957). One rationale for data transformations is that the transformations make for a more precise analysis by bringing real data into greater conformity with the assumptions of the analysis, e.g., homogeneity of variance, normality, additivity, etc. However, clear and convincing indicators as to where, for example, a square root or a reciprocal, a log or an arcsin transformation might most profitably be applied does not seem evident from the literature. The user must resort largely to trial and error, and testing for lack of fit. Some general guidelines have been put forward (Myers, 1972). Given a situation where the variance is proportional to the means (i.e., $\sigma^2/\mu_i = K$) the square root function should be tried to achieve variance constancy. Application of a log function to observed data might be attempted in the case where the standard deviation is directly proportional to the mean (i.e., $(\sigma/\mu)^2 = K$) and the distribution is considerably skewed. Use of the arcsin function has resulted in reports of success in the case where the data observations are proportions (or percentages) with $\sigma^2 = \mu(1-\mu)$.

General Regression Model

$$Y = \beta_0 Z_0 + \beta_1 Z_1 + \ldots + \beta_n Z_n + e = \sum_{i=0}^{n} \beta_i Z_i + e \quad (1)$$

where $Z_i = f(X_1, X_1, \ldots, X_m)$

Regression Model with Dependent Variable Transformation

$$\ln Y = \sum_{i=0}^{n} \beta_i Z_i + e \quad (5)$$

This should be enough to make it clear that the possibilities with respect to use of dependent variable transofrmations are virtually limitless.

Compared to data transformations on continuous variables, there appears to be little published material concerned with the generation of new independent variables from a set of data where the independent variables are measured at the categorical level (Cohen, 1968). In general, this second group of alternative models can be characterized by the fact that they are the result of forming new independent variables by taking the cross-products of the basid set of independent variables in some cases referred to as moderate variables. Thus, for example,

$$\text{General Regression Model} \quad Y = \sum_{i=0}^{n} \beta_i Z_i + e \quad (1)$$

where $Z_i = f(X_0, X_1, \ldots, X_m)$

and X_0 through X_m make up the original data base. Typically we would simply have the equality $Z_i = X_i$; however, in the case of cross-products we might have

$$\text{Interaction (Variable) Between } X_{i-1} \text{ and } X_i \quad Z_i = X_{i-1} \cdot X_i \quad (6)$$

Given as few as, say, three original data variables there are, as can be quite easily shown, several interaction variables which can be computed (which is not to say that they should be!). Whether these interaction variables account for any variance of consequence can be tested by the usual multiple R^2 ratio following the F distribution:

$$F = \frac{(R^2_{y\cdot 12} - R^2_{y\cdot 1})/Df}{(1-R^2_{y\cdot 12})/Df} \quad (7)$$

REFERENCES

Bartlett, M. S. The use of transformations. *Biometrics*, 3, 1957.

Box, G. E. P., and Tidwell, P. W. Transformation of the independent variables. *Technometrics*, 4, 1962

Cohen, J. Multiple regression as a general data analytic system. *Psychological Bulletin*, 1968, 70.

Draper, N. R., and Smith, H. *Applied Regression Analysis*. New York: Wiley, 1966.

Hayes, W. *Statistics for the Social Sciences*. New York: Holt, Rinehart and Winston, 1973.

Kerlinger, F., and Pedhazur, E. *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart and Winston, 1973.

Myers, J. *Fundamentals of Experimental Design* (2nd ed.). Boston: Allyn and Bacon, Inc., 1972.

Overall, J., and Klett, C. *Applied Multivariate Analysis*. New York: McGraw-Hill, 1972.

Tukey, J. W. On the comparative anatomy of transformations. *Ann. Math. Statistics*, 28, 1957.

Wolf, G., and Cartwright, B. Rules for coding dummy variables in multiple regression. *Psychological Bulletin*, 81, 1974.

WHITE RESIDENTIAL MOBILITY IN A RACIALLY CHANGING AREA

Lloyd A. Turner, University of Pennsylvania

Past research on neighborhood racial transition and residential mobility has generated few propositions about the impact of naturally-occurring change processes or proposed policy changes on the social composition of urban neighborhoods. A major difficulty in the development of policy-oriented models of neighborhood occupancy change is the assessment of the various situational factors that influence household mobility behavior. One of the major obstacles to systematic study of the relationships between the so-called "neighborhood" effects and residential mobility has been the absence of data sources that permit inferences about micro-level, short-run behavior.[1]

The development of disaggregated, longitudinal data files opens up several new possibilities for research on intrametropolitan migration, including the investigation of these "neighborhood" factors. The principal objective of this paper is to indicate how data from such files can be utilized in developing formal models of residential mobility. More specifically, we examine several models of white outmigration from selected intraurban areas using detailed characteristics of individual households and characteristics of blocks in which these households live.

1. Procedure

The research strategy employed in this study involves fitting a variety of models of white outmigration for two census tracts, one of which is currently undergoing white-to-black transition, the other remaining racially stable and predominantly white.[2] The justification for identifying two areas with similar population and housing stock characteristics is that this procedure facilitates not only the recognition of intertract differences in mobility rates but also the extent to which such differences are associated with block-level environmental factors.

The data base required for making such comparisons, then, must meet two specifications: first, it must contain detailed socioeconomic characteristics of individual households in suitable test-control areas; and second, it must include information about the mobility behavior of these separate households. (The decennial census clearly fails to satisfy these criteria, as it permits inferences only about net migration by various types of households within a particular area.) At present one of the strongest data sets that meets these two criteria is the Wichita-Sedgwick County (Kansas) Annual Enumeration of Households, a complete census of the Wichita metropolitan area that has been conducted for each of the past four years.

According to U.S. Census data five census tracts in Wichita had a marked increase in black households between 1960 and 1970. Of these five tracts, tract 18 best met the criteria of (1) recency and extent of black inmigration, (2) having a sufficiently large number of long-term white residents, and (3) being suitably matched to some all-white tract. Table 1 lists the indicators that were used in the matching process.

The all-white tract that matched tract 18 on the largest number of indicators was tract 28. Tract 18 lies about half a mile east of center city, whereas tract 28 lies about the same distance west of the center and across the Arkansas River. Table 1 indicates that the tracts are very similar with respect to household size, population composition and density, education levels, occupational and employment characteristics, and average rent. Both tracts had a large net outmigration of whites between 1960 and 1970, but only tract 18 had a sizeable increase in blacks. Aside from racial composition these tracts differ noticeably in four respects: median income ($4411 for tract 18, $5431 for tract 28); average value of single-family, owner-occupied units ($8300 for 18 as compared with $10,300 for tract 28); percent of year-round vacancies (19.2% versus 6.3%); and percent at same residence in 1965 (39.4% compared with 50.6%). The high vacancy portion of tract 18 is concentrated in the southwest corner of the tract and may be attributable to the "natural" growth of the center-city commercial district. The variable "percent at same residence in 1965" is highly correlated with the propensity to move for white households and hence was not used as a matching characteristic; it is listed in Table 1 merely for reference.

A convenient method for estimating mobility rates involves the use of the logit model, which specifies (for the case in which the dependent variable is the propensity to move) that the natural logarithm of the expected odds that a household moves during a given time period is a linear function of some set of independent (or predictor) variables.[3] Bishop (1969) has demonstrated that Dyke and Patterson's (1952) somewhat cumbersome method of logit estimation is analytically equivalent to an appropriately-chosen log-linear model, which can be estimated by several procedures.[4]

The method of iterative proportional fitting [see Goodman (op. cit.) and Fienberg (op. cit.)] is used in this paper to estimate the parameters of the logit model because of the flexibility of this approach and its efficiency in model specification and hypothesis testing. The data input for iterative proportional fitting is an observed contingency table, and the output is a fitted table that is obtained by setting certain marginals in the fitted table equal to the corresponding marginals in the observed table. The likelihood-ratio chi-square statistic and degrees of freedom are computed for the models, and the significance of any k-way interaction is determined by comparing whether the difference in value of the likelihood-ratio statistic between this model and the same model with only (k-1)-way interactions compensates for the difference in degrees of freedom, where the critical points are based on the usual chi-square test. The significance level (or "p-value") of a chi-square test statistic is the probability that that statistic will exceed the observed value under the assumption that a random sample is taken from a population satisfying the hypothesized model. [See Fienberg (op. cit.) for a discussion of these test statistics.]

2. Findings

The first set of models that were tested examined the role of several block-level characteristics in the outmigration of white households from tract 18, the racially changing tract. It was hypothesized that such neighborhood factors as racial composition, vacancy rate, extent of substandard housing, and average rent would be related to white mobility rates. Three household characteristics--stage in life cycle, duration of residence, and tenure--were used as control variables in these models. Although there are undoubtedly many possible ways of defining life cycle variables, the approach taken by Goldstein (1973) is particularly well suited to research in residential mobility in that it is based on characteristics that in some sense relate to housing needs. Goldstein argues that life cycle variables should consisit of not only age of head of household but also marital status, size of household, and possibly other factors. His paper used data from the 1965 Bay Area Transportation Study Commission to estimate the effects of four variables on subsequent mobility --life cycle, education, years at current job, and previous tenure. His table 2.2 listing coefficients for each of the dummy variables is reproduced below for reference. The fifteen estimated regression coefficients in this table compare the propensity to move for difference types of households with the mobility of the reference group, which is single persons under thirty years of age. A negative coefficient for a particular cell means that households in that cell are less likely to move than are households in the reference group. The coefficients in this table fall conveniently into three groups (0.03 to -0.15, -0.16 to -0.25 and -0.26 to -0.40) and form the basis for our trichotomous life cycle variable. The levels of this variable are listed in Table 3, in which the row variable is "persons in household" rather than "marital status", which is not recorded in the 1971 Wichita enumeration.

The other two control variables, duration of residence and tenure, are less ambiguous than life cycle. A three-level variable was created for the former variable, with short duration being defined as two years or less and long duration as ten or more years at the same address. Tenure was dichotomized as own and rent. All the household characteristics are based on the 1971 enumeration except for mobility, which is based on the time period between the 1971 and 1972 censuses.

