RELATION OF RESPONDENT SOCIAL STATUS TO INTERVIEWER BEHAVIOR

Barbara Snell Dohrenwend, The City College of New York*

Let me start by asking a naive question: How can we standardize interviews with respondents who differ markedly in background or social status? How, for example, can we interview poorly educated lower class respondents and well educated middle class respondents so as to be sure that their answers on a given topic are comparable? The well established answer to this problem is that we write and pretest an interview schedule so that respondents at lower as well as higher levels of education will understand the questions, and then train interviewers to adhere to this interview schedule.

The general problem I would like to discuss is how this procedure actually works in practice. While this general problem translates into a number of empirical questions concerning how interviewers behave and how respondents perceive their behavior, I will deal with only one such question, to wit: How well does the traditional scheduled interview succeed in standardizing the behavior of interviewers regardless of respondent status?

As Cannell and Kahn [1] pointed out in their chapter on interviewing in the 1968 edition of The Handbook of Social Psychology, examination of verbatim transcripts of interviews reveals that interviewers routinely employ probes that are not part of the prepared schedule of questions (pp. 572-573). What I will report today are preliminary results of a study designed to determine whether these probes and other additional interviewer behavior vary as a function of the status of the respondent.

This study is based on a set of interviews that were conducted with approximately 200 residents of a section of New York City as part of a program of research in psychiatric epidemiology. The interviews were done by psychiatrists who had been specially trained in the use of the two interview schedules employed in the study. Both schedules covered psychological symptomatology, role functioning, recent stressful life events, and background information, and each contained about 125 questions. One schedule was more open-ended in style than the other, but the specific differences between them are not important for the present discussion, since the results relating to respondent status did not differ for the two schedules.

For the analysis that I will report respondents were divided into three levels of education and two types of ethnicity. The educational levels are less than eight years, eight to 11 years, and 12 or more years of schooling. The ethnic types are relatively advantaged groups, predominantly Jews and Irish, as against relatively disadvantaged groups, that is, Negroes and Puerto Ricans. The interviewers were all white and members of relatively advantaged ethnic groups.

Typed transcripts of the tape recorded interviews were coded to describe the number and types of interviewers' spontaneous probes and other spontaneous behavior, that is, all behavior not dictated by the interview schedule. Measures derived from this coding are presented in Table 1.

Our hypotheses about how these measures will be influenced by respondent educational level are based primarily on a report by Strauss and Schatzman [3] of differences in the way interviewers questioned better educated and less educated respondents in a set of recorded non-scheduled interviews. Their results suggest three hypotheses: first, there will be more difficulty or confusion in communication with less educated respondents compared to better educated respondents; second, there will be more probing of answers given by less educated respondents; and, third, the probing of answers given by less educated respondents will include more closed and suggestive questions than the probing of answers given by better educated respondents.

The recent work of Marquis and Cannell [2] suggests similar hypotheses concerning the effects of respondent ethnicity on the questioning process. While in general their analysis of interviews conducted with a sample of employed white and Negro males did not reveal significant differences related to race of respondent, the authors suggested that certain trends were worth further attention. Specifically, they pointed to trends indicating more difficulty in communication and more extended probing by their white interviewers with Negro respondents than with white respondents. We will, therefore, look for these effects in our interviews.

The analysis, utilizing multivariate analysis of variance, showed, first, that there was no significant effect of the interaction of respondent ethnicity and education, and that only one of the 17 variables in Table 1 differed as a function of ethnicity. This single significant difference related to ethnicity indicated that the interviewers repeated more questions with respondents of disadvantaged than with respondents of advantaged ethnicity. This may, however, be largely due to Puerto Rican respondents for whom the language of the interview, English, was often a second language. This possibility will have to be tested before any other interpretation is considered.

In contrast to ethnicity, the respondents' education affected seven of the variables described in Table 1. The means for these seven variables according to respondents' education are shown in Table 2. We note first that the major difference in every case is between respondents with less than eight years of education and all others. On two variables there is even a reversal between those with some high school and those who graduated from high school, with the latter more nearly resembling the group with less than eight years of schooling.

The second point about Table 2 concerns the nature of the variables that show significant effects due to respondent education. All four of the variables classified in Table 1 as indicators of difficulty in communication are

Table 1

Measures of Spontaneous Interviewer Behavior

1. Indicators of amount of probing
   a. Number of spontaneous probes
   b. Number of different topics included in spontaneous probes

2. Indicators of style of probing - present of spontaneous probes that are:
   a. Open questions
   b. Closed identification question, e.g. Who, Where, When
   c. Closed fixed alternative questions
   d. Closed questions calling for Yes or No as response
   e. Informed suggestions - questions suggesting a particular answer in which the suggestion is based on information previously given by the respondent
   f. Uninformed suggestions - questions suggesting a particular answer in which the suggestion is not based on information previously given by the respondent
   g. Questions with informed premises, i.e. based on a premise that is supported by information previously given by the respondent
   h. Questions with uninformed premises, i.e. based on a premise that is not supported by information previously given by the respondent

3. Indicators of difficulty in communication with respondent
   a. Number of times that a question previously asked is repeated with no change in meaning
   b. Number of times that response is repeated in part or in whole exactly or nearly exactly in the respondent's words
   c. Number of volunteered statements explaining a question or procedure
   d. Number of statements in answer to respondents' comments or queries about questions or procedure

4. Expressions of positive affect or understanding
   a. Number of instances of contentless positive feedback, e.g. "Um hmm"
   b. Number of polite or sociable remarks

5. Indicator of disruptions from outside the interview
   a. Number of statements or questions about events external to the verbal interaction between interviewer and respondent, e.g. statements about the tape recorder or a third person

Table 2

Mean Frequencies per Interview of Activities that Varied Significantly with Respondent's Education

Interviewer activity	Respondent's education		
	Less than 8 years	8 - 11 years	12 or more years
Spontaneous questions**	75.04	53.51	47.83
Repetition of question**	25.22	13.07	10.49
Repetition of response**	8.07	4.89	3.98
Response to query about interview procedure**	13.48	6.51	7.58
Volunteered statement about interview procedure **	10.89	6.24	6.02
Contentless positive feedback+	17.00	7.14	8.30
Sociable remark*	2.07	1.20	0.99
(Number of Respondents)	(27)	(55)	(125)

+p(univariate F) = .055

*p(univariate F) $<$ .05

**p(univariate F) $<$ .01

significant, thus strongly confirming the prediction of greater difficulty in communication with less educated respondents. Consistent with this apparent difficulty is the finding that the number of spontaneous probes was greater with less educated respondents, while the number of topics probed did not differ with respondents education. That is, while Table 2 shows that the interviewers, on the average, asked half again as many spontaneous probes of poorly educated respondents as they did of better educated respondents, there was little variation in the mean number of topics probed. Specifically, this mean was 21 for the lowest educational group, 18 for the middle group and 17 for the highest group. The implication is that more questions were needed to get the same amount of information from less educated respondents.

While the level of polite and sociable remarks made during the tape recorded interviews was low, it was highest with the least educated respondents. Furthermore, the number of instances of contentless positive feedback, our second indicator of expression of positive affect or understanding, yielded a difference of borderline significance, just slightly greater than five percent, which is consistent with the result for sociable comments. Thus, encouraging and reinforcing behavior occurred most often with the type of respondent with whom the interviewers experienced greatest difficulty in communication, an understandable but perhaps unfortunate pattern of behavior. It may be unfortunate since it calls to mind the finding of Marquis and Cannell [2] that their interviewers tended to reinforce many behaviors other than adequate answers (p. 30), that is behaviors that should not have been reinforced. We wonder whether further analysis will reveal this tendency among our interviewers as well, and perhaps even reveal that it is particularly strong in interviews with poorly educated respondents.

Strikingly absent from Table 2 are any of the variables indicating differences in style of probing. Thus, the prediction of more directive probing with less educated respondents is not supported in these interviews in which the interviewer's behavior is constrained to some extent by a prepared schedule of questions.

In general, then, our results indicate greater difficulty in communication with the least educated respondents together with more probing of the answers of these respondents. Whether this extra probing is a good or a bad thing for the standardization of interview responses remains to be determined. That is, it is possible that these probes serve to equalize across educational groups the respondents' level of understanding of the scheduled questions, or it is possible that they introduce systematic bias into the answers of the least educated group. Further analysis will be designed to answer this and other questions raised by our findings to date.

References

[1] Cannell, C.F. and Kahn, R.L., "Interviewing" in G. Lindzey and E. Aronson (Eds.), The Handbook of Social Psychology, Addison-Wesley Publishing Company, 1968, 526-595.

[2] Marquis, K. H. and Cannell, C.F., A Study of Interviewer Respondent Interaction in the Urban Employment Interview, Survey Research Center, The University of Michigan, 1969.

[3] Strauss, A. and Schatzman, "Cross-class interviewing: An analysis of interaction and communicative styles," in R.N. Adams and J.J. Preiss (Eds.) Human Organization Research: Field Relations and Techniques, Dorsey Press, 1960, 205-213.

Footnote

*This work was supported in part by National Science Foundation Grant GS-2896.

SPLIT-PANEL TESTS OF CENSUS AND SURVEY QUESTIONNAIRES

Thomas B. Jabine and Naomi D. Rothwell, U.S. Bureau of the Census

Someone remarked at last year's ASA meeting that the Census Bureau and other survey organizations are doing a lot of methodological research that isn't being reported. The Census Bureau does report on and/or list much of its research output in the annual Census Bureau Methodological Research List, the technical and working papers series, the technical notes series, census procedural histories, and articles published in professional journals. Nevertheless, there was some justification for the remark. We have done many things that are not readily accessible even within our own organization. Some results are not reported at all, others are widely scattered, and most people designing surveys don't have time to make a comprehensive review of the literature.

This paper constitutes an effort to make results of our experience more readily available. In preparation for it we identified, reviewed, and summarized all split-panel tests of questionnaires conducted by the Census Bureau. We identified 15 such tests. To learn how these tests were conducted and what their results were, we assembled available documentation and spoke to the principal investigators. We then prepared a summary description of each test in a standard format. This proved to be so rewarding, both in terms of giving us a clearer picture of what was learned from the tests and in suggesting ways of improving the design of future experiments, that we plan to present the full results in a Census Bureau technical paper, to be issued within a few months.

To define more precisely what is included in this review, we mean by a split-panel test a controlled experiment in which the treatments to be compared and analyzed consist of versions of a questionnaire which differ in one or more respects but have the same data objectives. The variations to be tested may involve the wording of individual questions, the answer format (check-boxes for categories vs. open end), the order of questions, the general format, including color of paper and printing, size, arrangement of questions on the page, and other features of the questionnaire.

Split-panel tests need not be restricted to questionnaires -- they may cover any aspect of survey methodology, including collection procedure --mail vs. telephone vs. personal interview; number and spacing of mail followups; use of information from prior rounds in a panel survey; length of reference period; and respondent rules in interview surveys. We have, in fact, conducted split-panel experiments concerned with many aspects of the data collection process; here, however, we limited the review to tests involving multiple versions of questionnaires.

Exhibit I is a summary of the 15 tests we identified and reviewed. The earliest test was made in 1957; the latest and possibly the most sophisticated from the point of view of questionnaire design (Number 4: 1969 Questionnaire Variation Study) is still in progress. At least two additional tests, now in the planning stage, have not been included on this list. We have also excluded at least two tests for which the necessary experimental controls were obviously inadequate.

Most of the tests are characterized either as "built-in" tests or pretests. The built-in test is conducted as an integral part of a census or survey. The built-in tests have obvious advantages -- they are carried out under realistic conditions, their costs are marginal since most of the data are going to be collected and processed anyway, and sample sizes can be large enough to minimize the contribution of sampling error and permit clear emergence of treatment differences.

In a pretest, on the other hand, while costs are higher, radical innovations can be tested without fear of upsetting an on-going system. Furthermore, pretests provide the only opportunity to experiment with census materials and procedures while the census is not going on.

"Methodological test" covers everything that does not qualify as a built-in test or pretest, e.g., Number 7, the Current Population Survey Methods Tests, where alternatives to standard procedures in the regular monthly survey are tested in a separate group of sample areas, specifically set aside for experimental purposes.

Ten of the 15 tests used mail as the principal collection method; this reflects the Census Bureau's increasing reliance on self-enumeration.

Seven of the tests were nationwide. Thus, potential bias attributable to regional differences could be eliminated. When personal interviews were required, however, it was not always possible to conduct nationwide tests and it became necessary to exercise judgments about general applicability of results observed in a geographically limited study population.

The built-in tests usually provided larger samples than the pretests and methodological tests. The sample size requirements depend on the kinds of comparisons to be made and the sensitivity needed in measuring treatment differences. Sample sizes of from 1,000 to 5,000 cases for each version have been adequate for nearly all purposes, although when dealing with low frequency items, such as unemployment, or with aggregates, larger samples may be needed.

Validity checks are needed unless the only concern is to learn which of two questionnaires will produce a higher mail return rate, or lower item nonresponse. In most cases, however, we also want to know which alternative leads to less response error. The principal methods used to obtain direct measures of validity are reinterviews and record checks. Test Number 15 illustrates a direct way of checking validity. Retrospective data on occupation and industry five years ago were compared with current data collected for the same persons five years earlier.

External sources of information about income, morbidity data, and some other subjects indicate that standard survey techniques produce underestimates of known aggregates. Such information is the basis for indirect validity checks; i.e., we have been willing to conclude that an experimental version which produces a higher figure is better.

Exhibit II consists of standard descriptions of 3 of the 15 tests, and shows the relevant parts of the questionnaires. The last topic in the standard description, "Possible Generalizations," goes beyond the results of the particular test described. It contains one or more principles for questionnaire design <u>suggested</u> by the results of the test. Some of these are summarized below. Although they are not startling or novel, a step forward has been taken when ideas about questionnaire design are supported by experimental evidence.

1. Probing questions can improve accuracy of response. A version of the Current Population Survey questionnaire which included probes about hours worked and about self-employment provided improved statistics for these subjects. Marginal improvement in reports about size of farm were also achieved through probing questions. There are, of course, limits to effectiveness of probing. Additional probing about labor force status in test Number 7 produced no significant changes.

2. It is usually better to break complex classifications into simple elements and ask questions about each element than to ask interviewers to make the classifications. A test for the 1960 census which required classification of living places as dwelling units or quasi-dwelling units indicated that it could be done better if the enumerators recorded information about the criteria they used in making their classification than if they did not.

3. When a rough estimate or guess of an amount is needed, the checkbox alongside of a class interval obtains a higher response rate on a self-enumerative questionnaire than a question requiring a write-in amount. This was observed in an agriculture census pretest for a question about value of land and buildings.

4. In self-enumerative questionnaires, when one choice is made easier than another, responses may be biased toward the easier choice. Two tests indicate such bias. One showed that respondents, when asked to recall their occupation five years ago, were more likely to report it inaccurately as "same" if a checkbox was provided for such a response than if there were no such precoded category. A second test showed fewer reports of mixed ethnic origins when checkboxes could be marked for single ethnic but write-in entries were required for mixed origins than when write-in entries were required for <u>all</u> reports of ethnic origins.

5. In a voluntary survey, an explicit statement that cooperation in the survey was voluntary reduced returns by a small percentage but did not affect completeness of response (even to a detailed income question) by those who returned questionnaires.

We cannot, of course, guarantee that the ideas just presented will be useful in all situations. The particular study population, the collection procedures, the various kinds of subjects covered in the survey, the sponsorship of the survey, the time at which the survey is taken -- all these and many other factors may interact with the features we have discussed to produce results differing from those observed.

Papers like this traditionally end with a plea for more research -- in this case we think it appropriate to follow tradition and, further, to urge that experimenters take pains to report their findings in at least as much detail as we have provided for these 15 tests. All such efforts will contribute to the development of a much-needed set of tested principles for questionnaire design.

Exhibit I

SPLIT-PANEL TESTS OF QUESTIONNAIRES CONDUCTED BY THE U.S. BUREAU OF THE CENSUS

Name of test	Parent program	Type of test	Primary collection procedure	Study population	Number of questionnaire versions	Approximate sample size per version	Validity checks
		FARM CENSUSES AND SURVEYS					
1. 1961 Experimental Farm Survey	1964 Census of Agriculture	Methodological test	Mail	Farms in selected counties	4	400 farms	Intensive interview for a subsample
2. 1963 Farm Panel Pretest-Round 2	Panel Evaluation Survey, 1964 Census of Agriculture	Pretest	Mail	Farms in selected counties	2	230 farms	None
3. 1968 Agriculture Census Pretest	1969 Census of Agriculture	Pretest	Mail	Farms in two States	4	4,500 farms	None
4. 1969 Questionnaire Variation Study*	1969 Census of Agriculture	Built-in	Mail	Farms receiving standard census questionnaire, excluding very large ones	9	2,000 in-scope farms	Resurvey with alternate version of questionnaire and reconciliation of reported differences
		DEMOGRAPHIC SURVEYS					
5. 1967 Health Interview Survey Study	Health Interview Survey	Built-in	Interview	Civilian noninstitutional population	2	17,500 persons	None, but higher rates of illnesses reported assumed to be better
6. 1969 CPS Income Experiments	Current Population Survey (March Income Supplement)	Built-in	Interview (some by telephone)	Households	2	13,000 households	None, but larger amount of income reported assumed to be better
7. 1963 CPS Methods Test, First Series*	Current Population Survey	Methodological test	Interview (some by telephone)	Civilian noninstitutional population in selected counties	3	5,500 households	Intensive reinterview for subsample; employer record checks
8. 1969 Test of Explicit Statement on Voluntary Reporting	Survey of New Beneficiaries	Built-in	Mail	Persons recently awarded Social Security or Medicare benefits	2	3,400 persons	None

* See Exhibit II for further description of study design and results.

Exhibit I -- Continued

SPLIT-PANEL TESTS OF QUESTIONNAIRES CONDUCTED BY THE U.S. BUREAU OF THE CENSUS

Name of test	Parent program	Type of test	Primary collection procedure	Study population	Number of questionnaire versions	Approximate sample size per version	Validity checks
		POPULATION AND HOUSING CENSUSES					
9. 1957 Employment Pretest	1960 Census of Population and Housing	Pretest	Interview	Civilian noninstitutional population 14-years-old and over in Philadelphia SMSA	3	800 households	Intensive reinterview
10. 1958 Housing Pretest	1960 Census of Population and Housing	Pretest	Interview	Housing units in selected areas of Washington, D.C.	3	700 housing units	Independent reinterview and reconciliation of differences
11. 1966 (First) Content Pretest	1970 Census of Population and Housing	Pretest	Mail	Households in St. Louis Park, Minn. and Yonkers, N.Y.	2	2,350 households	Independent reinterview and reconciliation of differences
12. 1967 (Second) Content	1970 Census of Population and Housing	Pretest	Mail	Households in Gretna, La.	2	2,100 households	None
13. 1966 (First) Questionnaire Format Test	1970 Census of Population and Housing	Pretest	Mail	Households	2	1,150 households	None
14. 1967 (Second) Questionnaire Format Test	1970 Census of Population and Housing	Pretest	Mail	Households	4	1,120 households	None
15. 1968 Subject Response Study*	1970 Census of Population and Housing	Pretest	Mail	Civilian noninstitutional population 14-years-old or older	2	1,400 households	Comparison of current responses on occupation 5 years ago with information given 5 years ago.

* See Exhibit II for further description of study design and results.

Test No. 4

1969 QUESTIONNAIRE VARIATION STUDY

1. Parent Program: 1969 Census of Agriculture.

2. Type of Test: Built-in.

3. Date: January 1970 and January 1971.

4. Study Population: In general, the study population consists of farms with total value of products from $2,500 to $500,000, as indicated in the administrative records from which the initial census mailing list was formed. Due to changes as well as errors in the records, the actual population includes some out-of-scope cases (to be excluded from analysis) and some farms with total value of products outside the indicated range.

5. Questionnaire Versions Tested: There were nine versions, differing in wording and format of questions and in use of shading. The nine questionnaire versions combined seven alternative formats in a way that permits measurement of individual format effects and of interaction effects. The alternative formats were:

   (a) An alternative version of questions about acres-in-place and tenure.

   (b) The inclusion among the land-use questions of a space for transcription of acreage reported in a preceding series of questions.

   (c) The addition of checkboxes marked "None" to the answer spaces of questions about numbers of various kinds of machinery on the farm.