The four neighborhood characteristics that were investigated are described in Table 4. The break-points in this classification scheme were assigned so as to include approximately the same number of blocks in each category. For each of the four neighborhood characteristics a set of five-way contingency tables was made. Each of these tables was then analyzed by a stepwise logit procedure, which is similar to stepwise regression methods.

The results of this series of computations are given in Table 5. Eleven models are fitted for each of the four five-way tables. The numbers group together under the column "Fitted Marginals" specify which marginals in the fitted table are set equal to the corresponding marginals in the observed table. Variable five, move/stay, is treated as the "dependent" variable, and consequently each fitted logit model must agree with the observed table in the four-way marginal that includes all the independent variables [Bishop (1969)]. For example, model (1) in Table 5 specifies the logit model in which each of the four independent variables has a main effect on variable five, while model (6) indicates the logit model with a main effect for each independent variable and also an interaction effect for variables one and two on variable five. In the first contingency table (which is represented in the (a) columns of Table 5) the significance test for the main effect of the variable MPBL on the logit of a move is (70.75-53.17), or 17.58, with (30-29), or 1, degree of freedom, which is significant at the 1% level. We therefore conclude that MPBL is associated with the propensity to move for white households in tract 18. From the (b), (c), and (d) columns it can be seen that the main effects of MSUB and RENT are significant at the 1% and 10% levels, respectively, whereas the main effect of VAC is not significant even at the 10% level. The moderately low p-values found in most of the models suggests, however, that one or more key variables may have been omitted. From the (a) columns it is clear that although the effect of the race variable MPBL on mobility is highly significant, none of the eleven models provides an adequate fit to the data under the 10% significance level criterion. In the (c) columns it is seen that the extent of substandard housing (MSUB) is strongly related to the propensity to move. The (d) columns indicate a possible relationship between average rent level (RENT) and mobility. The only interaction effect that is significant in all of the columns is the one between life cycle and tenure, which suggests that certain types of households (young renters, perhaps) are more sensitive to neighborhood factors than are other types.

The second set of models investigated in this study uses only household characteristics to explain subsequent mobility. The variables that are used in these models are sex of head of household, stage in life cycle, household income, duration of residence, tenure, census tract, and mobility behavior (i.e., move or stay). These models were originally fitted for tracts 18 and 28 separately and then were combined into one larger table.

The separate analyses for each tract, which are not presented here due to space limitations, yielded seven principal findings. First, an examination of the relationship between sex of head of household and the propensity to move failed to discover any differences between male- and female-headed households in either tract. Second, household characteristics explain a larger fraction of the variation in mobility rates for tract 28 than for tract 18. Third, tenure and life cycle have significant main effects in both tracts. Fourth, a comparison of several alternative life cycle indices indicates that the goodness of fit of the models is highly sensitive to the particular index chosen. Fifth, income appears to be related to mobility in tract 18, and the effect of this variable in tract 18 is somewhat less than the effects of tenure and life cycle. Sixth, although duration of residence is strongly related to mobility in tract 18, it is not clear whether or not there is any association between these two variables in tract 28. Finally, the most consistent interaction effect (on mobility rates) is between life cycle and tenure variables.

In the combined six-way table (in which tract is treated as a dichotomous predictor variable) all five independent variables (life cycle, income, duration, tenure, and tract) are significant at the 1% level except income, which is significant at 5%. The strongest main effect is from tenure, followed by tract, life cycle, duration of residence, and income.

3. Estimation of effects

In this section we estimate values for the main and interaction effects of the two best models in the six-way table that was described in the previous section. Letting X_{1ijklm} (X_{2ijklm}) be the observed number of movers (stayers) in categories i,j,k,l, and m of the five independent variables, we define the logit for a move (in the subsequent year) by households having this set of characteristics as

$$L_{ijklm} = \log (X_{1ijklm}/X_{2ijklm})$$

and write

$$L_{ijklm} = \alpha + \beta_i + \gamma_j + \delta_k + \epsilon_l + \mu_m$$

where we adopt Theil's (1970) parameterization:

$\beta_1 = \gamma_1 = \delta_1 = \epsilon_1 = \mu_1 = 0$

β_i = differential effect of i^{th} life cycle class

γ_j = differential effect of j^{th} income class

δ_k = differential effect of k^{th} duration-of-residence class

ϵ_l = differential effect of l^{th} tenure class

μ_m = differential effect of m^{th} census tract

The best logit model in the six-way table includes the main effects of the five independent variables plus an interaction between life cycle and tenure. Table 7 gives parameter estimates for this model, whereas Table 6 gives estimates for the logit model with only the five main effects. [In obtaining these estimates .5 has been added to each cell of the table to minimize the bias; cf. Gart and Zweifel (1967).]

Several interesting comparisons can be drawn from Table 6. First, the effects seem quite reasonable. In comparison with the reference group (young, low-income, short-duration owners in tract 18), the probability of a move during the subsequent year is greater for renters and high-income households and lower for older, longer-duration, and middle-income households. Furthermore, households in tract 28 are less likely to move than are those in tract 18, even after controlling for the effects of life cycle, income, duration of residence, and tenure. The curvilinear relationship between mobility and income may be the result of two opposing effects: household instability is associated with low income and high mobility, whereas the direct income effect on mobility is positive for households living in low- to moderate-priced dwellings. The intertract difference in propensity to move is attributable to both environmental ("neighborhood") factors and to recent commercial development in tract 18.

The interaction model (Table 7) yields two additional findings. First, the low mobility attributed to the second life cycle class (thirty to fifty-nine year-old heads of household who do not live alone) in Table 6 is characteristic of owners only, as renters in this life cycle category are more highly mobile than are reference group households. Second, the lack of significance of β_{32} implies that households in the third life cycle category who rent have approximately the same propensity to move as do households in the reference group.

4. Conclusions

Five conclusions summarize the results of this study. First, at adequate model (in the sense of high significance level) of residential mobility can be constructed using life cycle, income, duration of residence, and tenure variables. Second, the larger unexplained variation in mobility rates in the racially changing tract suggests that neighborhood factors influence mobility decisions to a greater degree in that area than in the comparison area. Third, block-level housing stock and social characteristics seem to have some explanatory power. Three of the four block characteristics investigated--racial composition, extent of substandard housing, and average contract rent--were significantly related to subsequent mobility. Future study will investigate these factors and other environmental variables constructed at the street-front rather than block level. Fourth, the findings suggest that there are no significant differences in propensity to move between female- and male-headed households for whites. It would be desirable to examine the sex-mobility relationship for both whites and nonwhites in other areas. Finally, although sex seems to have no direct effect on mobility, attention should be given to the construction of life cycle variables that include age of head, size of household, and sex of head of household. The alternative life cycle variables considered in this study give rise to models that differ considerably in explanatory power.

Subsequent research in this project will extend the areal scope of the study in an attempt to overcome some of the major biases introduced by examining individual census tracts. On the basis of a social area analysis we hope to define two neighborhoods that will both increase the sample size and provide more meaningful units of analysis.

Footnotes

*This paper was prepared with the support of the Wichita-Sedgwick County Metropolitan Area Planning Department; the National Institute of Mental Health, PHS Research Grant No. 1 R01 MH25096-01 from the Center for Studies of Metropolitan Problems; and the Fels Center of Government, University of Pennsylvania. The author wishes to thank Professors Stephen Gale and Ralph Ginsberg for their critical readings of an earlier version of the manuscript.

[1]Two studies of "neighborhood" effects that would have been significantly improved with a more extensive data base are, for example, Rapkin and Grigsby (1960) and Molotch (1969).

[2]The mobility behavior of nonwhite households is not investigated in this study because of sample size limitations.

[3]For recent discussions of logit models see Goodman (1971 and 1972), Fienberg (no date), Theil (1971), and Speare (1971).

[4]See, for example, Goodman (1971 and 1972), Ku and Kullback (1968), and Grizzle, Starmer, and Koch (1969).

References

Bishop, Yvonne M. M. (1969). "Full Contingency Tables, Logits, and Split Contingency Tables," Biometrics 25,383-9.

Dyke, G. V. and Patterson, H. D. (1952). Analysis of Factorial Arrangements when the Data are Proportions, Biometrics 8, 1-12.

Fienberg, Stephen E. "The Analysis of Cross-Classified Data," Dept. of Applied Statistics, University of Minnesota.

Gart, J. J. and Zweifel, J.R. (1967). "On the Bias of Various Estimators of the Logit with Applications to Quantal Bioassay," Biometrika 54 (1 and 2), 181-87.

Goldstein, Gerald S. (1973). "Household Behavior in the Housing Market: The Decision to Move and the Decision to Buy or Rent Housing," in Moore, E.G., Models of Residential Location and Relocation in the City, Dept. of Geography, Northwestern University.

Goodman, Leo A. (1972). "A General Model for the Analysis of Surveys," American Journal of Sociology 77, 1035-86.

________ (1971). "The Analysis of Multidimensional Contingency Tables," Technometrics 13, 33-61.

Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). "Analysis of Categorical Data by Linear Models," Biometrics 25.

Ku, H. H. and Kullback, S. (1968). "Interaction in Multidimensional Contingency Tables," Journal of Research of the National Bureau of Standards, 72B, 159-99.

Molotch, Harvey (1969). "Racial Change in a Stable Community," American Journal of Sociology 75, 226-38.

Rapkin, Chester and Grigsby, William (1960). The Demand for Housing in Racially Mixed Areas, University of California Press, Berkeley and Los Angeles.

Speare, Alden, Jr. (1971). "Alternative Models of Individual Mobility," these Proceedings, 364-68.

Theil, Henri (1970). "On the Estimation of Relationships Involving Qualitative Variables," American Journal of Sociology 76, 103-54.

________ (1971). Principles of Econometrics, New York: Wiley.

Table 1: Comparison of Tracts 18 and 28

Characteristic		Tract 18	Tract 28
Percent nonwhite:	1960	10.9	0.02
	1970	20.1*	0.4*
Racial composition (black-white-other):	1960	578-4663-58	1-4624-16
	1970	718-2807-42*	17-4099-36*
Percent one-person households		25-40+	25-40+
Percent female-headed households		18.3*	14.3*
Average persons per household		2.38*	2.63*
Median income		$4411*	$5431*
Median years school completed		11.3*	12.2*
Percent of dwellings overcrowded		5.9*	5.3*
Percent civilian labor force unemployed		7.9*	8.7*
Percent single-family dwellings		50-74.9+	50-74.9+
Percent 2-4 family dwellings		30 or above+	15-29.9+
Percent 1-family, owner-occupied dwellings		70-84.9+	50-69.9+
Median value, 1-family owner-occupied dwellings		$8300*	$10,300*
Median rent		$71*	$74*
Percent dwellings vacant year-round		19.2*	6.3*
Percent dwellings substandard		17-39.9+	5-16.9+
Percent at same residence in 1965		39.4*	50.6*
Percent residing outside SMSA in 1965		19.5*	16.1*
Percent professional and kindred		12.9*	13.8*
Percent clerical and kindred		18.2*	18.8*
Percent nonfarm laborers		4.7*	2.3*
Percent families below poverty level		15.4*	10.5*

Sources: *U.S. Census of Population and Housing, Census Tract Reports, 1970.
+Wichita Profile 1970, Metropolitan Area Planning Department, Wichita, 1971.

Table 2.2: The Decision to Move*

Marital Status	Age of Head: Less than 30	30-44.9	45-59.9	60 and over
Single	Reference group	-0.2978 (-4.4472)	-0.3785 (-5.0034)	-0.2942 (-3.0912)
Married, no children	0.0300 (0.5676)	-0.2070 (-3.6855)	-0.2223 (-4.4359)	-0.2912 (-5.5666)
Married, with children	-0.0541 (-1.1414)	-0.1809 (-4.0595)	-0.2492 (-5.1970)	-0.3000 (-3.6559)
Other	-0.0310 (-0.4123)	-0.2043 (-3.5868)	-0.3992 (-7.2051)	-0.3831 (-6.6604)

Note: t-values are in parentheses. N = 2246, $R^2 = 0.3424$, F = 64.4292, $S_e = 0.3757$

*Adapted from Goldstein (1973).

Table 3: Definition of Life Cycle Variable

Persons in household	Age of Head of Household: Under 30 yrs.	30-44 yrs.	45-59 yrs.	60 and over
1	1	3	3	3
2	1	2	2	3
3 or more	1	2	2	3

Table 4: Definition of Neighborhood Characteristics*

Variable name	Description	Level 1	Level 2	Level 3
(a) MPBL	The maximum percent of the population that is black in a five-block area including the block on which the dwelling is located and the four adjacent blocks	0-9%	10-100%	--
(b) VAC	The percent of year-round dwellings on the block that are vacant	0-9%	10-100%	--
(c) MSUB	The maximum percent of substandard units on a five-block area including the block on which the dwelling is located and the adjacent blocks	0-4%	5-100%	--
(d) RENT	The average contract rent for the block	$0-69	$70-79	$80 and above

*Source: U. S. Census of Population and Housing, Census Tract Reports, 1970.

Table 5: Life Cycle by Block Characteristics by Duration by Tenure by Move/Stay (Tract 18)

Model Fitted Marginals	Degrees of Freedom: (a)-(c)	(d)	(a) MPBL L.R.χ^2	p-value	(b) VAC L.R.χ^2	p-value	(c) MSUB L.R.χ^2	p-value	(d) RENT L.R.χ^2	p-value
(1) 1234 15 25 35 45	29	46	53.17	.004	36.48	.160	35.98	.174	56.26	.143
(2) 1234 15 25 35	30	47	68.74#	.000	54.96#	.004	50.22#	.012	74.40#	.007
(3) 1234 15 25 45	31	48	92.82#	.000	76.57#	.000	75.59#	.000	98.22#	.000
(4) 1234 15 35 45	30	48	70.75#	.000	37.60	.161	46.98#	.025	61.83*	.087
(5) 1234 25 35 45	31	48	71.68#	.000	57.14#	.003	57.83#	.003	75.77#	.007
(6) 1234 35 45 125	27	42	40.79#	.043	30.52*	.291	34.35	.156	49.91	.188
(7) 1234 25 45 135	25	42	46.99	.005	29.62	.239	29.77	.233	50.12	.183
(8) 1234 25 35 145	27	44	46.59+	.011	30.46+	.294	29.48+	.338	50.93*	.220
(9) 1234 15 45 235	27	42	42.54#	.029	28.87+	.367	33.89	.169	54.13	.100
(10) 1234 15 35 245	28	44	45.59#	.019	36.46	.131	35.78	.148	53.95	.145
(11) 1234 15 25 345	27	44	49.60	.005	33.52	.181	32.86	.202	52.82	.170

Key: L.R. difference χ^2 significant at (*) 10% level, (+) at 5% level, or (#) at 1% level.

Table 6: Estimated Effects in the Logit Model Having Main Effects Only
$(L_{ijklm} = \alpha + \beta_i + \gamma_j + \delta_k + \epsilon_l + \mu_m)$

parameter	interpretation	est. effect	std. error	parameter	interpretation	est. effect	std. error
α	constant	-0.070	0.382	δ_2	med. duration	-0.594	0.243
β_2	life cycle 2	-0.658	0.255	δ_3	long duration	-1.032	0.259
β_3	life cycle 3	-1.058	0.269	ϵ_2	renter	1.168	0.181
γ_2	med. income	-0.308	0.236	μ_2	tract 28	-0.600	0.181
γ_3	high income	0.368	0.262				

Table 7: Estimated Effects in Logit Model with Life Cycle-Tenure Interaction
$(L_{ijklm} = \alpha + \beta_{il} + \gamma_j + \delta_k \ \mu_m)$

parameter	interpretation	est. effect	std. error	parameter	interpretation	est. effect	std. error
α	constant	-0.066	0.384	γ_2	med. income	-0.282	0.238
β_{21}	l.c. 2, owner	-1.012	0.367	γ_3	high income	0.432	0.263
β_{31}	l.c. 3, owner	-0.730	0.381	δ_2	med. duration	-0.610	0.243
β_{12}	l.c. 1, renter	1.160	0.341	δ_3	long duration	-1.096	0.261
β_{22}	l.c. 2, renter	0.712	0.389	μ_2	tract 28	-0.620	0.181
β_{32}	l.c. 3, renter	-0.054	0.422				

A SIMPLE METHOD FOR EVALUATING THE SPATIAL ACCESSIBILITY OF PUBLIC SERVICES

Andrew N. White, University of Pennsylvania

Robbieburr Berger, University of Pennsylvania

1. Introduction

In an urban society, public services and public facilities play an increasingly influential role in shaping the quality of urban life. For the planner, the location, size, and composition of public facilities present some of the most serious questions regarding service quality and service user welfare: What kinds of services should be provided at a facility? At what scale should the facility be constructed? What kinds of factors should be considered in determining facility location (e.g., cost, neighborhood opposition, time frame, economies of scale and agglomeration)? From a more pragmatic point of view, it is also crucial to deal with the effectiveness of existing facilities--i.e., to reevaluate the existing public facility locations relative to changing patterns of accessibility.

Where a public service is intended to be directly available to consumers or users, it has been argued that the problem of <u>accessibility</u> plays a dominant role in the decision of where to locate the facility and its subsequent utilization (Harris, 1966; Hansen, 1959). This concern with accessibility has, in turn, given rise to a class of models using distance minimization as a basis of a decision criterion for determining the optimal location of a set of facilities to serve an urban region. These models share a common mathematical programming formulation which serves both to <u>locate</u> facilities and to <u>allocate</u> consumers to them (thereby determining the boundaries of the service areas for each facility), given a concomitant minimization of the cost of assigning consumers to given facilities.

Though these models have shortcomings, they can prove useful for evaluating the accessibility of existing public services when information is available on user needs and characteristics, as well as on general population characteristics. The present paper attempts to show how the use of these models can help planners develop strategies for changing or improving public services when they are combined with localized, small-area, continuously provided information. Note that the point we wish to stress does not concern the development of new methodologies for analyzing public facility locations, but rather that the use of more detailed information within existing methodologies increases their effectiveness as inputs to decision-making procedures.

2. Public Facility Location Models

Scott (1971) has characterized the facility <u>location-allocation</u> problem as follows: given <u>n</u> points distributed in a plane and <u>m</u> centroids to be located in the plane, find locations for the centroids and an allocation of each point to a centroid in order to optimize an objective function (e.g., minimize cost or maximize net benefits). Several authors have addressed themselves to this problem and have developed a number of algorithms, with variations accounting for facility capacity, maximum distance constraints, and pre-specification of the number of centroids to be located (e.g., Hakimi, 1964; Hakimi, 1965; Cooper, 1967; Scott, 1971; Toregas, Swain, ReVelle, and Bergman, 1971).

When we turn to realistic problems of evaluating the existing pattern of locations of public facilities, however, we note several drawbacks to the mathematical programming methods. First, consideration of census tracts as user nodes (or even a small proportion of those tracts as possible centroid locations) escalates the dimensions of the problem into a major and costly undertaking. Furthermore, the programming approach usually involves a methodological framework based upon classical, static optimization theory. For problems concerning public facility location, it is not at all clear that an "optimal location," in this sense, is the most or only important objective. The distribution of demand for such facilities will almost certainly vary over time (though not necessarily in a predictable manner). And the introduction of stochastic optimization functions and estimates of future distribution only serves to complicate an already complex problem.