   (d) The use of two answer columns for questions on cost of chemicals used; one for the cost of chemicals only and the other for the cost of application. The standard format had one answer column, asking for cost of chemicals, excluding cost of application.

   (e) The use of two answer columns for questions on income and expenditures; one for the farm operator's share of income or expense and the other for landlords' or contractors' shares. (Standard format had one column asking for these shares combined in one figure.)

   (f) The deletion of many explanatory notes from the questions on income and expenditures.

   (g) The omission of shading. The standard census questionnaire was lightly shaded everywhere except the answer spaces.

6. Design of Study: A national systematic sample of cases was selected from the census mailing list and randomly divided into nine subsamples, one to receive each of the nine questionnaire versions. The sample size was selected to yield about 2,000 in-scope farms in each subsample.

7. Collection Procedure: Mail, with a reminder card and four mail followups.

8. Validity Check: Plans call for a January 1971 mailing to a subsample of respondents. They will receive the alternative version of the acreage and tenure questions from the one mailed to them in January 1970. The two forms will be matched and when differences are observed, reinterviews will be conducted to determine whether changes in acreage or tenure occurred or, if not, which version obtained the more valid information.

9. Principal Items Analyzed by Version: Return rates, acceptability of entries, item nonresponse, and selected agricultural characteristics.

10. Limitations: Very small and very large farms are not included; newly established farms are not included.

11. Principal Findings: No difference in return rates between versions. Other results not yet available.

12. Consequences: None to date.

13. Possible Generalizations: None to date.

Sources of material: Outline prepared by John Forsythe, using the following sources:

(a) Memorandum A-69-A-15 of June 30,1969, "Plans for Questionnaire Variation Study, 1969 Census of Agriculture"

(b) U.S. Bureau of the Census Advance Report: "Data-Collection Forms and Procedures for 1969 Census of Agriculture, Irrigation, and Agricultural Services"

ILLUSTRATION. Standard Census Format

Section 34 – Production EXPENSES for this place in 1969.

Include your best estimate of expenses paid by others – your landlord, contractors, buyers, etc. – for crops, livestock or livestock products produced on this place. *(See Leaflet, section 34.)*

CENTS NOT REQUIRED

		Dollars	Cents
1. Livestock and poultry purchased – cattle, calves, hogs, pigs, sheep, lambs, goats, horses, baby chicks, poults, started pullets, etc.	630	$	
2. Total feed purchased for livestock and poultry – grain, hay, silage, mixed feeds, concentrates, etc. *(Total of dollars for a, b, c, and d)*	631	$	

		Tons	Tenths		Dollars	Cents
a. Commercially mixed formula feeds purchased – complete, supplement, concentrates. *(Do not include ingredients purchased separately, such as soybean meal, cottonseed meal, and urea.)*	632		/10	633	$	
b. Ingredients purchased – such as soybean meal, cottonseed meal, urea, etc., millfeeds or other milling byproducts. *(Do not include whole grains.)*	634		/10	635	$	
c. Whole grains purchased – such as corn, oats, barley, grain sorghum, wheat, rye, etc. Include cracked grain. *(Do not include millfeeds or other milling byproducts, or green chop.)*	636		/10	637	$	
d. Hay, green chop, silage, etc.	638		/10	639	$	

		Dollars	Cents
3. Seeds, bulbs, plants, and trees purchased	640	$	
4. Commercial fertilizer purchased – all forms, including rock phosphate and gypsum	641	$	
5. Total gasoline and other petroleum fuel and oil purchased for the farm business – Diesel fuel, LP gas, butane, propane, piped gas, kerosene, fuel oil, motor oil, grease, etc. *(Total of a, b, c, and d)*	642	$	

		Dollars	Cents
a. Gasoline for the farm business	643	$	
b. Diesel fuel for the farm business	644	$	
c. LP gas, butane, and propane for the farm business	645	$	
d. Motor oil, grease, piped gas, kerosene, and fuel oil for the farm business	646	$	

		Dollars	Cents
6. Hired farm labor – *Include all money paid in cash for farm labor including payments to family members, and for Social Security taxes. (Do not include housework, customwork, and contract work.)*	647	$	
7. Contract labor – *Include expenditures primarily for labor, such as harvesting of fruit, vegetables, berries, etc., performed on a contract basis by a contractor, a crewleader, a cooperative, etc.*	648	$	
8. Machine hire and customwork – *Include expenditures primarily for use of equipment, and for customwork such as grinding and mixing feed, plowing, combining, corn picking, silo filling, spraying, dusting, etc.*	649	$	
9. Agricultural chemicals purchased – *Add dollars reported in section 33 and enter total here*	650	$	
10. All other production expenses – *Include current operating expenses, and depreciation, taxes, interest, cash rent, insurance, repairs, etc., for the farm business. (See Leaflet, section 34.)*	651	$	
11. Total production expenses – *Add dollars for items 1 through 10 and enter total here* →	652	$	

ILLUSTRATION – Continued. Variations as described in 5E and F combined

Section 34 – Production EXPENSES for this place in 1969.

If all expenses were paid by you, fill only the first column.

If any expenses for crops, livestock or livestock products produced on this place were paid or furnished by others – your landlord, contractors, buyers, etc.– enter your best estimate of these expenses in the second column and yours in the first column. *(See Leaflet, section 34.)*

	Tons	Tenths	Total production expenses – Amount paid by you – CENTS NOT REQUIRED – Dollars	Cents	Amount paid or furnished by landlords or contractors, if any *(See Leaflet)* – CENTS NOT REQUIRED – Dollars	Cents
1. Livestock and poultry purchased			630 $		930 $	
2. Feeds purchased for livestock and poultry:						
a. Commercially mixed formula feeds	632	/10	633 $		933 $	
b. Ingredients *(Do not include whole grains.)*	634	/10	635 $		935 $	
c. Whole grains	636	/10	637 $		937 $	
d. Hay, green chop, silage, etc.	638	/10	639 $		939 $	
3. Seeds, bulbs, plants, and trees purchased			640 $		940 $	
4. Commercial fertilizer			641 $		941 $	
5. Gasoline and other petroleum fuel and oil purchased for the farm business:						
a. Gasoline			643 $		943 $	
b. Diesel fuel			644 $		944 $	
c. LP gas, butane, and propane			645 $		945 $	
d. Motor oil, grease, piped gas, kerosene, and fuel oil			646 $		946 $	
6. Hired farm labor			647 $		947 $	
7. Contract labor			648 $		948 $	
8. Machine hire and customwork			649 $		949 $	
9. Agricultural chemicals – *Total of this line should equal total of dollars column in section 33*			650 $		950 $	
10. All other production expenses – *Include current operating expenses, and depreciation, taxes, interest, cash rent, insurance, repairs, etc., for the farm business. (See Leaflet, section 34.)*			651 $		951 $	
11. Total – *Add dollars for all items and enter totals here* →			652 $		952 $	

Test No. 7

1963 CURRENT POPULATION SURVEY METHODS TEST, FIRST SERIES

1. Parent Program: Current Population Survey.

2. Type of Test: Methodological.

3. Date: April 1963 to December 1964.

4. Study Population: Households in Boston, Mass.; Charolotte, N. C.; and Marion County, Ohio.

5. Questionnaire Versions Tested: Three versions and one supplementary form as follows:

   CPS-1: Standard form used for survey.

   CPS-X2: Questionnaire with more detailed inquiries about labor force participation, and, for employed persons, hours worked and class of worker.

   CPS-X3A & X3B: Advance form covering key items to be filled by respondent; followup form for interviewer's use.

   CPS-X45: Supplementary form summarizing information given in previous interview to be checked against and reconciled with current month's data.

6. Design of Study: About 1,300 interviews were conducted every month during the experiment. Six interviewers in each of the three areas each conducted about 24 household interviews a week for three weeks of each month. Each interviewer used a different one of the following procedures in each of the three survey weeks of the month:

   Procedure 1: Standard questionnaire (CPS-1) used and independent interviews conducted each month.

   Procedure 2: Detailed questionnaire (CPS-X2) used and independent interviews conducted each month.

   Procedure 3: Advance form (CPS-X3A) mailed with request that it be filled and held for interviewer. Interviewer transcribed data to the CPS-X3B. The procedure was later changed to request mail return of the completed form by a specified date.

   Procedure 4: Detailed questionnaire (CPS-X2) used for the first step in the interview. The previous month's information shown on the CPS-X45 was consulted after the first step. Reconciliation performed as needed.

   Procedure 5: Same as Procedure 4, except standard questionnaire (CPS-1) was used for the first step.

   Assignments were randomized over an entire city or county by interviewers and week of the month. Households were initially interviewed for four consecutive months. The rotation cycle, however, was later reduced to three months.

7. Collection Procedure: For Procedure 3, there was an advance mail form followed by personal or telephone interview; for the other four procedures, information was collected by personal interview.

8. Validity Check:

   (1) Comparisons were made with reinterview results for a subsample of cases.

   (2) Checks were made against independent lists of corporations and establishment payroll reports to validate the classification of persons as self-employed or as employees.

9. Principal Items Analyzed by Version: Comparisons were made among experimental procedures for total population in labor force, employed and unemployed persons by sex, gross changes in employment, employed persons by hours worked and class of worker, part-time workers by reason for part-time employment.

10. Limitations: Results apply only to the selected study areas.

11. Principal Findings: "Perhaps the most interesting finding of this experiment is that, in spite of the sharp differences in approach, there do not appear to be any major differences among the procedures in the distributions of the sample by employment status ... While findings have been inconclusive with regard to the basic employment status classifications, some clearcut improvements have been evident in the detailed questionnaire ... Most striking of these is the case of the information on hours worked."*

12. Consequences: After further testing in a national experimental sample, changes were made in the standard questionnaire for CPS based on the results of the test. These changes were in the detailed questions about hours worked, duration of employment, class of workers, etc.

13. Possible Generalizations: Probing questions can improve accuracy of response for some items. However, adding questions to an already fairly comprehensive inquiry may have very little effect.

Source of material: Prepared by N. D. Rothwell in consultation with Daniel Levine and Joseph Waksberg. The only written source is cited in the footnote below.

---

* Quotation from "New Methodological Research on Labor Force Measurement," Joseph Waksberg and Robert B. Pearl, September 1965.

ILLUSTRATIONS

HOURS WORKED

CPS-1 and CPS-X2

20. Did ... do any work at all LAST WEEK, not counting work around the house?

Yes O   No O *(Skip to 22)*

21. How many hours did ... work last week (at all jobs)?

0 O O
1 O O
2 O O
3 O O
4 O O
5 O O
6 O O
7 O O
8 O O
9 O O

If 1 - 34 hours in 21, fill 21A and 21B or 21C.

If 35 or more hours in 21, skip to 26 and enter job worked at last week.

IF WORKED LESS THAN 35 HOURS LAST WEEK (1-34 in Item 21)

21A. Does ... usually work 35 hours or more a week at this job?

Yes O   No O

21B. What is the reason ... worked less than 35 hours LAST WEEK?

21C. What is the reason ... USUALLY works less than 35 hours a week?

Mark one circle below

Slack work .... O
Material shortages ... O
Plant or machine repairs .... O
New job started during week .... O
Job terminated during week .... O
Could find only part-time work ...... O
Holiday (legal or religious).... O
Labor dispute .... O
Bad weather .... O
Own illness .... O
On vacation .... O
Too busy with housework, school, etc ..... O
Did not want full-time work .... O
Full-time work week under 35 hours........ O
Other reason *(Specify)*.. O

If entry in 21B or C, go to 26 and enter job worked at last week.

CPS-X2 only

IF WORKED 35-48 HOURS LAST WEEK

21D. Did ... lose any time or take any time off from work LAST WEEK for any reason, such as illness, holiday, or personal business?

Yes O   No O

How many hours did ... take off?

(NOTE: Correct item 21 if lost time not already deducted; if item 21 reduced below 35 hours, ask 21 A-C.)

21E. Did ... work any overtime or extra hours LAST WEEK?

Yes O   No O

How many extra hours did ... work?

(NOTE: Correct item 21 if extra hours not already included.)

21F. Did ... work at more than one job or for more than one employer LAST WEEK?

Yes O   No O

(Find out whether hours on extra jobs were included in item 21; otherwise correct.)

(After completing 21F, skip to item 26 and enter job worked at last week.)

CLASS OF WORKER

CPS-1 only

D. Was this person-

An employee of PRIVATE company, business, or individual for wages, salary, or commission ....P O
A GOVERNMENT employee (Federal, State, county, or local) .....G O
Self-employed in OWN business, professional practice, or farm ....O O
Working WITHOUT PAY in family business or farm ....WP O
NEVER WORKED ....NEV O

CPS-X2 only

D. CLASS OF WORKER   Was this person

An employee of PRIVATE company, business, or individual for wages, salary, or commission... P O
A GOVERNMENT employee (Federal, State, county, or local)................................ G O
Self-employed in OWN business, professional practice, or farm................................ O O
(Ask if not a farm.) **Is the business incorporated?**   Yes O   No O
Working WITHOUT PAY in family business or farm .... WP O
NEVER WORKED.. NEV O

LABOR FORCE STATUS

CPS-1 and CPS-X2

22. *(If LK in 19, skip to 23.)*
Was ... looking for work--or on layoff from a job?   (Include special cases)

Yes O   No O *(Skip to 24)*

23. How many weeks has ... been looking for work?

_______________ *(Continue with 23A-E)*

23A. Has ... been looking for full-time or part-time work?

Full O   Part O

24. *(If J in 19, skip to 25.)*
Even though ... did not work LAST WEEK, does he have a job or business?

Yes O   No O *(Skip to 27)*

25. Why was ... absent from work LAST WEEK?

Own illness .... O
On vacation .... O
Bad weather .... O
Labor dispute.................. O
Temporary (less than 30-day) layoff O

How many weeks ago was ... laid off? *(Enter in 23)*

New job to begin within 30 days ... O

How many weeks ago did ... start looking for a job? *(Enter in 23.)*

Other *(Specify)*................. O

25A. Is ... getting wages or salary for any of the time off last week?

Yes......... O
No.......... O
Self-employed O

(Go to 26 and enter job held last week.)

CPS-X2 only

23B. What has ... been doing to try to find work? *(Do not read list; mark each applicable circle.)*

Checking with public employment agency.......... O
Checking with private employment agency .... O
Checking directly with employers................ O
Placing or answering advertisements............. O
Checking with friends or relatives............... O
Other ..... O

23C. When did ... last work at a job or business?

Month, day, and year _______________   Over 6 months ago O   Never worked .... O *(Skip to 23E)*

(Convert to ______ weeks ago **or mark**)

23D. INTERVIEWER CHECK: If number of weeks in 23 and 23C differ by 3 or more, review and correct whichever is in error or give reason for difference below.

23E. (Go to 26 and enter last full-time civilian job; if waiting to start new job, enter description of new job in 26.)

27. INTERVIEWER CHECK ITEM

O Household not in 1st month of enumeration
O Person listed prior to current month *(Skip to 27D)*
O Person listed for 1st time this month *(Ask 27A)*
O Household in 1st month of enumeration

27A. When did ... last work at a full-time job or business?

1958 or later O   Before 1958.......... O   Never worked full time O *(Skip to 27C)*

Month and year _______________ *(Ask 27B)*

27B. Why did ... leave that job?

(Go to item 26 and enter last full-time civilian job; return to 27C.)

27C. Does ... intend to look for any kind of work--full-time or part-time--in the next 12 months?

Yes-definitely.. O
Yes-probably... O
Maybe-depends. O

No........ O
Don't know O

(Enter "2" for group no. in C.C. item "X" and ask)
Is there any particular reason ... is not looking for work now? *(Go to 27E)*

(Enter "1" for group no. in C.C. item "X"; go to 27E)

27D. INTERVIEWER CHECK ITEM

Transcribe from Control Card item "X"

Group 3 O *(Ask 27E)*
Group 2 O
Group 1 O *(Omit 27E)*

27E. Many people earn money from time to time even though they don't consider themselves as employed. Would you please look over this list and tell whether ... has been doing or looking for any work of this kind. (Give respondent checklist.)

Yes O   No O *(go to next person)*

Was ... doing this last week?

No........ O *(go to next person)*
Yes-working O *(Change "No" to "Yes" in item 20 or 22 and proceed accordingly.)*
Yes-looking O

Test No. 15

1968 SUBJECT RESPONSE STUDY: OCCUPATION FIVE YEARS AGO

1. Parent Program: 1970 Census of Population.

2. Type of Test: Pretest.

3. Date: August 1968.

4. Study Population: Civilian noninstitutional population 14-years-old and over as of July 1963.

5. Questionnaire Versions Tested: There were two versions, SR-15 and SR-16, of questions about occupation five years ago. Question wording was identical but response categories differed as follows:

   (a) SR-15 permitted respondents to mark a circle if their business or industry and kind of work five years ago were the same as reported in an earlier question about current employment.

   (b) SR-16 required a write-in entry of all respondents employed five years ago to show kind of business or industry and kind of work.

6. Design of Study: The sample for the study was selected from a list of households in which interviews had been conducted in the July 1963 Current Population Survey five years before the test.

   Approximately 2,800 households were selected and questionnaires were mailed to all household members who had been 14-years-old and over as of July 1963 in the selected households. The 2,800 households initially selected were sorted into two groups on an alternating household basis. 6,401 questionnaires were mailed.

7. Collection Procedure: Mail; reminder cards sent a week later; personal followup.

8. Validity Check:

   (1) The use of the 1963 survey panel as the sample provided persons whose occupation five years before the test had been reported five years earlier as their current occupation.

   (2) The level of occupational mobility is measured in the Current Population Survey. This served as a standard for the pretest results.

9. Principal Items Analyzed by Version:

   (a) Return rates.

   (b) Item nonresponse rates.

   (c) Differences between employment status, occupation and industry as reported on a current basis in 1963 and retrospectively in 1968.

10. Limitations: Analysis of differences was made only for those who responded by mail.

11. Principal Findings: There were no significant differences between versions in mail return rates and item nonresponse rates.

    For persons who reported being employed on both occasions, but reported different occupations, a higher proportion of those reporting on SR-15 said (incorrectly) in 1968 that their previous occupation was the same as their current one (27.8 percent vs. 22.6 percent).

12. Consequences: Because of the high difference rates observed for both versions, quiries about specific occupation or industry five years ago were included only on the 5-percent sample version of the 1970 census questionnaire.

13. Possible Generalizations: In self-enumerative questionnaires, when one choice is made easier than another, responses may be biased toward the easier choice.

    Sources of material: Outline prepared by N.D. Rothwell, using following sources:

    Memoranda from Thomas G. Walsh:
    (1) E15 No.2 -- "Data Results from the 1968 Subject Response Study"
    (2) E15 No.3 -- "Accuracy of Retrospectively Reporting Work Status and Occupation Five Years Ago"

    Memorandum from Thomas G. Walsh and Paula J. Buckholdt:
    (3) Results Memo No.1 -- "Mail Return Results of the Subject Response Study"

TOWARD A THEORY OF RESPONSE EFFECTS IN SURVEY RESEARCH

Seymour Sudman, Survey Research Laboratory, University of Illinois

It is a wise experimenter who knows his artifact from his main effect; and wiser still is the researcher who realizes that today's artifact may be tomorrow's independent variable.

William J. McGuire

The importance of response factors on survey results is now widely recognized by survey users and practitioners. There are literally hundreds of studies that demonstrate some kinds of response effects or non-effects. Interest in the problem has been heightened by the work of Rosenthal and others in experimental psychology on the presence of experimenter effects and other artifacts in behavioral research.

As McGuire wittily points out, there are three stages in the life of an artifact.

A. The Ignorance Stage

B. The Stage of Coping

C. The Exploitation State where:

> It is rather heartwarming to observe that in the final stage in the career of an artifact, the variable comes into its own. The Ugly Duckling becoming the Prince Charming which gives rise to a new line of research.

We are now well past the ignorance stage, and are into the stage of coping. Although there have been some examples of exploitation such as the notion of acquiescence or yeasaying, in general, survey research is still quite far from exploiting response factors. One limit to coping and exploitation has been the lack of general theoretical structure of response factors. Most studies have demonstrated response factors in highly specific situations that are difficult to generalize.

In this monograph, we attempt to present a more general framework than has been available up to now. It would be presumptuous to claim that we have developed a general theory of response, but this is the direction of this research, and hopefully the results presented here are the first stage in the development of such a theory.