The principal problem associated with the use of the usual mathematical programming methods for the location of public facilities may be traced to its generalized framework in which the location and service area of all facilities are solved simultaneously. While such methods might be warranted in a new urban area designed for a proscribed population, it is not suited to more heterogeneous urban contexts in which private and public planning coexist and many of the facilities are already fixed. In such cases, the consideration of the location of public facilities must account for a temporal sequence of decisions constrained by predetermined, bounded catchment areas which are relatively independent of each other.

We therefore conclude that (i) optimal facility location cannot be treated as an end in itself, (ii) the usual kinds of approaches may not be worth the effort expended, and (iii) individual facility locations may be treated independently in present urban contexts. The reality of public facility location, then, is a dynamic, incremental process.

Now, given all these disclaimers and provisos, what sort of procedure can be employed to determine public facility locations? After all, some method needs to be (and will be) employed, if only to provide a <u>benchmark</u> <u>index</u> against which to evaluate and compare the effects of existing patterns of facilities. In this context, optimization procedures have, in fact, proved to be quite useful---particularly where user information is available over closely spaced temporal intervals and a trend of "optimal locations" for a given facility may be obtained. It is in this context that optimization methods will be employed in this study: given the yearly data on community health facility utilization collected in Wichita, Kansas, (Sedgwick County Department of Mental Health, 1974), "optimal locations will be employed as a basis for calculating the locational deviation of the actual facility locations from an optimal facility location; the trend itself is then used to provide information on the rate of change in user accessibility. Our contention is that, used in this way, optimization methods can offer a basis for comparing alternative planning strategies vis à vis public facility location.

3. A Simple Facility Location Technique

Planners wishing to keep records of and derive trends from the optimal location of a given public facility relative to its actual location first need to identify the criteria by which "optimality" is judged. Since public services are ideally planned and located with the "public good" in mind, we propose here to employ a singular, consumer-oriented criterion for facility location; note that a more realistic treatment would also include considerations of professional and staff convenience (accessibility), scale and agglomeration economies, and other location factors (e.g., competing land uses).

The consumer or user orientation is based explicitly on the distance between the users of the service and the service facilities. Two criteria may be employed: the minimization of the aggregate distance traversed (i.e., the median point in a network of nodes and arcs) or the average distance between users and the facility (i.e., the mean point or center of gravity of a network).[1] For an urban region, areas (such as census tracts) are thus reduced to nodes and their position plotted in relation to an arbitrary orthogonal coordinate system. Distances of nodes from each other and from the coordinates' zero-value are then required for calculation of the median and mean points, respectively, which can be based on one of two metrics: (a) a euclidean distance metric (which assumes that users travel directly and in a straight line between two nodes), or (b) a metropolitan metric which incorporates the kind of resistance introduced by an orthogonal street network. Each node is then weighted by the number of users originating there and the median and mean distances are calculated as follows:[2]

$$\text{Median Distance} = \text{Min} \ \left(\sum_{j=1}^{n} w_j d_{ij} / \sum_{j=1}^{n} w_j\right) \text{ for each } i = 1,\ldots,n$$

where w_j = the weighting of the user node j, and d_{ij} = the distance between two user nodes i and j

$$\text{Mean} = \left(\sum_{j=1}^{n} w_j d_{ij} / \sum_{j=1}^{n} w_j\right) \text{ for each } i = 1,\ldots,n$$

where d_j = the distance of user node j from the zero-value of the reference coordinate system

Now, in order to depict the trend in optimal location and in the deviation of optimal from actual location, the mean or median point must be calculated at regular intervals over a period of time. As noted above, the information requirements include the availability of data on number of service users by origin node and destination point (public facility location) over several years. Furthermore, due to the relative ease with which the deviation of the median location from the actual location of a given public facility can be interpreted as the difference in aggregate user distance travelled and the immediate implications for user accessibility understood, the median point location method was selected as the more appropriate measure. An extension of this general method is also employed: weighting of user nodes. In effect, the number of users originating at a given node (census tract) can be considered as a form of weight and, by further differentiating the population by other properties, additional kinds of weights may also be obtained. Our argument for applying these alternative weighting functions is as follows: in public facility location, concern with consumer distance and accessibility inevitably raises the question of relative ability to overcome distance, i.e., the relative accessibility of the user population. For example, the poor and the elderly, who have fewer resources to spend on transportation, generally can be considered to have low relative accessibility to public services at any given location. Availability of conventional public transportation may ameliorate the problem but it cannot change the essential inequity. Planners who are faced with determining facility locations or evaluating the effects of relative accessibility to presently located facilities must therefore bear in mind such considerations and should, where possible, weight the location problem in favor of those people with poor accessibility.

4. An Example: Outpatient Mental Health Facilities

Wichita, Kansas has a population of about 350,000 people and thus is representative of medium-sized American cities. While the city and the surrounding county as a whole define a single catchment area for community mental health services, inpatient and outpatient services have been partitioned into a north district and a south district with one center for each. Their locations are marked on the accompanying map. (See Figure 1.) The derivation of optimal locations of outpatient centers and comparison with existing facilities was therefore carried out for the two districts separately in order to determine the trend in the accessibility of services relative to users over several years of operation of the centers.

Using a metropolitan metric in order to reflect actual distances over an urban street pattern and its influence on intra-urban travel, and an orientation of the reference coordinate system in conformity with the north-south, east-west directions of the grid pattern of streets in Wichita, the median location of outpatient centers corresponding to each district's user distribution for the years 1971, 1972, and 1973 was calculated. The first set of calculations was based upon the use of weights reflecting only the number of users of outpatient services originating at a given census tract. The second set of calculations involved a form of "welfare weighting" of the nodes of origin.[3] While such weights were necessarily averages reflecting the properties of the entire population of the given census tract, they serve to index the differential accessibility by user resources as well as distance.

Employing a unique set of census records that are updated yearly in Wichita, the welfare weighting was devised to take account of income and vehicle ownership in a multiplicative relationship such that the disadvantage of the poor with no vehicles as opposed to the poor with vehicles and the discrepancy between the poor and the well-off with one or more vehicles would be emphasized. Where a_j = the service users at node j, Y_j = the percentage of the population at node j with income less than $5000, and v_j = the average number of vehicles owned per household in node j, the welfare weighting of any given node is $W_j = a_j/(100.0 - Y_j)v_j$.

The weighted and unweighted sets of optimal facility locations over time are plotted in Figure 1. The unweighted optimal location is shaded with horizontal lines and the weighted location with vertical lines. (Since the weighted locations for the south district change for the two years, the location for 1973 is shaded diagonally.) The actual location is cross-hatched. The deviation calculation involves the determination of the difference between the aggregate distance travelled by all users to the actual outpatient location minus that of the optimal location. The trend in the distance deviation from optimal location (both weighted and unweighted), as well as the trend in the masses at those locations, is provided in Table 1.

As can be seen from Table 1, the unweighted optimal location in both the north and south districts remains constant over the three years. However, there is a significant difference in the deviations of the optimal location from the actual location: users in the south district face a distance deviation nearly 2½ times that faced by users in the north district. However, the picture with respect to the weighted optimal locations is less clear. First, while the optimal location in the north district is the same for the two years, this is not true for the south district. The optimal location in the latter case shifts from census Tract 32 to Tract 39 (i.e., the unweighted optimal location). This, of course, leaves us with the question: Does this observation result from an actual shift in the user population (or characteristics of that population), or does it reflect other possible biases in the data? The answer can only come with further analysis of public service use based on data which is not restricted to the areal aggregations provided by the decennial census. Also, it should be noted that in the north district, the distance deviation for the weighted location is less than that for the unweighted one. By taking the welfare of users into consideration, we decrease the distance between the optimal and actual locations, i.e., we move the optimal location closer to the actual one. With respect to the south district, the deviation increases for 1972 and remains the same for 1973. Finally, notice that for both districts and for both the unweighted and weighted locations, the masses (or "weights") increase from year to year. This implies that, in proportion to the total number of users, the number of service users in more distant areas is getting larger.[4] This probably reflects population growth (and therefore health service user growth) on the metropolitan fringes.

5. Conclusions and Policy Implications

Using outpatient mental health services in Wichita, Kansas as an example, it has been shown that elaborate mathematical programming methods need not be resorted to in order to calculate optimal facility locations, especially when facilities can be treated independently. A relatively simple location technique was employed in order to derive optimal facility location with respect to user accessibility and to measure the trend in user accessibility over time via location deviation measures.

In effect, by using methods akin to that described above, public service planners are in a position to know whether or not a service facility at a given location is readily accessible to the people who need the service the most and those who have the most difficulty overcoming spatial separation. Moreover, with the availability of data over time, the trend in accessibility (indexed by the location deviation trend) can be determined. In cases where accessibility continues to decrease over time, the planner may be faced with a number of response options, including: relocating the facility, locating more facilities, changing the boundaries of user areas, or helping to increase the mobility of either the service (e.g., via mobile home trailers, van, or the like) or the users, perhaps by subsidizing their trips to the facility or by providing a shuttle bus system (Lankford, 1974). These options can only be evaluated in the light of information on the relative accessibility of the user population to public facilities over time. The methodology provided here is one means of obtaining that information quickly and cheaply.

Footnotes

1. Properties of the median point include the representation of the point of minimum aggregate travel, location with relatively little influence from extreme user node locations, and location that is not invariant with respect to the coordinate system to which nodes are referred.

The median point is calculated by finding that user node which has the smallest total sum of distances from all other user nodes. Properties of the mean point include an invariance with respect to the coordinate system used for positional reference, but greater influence from extreme user node locations and less intuitive meaning attached to the interpretation of the optimal location, which is calculated by taking the average distance of all user nodes to the zero-value of an arbitrary reference coordinate system (Neft, 1966; King, 1969; Scott, 1971).

2. Note that since the calculation of the mean results in a single number, a distance which is either the constant sum of the x- and y-distances in metropolitan metric or the euclidean distance between the mean point and the zero-value, two calculations must be made utilizing a different coordinate system (e.g., shift in measurement of x-coordinate distances) in order to yield a "complementary" distance value whose line plotted with the former will intersect at the mean point. Alternatively, for the metropolitan metric, the mean value of each coordinate can be determined separately and then paired to determine the coordinates of the mean location. The determination of the median point is made simply by keeping track of that user node which achieves the minimum aggregate distance.

3. Data were available for 1972 and 1973 only. In addition, the data on income for 1973 were incomparable with those of 1972, so that the percentage of the population having incomes from $4000 to $7000 was calculated, and one-third of this percentage was added to the percentage having income under $4000. This might introduce a bias in the findings.

4. A simple increase in the total number of users does not increase the mass. Hence, since the distances remain constant over time, the changes in masses are due to proportionate changes in the user population.

REFERENCES

Cooper, L. (1967), "Solutions of Generalized Locational Equilibrium Models," Journal of Regional Science, 7:1-18.

Hakimi, S. (1964), "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," Operations Research, 12:450-459.

Hakimi, S. (1965), "Optimum Distribution of Switching Centers in a Communications Network and some Related Graph Theoretic Problems," Operations Research, 13:462-475.

Hansen, W. (1959), "How Accessibility Shapes Land Use," Journal of American Institute of Planners, v. 25n.2(May):73-76.

Harris, B. (1966) "Notes on Accessibility," Department of City and Regional Planning, University of Pennsylvania (mimeo).

King, L. (1969), Statistical Analysis in Geography, (Englewood Cliffs, New Jersey: Prentice-Hall).

Lankford, P. (1974), "Physician Location Factors and Public Policy," Economic Geography, 50 (July):244-255.

Neft, D. (1966), "Statistical Analysis for Areal Distributions," Monograph Series No. 2, Regional Science Research Institute (Philadelphia, Pennsylvania).

Scott, A. (1971), Combinatorial Programming, Spatial Analysis, and Planning (London: Methuen).

Sedgwick County Department of Mental Health (1974), "Utilization of Direct Service by Census Tract and Unit for 1971-1973."

Toregas, C., Swain, R., ReVelle, C., Bergman, L. (1971), "The Location of Emergency Service Facilities," Operations Research, 19 (October):1363-1373.

TABLE 1

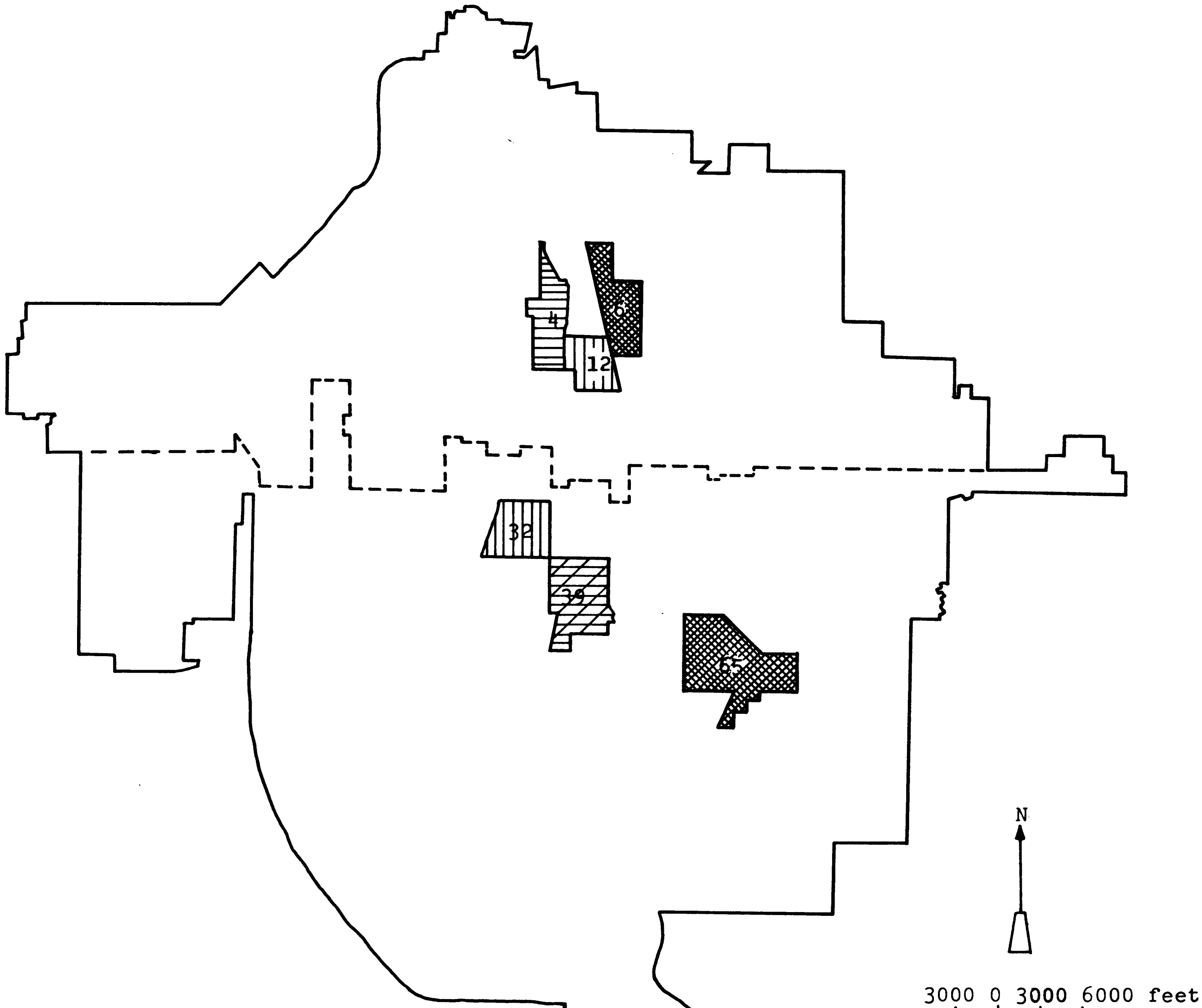
South District

		optimal location (CT)	actual location (CT)	deviation distance (ft)	mass
	1971	39	65	14362	14767.0
unweighted	1972	39	65	14362	15448.76
	1973	39	65	14362	15623.82
weighted	1972	32	65	22361	12229.17
	1973	39	65	14362	14739.94

North District

		optimal location (CT)	actual location (CT)	deviation distance (ft)	mass
	1971	4	6	5817	13124.98
unweighted	1972	4	6	5817	13416.66
	1973	4	6	5817	13779.83
weighted	1972	12	6	5454	9665.71
	1973	12	6	5454	11302.05

Figure 1



Source: 1970 Census of Population and Housing

Wichita

ON THE ANALYSIS OF FACTORIAL EXPERIMENTS WITH TYPE I CENSORED DATA

Douglas A. Zahn, The Florida State University

1. Introduction

Consider a factorial experiment with n replications per treatment combination in which the observations are subject to Type I (time) censoring. Under this type of censoring the observations are known exactly only if they are less than or equal to an *a priori* selected value T_0. All that is known about other observations is that they exceed T_0. Only censoring on the right is considered here, since the extension to censoring on the left is straightforward. The problem considered in this paper is how to estimate and test the parameters in a typical linear model for the mean observation at each treatment combination using Type I censored data. The models for the distribution of the dependent variable considered here are the exponential, Weibull, Type I extreme-value, normal, and lognormal distributions.

The above problem arose in the course of current research by West, *et al.* (1974) in social psychology investigating the phenomenon of helping behavior and what variables influence helping behavior. A 2^4 factorial experimental design was used to investigate the relationship between the dependent variable Y and four independent variables. Here Y=elapsed time until a passing motorist stops to help a confederate, i.e., an employee of the experimenter, standing at the roadside next to a car with a raised hood. The four two-level factors are the sex and race of the confederate, the racial composition of the neighborhood in which the experimental trial occurred, and the proximity of the trial to a college campus. There were eight trials per treatment combination. For economic reasons each trial was halted at 15 minutes if no help had been given. There was no replacement if help was given prior to 15 minutes. Note that in this problem, the Type I censoring occurs at T_0 = 900 seconds.

The similarity of this problem to problems encountered in life testing situations is apparent. In life testing situations the dependent variable is typically time until failue or death, whereas the dependent variable here is time until helping or time until success. Clearly, this poses no difficulty.

Section 2 summarizes results available in the statistical literature relating to parameter estimation and hypothesis testing for various experimental designs when the data are Type I, censored on the right, and when the model being considered is exponential, Weibull, Type I extreme-value, normal, or lognormal. Research relating to nonparametric approaches to this problem is not considered. Section 3 describes a general procedure utilizing linear models and maximum likelihood estimatiors (MLE's) for the analysis of factorial experiments with censored data. Section 4 illustrates the application of this procedure and Section 5 presents concluding remarks.

2. Literature review

After the author was first confronted by the social psychological problem given in Section 1, an extensive review of the life testing literature was undertaken. Though the literature on the analysis of Type I censored data contains much work on components of the problem considered here, a complete solution was not found. This section reviews results for problem components, and Section 3 presents a procedure which synthesizes these components into a complete solution to the problem. This section is organized as follows. Results are presented for five parametric models mentioned in Section 1. For each model results relating to the MLE's of the parameters in a one-sample problem are presented. This is a problem in which all observations are gathered under one set of conditions. Then results relating to best linear unbiased estimates (BLUE's) are given. Most of the work relating to the problem posed in Section 1 has been done for a one-sample problem. Progress on estimation and testing problems in one-, two-, and k-sample problems is presented. The review then considers any work done on testing and estimation relative to linear models for the parameters of the distributions.

Bibliographies on life testing and related results have been published by Mendenhall (1958) and Govindarajulu (1964). A new book which summarizes much work in this area is Mann, Schafer, and Singpurwalla (1974).