Methodology

The first stage of the project consisted in preparation of a bibliography of books and articles related to the topic of response effect. This bibliographic search was greatly aided by a more general bibliographic file on survey methods that had been compiled under Charles Cannell's direction at the Survey Research Center, University of Michigan. As of now, the bibliographic file for this project consists of about 1,000 items and new sources are still being added. This bibliography is included in this report in its entirety, although about half the items could not be coded using our coding system.

Throughout this study, the dependent variable remains the relative response effect. Many studies gave information that demonstrated response effects, but where the magnitude and direction of these effects could not be determined. Since we were attempting to measure the effects of a large number of independent variables on response, we needed sufficient data on both the actual figures reported and some validation measure to establish a reasonably accurate estimate of the magnitude of the relative response effect.

The "field work" for this project consisted of obtaining copies of the studies and coding them in a uniform format to be described in detail below. The sample then consisted of the information coded from these studies. The number of items of information from a single survey varied from one to several hundred. Thus, when one uses unsummarized results, these are weighted towards the studies that provided more items of information. In some of the analyses, all results from a given study are combined so that the studies are given equal weight.

Where the study dealt with actual behavior such as purchasing, voting, owning a driver's license or a library card and outside validation sources were available it seems appropriate to call the differences response error or bias. Where the study related to attitudes and there were no outside validation sources, it seems better to use the more neutral phrase "response effect." Since many different types of studies were included, absolute errors become unworkable since there is no way of combining studies. A measure of relative effect was adopted and defined to be:

$$RE = \frac{\text{(Actual-Validating)}}{s}$$

where s is the standard deviation of the population, obtained from the validation information, if possible. Where no data were given on the size of s, an estimate of RE was made using $\frac{\text{(Actual-Validating)}}{\text{Validating}}$. This estimate is satisfactory for populations where the coefficient of variation is near to one, but becomes quite poor as the validating mean approaches zero. In those cases, the results were omitted.

Where the studies reported about additudinal information and no outside validation was possible, the weighted mean of all observations was used for validation. Thus, for example, in a study which contrasted responses of blacks to black and white interviewers, the grand mean was found by combining the responses to black interviewers with those to white interviewers weighting by the sample sizes. The relative effect was found by computing the difference between the response given to black (or white) interviewers and the grand mean, with s being computed from the grand mean.

Analysis

The basic (perhaps simple-minded) approach in analyzing the results is to treat the several thousand items of information as one would treat responses to a typical survey. One observes differences in the dependent variable (here response effects) and one proceeds to search for the combination of independent variables that best explain the results. One never finds complete consistency in the real world. Thus, in studies of prejudice, it is generally the case that prejudice by whites against blacks declines with increasing education, but regional factors are also critical. College graduates in the South have higher prejudice scores than do Northern respondents with only an elementary school education. Nevertheless, there are some Southerners with low education who show no prejudice.

Similarly in this analysis of response error, we are unable to find absolute truths. We may make generalizations such as that the presence of another adult during the interview causes negative response effects if the respondent is reporting about threatening personal information, but there will be counter-examples reported so that we deal with probabilities less than one.

There are two chief differences between our analysis and that of a typical survey. For a typical survey, one could have the complete range of information on independent variables for all respondents. That is not the case for the studies we have reported. Most of these studies have been concerned only with the relation between one or a few independent variables and response effects. Sometimes it is possible by careful reading to determine the characteristics of other variables that are not analyzed in the study. For most studies, however, there are a large number of independent variables for which no information is available. Thus, our large sample size of items and studies is misleading, since much of the information is missing. Most of the generalizations made are based on samples much smaller than the total. This, unfortunately, limits our ability to discuss interaction effects in great detail. As the reader will observe, some of these interaction effects appear to be of great importance. At least, these results point to the gaps in information about response effects that may be fruitfully explored in future research.

The other difference between a typical survey and this meta-study is that on a survey, respondents are considered to be equally reliable. In this study, a cursory reading of a few articles is enough to persuade anyone that there are large differences in the quality of the research. We have attempted to quantify the quality of the research by considering the following factors:

a. Researcher's reputation

b. Type of sample

c. Methodological details given in report

d. Type of validating information

Some of the analyses weight the results by this quality measure. In general, however, we have not found that these weighted results differ from the unweighted ones. We also coded the journal in which the report was found, the year of the study, the researcher's professional background, and the size of sample. Our analyses to date have not indicated that these are important variables in explaining differences in response results.

Independent Variables

Effects of Question, Questionnaire Design and Interviewing Situation

Length of interview
Location of interview, presence of others
Subject of report
Threat
Saliency
Method of administration
Structured or unstructured questions
Position of question in questionnaire
Position of question relative to related questions
Deliberate bias in questionnaire wording or deception in experiment
Number of words in question
Difficulty of words in question
Social desirability of answer

Time and Memory Factors

Time period
Records available
Aided recall

Characteristics of Respondents

Age
Sex
Occupation
Income
Education
Race
Religion
Political preference
Household size
Anxiety
Hostility
Conformity
Personal effectiveness
Yeasayer
Mobility

Characteristics of Interviewers

- Age
- Sex
- Education
- Occupation of household head in interviewer's household
- Race
- Religion
- Social class
- Experience
- Training
- Expectations
- Hostility
- Anxiety

Type of Data

INTERACTION IN THE RESEARCH INTERVIEW: THE EFFECTS OF RAPPORT ON RESPONSE

Carol H. Weiss, Bureau of Applied Social Research

For over forty years, researchers have realized that the interview is a somewhat leaky receptacle to take questions out into the field and bring answers back. They have found that much can go awry during the interviewing. Accordingly, a good deal of effort has been devoted to studying who and what cause errors in reporting.

We can identify three main traditions in this research. The first tradition looks at the people involved: the interviewers and the respondents. It has studied their demographic characteristics (age, sex, race, etc.), knowledge, attitudes, and personality characteristics, in an attempt to locate the culprits on both ends of the interview tandem.

The second tradition of research on interviewing recognizes that the interview is an interaction. Each individual is only one party in a two-party set. The emphasis, therefore, has been on the match between interviewer and respondent--on much the same kinds of characteristics that the first school has examined. Much of this research has focused on similarity and dissimilarity between interviewer and respondent on race, but there have been studies, too, of the effects of matching and non-matching in age, sex, social class, and religion.

The third major trend in interview research developed from the realization that, whatever the objective characteristics of the parties to the interview, the interviewer may be able to alter the respondent's perception of the situation by his behavior. He can provide cues that mediate the definition of the interview and establish the role that the respondent is to perform.

This line of investigation--looking at what actually goes on in the interview--appears particularly promising for a number of reasons. One is that studies of interviewer and respondent keep coming up with contradictory results. (This is partly because all other things are almost never equal. Aside from interviewer behavior, other important factors affect response, such as the topic of the interview and the degree of threat that the questions present.) But further, the study of the dynamics of interaction offers the promise of insight into behaviors that can be manipulated by the study director. If styles of interaction have an effect, then once he can define optimal interviewer behaviors, the study director can train interviewers to act in appropriate ways. Thus he increases his control over features of the interview that affect completeness and accuracy of response. There are indications that some status differences are so large--e.g. between high status whites and low status blacks--that nothing that goes on in the interview can affect the outcome, but there is a large intermediate range where interviewer behavior probably has important effects on the accuracy of response.

Few studies so far have looked directly at interviewer and respondent behaviors in the interview. Most of the research that has been done along these lines to date has been inferential. Thus, Ehrlich and Riesman (1961) found that young women interviewers received more "socially unacceptable" answers from 16-18 year old girls on parental norms and independence than did middle-aged or older women interviewers; but when middle-aged women were empathetic (as judged by personality test scores), they could bridge the age and authority gap and get more socially unacceptable answers, too. Williams (1968) and Dohrenwend, Colombotos, and Dohrenwend (1968), and Weiss (1968-69) have all found signs that the interviewer who takes a personal approach to respondents, or is friendly, or has good rapport, gets different responses from respondents than other interviewers. All of these studies, however, judged the interviewers' behavior at second hand--from personality inventories, interviewers' answers to questionnaires, or interviewer ratings of their rapport with the respondent.

You will notice that the word "rapport" has surfaced. Because there has been so little study of what interviewers and respondents say and do during the interview, the discussion of interaction in the research literature has tended to focus on the concept of rapport, and it is rapport that I intend to talk about in the rest of this paper.

First, let me get some assumptions out of the way. For standard survey interviews on routine topics with middle-class respondents, there probably is little difficulty in getting candid answers. Under two conditions, style of interaction is significant--and rapport has been assumed to be necessary:

> <u>first</u>, on questions that pose some degree of threat, either because they involve attitudes or behaviors that violate social norms (abortions, participation in riots, drug use), or because they threaten the respondent's self-image (not contributing to charity, making loans from finance companies.
>
> <u>second</u>, with special kinds of respondents--blacks, low socio-economic groups, political elites, addicts, etc.

Another set of assumptions is that there are types of information that some respondents will not reveal under any conditions to any interviewer; the reasons vary from lack of knowledge or lapse of memory to suppression and repression. And of course, the conceptualization of questions, their wording, and structure will affect the accuracy of reporting. Let us assume that all these things are true.

Now back to rapport. Much of the discussion in the research literature takes for granted that rapport is good. It motivates the respondent to talk. It makes anything he says acceptable. According to this prevalent view, it encourages completeness and accuracy, and for good measure, it rewards him for the effort of reporting.

The term "rapport," so far as I can tell, was imported into survey research from psychotherapy. There its main function is to overcome the patient's resistance to revealing himself, and to encourage him to pour out even undesirable and painful information in its accepting and supportive atmosphere. There is recognition in therapy of the dangers of over-rapport, but this is a matter of patient transference or of therapist over-involvement which gets in the way of the therapeutic task.

In survey interviews, over-rapport has been recognized as a risk, too. Hyman (1954) warned about the hen-party, the over-social relationship between interviewer and respondent that would lead the respondent to maintain the norms of polite social discourse and avoid unpleasantness, and might even influence the respondent to tailor responses to fit the perceived opinions or expectations of the interviewer.

Let us turn briefly to a few recent studies that have investigated the effects of rapport on response accuracy. The authors of these studies did not all use the term rapport; at least two purposely avoided it. But they all dealt with factors that others have construed as indicating rapport. An interesting fact that will emerge is that the results of these studies have been contradictory.

Williams, analyzing data from a 1960 study of Negroes' social and political attitudes in North Carolina, found that his first measure of rapport did not affect response. The measure was based on interviewers' scores on a personality test of "personal relations" and presumably tapped a dimension akin to friendliness. When he added another score--this one on "objectivity," there was a response effect, at least for Negro interviewers and low status Negro respondents. High friendliness/low objectivity can be interpreted as personal rapport. Negro interviewers who scored high on personal relations and low on objectivity (which meant to Williams that the interviewer was friendly and let his own opinions show) were more likely to receive liberal responses from low status Negro respondents--responses which were less "safe" in North Carolina and presumably more honest. The same relationship did not hold for white interviewers.

Dohrenwend, Colombotos, and Dohrenwend analyzed responses given to white interviewers by black and white respondents in a 1960-61 community health survey. The questions involved reports of neuropsychiatric symptoms. The measure of rapport was based on interviewers' reports of embarrassment in asking personal questions; embarrassment indicated that the interviewer had established a personal rather than a professional relationship with the respondent. The data showed that these "personal-relationship" interviewers received more (and presumably truer) symptom reports from low-income and high-income white respondents, but fewer from middle-income whites, who were seen as most similar to the interviewers. There were no differences in the answers of Negro respondents.

Hill and Hall (1963) studied upper-middle and upper class whites in an adult education program. The measure of rapport was the interviewer's ratings on a 3-item index asking about the frequency with which the respondent or the interviewer felt ill at ease during the interview and the enjoyability of the interview. Validity was measured by agreement of interview responses with those given on a questionnaire. Results showed that high rapport was associated with a high rate of question response but also with <u>low</u> validity.

My study of black welfare mothers in New York City, who were interviewed in 1966 by black interviewers, used the interviewers' ratings of rapport for each interview. Interviewers rated respondents on a 5-point scale from "confiding" to "hostile." Answers on factual items were checked against official records of registration, voting, welfare, and children's school performance. The higher the rapport, the less valid were the answers. Similarly, on attitudinal items, higher rapport was associated with more "socially acceptable" responses.

Cannell and his associates have published two studies on actual interview behaviors (1967, 1968). In the first, they sent observers along to record events; in the second they analyzed tape recordings. These were health interviews, where the respondent was asked about illnesses and health conditions. In the second of the studies they trained interviewers to reinforce each answer that reported a symptom, condition, or illness by saying things like "Yes, that's the kind of information we need," or "We're interested in that." This reinforcement technique increased reporting an average of 25%, which on a topic noted for underreporting no doubt improved its accuracy. Respondents who scored high on a scale measuring need for social approval showed an interesting reaction. They evidently found themselves in a conflict between reporting more conditions as a response to the interviewer's approval and reporting embarrassing conditions that might decrease approval.

They resolved the conflict by reporting more illnesses for family members for whom they were reporting by proxy and slightly fewer for themselves. In that way they satisfied the interviewer with greater amounts of information, while displacing the embarrassment on to the proxy relative.

From this brief review, at least two important things begin to emerge. The first is, as promised, that with the best will in the world it is difficult to reconcile the findings. Within the same-race interviewer-respondent pairs, Williams finds that rapport reduces bias. Dohrenwend, Colombotos, and Dohrenwend find that it has a curvilinear effect, reducing bias for higher and lower status groups but increasing bias for the middle status group most similar to the interviewers. Weiss, and Hill and Hall find that rapport increases bias. Cannell finds that interviewer "approval" improves reporting, but for some respondents biases the shape of the answers.

A second conclusion may be even more important. Not only are the measures of rapport so different that they make the whole concept ambiguous; in at least one case, they are directly in conflict. Dohrenwend, Colombotos, and Dohrenwend use the interviewer's embarrassment in asking certain questions as an indication of the kind of personal relationship that many people consider rapport. Hill and Hall, on the other hand, use low interviewer and respondent embarrassment to indicate rapport. When a concept is as muddy as this, it obviously needs drastic overhaul.

I recently went through the research literature on rapport and found a wide variety of definitions and operational measures. Some of the measures of rapport that have been used are:

rate (high) of eye contact between interviewer and respondent

frequency (high) of interviewer smiles, nods, gestures

frequency (high) of non-task conversation (i.e. irrelevant to interview questions and answers)

frequency (low) of "no answers" to questions

frequency (high) of interviewer reinforcement of responses (e.g. "I see," "Mm-hmm")

degree (low) of interviewer embarrassment in asking sensitive questions

degree (high) of interviewer embarrassment in asking sensitive questions

interviewer ratings of liking for the respondent

respondent ratings of liking for the interviewer

amount (high) of respondent verbal output

respondent rating of pleasurability of the interview

respondent willingness to be reinterviewed

interviewer ratings of degree of rapport

respondent ratings of degree of rapport

interviewer scores on personality tests of ascendance, objectivity, personal relations, etc.

Since there is so little agreement, I would suggest that we abandon the concept of rapport at this point and concentrate on the specific attitudes and behaviors of which--in some formulation or other--it is constituted. To advance the state of research on interaction in the interview, researchers might be well advised to examine friendliness, liking, verbal reinforcement of responses, smiles, nods, irrelevant conversation, and whatever else goes on in the interview. Let us see what effect each of these kinds of behavior has on the completeness and accuracy of response. In that way, we can learn how to improve the conduct of interviews.

Let me throw in one other study that has some relevance to this point. Rosenthal, Fode, Friedman, and Vikan (1960) conducted an experiment in which experimenters were purposely induced, through instructions and higher pay, to bias their subjects to give more favorable responses on a picture rating assignment. After the experiment, the subjects--who were summer school students--rated their experimenters. The experimenters who were most successful in biasing their subjects' answers were rated significantly higher on these characteristics: more likable, personal, interested, slower speaking, and more given to the use of hand, head, and leg gestures and movements.

I am tempted to go out on a limb and speculate that rapport has been overvalued. To the extent that it encompasses friendliness, pleasantness, a personal approach, and kindred elements, it may be as much of a danger to validity as a help. There are indications that it is inexperienced interviewers who place the greatest weight on rapport and spend the most time trying to build a pleasant relationship. This may be due in part to their fear that an unhappy respondent will break off the interview, but perhaps more to their sense of apology for imposing on people and their desire for a comfortable conversation. With experience, they apparently can concentrate more effectively on the business of the interview.

Rapport seems to be necessary for one function: to motivate respondents to work hard at the business of supplying complete and accurate information. This, of course, is a crucial function. By and large the interview is not a salient experience for the respondent, and he has little incentive to expend the energy necessary to understand, remember, and report fully. Studies have identified underreporting as the major problem in such fields as consumer expenditures, savings, health conditions, etc. If a friendly interviewer can coax better information-giving from the respondent, this is all to the good.

But we are becoming aware that such a benefit has its costs. On topics with a component of social desirability (e.g. voting, level of job responsibility, drinking, child rearing), biased reporting often presents more problems than underreporting. If rapport increases the peril of bias, we will have to learn to be more selective in our use of "rapport" and its component behaviors. We will have to gear the level of rapport to the topic under inquiry and the type of respondent group surveyed.

Perhaps interviewers who listen attentively and show that they understand and value the answers they receive are building as much rapport as they need. The important factor for securing valid answers is the respondents' understanding of his role as information-giver. Good professional performance by the interviewer, rather than personal comraderie, may do the job. Abetted by good question construction, the professional interviewer may be able to convey both the nature of the respondent role and its importance without getting caught up in other games (e.g. ingratiation, ego enhancement) that respondents play. The respondent may be able to get his rewards less from the interviewer as a person and more from the opportunity to talk and be listened to and understood. Sometimes he may enjoy the intellectual interest of the interview, and even on occasion (although this is apparently rare) derive satisfaction from the social value of the contribution he is making.

In the past, rapport has also served the unacknowledged function of rewarding the interviewer, who goes into the job because she likes people and enjoys pleasant relationships. Where rapport incurs the cost of biased response, we may be better advised to reward interviewers in other ways, e.g. by better pay or by greater involvement in the intellectual aspects of the research.

References

Cannell, Charles F. and Kent H. Marquis, Effect of Some Experimental Interviewing Techniques on Reporting in the Health Interview Survey, Institute for Social Research, University of Michigan, November 1967.

Cannell, Charles F., Floyd J. Fowler, and Kent H. Marquis, "The Influence of Interviewer and Respondent Psychological and Behavioral Variables on the Reporting in Household Interviews," Vital and Health Statistics, Series 2, Number 26, March 1968.

Dohrenwend, Barbara S., John Colombotos, and Bruce P. Dohrenwend, "Social Distance and Interviewer Effects," Public Opinion Quarterly, 1968, vol. 32, 410-422.

Ehrlich, June S. and David Riesman, "Age and Authority in the Interview," Public Opinion Quarterly, 1961, vol. 25, 39-56.

Hill, Richard J. and Nason Hall, "A Note on Rapport and the Quality of Interview Data," Southwestern Social Science Quarterly, 1963, vol. 44, 247-255.

Hyman, Herbert H., et al., Interviewing in Social Research, Chicago: University of Chicago Press, 1954.

Rosenthal, Robert, Kermit L. Fode, C. Jack Friedman, and Linda L. Vikan, "Subjects' Perception of Their Experimenter under Conditions of Experimenter Bias," Perceptual and Motor Skills, 1960, vol. 11, 325-331.

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

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

SOME PROBLEMS OF STATISTICAL INFERENCE IN DYNAMIC SOCIOLOGICAL MODELS

Kenneth C. Land
Russell Sage Foundation and Columbia University

Abstract

The most uniquely sociological context of dynamic analysis is that of the longitudinal model of several time points of observation on multiple units. In this context, the present article emphasizes that valid statistical inferences are possible only if the model of the process under consideration is precisely and parsimoniously specified. Assuming a lag-1 stationary stochastic process with serially independent, constant variance disturbances and additive, serially independent, constant variance measurement errors to be an appropriate model, a test is given for the hypothesis that the stability parameter is constant across two or more time intervals. Given the more parsimonious model which this result provides and an appropriate *a priori* specification of a correlation process, a procedure is derived for estimating the degree of correlation among the measurement errors or among the disturbances. The latter result requires at least four time points of observation.