2.1. The exponential model

Let $t_1, t_2, \ldots, t_n$ denote observations made on a random variable T having the (negative) exponential distribution with cumulative distribution function

$$F_T(t) = 1 - \exp(-t/\theta),\ 0 \le t < \infty,\ 0 < \theta < \infty,$$
$$= 0,\ t < 0 .$$

If the data are Type I censored at time T_0, the exact values of the t_i are known only if $t_i \le T_0$. Define

$$x_i = t_i \text{ , if } t_i \le T_0 \text{ ,}$$
$$= T_0 \text{ , if } t_i > T_0 \text{ , for } i = 1,2,\ldots,n, \text{ and}$$

let r denote the number of $t_i \le T_0$. Using life testing terminology, T is the time to failure and r is the number of failures before the censoring time. The parameter θ is the mean of T and, in a life testing context, is referred to as the mean time between failures. Of course, the exponential density can also be parametrized using the failure rate $\lambda = 1/\theta$. The failure rate is defined to be $f(t)/[1-F(t)]$, where f(t) is the probability density function of T.

Various authors, including Bartlett (1953), Bartholomew (1957), Deemer and Votaw (1955), and Littel (1952), have shown that the MLE of θ when the data are Type I censored is $\hat{\theta} = \sum_{i=1}^{n} x_i/r$.

Deemer and Votaw also showed that the asymptotic variance of $\hat{\theta}$ is

$$\theta^2[n(1 - \exp(-T_0/\theta)]^{-1} .$$

Since $\hat{\theta}$ is an MLE, and certain mild regularity conditions (Mann, *et al.*, 1974, p. 82) are satisfied here, the sampling distribution of $\hat{\theta}$ is asymptotically normal.

For small samples, Bartholomew (1957) found that $\hat{\theta}$ was biased and provided an exact expression for the bias. Mendenhall and Lehman (1960) gave tables to aid in computing the exact mean and variance of $\hat{\theta}$. Bartholomew (1963) derived the exact distribution of $\hat{\theta}$, which is too cumbersome unless n is very small. He suggested that n be greater than 40 for $\exp(-T_0/\theta)=0.10$ or n be greater than 80 for $\exp(-T_0/\theta) = 0.25$ before assuming that $\hat{\theta}$ is approximately normal. He derived another statistic, x in his notation, which is approximately normal for n as small as ten, where

$$x = r\left(\frac{\hat{\theta} - \theta}{\theta}\right)\{n[1 - \exp(-T_0/\theta)]\}^{-\frac{1}{2}} .$$

Note that Bartholomew's Equation (7) for x contains a typographical error. This statistic can be used to test hypotheses about θ or to construct confidence intervals for θ in one-sample problems.

For the exponential model, results for BLUE's comparable to those cited above for MLE's have not been published, to the best of the author's knowledge. This is not surprising since the MLE's can be expressed in closed form in this situation. Of course, it would be possible to construct BLUE's using results published for the Weibull distribution, since a two-parameter Weibull random variable with shape parameter equal to one is an exponential random variable.

No results were found relating to hypothesis testing or confidence interval estimation for the exponential parameter θ in two- or k-sample problems with Type I censoring. Mann, *et al.* (1974) give a test for θ in a two-sample problem, but their model assumes Type II censoring. Zelen (1959) discusses the analysis of a factorial experiment when the dependent variable is assumed to be exponentially distributed and the data are Type II censored. The data at each treatment combination are said to be Type II censored if the replications are observed until a preselected number r fail. At the time of the r'th failure for a particular treatment combination, all remaining unfailed replications are censored.

A number of authors have considered the problem of estimating the parameters of a linear model for θ in the exponential distribution. A typical situation is one in which the random variable is survival time in a medical study and the data analysis is to examine whether the survival time varies by some covariate, such as age. However, none of the work in this area provides a solution to the problem posed in Section 1 because either the articles do not consider censored data (Fiegl and Zelen, 1965; Glasser, 1967) or they do not consider linear models using Type I censored data (Cox, 1964; Sprott and Kalbfleisch, 1969; Cox, 1972; Prentice, 1973). Zippin and Armitage (1966) generalize the work of Fiegl and Zelen (1965) by considering censored data. The method of Zippin and Armitage could be used to solve the Section 1 problem in the case of one factor at two levels. For more complicated experimental designs, their method would be much more cumbersome and time-consuming than the method proposed in Section 3. For instance, one part of their method would require the iterative solution of k simultaneous non-linear equations in k unknowns, where k is the number of treatment combinations in the experimental design under consideration.

2.2. The Weibull and Type I extreme-value distributions

Let $t_1,t_2,\ldots,t_n$ denote observations on a random variable T having a two-parameter Weibull distribution with cumulative distribution function

$$F_T(t) = 1 - \exp[-(t/\alpha)^{\beta}],\ 0\le x<\infty,\ 0<\alpha,\ \beta<\infty .$$

The parameter α is the scale parameter and β is the shape parameter of the Weibull distribution. If β=1, the distribution has a constant failure rate and T is exponential with parameter α. If β<1, the distribution has a decreasing failure rate, while if β>1, it has an increasing failure rate. It is well known that $Y=\ell n T$ has the Type I (smallest) extreme-value distribution with cumulative distribution function

$$F_Y(y) = 1 - \exp\{-\exp[(y - u)/b]\},\ -\infty < y < \infty ,$$

where b=1/β, $u=\ell n\alpha$. (All logarithms used in this paper are natural logarithms.) Thus, methods developed for one of these models can also be used for the other.

For Type I censored data, Cohen (1965) has developed a procedure for determining the MLE's of the parameters α and β. Let r denote the number of failures before the censoring time T_0. The first step is to obtain the MLE $\hat{\beta}$ by solving the following equation for $\hat{\beta}$:

$$\frac{\sum_{i=1}^{r} t_i^{\hat{\beta}}\ell n t_i + (n-r)T_0^{\hat{\beta}}\ell n T_0}{\sum_{i=1}^{r} t_i^{\hat{\beta}} + (n-r)T_0^{\hat{\beta}}} - \frac{1}{\hat{\beta}} - \frac{1}{r}\sum_{i=1}^{r}\ell n\ t_i = 0 .$$

Then the MLE $\hat{\alpha}$ is obtained by solving the following equation for $\hat{\alpha}$:

$$\hat{\alpha}^{\hat{\beta}} = \frac{1}{r}\sum_{i=1}^{r} t_i^{\hat{\beta}} + (n - r)T_0^{\hat{\beta}} .$$

Cohen also derived the asymptotic variance-covariance matrix of $(\hat{\alpha},\hat{\beta})$.

Methods for estimating the parameters of a linear model for a percentile of the distribution of T when the data are subject to Type II censoring have been presented by Lieblein and Zelen (1956). Their approach is similar to that of Nelson and Hahn (1972, 1973). Additional results relating to linear estimation are given by Mann, *et al.* (1974).

2.3. The normal and lognormal models

Let $t_1,t_2,\ldots,t_n$ denote n independent observations of the random variable T which is normally distributed with mean μ and variance σ^2. It is well known that, if $T=\ell n V$, then V is lognormally distributed. Thus, methods developed for one of these models can be used for the other. For Type I censored data, Cohen (1961) gives formulas and tables for evaluating the MLE's of μ and σ^2 and gives asymptotic formulas for the variances of the MLE's. The author is not aware of any published results relating to BLUE's for μ and σ^2 in the Type I censoring model.

Nelson and Hahn (1972, 1973) develop methods for estimating the parameters of a linear model for μ when the data are Type II censored. They consider methods which utilize BLUE's for μ at each setting of the independent variables and thereby ignore the information in the censored observations. They maintain that the distinction between Type I and Type II censoring can be ignored in practical situations. This contention and a comparison of their method to the method of Section 3 are topics of continuing research. Sampford and Taylor (1959) considered the analysis of Type II censored data in a randomized blocks experiment using a normal model.

3. A general procedure for testing and estimation in the analysis of factorial experiments with Type I censored data

This section describes a general procedure which can be used for the analysis of factorial experiments with Type I censored data in the many situations noted in Section 2 for which there currently is no analytic method available. The procedure will be described using a two-factor experiment. Extensions to experiments containing more factors are obvious.

Consider a factorial experiment in which factor A at a levels is crossed with factor B at b levels with n replications at each treatment combination. Label the treatment combinations (i,j), i=1,2,...,a; j=1,2,...,b. The data consist of abn observations on the dependent variable T. There is Type I censoring at the value T_0 at each treatment combination. Assume that interest centers on θ_{ij}, a parameter of the distribution of T in treatment combination (i,j), and how θ_{ij} varies across the treatment combinations. This section develops a procedure for investigating the θ_{ij}'s without making any restrictive assumptions about the specific form of the distribution of T.

To investigate the relationship between θ_{ij} and the factors A and B, consider the following linear model for the θ_{ij}:

$$\underline{\theta}_{ab\times 1} = \underline{X}_{ab\times ab}\,\underline{\beta}_{ab\times 1} ,$$

where $\underline{X}$ is the design matrix of the factorial experiment in reduced form so as to be nonsingular. For example, if a=2, b=3,

$$\underline{X} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} .$$

A model of this type is illustrated by Draper and Smith (1966, p. 257).

Interest centers on the last ab-1 parameters in $\underline{\beta}$, since they represent, respectively, the main effects of A, the main effects of B, and the AB two-factor interactions. Typically, the first parameter in $\underline{\beta}$ is of little interest since it represents the grand mean.