1. Introduction.

Although sociologists sometimes analyze observations consisting of time series of measurements on a single social entity such as an organization or a population [7], the more uniquely sociological context of over-time statistical analysis seems to be that of a time series of observations on a number of social units which is often referred to as a panel or longitudinal analysis. Furthermore, although not a necessary condition of such designs, longitudinal studies are typically characterized by the number of time periods of observation on each social unit being smaller than the number of units observed at each point. Thus, the statistical context of dynamic sociological analysis is somewhat different from that of the time series situation which typically is treated in the statistical and economic literature. At best, the context of dynamic sociological analysis could be characterized as a situation of multiple time series [9].

Within this context of dynamic sociological analysis, questions of a statistical nature immediately arise. Several recent methodological papers have been addressed to such problems [1,6,10]. In this brief paper, I shall review and analyze some of the problems of statistical inference which have been raised by these and other authors. The emphasis will be on presenting problems and posing possible approaches to the problems rather than on the systematic development of definitive solutions.

2. Specification of Models.

Consider the longitudinal model represented in Figure 1 where, for convenience, the variables are assumed to be measured in standard deviation units. Notationally, the x_{tn} are the "true" scores of a variable on a sample of $n = 1,\dots,N$ individuals measured at $t = 1,2,3$ time points, the x'_{tn} are the corresponding obtained measurements on the variable, the u_{tn} are the random disturbances or shocks affecting the true values of the variable at the time intervals, and the e_{tn} are random variables representing the errors of the measurement in the x_{tn}. This is the model for longitudinal analysis considered by Heise [6]. In brief, it represents an attempt to separate the stability and error components of observed scores on a sample of N individuals at three time points.

Heise [6, p. 98] specified the model in Figure 1 according to the following assumptions:

(i) the relationship between the true variable x_{tn} and the index x'_{tn} is constant over time, i.e., the reliability coefficient p_{xx} is constant over time periods;

(ii) the measurement errors e_{tn} are uncorrelated with the true scores x_{tn};

(iii) the random shocks u_{tn} are serially uncorrelated;

(iv) the measure errors e_{tn} are serially uncorrelated;

(v) the rate of change in the true scores x_{tn} is approximately constant within the measurement interval.

Assumption (i) has recently been criticized by Wiley and Wiley (1970). They point out that the assumption of constant reliability requires, in general, that both true score and error variance be constant. This latter condition will not hold for any social process which results in an increase or a decrease in the variability of true scores without necessarily affecting the characteristics of the measuring instrument. Furthermore such an assumption will not allow comparisons across populations. On the other hand, many populations possess aggregate equilibrium with respect to a given variable over relatively short time intervals. That is, although the value of the variable may change for certain individuals

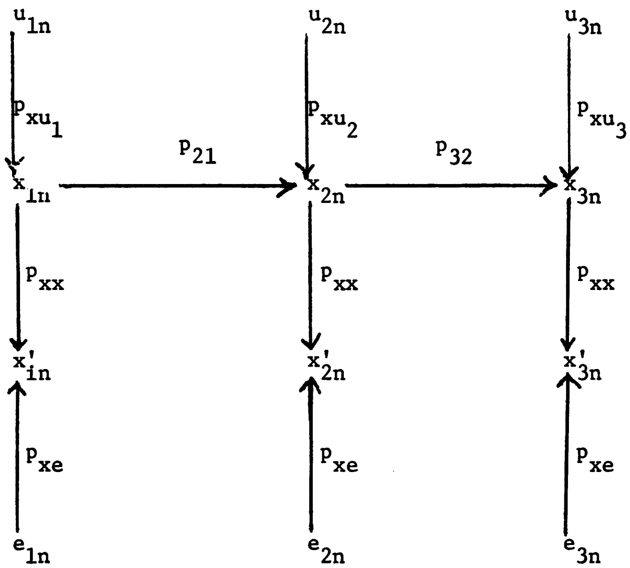


Figure 1

in the population, the aggregate result of individual changes is such that the population variances and covariances are constant. For populations governed by such processes, Heise's assumptions do not appear to be too strong.

Under the assumption outlined in the previous paragraph, Heise [6, pp. 96-97] showed that one could solve the following estimation equations

$$
\begin{aligned}
r_{12} &= p_{xx}^2 \, p_{21} \\
r_{23} &= p_{xx}^2 \, p_{32} \qquad (2.1) \\
r_{13} &= p_{xx}^2 \, p_{21} \, p_{32}
\end{aligned}
$$

for the following estimates of the reliability and stability coefficients

$$
\begin{aligned}
\hat{p}_{xx}^2 &= r_{12} \, r_{23} \, / \, r_{13} \\
\hat{p}_{21} &= r_{12} \, / \, p_{xx}^2 = r_{13} \, / \, r_{23} \qquad (2.2) \\
\hat{p}_{32} &= r_{23} \, / \, p_{xx}^2 = r_{13} \, / \, r_{12}
\end{aligned}
$$

where the symbol "^" over the coefficients denotes an estimate of the corresponding population quantity based on a sample and r_{ij} is the sample estimate of the population correlation ρ_{ij} of the variables x_{in} and x_{jn}, i,j = 1,2,3. However, Heise [6, pp. 99-100] also found that it is not possible to estimate the parameters of the model if either the assumption of serially uncorrelated random shocks is relaxed even if one adds additional time points of observations on the sample. Specifically, for the specified model, Heise [6, p. 100] could derive only a consistency relationship among the bivariate correlation coefficients for four observation points which must hold within sampling errors if the data were drawn from the specified model:

$$r_{14} \, r_{23} = r_{13} \, r_{24}. \qquad (2.3)$$

In this paper, we shall show that Heise's inability to relax the assumptions of his model stems from the incomplete and unparsimonious specification of the model. In particular, we shall present methods for testing the hypotheses that (1) the coefficients p_{21} and p_{32} are identical and (2) the measurement errors are uncorrelated. These results depend upon the more complete specification of the model as a stationary stochastic process.

3. Testing for Equality of Stability Coefficients for Two or More Time Periods.

We would like to specify that the causal model governing the relationship among the true scores in Figure 1 is of the form

$$x_{tn} = p_x x_{(t-1)n} + p_{xu_t} u_{tn} \qquad (3.1)$$

where we assume that

(i) the random shock variance is time-homogeneous for all observation points after the initial observation $[v(u_2) = v(u_3)]$, i.e., the process is stationary;

(ii) the random shocks u_{tn} are serially and contemporaneously independent;

(iii) the process is lag-1 (or Markov) and the parameter p_x is constant across time intervals of equal length.

With respect to a measurement model for Figure 1, we would like to assume that

(iv) the measurement errors e_{tn} are independent of the true scores x_{tn};

(v) the measurement errors are serially independent; and

(vi) the measurement error variance is time-homogeneous

$$[v(e_1) = v(e_2) = v(e_3)].$$

Assumptions (i) and (vi) together suffice to determine time-constant reliability. In brief, this model is the most parsimonious model for longitudinal sociological analysis which can be specified. It can be called a constant lag-1 stationary stochastic process with serially independent, constant variance disturbances and additive, serially independent, constant variance measurement errors.

The most important differences between this specification and that given by Heise are that the stationary property of the process is explicit, that the model is specified as Markovian with constant stability parameter p_x, and that the disturbances and measurement errors are assumed to be serially independent rather than serially uncorrelated. The constant value of p_x typically has not been assumed in longitudinal sociological models of [cf., 6, 10, 1]. However, given equal time intervals between observations and the plausibility of the remaining assumptions, it would seem natural to assume that p_x is constant.

Since the remainder of the assumptions are as specified by Heise, we should like to inquire here as to how one could statistically evaluate the assumption that p_x is constant. One possible approach to an evaluation of the assumption of a constant stability parameter is to utilize a large-sample modification of a dummy variable method based on the analysis of covariance which was recently presented by Gujarati [5]. For N observations at three time points, the test can be represented as follows:

$$x_{tn} = a_1 x_{(t-1)n} + a_2 (Dx_{(t-1)n}) + u_{tn}, \quad t = 2,3,\ n = 1,\ldots,N \tag{3.2}$$

where $D = 1$ if $t - 1 = 2$,
$= 0$ if $t - 1 = 1$.

The matrix representation of (3.2) is

$$\begin{bmatrix} x_{21} \\ \cdot \\ \cdot \\ \cdot \\ x_{2N} \\ x_{31} \\ \cdot \\ \cdot \\ \cdot \\ x_{3N} \end{bmatrix} = \begin{bmatrix} x_{11} & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ x_{1N} & 0 \\ x_{21} & x_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ x_{2N} & x_{2N} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} u_{21} \\ \cdot \\ \cdot \\ \cdot \\ u_{2N} \\ u_{31} \\ \cdot \\ \cdot \\ \cdot \\ u_{3N} \end{bmatrix} \tag{3.3}$$

which is an aid to visualizing the test. The test of equality of the regression coefficients for the two time intervals is as follows: one computes the least squares estimates of a_1 and a_2. Under the additional assumption that the disturbances are drawn from a normal distribution, the least squares estimators of a_1 and a_2 will be the maximum likelihood estimators. This insures that they will be the best asymptotically normal (BAN) estimators. Furthermore, the estimated approximate covariance matrix of the vector $[a, a_2]'$ is $\hat{\sigma}_u^2 (M_{xx})^{-1}$ where $\hat{\sigma}_u^2$ is the estimated variance of the disturbance term and M_{xx} is the moment matrix of the x variables in (3.2) [cf. 2, p. 386]. Since the estimators of a_1 and a_2 are asymptotically normal, it follows that one can form the ratio $\hat{a}_2/s(\hat{a}_2)$ and use it to test the hypothesis $H_o: \hat{a}_2 = o$ under the normal distribution where $s(\hat{a}_2)$ is the estimated standard deviation of $\hat{a}_2$. Of course this test is only approximately appropriate for small samples. If $\hat{a}_2$ is statistically significant, then $\hat{p}_{32} = \hat{a}_1 + \hat{a}_2$ while $\hat{p}_{21} = \hat{a}_1$ so that one must reject the null hypothesis that p_x is constant. On the other hand, if $\hat{a}_2$ is statistically insignificant, then $\hat{p}_x = a_1$ for both time periods.

This test can also be utilized to test the equality of constant terms in case the variables are not measured in deviation or standard deviation units. For such a test, one merely enters the dummy variable into the equation in an additive fashion. In the present context, the method can also be generalized to test for equality of the coefficients for an arbitrary number of time intervals. Thus, given that the other specifications of the model are maintained, one can statistically test the hypothesis that the coefficient p_x is constant across time intervals. Hopefully, sociologists will be encouraged to utilize this test, because of the more parsimonious representation which it facilitates.

Finally, it should be noted that the assumption of constant p_x has very little effect on the estimators of the reliability and the stability coefficients as computed by Heise. In particular, the formulae corresponding to those in (2.2) are

$$\begin{aligned} \hat{p}_{xx}^2 &= r_{12}^2 / r_{13} \\ \hat{p}_x &= r_{13} / r_{12}. \end{aligned} \tag{3.4}$$

Another implication of this model is that $r_{12} = r_{23}$. In other words, the correlations between adjacent time periods should be equal. In fact, for variables measured in standard deviation units, a test of this equality is equivalent to the above test of the equality of the stability coefficients.

4. Serial Correlation.

As we noted above, under his assumptions, Heise [6] was unable to derive estimates of the stability and reliability coefficients if either of the assumptions of serially uncorrelated shocks or of serially uncorrelated measurement errors was modified. Yet, Heise [6, p. 98] himself offers a strong argument for seeking a model in which serial correlation of measurement errors can be accounted for: "The assumption may be violated when respondents recall earlier answers and try to be consistent in their responses. In such a case, distortions occurring in early measurements will tend to be reproduced over time." Since the test of a constant stability coefficient among time intervals in the preceding section allows a simpler representation of the longitudinal model, we should like to inquire here as to whether this simplification can be utilized to allow the model to account for serial correlation of the disturbances or of measurement errors.

Before considering the estimation of serial correlation of measurement errors, we should possess some method of ascertaining whether or not serial correlation is a critical problem for a given set of observations. One possibility is to apply the Durbin-Watson d statistic [3, 4]. However, the validity of that statistic rests on the assumption that all of the explanatory variables in a regression equation are exogenous variables (as distinguished from lagged values of the dependent variable). Since this assumption is not valid in the present case, the statistic would be biased. Another possible approach is to regenerate the correlations between the variables on the basis of the estimated reliability and stability coefficients and to compare these with the correlation coefficients computed from the observations. The difference between the observed and the derived correlation coefficients is the maximum-likelihood estimator of the correlation between the residuals (consisting of both random shocks and measurement errors) as Land [8] has shown. Under suitable assumptions about the normality of the distributions of the residuals, this estimator can be subjected to the standard test of significance for correlation coefficients. However, an indication of positive correlation among the computed residuals by this test does not resolve the issue of whether the correlation is due to correlated measurement errors or to correlated disturbances. To resolve this issue, substantive theoretical considerations must be brought to bear on the interpretation. If the measurements consist of some social psychological test administered in relatively short time intervals and for which one would theoretically expect a strain for consistency, then one may interpret a positive correlation of residuals as a consequence of correlated measurement errors. On the other hand, if one suspects that there is some unmeasured variable that has stability over time and has a substantial effect on the true scores x_{tn}, then one may prefer to interpret a positive correlation of residuals as due to correlated disturbances.

Given that one has satisfied himself that the correlation among measurement errors is substantial enough so that one should take it into account, then the relevant question now becomes whether or not we can estimate the value of the correlation. The answer is that one can estimate the correlation among the measurement errors _if_ one obtains at least a fourth time point of observations on the sample _and if_ one places suitable _a priori restrictions_ on the process which is assumed to have generated the correlation among the measurement errors. To illustrate, suppose that we have a fourth set of observations on the variable x. Then we possess effectively only one additional piece of independent statistical information -- the observed correlation r_{14}. This severely restricts the class of processes which one can assume to have generated the correlation among the measurement errors. Specifically, one can consider only one-parameter processes for this set of observations.

One possible set of _a priori_ assumptions regarding measurement errors which can be fitted in the present context is to assume that the errors e_{tn} follow a first-order (Markov) autoregressive scheme

$$e_{tn} = p_e \, e_{(t-1)n} + v_{tn} \tag{4.1}$$

where $p_e < 1$ and v_{tn} satisfies the assumptions for all t

$$\begin{aligned} E(v_{tn}) &= 0 \\ E(v_{tn} v_{(t+s)n}) &= \sigma_v^2 \text{ is } s = 0 \\ &= 0 \text{ if } s \neq 0 \end{aligned} \tag{4.2}$$

where E denotes the expectation operator. In other words, we assume that the v_{tn} are serially uncorrelated. Substantively, this set of assumptions implies that any effect on the current response of memory of earlier responses is carried through the most recent response and is a scalar function of that response. For this set of assumptions, the estimation equations which must be solved for the estimates of the stability and reliability coefficients are

$$\begin{aligned} r_{12} &= p_{xx}^2 p_x + p_{xe}^2 p_e \\ r_{13} &= p_{xx}^2 p_x^2 + p_{xe}^2 p_e \\ r_{14} &= p_{xx}^2 p_x^3 + p_{xe}^2 p_e^3 . \end{aligned} \tag{4.3}$$

Recalling that $p_{xe} = 1 - p_{xx}$, we see that this is a system of three equations in three unknowns. Although the equations are nonlinear in the unknown, they can be solved by modern numerical procedures. For example, provided that all of the observed correlations are positive, one simple approach to a solution of (4.3) is to take logarithms of the equations which then become linear and can be solved by standard linear procedures.

The method of dealing with correlated measurement errors outlined here can be generalized. The crucial assumption is that p_x is constant. This gives one more degree of freedom which can be utilized to study correlated measurement errors. For observations at four time points, one must specify _a priori_ a one-parameter process as governing the correlation among the errors. The Markov scheme given above is only one possibility. Another likely candidate could be derived from an exponential decay model of memory. Of course, additional time points of observation would provide more degrees of freedom which could be utilized to fit more complex processes of correlation among measurement errors or among disturbances. Hopefully, sociologists will soon possess longitudinal surveys consisting of many time points of observation.

5. Concluding Comments.

This paper has been written on the assumption that models for longitudinal sociological analysis should consist of precise and parsimonious _a priori_ specifications of the processes which are assumed to generate the observations. From this perspective, we have shown that one may test the hypothesis that the stability coefficients relating the true scores on a variable are constant over time. Given the more parsimonious model which results therefrom, we have indicated how certain kinds of correlations among measurement errors can be estimated under appropriate assumptions. Although we have dealt only with measurements of a single variable over time, the procedures presented here can be generalized to situations of measurements on several variables over time if certain conditions of identifiability of the parameters are maintained.

References

[1] Blalock, H. M. Jr.
1970 "Estimating measurement error using multiple indicators and several points in time." American Sociological Review 35: 101-111.

[2] Christ, C. F.
1966 Econometric Models and Methods. New York: Wiley.

[3] Durbin, J. and G. S. Watson.
1950 "Testing for serial correlation in least squares regression. I." Biometrika 37: 409-428.

[4] 1951 "Testing for serial correlation in least squares regression. II." Biometrika 38: 159-178.

[5] Gujarati, D.
1970 "Use of dummy variables in testing for equality between sets of coefficients in two linear regressions: a note." The American Statistician 24: 50-52.

[6] Heise, D. R.
1969 "Separating reliability and stability in test-retest correlation." American Sociological Review 34: 93-101.

[7] Land, K. C.
1970a "A mathematical formalization of Durkheim's theory of the causes of the division of labor." in E. F. Borgatta (ed.), Sociological Methodology: 1970. San Francisco: Jossey-Bass.

[8] 1971 "Identification, parameter estimation, and hypothesis testing in linear static-stochastic simultaneous-equation sociological models." Presented at a Conference on Structural Equation Models at the University of Wisconsin at Madison, November 12-16, 1970 and to be published in a volume of the of the Conference papers. Mimeo. Russell Sage Foundation: New York.

[9] Quenouille, M. H.
1957 The Analysis of Multiple Time Series. New York: Hafner.

[10] Wiley, D. E. and J. A. Wiley.
1970 "The estimation of measurement error in panel data." American Sociological Review: 35: 112-117.

A LONGITUDINAL STUDY OF LABOR MARKET BEHAVIOR--ADVANTAGES AND SOME METHODOLOGICAL PROBLEMS IN ANALYSIS

Gilbert Nestel
Center for Human Resource Research
The Ohio State University

A. Introduction

I must confess that at the time I agreed to accept Dr. Sagi's invitation to prepare a paper which would dwell on selected problems in the analysis of longitudinal data, I could not help but to think that this assignment was not unlike the one presented on the TV show, Mission Impossible. Recall for the moment the format of that show: an impossible task is described; the urgency of a solution is stressed; the relevant background data are given; the sponsoring agency disassociates itself in case of a failure; and the agent always accepts his assignment, however "impossible."

The remainder of this one-hour show, for those of you in the audience who may not have seen it, consists of a series of episodes, a climax, and a final resolution of the problem. The entire program is staged with such skill, ingenuity, and insight that the viewer is often dazzled at the ability of the writers and actors to describe an urgent contemporary problem and to provide a logical, minutely detailed, albeit fantasized solution. Despite its success as entertainment the program fails as a description of human behavior simply because it does not treat in a realistic way two of the more important dimensions in all decision making: namely, the role of information and the extent to which the individual can interpret interrelated and often conflicting data. The program is staged as if the participants do not have to weigh alternatives at critical crossroads. The various forks in the road are chosen inexorably. Questions of expected costs and future benefits of different actions are ignored. Uncertainties are rarely admitted. The story develops in much the same way as a game of chess might if one of the players knew in advance all of his opponent's moves and was also sufficiently skilled to take advantage of this information. Even the most intensive analysis of such a game by the most skilled of analysts would provide at best only limited insight to the more typical games of chess. Similarly, a Mission Impossible script is hardly a prototype description of decision-making under uncertainty.

The parallel scenario describing my research experiences with panel data would overlap the script of Mission Impossible but it would also deviate from it in several substantive ways. Both the agent in the TV show and I initially face very difficult assignments. In my case, however, there are only scattered references in the social science literature where authors have addressed themselves to questions which are directly relevant to the analysis of panel data. Moreover, I do not face an identifiable single protagonist, and thus the notion of a confrontation has no empirical content. In addition, the sponsoring agency's considerable monetary investment insures that this group will not dissociate itself from the project.

There is one additional difference which distinguishes the two scripts. Whereas the agent in Mission Impossible is always successful in completing his assignment, I unfortunately do not have a comparable script writer to assure me of this success.