The first step in the estimation of $\underline{\beta}$ is the estimation of $\underline{\theta}$. The results quoted in Section 2 are used to compute $\hat{\theta}_{ij}$ and $\hat{V}(\hat{\theta}_{ij})$, the MLE of θ_{ij} and an estimate of its variance, for each treatment combination (i,j), using the data available on that combination. For each of the models considered in Section 2, $\hat{\theta}_{ij}$ is known to be asymptotically normally distributed with mean θ_{ij} and variance $V(\hat{\theta}_{ij})$. Since independent samples are observed at each treatment combination, $\hat{\underline{\theta}}\sim AN[\underline{\theta},\underline{V}(\hat{\underline{\theta}})]$, where $\underline{V}(\hat{\underline{\theta}})$ is an ab×ab diagonal matrix, whose (k,k) element is $V(\hat{\theta}_k)$, the variance of the k'th element of $\underline{\theta}$. Since $\underline{\beta}=\underline{X}^{-1}\underline{\theta}$, $\hat{\underline{\beta}}$, the MLE of $\underline{\beta}$, is $\hat{\underline{\beta}}=\underline{X}^{-1}\hat{\underline{\theta}}$. Also, $\hat{\underline{\beta}}\sim AN[\underline{\beta},\underline{X}^{-1}\underline{V}(\hat{\underline{\theta}})(\underline{X}^{-1})']$. Letting $\underline{C}=(c_{ij})=\underline{X}^{-1}$, it is easy to see that

$$\frac{\hat{\beta}_k - \beta_k}{\left[\sum_{j=1}^{ab} c_{kj}^2 V(\hat{\theta}_j)\right]^{\frac{1}{2}}} \sim AN(0,1) .$$

Use of this statistic is complicated by the fact that $V(\hat{\theta}_j)$ may depend on the unknown parameter θ_j. However, since $\hat{\theta}_j$ is a strongly consistent estimator for θ_j, $\hat{\theta}_j$ may be substituted for θ_j in the variance, yielding $\hat{V}(\hat{\theta}_j)$, the estimated variance of $\hat{\theta}_j$, and the following result still holds (Cramér, 1945, p. 254):

$$\frac{\hat{\beta}_k - \beta_k}{\left[\sum_{j=1}^{ab} c_{kj}^2 \hat{V}(\hat{\theta}_j)\right]^{\frac{1}{2}}} \sim AN(0,1) .$$

Define

$$V(\hat{\beta}_k) = \sum_{j=1}^{ab} c_{kj}^2 V(\hat{\theta}_j) .$$

The experimenter who wishes to test hypotheses about individual effects now has sufficient tools to proceed. A Bonferroni inequality can be used to develop simultaneous confidence intervals for the β_k's or to develop a test procedure which controls the experimentwise error rate (EER) at some prescribed level α. The EER is defined to be the probability of declaring at least one false positive in the analysis of the experiment, if in fact all effects are equal to zero. Note that β_1, the grand mean, is of no interest here, so that ab-1 β_k's are being tested. The Bonferroni inequality (Miller, 1966, p. 8) states that

$$P\Big(|\hat{\beta}_k-\beta_k|/[V(\hat{\beta}_k)]^{\frac{1}{2}} \le z\{1-\alpha/[2(ab-1)]\},\ \text{for } k = 2,3,\ \ldots,\ ab\Big)$$

$$\ge 1-\sum_{k=2}^{ab} P\Big(|\hat{\beta}_k-\beta_k|/[V(\hat{\beta}_k)]^{\frac{1}{2}} > z\{1-\alpha/[2(ab-1)]\}\Big)$$

$$= 1 - \alpha ,$$

where $z\{\gamma\}$ denotes the 100(1-γ)th percentile of the standard normal distribution. Thus, a procedure yielding EER less than or equal to α is to declare β_k significantly different from zero if

$$|\hat{\beta}_k|/[V(\hat{\beta}_k)]^{\frac{1}{2}} > z\{1-\alpha/[2(ab-1)]\},\ \text{for } k=2,3,\ldots,ab.$$

With probability greater than or equal to 1-α, all β_k are contained in the intervals

$$\hat{\beta}_k \pm z\{1-\alpha/[2(ab-1)]\}[V(\hat{\beta}_k)]^{\frac{1}{2}},\ \text{for } k=2,3,\ldots,ab .$$

4. An illustration of the use of the general procedure

This section presents the application of the procedure described in Section 3 to the data from the 2^4 factorial experiment discussed in Section 1. The extensions of the procedure required for the analysis of data from other designs are straightforward.

To begin, an exponential model for the "time to helping" random variable will be tentatively entertained. This model is reasonable, given the structure of the example being considered here.

Let P, N, S, R denote the four factors in the design, each of which appears at two levels. Let

t_{hijkm} = time to helping for trial m at level h of factor P (college proximity), level i of factor N (racial composition of neighborhood), level j of factor S (sex), level k of factor R (race), $m = 1,2, \ldots, M$,

denote M observations of the random variable T_{hijk}, $h,i,j,k=1,2$. In the example considered here "helping" occurs when someone stops to help the confederate and "replication m" is the m'th trial at a particular treatment combination. Assume

$$f(t_{hijkm}) = (1/\theta_{hijk})\exp(-t_{hijkm}/\theta_{ijkm}) ,$$
$$0 \le t_{hijkm} < \infty , \; 0 < \theta_{hijk} < \infty .$$

Now the t_{hijkm} are known exactly only if $t_{hijkm} \le T_0$=900 seconds, the censoring time. Define

$$x_{hijkm} = t_{hijkm} , \text{ if } t_{hijkm} \le T_0 ,$$
$$= T_0 , \text{ if } t_{hijkm} > T_0 ,$$
$$r_{hijk} = \text{number of } t_{hijkm} \le T_0 ,$$

= number of uncensored observations on treatment combination hijk, $h,i,j,k, = 1,2$.

Bartholomew (1963) noted that the MLE of θ_{hijk} is

$$\hat\theta_{hijk} = \sum_{m=1}^{M} x_{hijkm} \Big/ r_{hijk} ,$$

and that the sampling distribution of $\hat\theta_{hijk}$ is asymptotically normal with mean θ_{hijk} and variance

$$V^*(\hat\theta_{hijk}) = \theta^2_{hijk}\{M[1 - \exp(-T_0/\theta_{hijk})]\}^{-1} .$$

He also showed that for small M, y_{hijk} (x in his notation) is more nearly normal than $\hat\theta_{hijk}$, where

$$y_{hijk} = r_{hijk}(\hat\theta_{hijk} - \theta_{hijk}) \Big/ \left(\theta_{hijk}\{M[1 - \exp(-T_0/\theta_{hijk})]\}^{\frac{1}{2}}\right) .$$

The following material assumes that

$$\hat\theta_{hijk} \stackrel{.}{\sim} N[\theta_{hijk}, V(\hat\theta_{hijk})], \text{ where}$$
$$V(\hat\theta_{jijk}) = \theta^2_{hijk}\{M[1 - \exp(-T_0/\theta_{hijk})]\} \Big/ r^2_{hijk} .$$

To express the θ_{hijk} in terms of the usual reduced model for a 2^4 factorial experiment, write

$$\underline{\theta}_{16\times 1} = \underline{X}_{16\times 16}\, \underline{\beta}_{16\times 1} ,$$

where $\underline{X}$ denotes the design matrix for the 2^4 experiment. To illustrate the structure of $\underline{X}$, consider a 2^2 experiment. In this situation

$$\underline{X} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} .$$

For the 2^4 experiment β_1 represents the grand mean; β_2, β_3, β_4, β_5 represent the main effects of the four factors; $\beta_6, \beta_7, \ldots, \beta_{11}$, the two-factor interactions; β_{12}, β_{13}, β_{14}, β_{15}, the three-factor interactions; and β_{16}, the four-factor interaction. Note that $\underline{X}\,\underline{X}'=16\underline{I}$, where $\underline{I}$ denotes the 16×16 identity matrix. Hence, $\underline{X}^{-1}=(1/16)\underline{X}'$. This implies that $\underline{\beta}=\underline{X}^{-1}\underline{\theta}=(1/16)\underline{X}'\underline{\theta}$ and the MLE of $\underline{\beta}$ is given by $\hat{\underline{\beta}}=(1/16)\underline{X}'\hat{\underline{\theta}}$. Thus, the analysis first uses the individual observations to estimate the parameter θ for each treatment combination. Then the $\hat\theta$'s are used to estimate the parameters in the linear model. Individual parameters can be estimated by noting that

$$\hat\beta_i = (1/16)\underline{X}_i'\hat{\underline{\theta}} = (1/16)\sum_{j=1}^{16} X_{ij}\hat\theta_j ,$$

where $\underline{X}_i'$ denotes the i'th row of $\underline{X}$. Thus,

$$\hat\beta_i \stackrel{.}{\sim} N[\beta_i, V(\hat\beta)] , \text{ where}$$
$$V(\hat\beta) = (1/16)^2 \sum_{j=1}^{16} V(\hat\theta_j) .$$

A Bonferroni-type procedure as described in Section 3 may now be used to test hypotheses about the parameters in the linear model, while controlling the EER. Table 1 gives $\hat\beta_i$, $i=2,3,\ldots,16$, and a set of 95% confidence intervals for each of the parameters. These intervals correspond to the usual 0.05 level F-tests of individual effects and interactions performed on the results of a 2^4 factorial experiment using the conventional normal theory model. A second set consisting of 100(1-0.05/15)%=99.67% confidence intervals for each of the parameters can be constructed. The intervals in this second set are of the form $\hat\beta_i \pm 546.9$. Using these intervals to test hypotheses about the parameters yields EER≤0.05. An examination of Table 1 indicates that only the main effect of sex is significant when the individual parameters are tested at an 0.05 level. Using EER≤0.05, this effect is almost significant. Hence, this analysis of this experiment indicates that the only factor which significantly affected the mean time to helping was the sex of the confederate. Women were helped much faster than men.

A complete analysis of these data should also include a scrutiny of the assumption in the tentatively entertained model that the times to helping are exponentially distributed. Considerable work has been done on the question of model choice in this situation. Fercho and Ringer (1972) examined four tests of exponentiality and recommended the Gnedenko test as given by Mann, *et al.* (1974) when testing against a Weibull alternative in the presence of censored data. When this test was applied to these data at each of the treatment combinations, the hypothesis that the times are exponentially distributed was not rejected for any treatment combination.

It is exceedingly important to use a model which properly recognizes the censoring and the exponentiality of the data . To demonstrate this, an analysis using a conventional fixed effects analysis of variance model will now be presented. This approach is based on the assumption that the data are normally distributed and not censored. Not surprisingly this conventional model is employed by West *et al.* (1974) in their analysis of these data. The parameter estimates under this model, denoted $\hat{\beta}_1^N$, are given in the last column of Table 1. Comparing $\hat{\beta}_1^N$ to $\hat{\beta}_1$ indicates the dire consequences of ignoring the censoring and exponentiality. Many parameters are seriously underestimated. The normal model also seriously understates the variation in the parameter estimates. For the normal model, 95% and 99.67% confidence intervals are of the form $\hat{\beta}_1^N \pm 53.3$ and $\hat{\beta}_1^N \pm 78.9$, respectively. For the exponential model, they are $\hat{\beta}_1 \pm 367.1$ and $\hat{\beta}_1 \pm 546.9$, respectively. These difficulties stem from the following facts. Under the normal model ignoring the censoring, the treatment combination means are determined by dividing the total time on test (the sum of the eight times to helping) by eight, the number of subjects tested on that treatment combination. However, the correct estimate of the mean, as determined under the exponential model, is evaluated by dividing the total time on test by the number of failures. For instance, for the treatment combination RNP, the estimate of the mean time to helping using the normal model ignoring censoring is $\hat{\theta}^N_{RNP}=5940/8=742.5$ seconds. Using the exponential model which utilizes the information on censoring, the estimate is $\hat{\theta}_{RNP}=5940/3=1980$ seconds.

It should be noted that the purpose of this section has been to illustrate the use of the procedure described in Section 3. Had the purpose been to present an extensive analysis of the data set,additional variables, such as cars per minute and the race of the passing motorists, would have been considered. The complete analysis of these data is considered in another report.

5. Concluding comments

Two cautionary remarks should be made. The method proposed here is based on the asymptotic normality of MLE's. Thus, it should be used with caution when the number of replications at individual treatment combinations is small, as results by Bartholomew (1963) and Billman, Antle, and Bain (1972) indicate. Also, Zelen and Dannemiller (1961) have demonstrated the lack of robustness of procedures based on the exponential model when the true model is Weibull with shape parameter less than one. Hence, as in other applications, in the words of G. E. P. Box, one should not make the Pygmalian mistake of falling in love with the model. If the data indicate that the model is likely to be Weibull, the MLE's for the Weibull parameters should be used.

6. Acknowledgements

The author wishes to thank Ismail Shimi, Allan Sampson, and Frank Proschan for helpful conversations. He is also grateful to Susan Lattin and John Conlon for their assistance.

REFERENCES

[1] Bartholomew, D. J. (1957). A problem in life testing. J. Amer. Statist. Assoc. 52, 350-355.

[2] Bartholomew, D. J. (1963). The sampling distribution of an estimate arising in life testing. Technometrics 5, 361-374.

[3] Bartlett, M. S. (1953). Approximate confidence intervals. Biometrika 40, 12-19.

[4] Billman, B. R., Antle, C. E. and Bain, L. J. (1972). Statistical inference from censored samples. Technometrics 14, 831-840.

[5] Cohen, A. C., Jr. (1961). Tables for maximum likelihood estimates: singly truncated and singly censored samples. Technometrics 3, 535-541.

[6] Cohen, A. C., Jr. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and censored samples. Technometrics 7, 579-588.

[7] Cox, D. R. (1964). Some applications of exponential ordered scores. J. R. Statist. Soc. B 26, 103-110.

[8] Cox, D. R. (1972). Regression models and life-tables. J. R. Statist. Soc. B 34, 187-202.

[9] Cramér, H. (1945). Mathematical Methods of Statistics. Princeton: Princeton University Press.

[10] Deemer, W. L., Jr. and Votaw, D. F., Jr. (1955). Estimation of parameters of truncated or censored exponential distribution. Ann. Math. Statist. 26, 498-504.

[11] Draper, N. and Smith, H. (1966). Applied Regression Analysis. New York: John Wiley and Sons.

[12] Fercho, W. W. and Ringer, L. J. (1972). Small sample power of some tests of the constant failure rate. Technometrics 14, 713-724.

[13] Fiegl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. Biometrics 21, 826-838.

[14] Glasser, M. (1967). Exponential survival with covariance. J. Amer. Statist. Assoc. 62, 561-568.

[15] Govindarajulu, R. (1964). A supplement to Mendenhall's bibliography on life testing and related topics. J. Amer. Statist. Assoc. 59, 1231-1291.

[16] Lieblein, J. and Zelen, M. (1956). Statistical investigation of the fatigue life of deep-groove ball bearings. J. of Res. National Bureau of Standards 57, 273-316.

[17] Littel, A. S. (1952). Estimation of the T-year survival rate from follow-up studies over a limited period of time. Human Biology 24, 87-116.

[18] Mann, N. R., Schafer, R. E. and Singpurwalla, N. D. (1974). Methods for Statistical Analysis of Reliability and Life Data.

[19] Mendenhall, W. (1958). A bibliography on life testing and related topics. Biometrika 45, 521-543.

[20] Mendenhall, W. and Lehman, E. H., Jr. (1960). An approximation to the negative moments of the positive binomial useful in life testing. Technometrics 2, 227-242.

[21] Miller, R. G., Jr. (1966). Simultaneous Statistical Inference. New York: McGraw Hill.

[22] Nelson, W. and Hahn, G. J. (1972). Linear estimation of a regression relationship from censored data - Part I: Simple methods and their application. Technometrics 15, 247-269.

[23] Nelson, W. and Hahn, G. J. (1973). Linear estimation of a regression relationship from censored data - Part II: Best linear unbiased estimation and theory. Technometrics 15, 133-150.

[24] Prentice, R. L. (1973). Exponential survival with censoring and explanatory variables. Biometrika 60, 279-288.

[25] Sampford, M. R. and Taylor, J. (1959). Censored observations in randomized blocks experiments. J. R. Statist. Soc. B 21, 214-237.

[26] Sprott, D. A. and Kalbfleisch, J. D. (1969). Examples of likelihood and comparison with point estimates and large sample approximations. J. Amer. Statist. Assoc. 64, 468-484.

[27] West, S. G., Whitney, G. and Schnedler, R. (1974). Helping a motorist in distress: The effects of sex, race, and neighborhood. J. Personality and Social Psychol. 24. In press.

[28] Zelen, M. (1959). Factorial experiments in life testing. Technometrics 1, 269-288.

[29] Zelen, M. and Dannemiller, M. C. (1961). The robustness of life testing procedures derived from the exponential distribution. Technometrics 3, 29-50.

[30] Zippin, C. and Armitage, P. (1966). Use of concomitant variables and exponential survival parameter. Biometrics 22, 665-672.

TABLE 1 - *Effect estimates under the exponential and normal models and 95% confidence intervals for the effects under the exponential model.*

Effect β_i	$\hat{\beta}_i$	95% Confidence Interval	$\hat{\beta}_i^N$
R	-25.9	(-393.0, 341.2)	26.1
S	-495.3	(-862.4, -128.2)	-108.0
N	29.1	(-337.9, 396.2)	17.9
P	182.4	(-184.7, 549.4)	41.6
RS	94.6	(-272.5, 461.6)	0.5
RN	106.5	(-260.5, 473.6)	8.6
RP	-62.3	(-429.4, 304.7)	-8.6
SN	-51.4	(-418.4, 315.7)	-20.8
SP	-173.3	(-540.4, 193.8)	-11.8
NP	-169.1	(-536.2, 197.9)	-27.9
RSN	-61.5	(-428.6, 305.6)	13.0
RSP	2.4	(-364.7, 369.4)	-10.5
RNP	263.6	(-103.4, 527.3)	54.6
SNP	192.9	(-174.2, 560.0)	31.2
RSNP	-122.4	(-489.5, 244.6)	17.2

$\hat{\beta}_i$ = estimate of β_i under the censored exponential model.

$\hat{\beta}_i^N$ = estimate of β_i under the normal model, ignoring the censoring.

MINUTES OF THE ANNUAL MEETING OF THE SOCIAL STATISTICS SECTION

St. Louis, Missouri, August 26, 1974

The meeting was opened by Charles Nam, Chairman, at 5:30 p.m., attended by approximately 60 members. Officers for 1975 were announced:

Chairman	Joseph Waksberg
Chairman-Elect	Daniel G. Horvitz
Vice Chairman (1974-75) (Program Chairman)	Mollie Orshansky
Vice Chairman (1975-76) (Asst. Program Chairman)	Robert Parke
Secretary (1974-75)	Mary G. Powers
Representative on the National Board of Directors (1973-75)	Evelyn M. Kitagawa
Representative on the National Council (1974-75)	Eva L. Mueller
Representative on the National Council (1975-76)	Conrad Taeuber
Publications Liaison Officer (1975-76)	Harry Rosenberg

Mollie Orshansky, Program Chairman for 1975, reported that she is considering cutting the number of sessions and encouraging new sources of papers.

A discussion of the editorial policies of the JASA was conducted by the editor. It was brought to the attention of members of the Social Statistics Section that submission of manuscripts on social statistics are encouraged by the editorial board. Manuscripts on subjects such as social indicators, migration, education, and other substantive topics which offer new innovations on social statistics have not been submitted in large numbers during recent years. Members were reminded that the Proceedings are only minutes of the meetings and should not be considered the only methods for publication of all papers read at the meetings. Methods were discussed for screening papers presented at the meetings to encourage that they be revised and re-submitted to JASA for publication. It was pointed out that expository articles will be prepared for The American Statistician which will synthesize previous research into a set of developments. These articles may be useful as statements of the status of the art. Methods for encouraging new members to join the Social Statistics Section were discussed.

An active discussion was held on the confidentiality of statistics and the possible future impact of new legislation on the use and release of statistics. Members agreed that future meetings of the Social Statistics Section should be concerned with specifying the problem of keeping data confidential, maintaining the privacy of individuals, and also outlining the specific problems involved with the growing use of administrative records (such as job applications). It was agreed that an ad hoc session be organized to establish the range of possibilities.

L. Suter, Acting Secretary