What I will attempt to do in this paper is to paint some of the broad contours of the subject and to illustrate its terrain with data from an existing longitudinal study. An accurate painting of this canvas and the broad dissemination of the experiences of researchers working actively in this area is a necessary condition, I submit, for subsequent analytical progress in this field.

The data referred to in the illustrations were obtained from four National Longitudinal Surveys of the civilian noninstitutional population in the United States. The Bureau of the Census administered these surveys under a contract with the Department of Labor for The Center for Human Resource Research at The Ohio State University. A sample of approximately 5,000 people was selected by The Bureau from each of the following age-sex universes: young boys and young girls between the ages of 14-24; mature women 30-44; and older men who were between the ages of 45-59 when first interviewed. These surveys were designed so that nonwhites were overrepresented and comprised about 30 percent of the total sample. Each of the cohort groups are interviewed for five consecutive years after the initial interview. The specific groups that were chosen were done so because of the interests of the staffs of the Department of Labor and The Center for Human Resource Research in questions such as the following: (1) What are the occupational and educational aspirations of young boys and girls as they filter through the education system? Do these aspirations change in any systematic way over time and, if so, in what directions? To what extent are these aspirations realized and are there significant differences among the races? (2) To what extent is a woman's investment in her education and early job experiences fully depreciated when she decides some years later to return to the labor force? Does the depreciation of her human capital help to explain her subsequent experiences in the labor market? (3) Are the retirement expectations of the older men consistent with their actual behavior and to what extent are these expectations stable from one year to the next? What are the determinants of these expectations? The content of these questions and others that were identified strongly dictated that a prospective longitudinal design be used to collect the relevant data.

The remainder of this paper is divided into two sections. In the first, I discuss the major differences between a longitudinal and a cross-section design and evaluate some of the advantages of the prospective longitudinal approach. The second section is restricted to a discussion of some of the problems that are typically present in a longitudinal study. Both sections provide an accounting framework within which this subject and some of its dimensions can be discussed.

B. Definition and Advantages

The unique properties of a longitudinal design help to differentiate it from still another survey instrument--the cross-section sample. In the latter design the sample members are interviewed once and their responses provide a basis to characterize the universe or to extablish behavioral relationships among various variables at a single point in time. Any inter-temporal changes in behavior can only be inferred from these data and only then on the basis of average relationships under ceteris paribus conditions. The selection of a prospective longitudinal design presupposes that the researcher is primarily interested in the inter-temporal changes in behavior and that he has some theoretical reasons to believe that this behavior has a temporal ordering which can only be established or tested with data from a replicated experiment. The availability of repeated observations does not distinguish the two designs, however. The composition of the samples in the two designs is of far greater significance. In a prospective longitudinal survey the individuals who are selected at the time of the first sample comprise the sample of observations in all subsequent surveys. This control of the selection process is the essential characteristic of all longitudinal designs, and it provides a significant contrast with cross-section studies even when replicated.

It is also worth mentioning that a longitudinal survey does not necessarily require that the data be collected by repeated interviews. Individuals in the sample could be asked to answer questions about their past as well as current behavior. The major reservation about this procedure remains whether these people can accurately recall their past behavior particularly when asked about their plans, attitudes, or values. To the extent that they are unable to do this, errors of measurement are introduced which can seriously bias the statistical findings. These errors have very serious implications particularly in a study where the intent of the analysis is to measure change and to study its determinants. We shall return to this matter shortly.

The data from the young girls and women surveys provide an illustration of how longitudinal data from two different cohort groups can be used to establish possible causal paths. Consider the question of whether a woman's attitude about the propriety of employment of mothers with young children is related to her own labor force participation. The cross-section data from the women's survey show that the more favorable a woman's attitude about the world of work the greater the likelihood that she will be a member of the labor force. This association is not sufficient, however, to establish whether it is the attitude that determines the behavior or whether the behavior has conditioned the attitude.

The periodic monitoring of these attitudes and the subsequent behavior of the young girls over time could very well shed some light on this question. If it could be established that the girls who have the most positive attitudes toward working mothers prior to their marriage and entry into the labor force turn out also to have a very high participation rate after marriage and the birth of their children then the direction of the relation would be established. Unfortunately, this hypothesis cannot be tested at this time since we have not yet received the requisite data. Nevertheless we intend ultimately to test this hypothesis for its empirical content.

The data from a longitudinal study can be used for purposes other than to record and measure changes in behavior, or to assess the adequacy of alternative causal models. The measurements over time on selected characteristics of the cohort members provide one basis to study the distributional properties of these variables, and to identify the extent of the possible measurement errors whatever their origin. In a cross-section analysis a response is considered an outlier if the likelihood of the variable assuming this value is very small. However, the identification of the critical regions of the distribution is somewhat arbitrary at best. In any event the decision to include or exclude an observation is typically made without the aid of any additional measurements of the respondent's behavior in other time periods.

One is in much better position in a longitudinal study to evaluate the legitimacy of the various measurements. Thus if a variable takes on a value which deviates considerably from past behavior and if the magnitude of the response cannot be explained by changes in factors that are known or assumed to be associated with this variable, then these are two strong conditions for questioning the accuracy of the response. Thus the repetitive measurements on the same respondent that is typical of a longitudinal design also provide a mechanism for identifying extreme values and assessing their probable legitimacy.

The unique features of the longitudinal design are also of help when the interviewee fails to provide an answer to a specific question.

In a cross-section analysis where one does not want to exclude an observation for which there is not complete information, what is typically done is to stratify the universe according to several dimensions (thereby attempting to control for the factors known to be related to the variable where the nonresponse occurs) and then to impute a value for the missing datum that is computed from the responses provided. The quantity that is typically imputed is the mean or median value. With longitudinal data, however, one may not have to resort to a measure that is estimated from the responses of individuals who are assumed to share common characteristics. To the extent that the respondent has provided earlier or later measures for this variable then one can impute a mean or median value that is computed from his responses. Even if this is not possible, to the extent that there is a close and stable relation between the omitted variable and other factors for which we have complete information, an average relation among these variables could be estimated--say, by the method of least squares--and the reported characteristics of the individual then used to predict either the mean value of the omitted variable or the value itself.

The availability of longitudinal data also makes it possible to apply a more stringent test of whether a hypothesized behavioral relation is stable over time. Once again the superiority of this design can be traced to its unique properties. The repeated measurements on the same respondent mean that the data do not contain a source of variation that is introduced whenever a new sample is selected. This variation is present, however, in the case of the replicated cross-section sample, and it may well explain some portion of the observed intertemporal variation in the parameter estimates. In any event the magnitude of this additional variation needs to be assessed and tested for its statistical significance.

This concludes the discussion of some of the advantages of the longitudinal survey and how this design differs from a cross-section design. I turn next to a discussion of a selected number of problems.

C. Some Problem Areas

1. Panel Mortality

Perhaps the single most serious limitation of the longitudinal design is the loss of sample cases (panel members) over time, or what is referred to as the attrition or mortality problem. These noninterviews are a legitimate concern to any researcher particularly when the losses are large, or where there is some evidence that the characteristics of the members who disappear from the sample are significantly different from those who remain as cohort members.

Whether the loss of observations seriously biases the statistical analysis is a more subtle question and one that cannot be answered definitively once and for all. The longitudinal design does make it possible to use the data generated in one of the years, generally the first, to study the characteristics of the noninterviews in a subsequent survey. Although this comparison becomes a less useful test as the time interval between surveys increases it may provide suggestive evidence when used properly to allay some of the more lingering doubts.

Our experience thus far in retaining the cohort members in the four age-sex groups have been very encouraging. The first surley of the men began in 1966 when 5,034 individials were questioned. by 1969 there were still 4,400 men, or 87.5 percent of the original group who were available to be interviewed for the fourth consecutive year. A study of the reasons for the loss of approximately 14 percent of the panel members indicated that 5 percent had either died or were institutionalized and the remaining 8 percent consisted of men who disappeared between surveys and could not be located.

The corresponding sample mortality rates for the young boys is not nearly as heartening, but part of the difference can be explained by the number of youths who entered the armed forces between survey dates. Interviews of the boys began in 1966 with 5,234 youths. Three years later (in 1969) the cohort consisted of 4,017 members or 76.7 percent of the original sample. Deaths and institutionalization accounted for less than 2 percent of these losses. Slightly more than one-half of the losses (13 percent) were to the military services. The residual classification--or, the losses that are considered to be "pure" attrition--comprised approximately 10 percent of the total sample.

We are not too surprised to find a greater attrition rate among the boys than among the men. Over time one would expect that the youths would show a greater degree of geographical mobility. Thus to the extent that these movements are positively associated with the number of disappearances, the greater the mobility potential the higher the likelihood of attrition.

At the time of this writing we have had three years of experience in interviewing the mature women and two years of experience with surveying the young girls. The initial survey of the women took place in 1967 when 5,083 women were interviewed. By 1969 the original cohort had been reduced to 4,547 members, or to 87.5 percent of its original size. Death or institutionalization of a panel member claimed less than 10 percent of these losses (or 1 percent of the total sample). Disappearances accounted for more than 90 percent of the total attrition, or 10 percent of the total sample. It might also be noted that the second interview of women was by a mail questionnaire and therefore the personal interviews were two years apart.

The final cohort group--the young girls--was first surveyed in 1968 when 5,197 female youths were interviewed. In the following year 4,971 members of the original panel were interviewed. Thus the Bureau of Census interviewers were able to locate and interview better than 95 percent of the original sample of females. The lack of any detail description of the sources of the noninterviews at this time prevent us from analyzing the reasons for these losses.

It needs to be emphasized that the low attrition rates in the different cohort groups may be explained by the comprehensive procedures that the

Bureau of the Census have adopted and followed in this study. Thus our experiences may not be representative of other logitudinal surveys.

It becomes increasingly more difficult with the passage of time to assume that the increase in the magnitude of the noninterviews has no effect on the findings of the statistical analysis. We have earlier documented that the size of the noninterviews in each of the age-sex cohort groups is not very large. Even though this in itself is very encouraging it is not evidence that the characteristics of the noninterviews are similar to those of the remaining members of the panel and that therefore no biases are introduced.

The 1970 Consuses of Population and Housing will provide an additional opportunity to evaluate implications of the noninterviews in the four age-sex cohort groups for selected characteristics. Admittedly the test will not be a perfect one since there will be the inevitable differences in the wording of the questions, who reports the information, the reference periods referred to, as well as when the interview takes place and how the variable is measured. Nevertheless this analysis will at least provide one benchmark for a potentially more intensive study of this problem area.

2. Measurement Errors

Response accuracy is a second aspect of the data analysis that has to be evaluated. Errors in measurement are not an inherent characteristic of a longitudinal design. However, if a panel study can be thought of as a series of cross-section surveys where the composition of the sample has been restricted, then the likelihood that a specific characteristic of the individual will be reported inaccurately at least once may be greater under this design than in a replicated cross-section survey where the composition of the sample changes over time. The presence of these errors thus depends on the length of the survey period, but it is also conditioned by the types of questions asked, the period of recall, and the evolving attitude of the respondent and his relationship to the person doing the interviewing.

The significance of these errors for the statistical analysis is also not a priori determined but is related to the focus of the analysis and to the distributional properties of the different error terms. For example, if the purpose of the cross-section analysis is to estimate a universe parameter (e.g., a mean value or total) and the aim of the longitudinal analysis is to study the intertemporal stability of these parameters, then the errors in measurement have one kind of implication. The same two sets of data could also be used to estimate a behavioral relation in which case any variables reported with error would have a different implication.

I can report in this context two of our many experiences which clearly highlight the potential seriousness of this problem.

A case in point has to do with the extent of intrafirm occupation mobility among blacks and whites between the first and second surveys of the men. The universe is restricted in this discussion to men who were employed in both years and a change in occupation is measured by the difference in their 3-digit occupation codes in the two time periods. The data as orginially tabulated by the Bureau of the Census show that 18.4 percent of the whites and 19.7 percent of the blacks appear to have made an occupational change. When, however, the sample is further restricted to include only men who provided a reason for their change we find that only 3.3 percent of the whites and 4.2 percent of the blacks who were eligible to change actually showed a change in their occupations.

Variation in rates of pay between 1966 and 1967 for the cohort of men 45-59 provide a second illustration of how response errors can affect the statistical findings. Between the first two years, 21 percent of the blacks and 23 percent of the whites reported a decrease of 10¢ per hour or more in their rate of pay. Even when the universe is restricted to men who have the same job status in each year the percentage of whites and blacks whose wage rate declined by 10 or more cents per hour did not change significantly. Thirty-one percent of the whites and 20 percent of the blacks reported a decline of at least 10¢ per hour between the two years. Moreover, the decline was 30¢ per hour or more for 14 percent of the whites and 11 percent of the blacks.

The sharp decline in reported rates of pay cannot be explained by any identifiable changes in the characteristics of the respondents or by changes in their environment. We chose therefore to select a sample of these men and study how the wage rate variable was calculated in each of the two years. Forty-nine men were selected at random or 10 percent of the total number of sample cases and the rate of pay associated with those individuals identified. An hourly rate of pay was reported in both years by only two respondents. In another three cases a coding error or misplaced punch was found. And of the 46 cases where no errors in punching or coding was present slightly more than one-half of the respondents reported their earnings in different time periods. In addition, one fourth of the men also reported a different number of hours worked in the two periods.

The unit of measurement differences, or intertemporal metric, has obvious implications for the computation of the wage rate and may help to explain some of the observed intertemporal variations in these rates. In the cases where the respondent fails to provide an hourly rate this rate was estimated by dividing his reported earnings on his current job in whatever time unit he chose to express this amount by an estimate of the total number of hours worked in the same time period. Thus the computed wage rate can be in error if the numerator, denominator, or both of these measures are reported inaccurately.

Before closing this discussion I might mention that the selection of the four age-sex cohort samples also makes it possible to study some aspects of the errors-in-variable question.

The Bureau of the Census has had to exercise considerable discretion in designing this study or run the risk of incurring prohibitive costs. In effect what they have done to contain these costs is to allow certain households to be represented in more than one cohort group. In the survey of the older men, for example, one out of every three households has at least one other member of the household represented in one of the four age-sex cohort samples. In approximately two out of every three households in the sample of women there is at least a second household member represented in one of the cohort groups. Only one out of every four households in the young boys and young girls samples can be classified as a single respondent household; the other households all share the common characteristic of having at least two household members represented. Finally, approximately 3 percent of the households in each cohort group were of sufficient size to have at least one member represented in every cohort group.

The overlapping of households (this may well explain the low attrition rates we observe over time) provides one way to study the frequency and magnitudes of the response errors for selected characteristics of the respondents. For example, these selected households can be studied to examine whether the family's total income in a calendar year or its income by source as reported by the wife in the women's survey is consistent with the responses reported by her husband in the men's survey. And to the extent that the household is represented in the surveys of young men and young women one can study the extent to which these members have accurate knowledge of the family's total income. The same kind of approach can also be followed in studying whether the number of weeks worked in a calendar year by the male head of the household, or the occupation he held for the longest period within the year are known by other family members in the household. The longitudinal nature of these studies also makes it possible to study these questions across a time domain as well as at a single point in time.

The research design that I have just described has its limitations, as one would expect. The different timing of the four interviews within the year means that only certain types of questions are amenable to a comparative analysis. Moreover, the differences in the content of the questions asked (particularly the time reference) also limit the kinds of comparisons which are possible. Finally, the distribution of response errors for different household members and their characteristics are likely to also vary over the lifetime of the longitudinal survey.

One suspects that the variations in responses among household members will be larger in the first year of the interviews but that these differences will be reduced in subsequent reinterviews. The repeated questioning of the same respondents could condition them to answer more consistently over time than they would have done in a single interview. In addition, the ease with which individual members of the same household can exchange information also suggests that they would be more apt in a later interview than in the first survey to give more consistent and accurate responses. Both of these factors point to smaller response errors with the passage of time. On the other hand, the respondents may lose interest in this survey and become annoyed at having to respond periodically to very similar types of questions. This may well lead to indifferent answers that are largely inaccurate.

3. Nonresponses

A third major problem area, and one that is again not unique to longitudinal studies has to do with the frequency of the nonresponses and the associated characteristics of the individuals who fail to provide answers to specific questions. As I suggested earlier what is typically done to neutralize this problem is to stratify the universe by various socioeconomic and demographic dimensions that are known to be associated with the characteristic that is to be imputed, and then to assign a value to the nonresponse that is based on the reported information of other people in the same stratum. The Bureau of the Census is currently following this policy in its March CPS survey which it uses to estimate the incomes of husband-wife families in the civilian noninstitutional population of the United States.

The Bureau has found that very often respondents fail to report all sources of their family's income. Rather than eliminating these families from the sample and readjusting the sample weights of the respondents who report complete information the Bureau has evolved a procedure which imputes a value to each source of income that is not reported. In practive the procedure followed is slightly different for the earnings component of income than for the "other sources." In the earnings imputation mutually exclusive classes are formed that are based on the sex, age, color, weeks worked last year, occupation, and class of worker of the respondent. The same set of controls and the earnings and employment status of the respondent are used to impute values for other sources of income. Each of the sample cases is then sequentially assigned to one of the available classes and when a member is found who does not provide information on a source of income he is assigned the value reported by the last person previously added to that class. The process continues until all assignments and imputations have been completed.

The same kind of reservations discussed earlier in the context of measurement errors holds here too. I suspect that the problem is further compounded if the variable is usually reported inaccurately. It might be true, for example, that there is a high degree of positive association between the rate of refusal and the likelihood of a measurement error. In these cases the imputated values may be in error for two reasons. First, the controls by which the various mutually exclusive classes are formed may not be sufficiently correlated with the characteristic that is being measured. But second, even if one can adequately control for this variation, a further error is introduced because of the inaccurate responses of those in the stratum who reply. In summary, the consumer of longitudinal data is faced with the dilemma of having to live with a sample which had been substantially reduced because of refusals or having to impute values for these nonresponses which may be seriously inaccurate. In either case a bias may be introduced into the analysis but the relative advantages and disadvantages of the two alternatives remain still to be evaluated.

The need for such an evaluation is very apparent from the results of the first round interviews of the various age-sex cohort groups. In the men's survey, for example, one out of every three whites and almost two out of every five blacks failed to provide complete information on his family's resources (net assets). The same set of questions was also asked in women's survey and here the responses of three out of every ten white females and one out of every five blacks were incomplete. In the young boys' survey the questions related to the net assets of a family were asked only of respondents who were heads of households. However, approximately one out of every five white male youths and slightly more than one out of every ten blacks who were not enrolled in school at the time of the interview and who also aspired to an occupation at age 30 which was different from their current occupation failed to provide a reason why they believed they will not attain their goal. These examples, and others which could be introduced, fully illustrate the potential range of the refusals.

4. Metric

A fourth problem area that has to at least be mentioned concerns the metric that is assigned to the changes in behavior that one observes intertemporally. Here again the longitudinal survey is not the only design where this question arises. Nevertheless its emphasis on measuring the same person's characteristics at several points in time means that the potential paths that this behavior may take are much more numerous.

The responses to the expected age at retirement (ERA) question is a case in point. In the first interview a panel member could have reported a specific age; he could have answered that he was already retired; he could have responded that he never expects to retire; he could have claimed that he did not know when he was retiring; or he simply could have refused to answer the question.

The same options were also available when he was reinterviewed in the second year. Thus with five alternatives in each of two years a total of 25 paths can be uniquely identified. These paths can be most easily summarized in a 5x5 matrix where the rows represent the various 1966 responses and the columns the responses in 1967. The cell frequencies when transformed to a relative base can be viewed as estimates of the universe transition probabilities, particularly if the cell frequencies are not very small.

Fortunately, the 25 transition probabilities are not of equal interest. Nevertheless, a metric may have to be assigned to one or more of the paths if the extent of changes in the expected retirement ages is the focus of the analysis. In studying the retirement decision how does one assign a measure to a man who provides an age in one year but when reinterviewed reports that he will never retire? Or how does one classify the behavior of a man who changes his expectation from a never retire response to an already retired?

5. Other Selected Problems

There are additional problem areas that are peculiar to a longitudinal design which should also be discussed. I will forego this temptation in the interest of keeping what is already a long paper within manageable proportions and simply provide a listing with only minimal comments. First, the intertemporal influence of the interviewer on the reported responses of the interviewee needs to be evaluated. Second, the extent to which the respondent recalls and reports an earlier response even though his circumstances or his environment has not changed also needs to be studied. Third, there is a need to study the extent to which the interviewing process itself conditions the change in behavior. Fourth, the use of a longitudinal design requires that the questions asked be comparable over time, that there be consistent coding of the responses, and that when subjective rules are adopted (e.g., in coding open-ended questions) that these rules be followed consistently. Finally, the prospective longitudinal design loses some of its significance if the environment which conditions the responses does not show considerable variation over the lifetime of the interviews.

D. Summary

My occasional references to the retirement decision of the older men was suggested to me by the strategy of the writers of the TV and movie previews--provide a setting and backdrop so that the viewer has an incentive to return at a later date. I had hoped to report on some of these findings at this session but I have already exceeded my allotted time and besides that study is another Mission Impossible. It might be mentioned in passing, however, that the problems discussed in this paper are general in content and carry over to that study with only minor changes.

SOME EFFECTS OF CAUSAL CONNECTIONS IN SIMULATED TIME-SERIES DATA[1]

Donald C. Pelz and Ray E. Faith

Abstract

By computer simulation, two autoregressive panels x and y are generated, defined recursively by the relations $x_{it} = p_{xx}x_{it-1} + u_{it}$ and $y_{it} = p_{yy}y_{it-1} + c_{yx}x_{it-g} + v_{it}$, where u and v are uncorrelated disturbances, $|p_{xx}| \leq 1$ and $|p_{yy}| \leq 1$ are the autoregressive coefficients, and c_{yx} represents the causal effect on y of a prior x over causal interval g.

With simulated data the correlogram of cross-correlations $r(x_{it}, y_{it+k})$, where k represents any measurement lag, exhibited an expected asymmetric shape: correlations were higher when y was measured later than x. However, it was also observed that greater stability in either variable (higher p_{xx} or p_{yy}) produced higher cross-correlations--an apparently stronger causal connection. Another counter-intuitive observation was that, for certain high levels of p_{xx} and p_{yy}, the cross-correlation was maximum not over causal interval g but rather after some delay beyond g.

By assuming stationary properties, it has been possible to apply the procedures of path analysis to derive expressions which are identical with previous derivations by Lew, at the same time making it easier to grasp the model and explain its peculiar features.

* * * *

Origin of inquiry

The present investigation originated in an inquiry more than six years ago by the senior author and a colleague at the Survey Research Center [5] on the use of lagged correlations between two variables measured at two points in time, as a possible means of detecting direction of causal influence between the variables. It seemed intuitively plausible that if a change in variable x at time 1 produced a change in the same direction in variable y at time 2, then the lagged correlation between x_1 and y_2 should be larger than the corresponding lagged correlation between y_1 and x_2. Independently, Campbell and his associates arrived at the same conjecture [1,2].

Efforts to find mathematical support for this conjecture in the time series literature were unsuccessful, although at a later point Goldberger (personal correspondence, 1969); [and 3, p. 33] indicated a rather simple way of doing so. As a heuristic device, therefore, the senior author decided to simulate artificial time series data having known characteristics for a population of hypothetical individuals, in order to investigate how the introduction of known causal influence would affect the lagged correlations between variables. Spyros Magliveras devised a program to generate two artificial variables x and y for a set of N individuals over 50 periods of time.[2]

A legitimate question is raised as to whether the same results might not have been generated directly by mathematical analysis. In answer, we hope that analysis can eventually do the entire job, but meanwhile the simulation has been useful in demonstrating certain properties that were not obvious even to time series authorities, and which have been demonstrated analytically only with considerable labor.[3]

The simulation model

The present paper will deal with the simplest situation with two variables. For N individuals, independent variables x is determined only by its own prior value and by random outside disturbances; dependent variable y is also generated as a function of its own prior value and a different set of random disturbances, and in addition x exerts a uni-directional influence over a subsequent y after a specified causal interval g.

For each individual i in a population of N individuals, two time series x_{it} and y_{it} are generated at times t = 1,2,3,... in successive operations of the computer program, by the recursion equations

(1) $$x_{it} = p_{xx}x_{it-1} + u_{it}$$

(2) $$y_{it} = p_{yy}y_{it-1} + c_{yx}x_{it-g} + v_{it}$$

where u and v are random disturbances uncorrelated across individuals and across time, $|p_{xx}| \leq 1$ and $|p_{yy}| \leq 1$ are the respective autogregressive coefficients, and c_{yx} is the causal coefficient representing the effect on y of the x value for that individual which occurred g time units earlier, g being the causal interval.[4]

The coefficients p_{xx} and p_{yy} are equal to the theoretical autocorrelations between adjacent values of the Markov series x_t and y_t [4, pp. 405 ff.]. They will be called coefficients of "short-term stability."

The model also permits incorporation of "long-term stability" consisting of individual constants for each individual ζ_{ix} and ζ_{iy}, which remain unchanged through time. There is not space here to discuss the latter; the reader is warned that effects of the two sources of stability are markedly different.[5]

Some unexpected features

From time series data thus generated, one can obtain three sets of correlations between pairs of variables across N individuals, where measurements of the variables in each pair is separated by an interval designated k. Two of the sets are the autocorrelations $r(x_{it},x_{it+k})$ and $r(y_{it},y_{it+k})$; the third set is the cross-correlation $r(x_{it},y_{it+k})$, where measurement lag k can be any positive or negative integer 0, 1, 2, ...

It is instructive to compare the behavior of these correlational functions with what was expected in advance. The time-series literature indicated that for independent variable x the autoregressive correlogram (i.e., the plot of autocorrelations $\rho(x_{it}, x_{it+k})$ as a function of measurement interval k, where the symbol ρ designates the theoretical correlation in contrast to the empirical value r) should be a declining exponential function of the form: p_{xx}^{k}. The literature gave no immediate expectation as to the autocorrelational function of y, although we assumed it would decline exponentially in much the same fashion as x. These expectations were confirmed with the simulated data.[6]

We could find no theoretical expectation for the cross-correlational function. An intuitive conjecture was that as measurement interval k approached causal interval g, the cross-correlation between x and y would rise from zero to a maximum value at g, after which it would decline again toward zero. (These conjectures are illustrated in [7, pp. 6-10.) Under certain circumstances, a cross-correlogram of this shape did appear in simulated data, as illustrated in the bottom curve in Figure 1, where p_{xx} and p_{yy} are both set = .70.

In such a case, the phenomenon we have called the "cross-lagged differential" (following a suggestion by Campbell) will appear. If one compares the cross-correlation when y is measured at g time units <u>after</u> x (in Figure 1, at the point k = +4) with the corresponding correlation when y is measured g time units <u>before</u> x (at the point k = -4) a distinct difference is observed. This differential may disappear, however, if the measurement interval departs substantially from true causal interval. (In the bottom curve of Figure 1, e.g., there is no difference in height of the curve at k = +15 and -15).

In advance of the simulation, we were uncertain what effect an increase in stability would have on the cross-correlations. It seemed likely [7, p. 45] that extremely high stability of <u>either</u> short-term or long-term character (see footnote 5) might cause the cross-correlogram to rise and fall

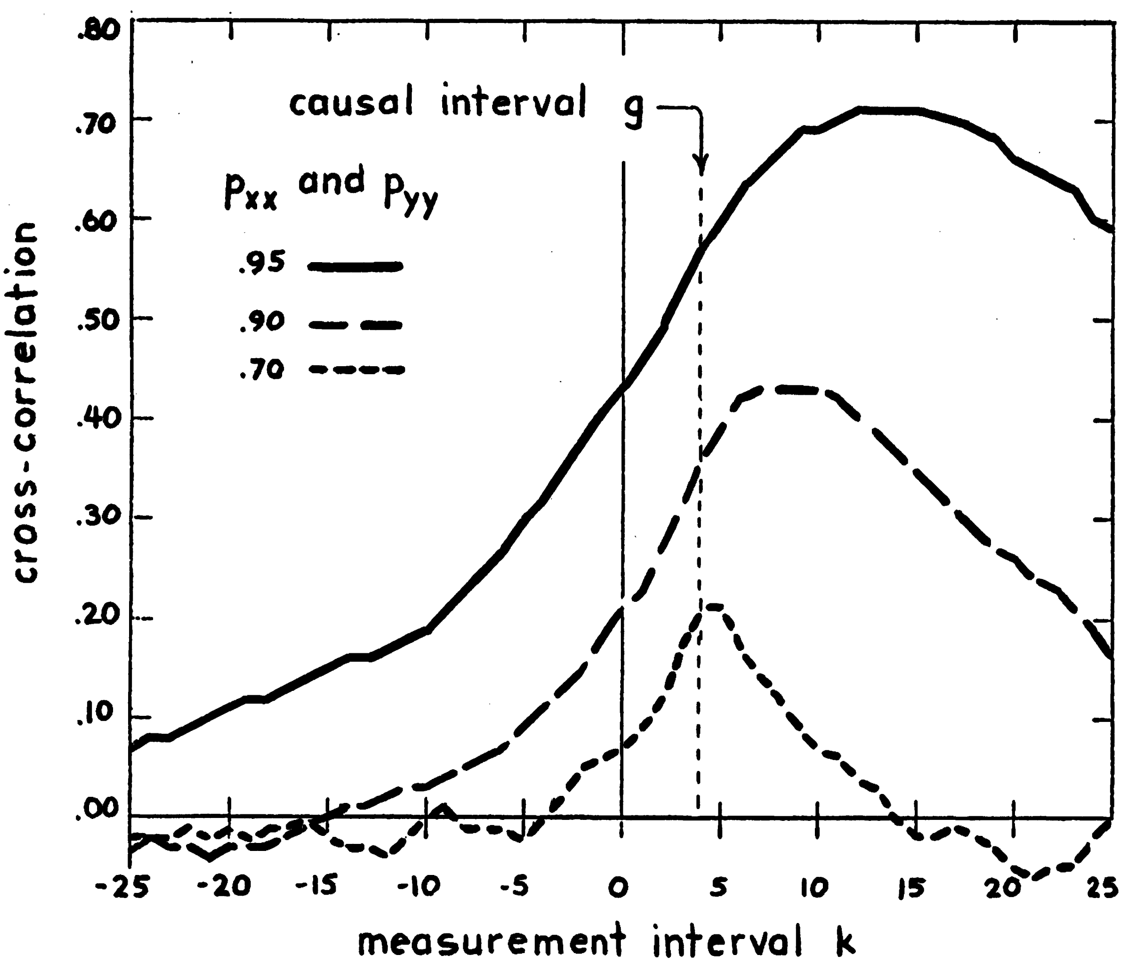


Figure 1. As autoregressive stability (p_{xx} and p_{yy}) increased for simulated variables x and y, the cross-correlations (1) became higher at point g, (2) increased in duration, (3) were increasingly delayed in reaching a maximum height beyond causal lag g. All curves used a constant causal coefficient (c_{yx} = .10). (From [7, p. 36.)

very slowly, and thus would obscure the cross-lagged differential.

When simulated data were generated and cross-correlations examined, some surprising departures from expectation appeared when high autoregressive coefficients p_{xx} or p_{yy} were used. The reasons for these departures were not intuitively obvious, but it is hoped that the mathematical analysis reported below will clarify them.

1. The higher the autoregressive coefficients (that is, the higher the short-term stability) the larger was the cross-correlation which appeared at causal interval g, and the sharper was the cross-lagged differential. Note in Figure 1 the increasing height of the cross-correlograms at $k = g = 4$, as p_{xx} and p_{yy} were increased from .70 to .95.

Thus in terms of the observed correlation between x and y, the causal influence appeared to become stronger, even though the actual coefficient of causal influence (c_{yx}) remained constant.

2. The higher the stability coefficients, the longer the cross-correlations persisted through time, over measurement lags several times longer than the causal interval. Thus in Figure 1 if one measures the cross-lagged differential--the difference in height of the correlogram at points +k and -k respectively--in the bottom curve the differential has disappeared at $k = \pm 15$. But in the top curve the differential is substantial even at $k = \pm 25$, and by extrapolation it appears to persist to $k = \pm 50$ or beyond.

Thus high autoregressive (or short-term) stability served to to magnify and to perpetuate the observable effects of causal influence.

3. Counter to intuitive expectation, the point of maximum cross-correlation was not necessarily at the point of actual causal impact (causal interval g). Rather, the higher the autoregressive coefficients, the greater was the delay between g and the point of maximum cross-correlation. In Figure 1, when the stability coefficients were extremely high (both = .95) the cross-correlation was maximum at an interval of from 10 to 15 time periods after the causal interval.

4. We had expected that a cross-lagged differential would appear only if it took time for the causal effect of x to appear in y--that is, only if the causal interval were several time units (one "time unit" being the interval for the autoregressive effect to occur). This expectation was confirmed in the case of variables with low stability. The effect of shortening g was simply to move the entire cross-correlogram to the left, without altering its shape. Thus, if the bottom curve in Figure 1 were moved 3 or 4 time units to the left, it would become almost symetrical around $k = 0$, and the cross-lagged differential would almost disappear.

Note however that with high stability (middle and upper curves in Figure 1), even if the curves were moved leftward so that $g = 0$ the asymmetry would persist, and so would the cross-lagged differential. An important property emerged: if the two time series were quite stable (in the short-term sense of high autoregressive coefficients), the causal influence of x on y remained apparent even when the causation was almost simultaneous.

Mathematical properties derived by Lew.

Robert Lew undertook the task of deriving mathematical properties of the simulation model. Like the model, he began with a finite starting point at which values for x_{it} and y_{it} were generated randomly for each of N individuals; uncorrelated random terms u_{it} and v_{it} were added; and after g time-periods had passed the causal component $c_{yx}x_{it-g}$ was added to the y variable. When the mathematical equivalent of this model was in operation over a long time certain terms vanished, and the equations could be expressed in asymptotic form.

The basic expressions for autocorrelations of x_{it} and y_{it}, and for the cross-correlations xy, are given in a technical note to Pelz and Lew [6] and in full detail in appendices of the interim report [7]. Lew's expressions generate theoretical autocorrelations and cross-correlations which parallel very closely the empirical curves produced by simulation such as those in Figure 1.

Lew also derived a prediction of phenomenon 3 noted above, that when autoregressive coefficients are high, the maximum cross-correlation will appear with some delay following causal interval g, as illustrated in Figure 2.

Theoretical properties derived from application of path analysis

Co-author Faith has conceptualized our time series model in terms of a path analysis model which[7] is fully described in the technical appendix. The path diagram is given in Figure 3. Variables x_0 and y_{g-1} are specified as the correlated inputs. From these, together with the residual disturbances u and v, new variables x_t and y_t are defined by the following recursive path formulas:

$$(3) \quad x_{t+1} = p_{xx}x_t + p_{xu}u_{t+1} + 0\cdot(\text{all variables } x_s,\ y_{s+t} \text{ where } s \quad t)$$

$$(4) \quad y_{t+g} = p_{yy}y_{t+g-1} + p_{yx}x_t + p_{yv}v_{t+g} + 0\cdot(\text{all variables } x_s,\ y_{s+g-1} \text{ where } s \quad t)$$

for $t = 0,1,\ldots$

These are similar to the basic simulation equations (1) and (2) given above, when operating over an extended time period.

Contrary to Lew's approach of assuming an arbitrary set of initial conditions which gradually stabilize, the path model assumes a system that is stationary from the outset. This assumption is achieved by a suitable restriction of the parameters of the model (notably the causal path coefficient p_{yx} in expression (4), and the correlation between the variables regarded as inputs.)

It is deduced that under the specified conditions, the autocorrelations and cross-correla-

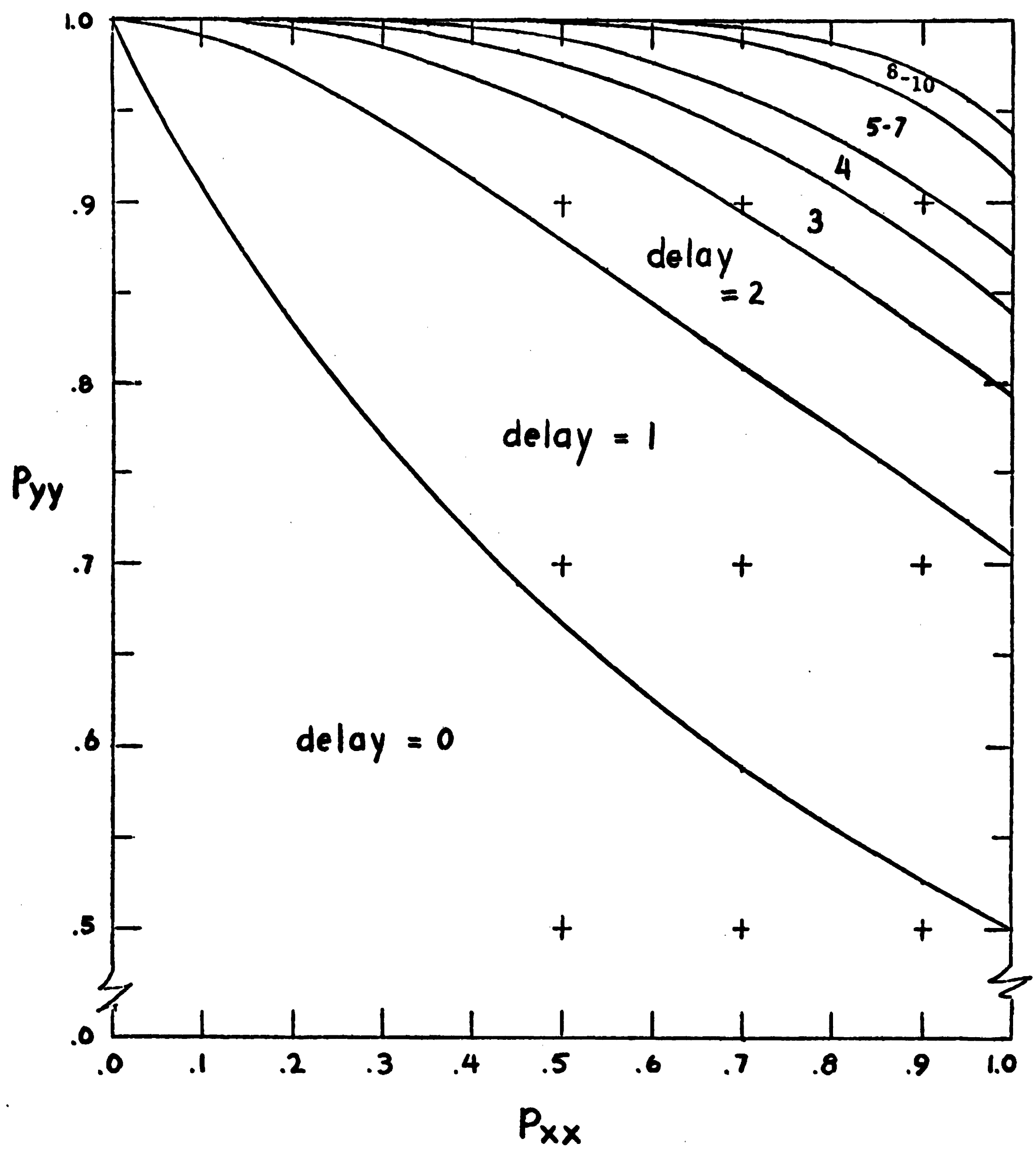


Figure 2. As autoregressive stability of the two variables increases, so does the theoretical expectation of delay between the causal lag and the measurement lag at which the xy cross-correlation is maximum. Thus, where p_{xx} and p_{yy} both = .9, a delay of 4 time units can be expected. (From [7], p. 41).

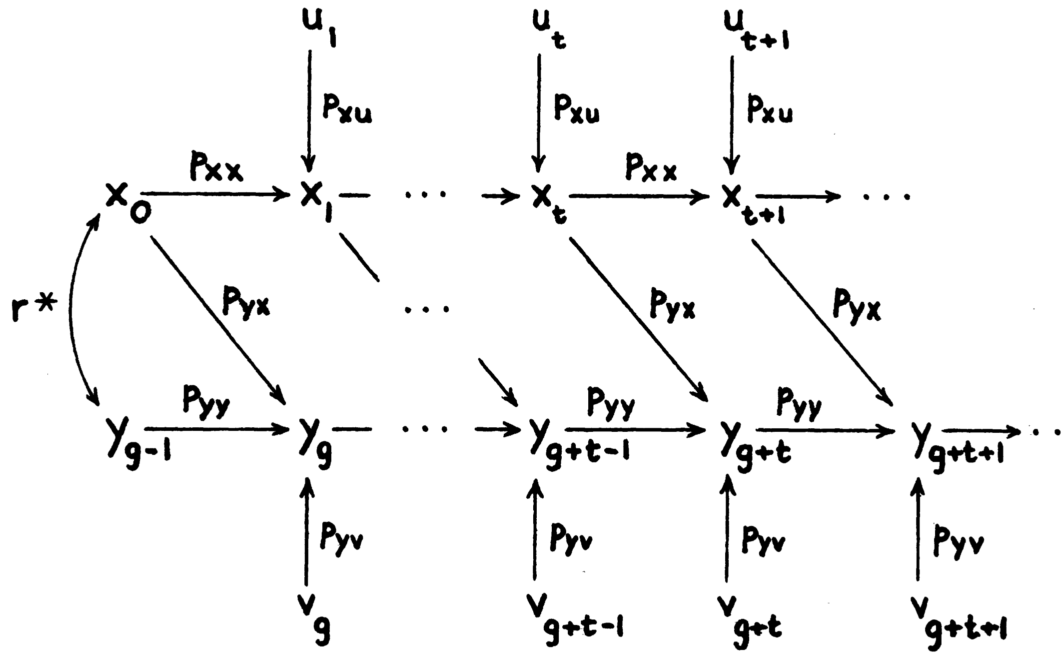


Figure 3. Path diagram of independent variable x and dependent variable y, governed by recursive relations (3) and (4) in technical appendix. Although y_g is plotted below x_1, g can take any value.

tions for the x and y variables are depéndent only on the time interval k over which the measurements are taken.

The values of these correlations are predicted from the path coefficients by the rules of path analysis. The computation of the cross-correlations $\rho(x_t, y_{t+k})$ is of particular interest. We find that these are governed not by a single formula but rather by two. One formula applies to cross-correlations for which measurement interval k is less than the causal interval g, and the other applies where k is greater than g, as illustrated in Figure 4. (The two expressions give the same correlation for k = g.) The derivation of the cross-correlation formulas is facilitated if one considers not the actual measurement interval k, but rather the amount by which k differs from g, defined as $k' = |k-g|$.

Consider first that part of the cross-correlation to the right of g (i.e., k > g). In the technical appendix it is shown that the expression for this part of the cross-correlation is given by appendix formula (7c):

$$r_{xy}(k') = p_{yx}p_{xx}r_{xx}(k'-1) + p_{yy}r_{xy}(k'-1)$$

Let us examine the two components after the equal sign in this expression. If the cross-correlation $r_{xy}(k')$ were determined only by the second component, it would be lower than the cross-correlation over the previous lag ($r_{xy}(k'-1)$), since, the latter term is multiplied by p_{yy} which is less than unity. However, the cross-correlation is increased at the same time by the first component, which depends on the size of the causal coefficient p_{yx} and also on the size of p_{xx}. When the first component adds more than the second component removes, the cross-correlation will rise from one time period to the next. Otherwise it will fall.

The technical appendix specifies the conditions (see expressions (16) and (17)) under which the cross-correlation will rise following g.

Let us return to that part of the cross-correlation to the left of g (i.e., k < g). Expression (13b) in the appendix shows that this portion is given by a relatively simple exponential function

$$r_{xy}(-k') = r_{yx}(k') = \frac{p_{yx}}{1 - p_{xx}p_{yy}} p_{xx}^{k'}$$

which rises steadily as measurement interval k approaches g.

<u>Relative influence of stability in x or y</u>

We saw in Figure 1 that as the autogreressive stability of both x and y increased, the xy cross-correlation became more asymmetrical around causal interval g. It is appropriate to ask whether this asymmetry is produced more by stability in x or by stability in y, or equally by both.

A partial answer was seen in Figure 2, showing the relative effect of the two autoregressive coefficients on the amount of delay in the maximum cross-correlation. The reader will observe that if variable y has only moderate stability (such as p_{yy} = .70), then even a very high stability in x will produce a delay no greater than 1 time unit. In contrast, if variable x has the same moderate

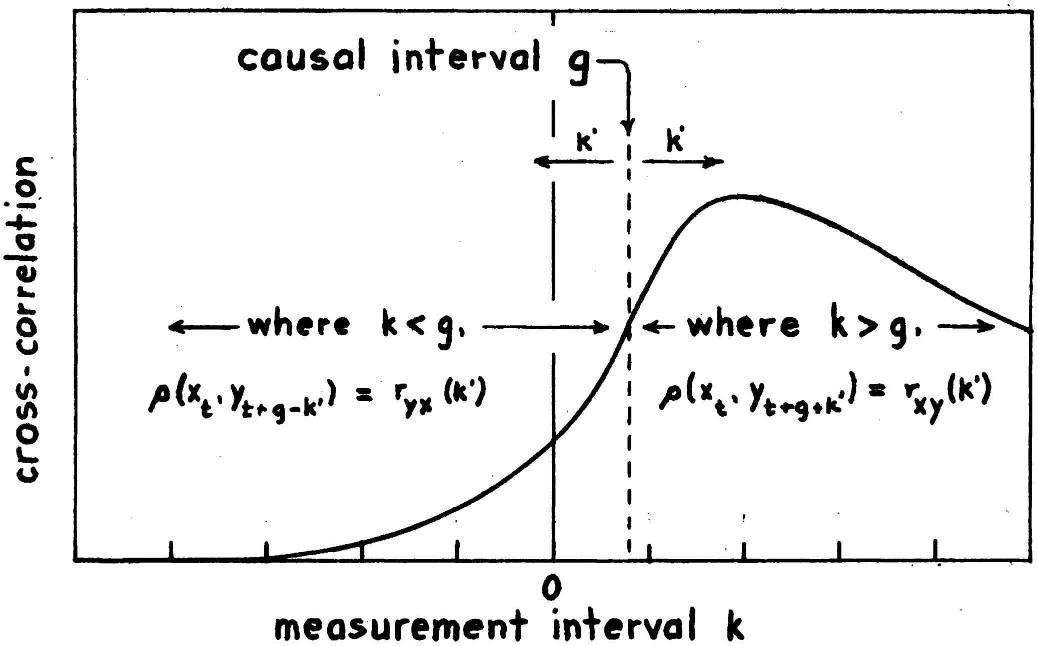


Figure 4. Formula for cross-correlation is composed of two parts, one where measurement interval $k <$ causal interval g, the other where $k > g$. Expressions for $r_{yx}(k')$ and $r_{xy}(k')$ are given in text and appendices.

stability (p_{xx} = .70), then increases in stability of y can produce delays up to 5 or more time units.

Hence it appears that the <u>stability of dependent variable y is more critical than that of independent variable x,</u> in producing asymmetry and delay in cross-correlations. (It is also true that for much longer delays of 8 or more time units, both x and y must be highly stable.)

A more complete answer can be given with the theoretical expressions derived in the appendix. These have been used to generate four cross-correlograms plotted in Figure 5. The bottom curve uses moderate stability in both variables (p_{xx} = p_{yy} = .70) and is very similar to the bottom curve in Figure 1 for simulated variables using the same parameters. The top curve, with very high stability in both variables (p_{xx} = p_{yy} = .95), is again similar to the top curve in Figure 1 with the same parameters.

The two middle curves show what happend when stability of the two variables is allowed to differ. When x is the more stable (p_{xx} = .95, p_{yy} = .80), the correlogram rises and falls gradually, with only slight asymmetry. When the two parameters are reversed and y is the more stable (p_{xx} = .80, p_{yy} = .95), the correlogram rises abruptly and is markedly non-symmetrical. Again it is clear that stability in the dependent variable is mainly responsible for the peculiar properties of asymmetry around g, and delay.

<u>In conclusion</u>

Simulated time series were created in which independent variable x was allowed to exert a causal influence on dependent variable y over causal interval g. Certain puzzling features of the simulated data are explained by application of a path analysis model. Specifically it is shown that when autoregressive stability of both variables is increased:

1. The cross-correlation between x and y becomes stronger[8] at the interval of causal influence, g.
2. The cross-correlogram becomes asymmetrical around g--that is, the correlations remain higher following g than preceding it, so that the cross-lagged differential persists over a longer measurement interval.
3. The point of maximum correlation between x and y is increasingly delayed beyond g. (Features 2 and 3 are found to result especially from stability in the dependent variable y.)

Finally, from the path equations, it is clear that when autoregressive stability is high (especially in the dependent variable):

4. The asymmetric effect of causal influence remains the same regardless of whether the causal interval is long or short; hence causal influence could be detected even when simultaneous.

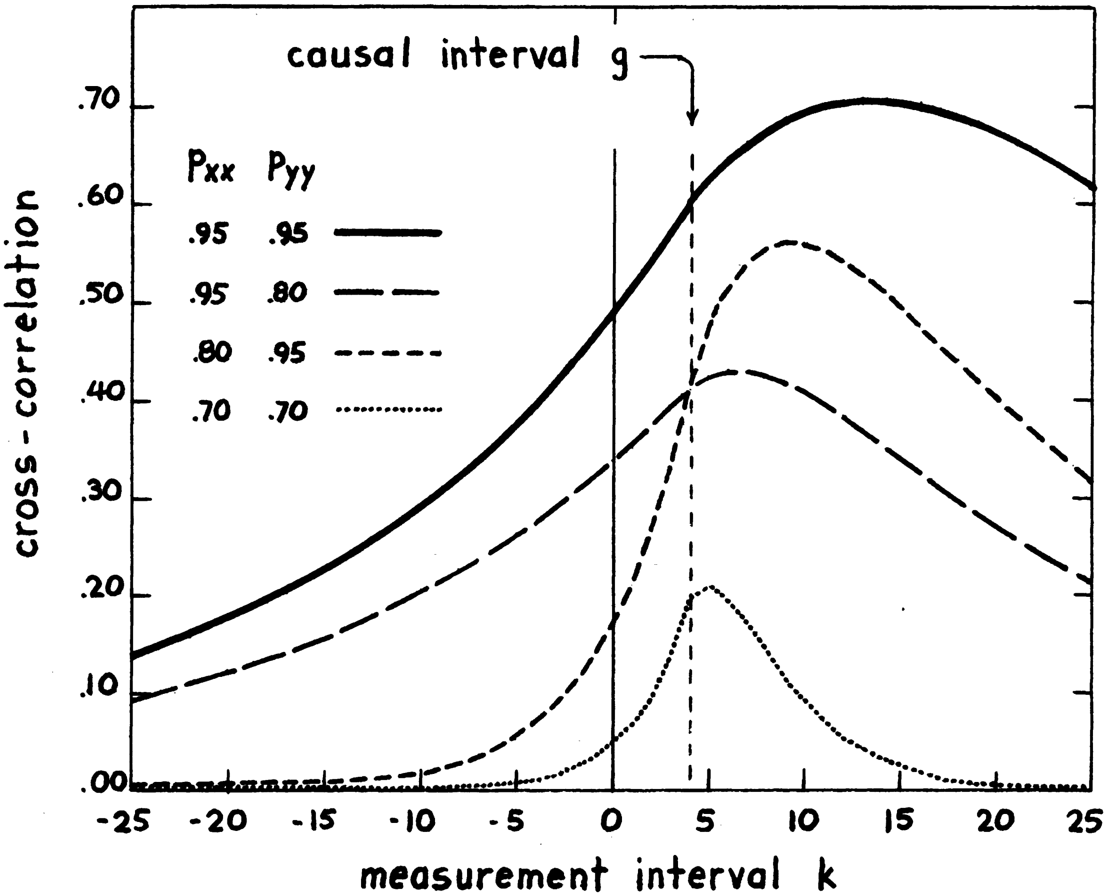


Figure 5. Theoretical cross-correlations based on path model. High stability in the dependent variable y (third curve) produces sharper asymmetry than does an equally high stability in the independent variable x (second curve). (The causal coefficients for the bottom three curves were set at $p_{yx} = .10$; for the top curve $p_{yx} = .06$.)

FOOTNOTES

[1] The initial simulation was performed under a grant from the National Science Foundation (GS-1873), with supplementary aid from the National Broadcasting Company. The original simulation program was written by Spyros Magliveras, and was revised by Robert A. Lew, assisted by George Gluski. Lew also derived the main mathematical properties of the model. The mathematical extensions reported here were supported by grants from the U.S. Office of Education (Grant No. OEG-5-9-239459-0076) and National Science Foundation (GS-2710). Omitted here is the technical appendix by Faith which was handed out at the ASA session. A copy of the full paper including appendix may be obtained by writing the senior author.

[2] Fruitful guidance was given by Graham Kalton, visiting lecturer in sociology and sampling statistics from the London School of Economics, in helping to structure the simulated model and determine some of its properties.

[3] Subsequently a more complex simulation program has been prepared which can generate up to 10 variables, each of which can exert a causal influence on any other, over any specified combination of causal intervals. This program was developed by Robert A. Lew with the assistance of George Gluski, primarily with support from the research department of the National Broadcasting Company.

[4] The u and v components are not the same as measurement error; unlike the latter they become incorporated in x_{it} and y_{it} respectively and hence enter into subsequent values. The two-variable simulation model does not allow for measurement error; all values are assumed to be perfectly reliable or "true scores." To maintain stationary variance in x_{it} over time, $|p_{xx}| \leq 1$, and variance of $u_{it} = (1-p_{xx}^2)\ \sigma_{x_{it}}^2$; corresponding limitations are imposed for p_{yy} and v_{it}.

[5] With many real social data it is inappropriate to assume a simple Markov process in which the value at time t depends only on the immediately prior value at t-1. Rather, each individual is likely to have an enduring tendency to persist at

the same level through time--equivalent to the effect of individual differences in intellectual ability, personality, or socio-economic background. The contrast between the "short-term stability" of the simple Markov process and "long-term stability" introduced by stable individual differences is demonstrated briefly in Pelz and Lew [6] and in more detail in [7].

[6] The autocorrelations were in fact higher for the dependent variable than predicted from the value of p^k_{yy}; the exact function was subsequently derived by Lew.

[7] The technical appendix is omitted here for lack of space. A copy may be obtained by writing the senior author.

[8] This property is governed equally by stability in either variable. The reader may observe in Figure 5 that the two middle curves cross at g. That is, a given pair of autoregressive coefficients assigned to either x or y will have the same effect on the size of cross-correlation at g.

REFERENCES

[1] Campbell, D.T. "From Description to Experimentation: Interpreting Trends as Quasi-experiments." Pp. 212-242 in C.W. Harris (ed.), Problems in Measuring Change. University of Wisconsin Press, Madison, 1963.

[2] Campbell, D.T. and Stanley, J.S. "Experimental and Quasi-experimental Designs for for Research on Teaching." Pp. 171-246 in N.G. Gage (ed.), Handbook of Research on Teaching. Rand McNally, Chicago, 1963.

[3] Goldberger, Arthur S. "Econometrics and Psychometrics: A Survey of Communalities." Paper EME 7013 (March, 1970), Social Systems Research Institute, University of Wisconsin, Madison. Also Psychometrica, 36 (June 1971).

[4] Kendall, M.G. and Stuart, A. The Advanced Theory of Statistics, Vol. 3, Hafner, New New York, 1966.

[5] Pelz, D.C. and Andrews, F.M. "Detecting Causal Priorities in Panel Study Data." American Sociological Review, 29 (1964), 836-848.

[6] Pelz, D.C. and Lew, R.A. "Heise's Model Applied." In Edgar F. Borgatta and George W. Bohrnstedt (eds.), Sociological Methedology--1970, 1970.

[7] Pelz, D.C., Magliveras, S., and Lew, R.A. "Correlational Properties of Simulated Panel Data with Causal Connections between Two Variables." Causal Analysis Project, Interim Report No. 1, 110 pp. Survey Research Center, University of Michigan, Ann Arbor, 1968.

DISCUSSION

Robert P. Althauser, Princeton University

These three rather different papers run the gamut from "theoretical" to "empirical" methodology: from Land with a different ending for Heise's [2] theoretical piece on reliability and stability to Nestel on the practical realities of a "real" panel study. For me, at least, it is the mixture of theory and simulated, if not real data, that makes the third paper by Pelz and Faith the most interesting of the three.

As most of my attention will be given to this third paper, let me begin with some brief comments on the other two papers.

Nestel's paper inadvertently points up the increasing need for a joint consideration of theoretical and practical matters in the methodology of panel studies. He suggests at one point that, in effect, variation in measurement error should decrease over time in successive remeasures (i.e., respondents would make their answers more consistent with their "true" feelings). Yet, in other, more theoretical papers by Heise [2] and Wiley and Wiley [8], we find that a constant error variance is assumed in the solution of equations which lead to estimates of reliability and stability coefficients.

Consider a second example of the same thing. One of the three additional assumptions which Land proposes as a way of finessing an identification problem which thwarts Heise late in his paper [2] is one of a Markov process among errors (the other assumptions being equal stability coefficients and a Markov process among "true" scores). Yet Charles Werts (of E.T.S.) has found that in certain simulations, this first assumption is most unsound. While realism in theoretical assumptions need not be our only criteria for their use, these two examples may suggest that we ought to think more about the empirical realism of our theoretical assumptions.

One other comment -- about Land's test for the equality of stability coefficients. It would seem at first more straightforward to statistically compare the different stability coefficients as calculated by Heise's method [2]. In contrast, Land's approach which regresses later on earlier measures must assume perfect measurement. If that assumption is, in fact, violated, his test would be quite misleading. It confounds less than perfect epistemic path coefficients between measures and true scores with path coefficients of relationships between true scores.

Yet, in a way, Heise is in the same boat, similarly making other assumptions (e.g., constant reliability of successive measures) to permit an estimate of (possibly differing) stability coefficients. The trick, as Wiley and Wiley [8] illustrate, is to make the weakest possible assumptions. As between equal reliability coefficients with imperfect measures (Heise) and perfect measures (Land), the former appears to be the weaker, hence preferable assumption.

Pelz and Faith

Whatever the ultimate practicality of this line of continuing investigation into the inference structure of cross-lagged correlations by Pelz and his associates (and there are some critics: see Heise [3]; Rozelle and Campbell [7] and Duncan [1]), it has certainly resulted in some intriguing work. Though what follows is primarily a response to their paper above, scattered references will be made to Pelz and Lew [6] and their first Interim Report [5]. (Note that their paper above constitutes the second Interim Report of their "Causal Analysis Project.")

There are two fundamental question for a critic of this line of work to consider. Both require us to keep in mind the simple distinction between the problem of recovering a causal model, or at least an original X - Y relationship built into simulated data, and the problem of making similar inferences from "real" data. About the second question -- whether the work so far which has been devoted to recovering already known relationships will have significant bearing on "real" inference -- more later. The prior question is how feasible does "recovery" of known relationships now appear to be, on balance?

I see three serious obstacles to the general applicability of this approach to recovery:

1) When Pelz and Lew [6] considered the effects of individual differences (which result in "long-term stability" -- autocorrelations persisting over time), they found that cross-correlations will lead to incorrect inferences about the direction of the causal relationship between X and Y. The question that remains is how large the relative variance of individual differences tends to be, for various types of data.

2) A second difficulty noted even in Pelz and Andrews [4] arises when the interval between measures (k) does not coincide with the causal interval (g). Working with X's and Y's with high to modest stability (.70 to .90), it is apparent in their paper above that when the difference between k and g is great, one may not be able to correctly recover the direction of causation. Thus, for p_{xx} and p_{yy} both equal to .70, we see that the cross-lagged differential may disappear

at k $\pm$ 15. Practically speaking (and depending on the type of data), the problem does not look too serious so far.

But suppose that the stability coefficients were even more modest, approaching .5 or .4. What would the correlograms' curves look like then? Let us assume the same low value for the causal relationship p_{yx} = .10 as in the figures above.

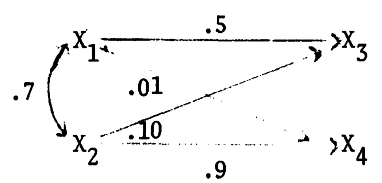
It is evident from equations 13 a - 13 d' in their technical appendix (which may not be reproduced above) that, first, the curves will retain the bell-shaped symmetry that the curves for $p_{xx} = p_{yy} = .70$ have. The cross-correlations will not persist over time, in other words. Second, the size of the maximum cross-correlations will drop, as they did previously when $p_{xx} = p_{yy}$ when down from .95 to .70.

But most importantly, the rate of increase in the cross-lagged correlations up to the interval g and the rate of decrease soon thereafter will greatly increase: the bell-shaped curve will now approach the shape of a needle.

Accordingly, it becomes much more crucial that our measurement interval k closely coincide with the true causal interval; otherwise, we cannot recover the (known) direction of causation from the cross-lagged correlations.

3) The third problem arises when we carry one step in the paper above a little further. The authors note that "with high stability... even if curves were moved leftward so that g = 0, the asymmetry would persist, and so would the cross-lagged differential." Consequently, "...the causal influence of X on Y remained apparent even when the causation was almost simultaneous." If the curves were moved still further leftward, so that g was less than zero, might not the cross-lagged differential still persist, leading us to the wrong causal inference?

The second question remains -- what sorts of connections are there going to be between recovery of a simulated X - Y relationship and causal inference from real data? That the authors are aware of this distinction is clear in their first Interim Report [5]. An example of the importance of the distinction has developed in a post-convention letter to me from Pelz, who wished to respond to my original discussion on this point. I borrowed Heise's argument [3] against cross-lagged correlations, using the following hypothetical example, with $r_{14} = p_{41} + p_{42}r_{12} = .01 + (.7)(.9) = .64$, and



$r_{23} = p_{32} + p_{31}r_{12} = .1 + (.7)(.5) = .45$. The effect of X_2 on X_3 is larger than that of X_1 on X_4 (.10 vs. .01), but the cross-lagged correlations would suggest the opposite. The result follows from the rather different stability coefficients shown (.5 vs. .9).

Pelz's response to this argument is quite interesting. He reports that the degree of stability in either X or Y does not reverse the cross-lagged differential in his simulations. To lead the way into his next point, further, he makes a trivial alteration in my example, assuming one-way causation by completely eliminating the relation between X_1 and X_4. This is necessary according to Faith for my example to correspond to a 'stable' time series.

Pelz then proposes a not-so-trivial alteration in the correlation between X_1 and X_2. This is required, following equation 14 in their technical appendix, so that stationary assumptions can be made (i.e., all autocorrelations and cross-correlations are independent of time, and the total correlations of the X's with themselves are unity). Adopting their equation 14 to my example, we have

$$r_{12} = \frac{p_{32}\ p_{42}}{1 - p_{31}\ p_{42}} = \frac{(.1)(.9)}{1 - (.5)(.9)} = .164$$

Thus, to have a stationary time series, given the path coefficients arbitrarily chosen in my example above, r_{12} must equal .164. If it does, the result is that "the cross-lagged differential is now consistent with the difference in causal coefficients." And so it is.

It would appear, then, that the nature of a stable time series precludes arbitrary examples like mine. Does their response answer Heise's [3] objections? If there was a stable time series hidden in some "real" data, his criticism would not apply. Cross-lagged differentials would be useful tools of causal inference.

But in real data, how do we know if there is a stable time series or not? Unless we know from past experience that certain types of data tend to embody stable time series, we are still faced with the problem Heise has raised. The estimates of paths or the interpretation of correlations still depend on the models postulated, as Duncan has vividly shown [1]. The problem is one of model testing, and cross-lagged differentials do not appear to be as trustworthy a tool of causal inference with "real" (as opposed to simulated) data, as is, for example, the use of instrumental variables.

In short, when Pelz and his associates later attack the basic problem of causal inference with real data, I suspect that their previous work may not be quite as useful as hoped. Yet there are enough surprises of methodological interest in their present work to give any skeptic pause!

REFERENCES

[1] Duncan, Otis Dudley, "Some Linear Models for Two-Wave, Two-Variable Panel Analysis," Psychological Bulletin, 72 (September 1969), 177-182.

[2] Heise, David, "Separating Reliability and Stability in Test-retest Correlation," American Sociological Review, 34 (February 1969), 93-101.

[3] Heise, David,"Causal Inference from Panel Data," in E.F. Borgatta and G.W. Bohrnstedt (eds), Sociological Methodology 1970, San Francisco: Jossey-Bass, 1970.

[4] Pelz, D.C., and Andrews, F.M., "Detecting Causal Priorities in Panel Study Data," American Sociological Review, 29 (December 1964), 836-848.

[5] Pelz, D.C., Magliveras, S., and Lew, R.A., "Correlational Properties of Simulated Panel Data with Causal Connections between Two Variances," Interim Report No. 1, Causal Analysis Project, Ann Arbor: Survey Research Center, University of Michigan, 1968.

[6] Pelz, D.C. and Lew, R.A.,"Heise's Causal Model Applied," in E.F. Borgatta and G.W. Bohrnstedt (eds), Sociological Methodology 1970, San Francisco: Jossey-Bass, 1970.

[7] Rozelle, Richard M. and Campbell, Donald T., "More Plausible Rival Hypotheses in the Cross-Lagged Panel Correlation Technique," Psychological Bulletin, 71 (January, 1969), 74-80.

[8] Wiley, David E. and Wiley, James A., "The Estimation of Measurement Error in Panel Data," American Sociological Review, 35 (February 1970), 112-117.

DISCUSSION

Donald F. Morrison, University of Pennsylvania

Dr. Nestel has touched on the problem of the loss of subjects from the original study cohort or panel. The subjects with incomplete data vectors can still be used to estimate population means and covariance matrices: maximum likelihood estimates based on the multivariate normal model have been given and discussed by a number of investigators (see, in particular, Anderson [1]). When all lost subjects are never interviewed again, the data matrix with its blocks of rows ordered from those subjects with complete data to those who dropped from the study at the earliest time is called monotonic, and the estimates of the parameters can be obtained explicitly. However, when the number of complete subjects is very small and the correlation among the responses is low some recent results [8] have indicated that the multivariate maximum likelihood estimates are less efficient than those obtained from the complete data alone.

Under the same assumptions of the multivariate normal model and a monotone data matrix Bhargava [2] developed generalized likelihood ratio tests for hypotheses on the mean vector, covariance matrix, and the multivariate general linear model. Percentage points of the mean vector test statistics have been computed by Mr. Dinesh Bhoj for a forthcoming doctoral dissertation [3].

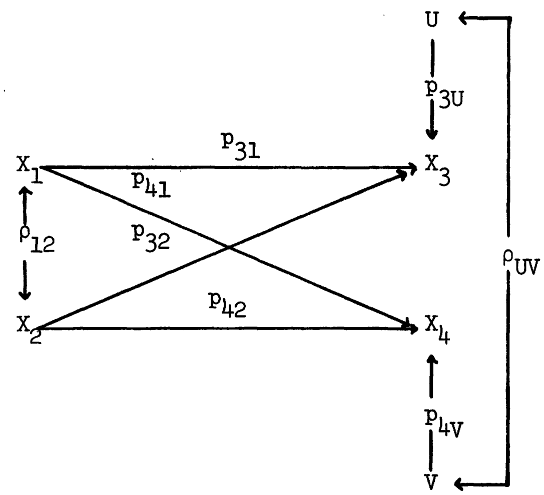
I would be curious to know what use Dr. Nestel plans to make of information on such exogenous factors as changes in the economy, labor supply, employment practices, the inauguration of training programs, and other environmental conditions that will influence his subjects' experiences in the labor market. Presumably some of those quantities would have to be incorporated as concomitant variables in the analysis of changes and trends in his data.

Now let us turn to the correlation models of Drs. Pelz, Faith, and Land. Their investigations suggest to me the following questions:

(1) Are their analyses of longitudinal studies solely concerned with dependence structure? As in the cases of growth functions and repeated measurements experiments, would not the more important conclusions be derived from the mean, or expectation vector, structure?

(2) What are the motivations and consequences of their models for correlation structure? Are their purposes the more parsimonious explanation of the dependence structure, the estimation of components of variance attributable to true scores and measurement error, or the development of hypothesis tests on the means or covariance structure which take advantage of the special properties of patterned covariance matrices?

It is my opinion that more results on sampling properties and hypothesis tests must be developed for the path coefficient models before they can be applied with confidence in the analysis of data. One of the simplest path coefficient models is amenable to statistical inferences about its coefficients: it is the four-variable model discussed by Heise [5] with the following path diagram:



If we assume that the X_i have a covariance matrix with general element σ_{ij}, the path coefficients are equal to

$$\underset{\sim}{B} = \begin{bmatrix} p_{31} & p_{41} \\ p_{32} & p_{42} \end{bmatrix}$$

$$= \frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}\begin{bmatrix} \sigma_{22}\sigma_{13}-\sigma_{12}\sigma_{23} & \sigma_{22}\sigma_{14}-\sigma_{12}\sigma_{24} \\ \sigma_{11}\sigma_{23}-\sigma_{12}\sigma_{13} & \sigma_{11}\sigma_{24}-\sigma_{12}\sigma_{14} \end{bmatrix},$$

or the matrix of regression coefficients of X_3 and X_4 on the pair of variates X_1 and X_2. When the X_i have the quadrivariate normal distribution we can test all hypotheses of the kind

$$H_o: \underset{\sim}{c}'\underset{\sim}{B}\underset{\sim}{a} = 0,$$

for all non-null two-component vectors $\underset{\sim}{a}$, $\underset{\sim}{c}$, by a variant of the Scheffe'-Roy simultaneous confidence intervals for all bilinear forms of the regression coefficients [7, Sec. 3.6]. Such hypotheses would include

$$H_o: p_{31} = p_{42}, \quad H_o: p_{41} = 0, \quad H_o: p_{32} = 0,$$

and other hypotheses of equality or significance of the coefficients. This approach errs, of course, on the conservative side, but conversely it provides an exact control over the Type I error probability for tests of all hypotheses generated by the bilinear compounds.

Most path models do not possess the simple one-to-one connection with regression coefficients of our illustration, and this appears to be the case for the Pelz-Faith bivariate model. However, estimation and testing for the general model might be handled by an approach due to Jöreskog [6].

Because the Pelz and Faith investigation has been principally concerned with the case of high autocorrelation, some simplification could have been achieved by beginning with the autoregressive process with unit coefficient. Then the tth realization is equal to the sum of t independent variates with common variance σ^2. The covariance matrix has the Wiener pattern

$$\sigma^2 \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2 \\ \vdots & \vdots & & \vdots \\ 1 & 2 & \cdots & p \end{bmatrix} .$$

The y_t process can be developed by the same additive model used by the authors, and its cross correlations can be computed in a less involved manner. The cross-correlation function is equal to zero for certain values of the x and y variates' indices. We note, of course, that the processes are no longer stationary.

Let us begin our discussion of Dr. Land's paper with the hypothesis

$$H_o: p_{32} = p_{21}$$

of constant path coefficients from x_1 to x_2 and x_2 to x_3. We shall restrict our attention to the case of no measurement error ($p_{xe} = 0$) and unit random shock coefficients ($p_{xu_2} = p_{xu_3} = 1$). An alternative test of H_o can be constructed by the generalized likelihood ratio principle. Following the author's model, we assume that the random shocks u_{t1}, u_{t2} are independently and normally distributed with zero means and common variance σ^2 for $t = 1, \ldots, N$. The likelihoods under H_o and the general alternative of unequal path coefficients can be obtained as products of the univariate normal likelihoods of the u_{tj} variates. The likelihood ratio statistic λ is merely a power of the estimates of the common variance under the null and alternative hypothesis, or

$$\lambda^{2/N} = \frac{\hat{\sigma}^2_\Omega}{\hat{\sigma}^2_\omega}$$

$$= 1 + \frac{\dfrac{(\Sigma x_{t1}x_{t2} + \Sigma x_{t2}x_{t3})^2}{\Sigma x^2_{t1} + \Sigma x^2_{t2}} - \dfrac{(\Sigma x_{t1}x_{t2})^2}{\Sigma x^2_{t1}} - \dfrac{(\Sigma x_{t2}x_{t3})^2}{\Sigma x^2_{t2}}}{\Sigma x^2_{t2} + \Sigma x^2_{t3} - \dfrac{(\Sigma x_{t1}x_{t2} + \Sigma x_{t2}x_{t3})^2}{\Sigma x^2_{t1} + \Sigma x^2_{t2}}}$$

Under H_o $- N \ln (\hat{\sigma}^2_\Omega/\hat{\sigma}^2_\omega)$ is asymptotically distributed as a central chi-squared variate with one degree of freedom. Modification of the statistic for unknown, non-zero means of the x variates is evident. The assumption of general random shock coefficients leads to unequal variances in the univariate likelihoods and a cubic equation for the maximum likelihood estimate of the path coefficient under H_o.

It is unclear to me how the equations (4.3) could be linearized by a logarithmic transformation. An alternative approach would consist of writing the covariance matrix of the measurement variates as

$$\underset{\sim}{\Sigma} = \alpha^2 \underset{\sim}{P}_1 + \beta^2 \underset{\sim}{P}_2 ,$$

where $\underset{\sim}{P}_1$, $\underset{\sim}{P}_2$ are the respective Markov covariance matrices of the true scores and measurement errors. When $\underset{\sim}{P}_2$ is the identity matrix an estimation technique due to Jöreskog [6] could be employed. More general forms of $\underset{\sim}{P}_2$ would introduce identification problems. If these can be resolved by appropriate constraints on the parameters the estimates might be obtained through an iterative solution of the almost certainly nonlinear maximum likelihood equations. Finally, we note that the equal-correlation matrix might be chosen as an alternative to $\underset{\sim}{P}_2$: such a matrix might be justified from a variance-components model for the observation variates.

In connection with Dr. Land's models and results we note that a family of tests for the degree of dependence of ordered multinormal variates has been given by Gabriel [4].

References

[1] Anderson, T. W., "Maximum Likelihood Estimates for a Multivariate Normal Distribution when some Observations are Missing," Journal of the American Statistical Association, 52 (June 1957), 200-203.

[2] Bhargava, R., "Multivariate Tests of Hypotheses with Incomplete Data," Technical Report No. 3, Applied Mathematics and Statistics Laboratories, Stanford University, 1962 (mimeographed).

[3] Bhoj, D., Personal communication.

[4] Gabriel, K. R., "Ante-Dependence Analysis of an Ordered Set of Variables," Annals of Mathematical Statistics, Vol. 33 (March 1962), 201-212.

[5] Heise, D. R., "Causal Inference from Panel Data," in E. F. Borgatta and G. W. Bohrnstedt, eds., Sociological Methodology: 1970, San Francisco: Jossey-Bass, 1970.

[6] Jöreskog, K. G., "A General Method for Analysis of Covariance Matrices," Biometrika, 57 (August 1970), 239-251.

[7] Morrison, D. F., Multivariate Statistical Methods, New York: McGraw Hill Book Company, 1967.

[8] Morrison, D. F., "Expectations and Variances of Maximum Likelihood Estimates of the Multivariate Normal Distribution Parameters with Missing Data." To be published in the Journal of the American Statistical Association, September 1971.

DISCUSSION

Bruce Hoadley, Bell Laboratories, Holmdel

Discussion of Paper by Bruce Hill

I will focus my discussion on the contingency table example because I believe it has the greatest practical value. In particular, I want to compare Hill's model with the superpopulation model.

Let R_j $j = 1,\ldots,K$ be the number of population units in cell j. So $\sum_j R_j = N$ which I assume is known. The vector $\underline{R} = (R_1,\ldots,R_K)$ can be regarded as the unknown parameter in the problem. The data consists of $\underline{y} = (y_1,\ldots,y_K)$, where y_j is the number of sample units in cell j. The sampling distribution of $\underline{y}$ is multivariate hypergeometric, so the Bayesian's problem boils down to the specification of a prior distribution $Pr\{\underline{R}\}$ for $\underline{R}$.

In the superpopulation model, the finite population is regarded as a sample from an infinite superpopulation for which the proportion of units in cell j is Q_j, and $\underline{Q}$ has some prior distribution. In this case, $Pr\{\underline{R}\}$ is a compound multinomial distribution. Furthermore, if $\underline{Q}$ has a Dirichlet distribution, then the posterior distribution of $\underline{R} - \underline{y}$ is compound multinomial and the problem is solved.

The question is: How does Hill's formulation differ from the above? Hill lets M be the number of cells with positive R_j, $\underline{X}$ be the vector of coordinate of $\underline{R}$ which have positive entries (e.g. if $\underline{R} = (0,5,4,0,2)$, then $\underline{X} = (2,3,5)$), and $\underline{L} = (R_{X_{(1)}},\ldots,R_{X_{(M)}})$. There is a one-to-one correspondence between $\underline{R}$ and $(M,\underline{X},\underline{L})$, so a distribution on $\underline{R}$ can be defined by a distribution on $(M,\underline{X},\underline{L})$. Hill does this by specifying $Pr\{M\}$, $P\{\underline{X}|M\}$, and $Pr\{\underline{L}|M,\underline{X}\} = Pr\{\underline{L}|M\}$, where $\underline{L}$ is exchangeable and independent of $\underline{X}$.

Hill pays particular attention to the case where $Pr\{\underline{X}|M\}$ is a uniform distribution. When this is true, $\underline{R}$ is exchangeable, so that the ER_j's are all equal. So in a sense, Hill's model is less general than the superpopulation model where the ER_j's can be different.

In the case where $Pr\{\underline{L}|M\}$ is also a uniform distribution, Hill works out useful expressions for the posterior distribution of $\underline{R}$. If in addition, it was assumed that

$$Pr\{M\} = \binom{K}{M}\binom{N-1}{M-1} \Big/ \binom{N+K-1}{K-1},$$ then $Pr\{\underline{R}\}$ would be uniform, which is a special case of the superpopulation model. This points out a sense in which Hill's formulation is more general; namely, for the superpopulation model, $Pr\{M\}$ has a very special form, whereas in Hill's model it is completely arbitrary.

Two questions that are unresolved in my mind are: (i) What is the practical value of being able to specify $Pr\{M\}$ arbitrarily? (ii) What are some other choices (besides uniform for $Pr\{\underline{X}|M\}$ and $Pr\{\underline{L}|M\}$) that lead to useable expressions for the posterior distribution of $\underline{R}$?

Discussion of Paper by M. R. Novick and P. H. Jackson

Useful insight into the Bayesian method for recovery of collateral information can be obtained by plotting the Bayes and least squares estimates of β_1 for the various colleges. By doing this we see that the Bayes estimates amount to a smoothing of the least squares estimates. The remarkable thing is that the smoothing is more pronounced for the 25% sample than for the 100% sample. One wonders whether this is a property of the method, or just a coincidence. Presumably the Bayesian method is a device for dividing the total between college variation in the least squares estimates of the $\underset{\sim}{\beta}$'s into a component due to estimation error and a component due to the between college variation in the $\underset{\sim}{\beta}$'s themselves. If so, then I would not expect that the second component would necessarily be underestimated just because the sample size is small.

Another insight from the plot is that the Bayes estimates of β_1 do not always track the least squares estimates. The reason for this is that the components of $\underset{\sim}{\beta}$ are being smoothed jointly.

The paper emphasizes that Bayes estimates are much better than the independent least squares estimates in the 25% sample case. The plot suggests that the Bayes estimates would not be much better than the pooled least squares estimates in this same case. Perhaps the Bayesian model should be compared with several alternative classical models.

(The papers by Hill and Novick, discussed above by Mr. Hoadley, were not sent for inclusion in this Proceedings volume.)

COSTS, RESPONSE RATES, AND OTHER ASPECTS OF DATA COLLECTION IN THE 1970 CENSUS

Robert B. Voight, Bureau of the Census

Introduction.—The Nineteenth Decennial Census of the United States, conducted as of April 1, 1970, was the first in which a mail-out/mail-back technique was employed to collect the information from the majority of the nation's population. Almost two-thirds of the households located in the larger metropolitan areas and some adjacent counties received their census questionnaires by mail containing a request to complete them and mail them back to a local census office in a postage-free envelope provided in the mailing piece. For the balance of the country, the traditional house-to-house canvass by enumerators was used, prefaced as in 1960 by the distribution to all households shortly before Census Day of a questionnaire containing the 100-percent population and housing items and a request that the householder fill in the answers and hold it for the enumerator's visit.

The extensive use of the mails was the culmination of a series of pretests which actually began with an initial experiment in the 1950 census and the experience with self-enumeration in the 1960 census. The 1970 mail collection technique was tested and developed through a series of six pretests conducted during the period 1961 through 1967 and given its final operational rehearsal in three dress rehearsals during 1968 representing the three basic data collection systems used in the 1970 census.

It is obvious that the feasibility of the mail system depends in large measure on the degree of public cooperation in filling out the questionnaires. In the five full-scale mail tests--Louisville (1964); Cleveland (1965); New Haven (1967); Madison (1968); and Trenton (1968), the percent of households who mailed back their census forms was, respectively, 88, 80, 78, 91, and 66. These rates cover both 100 percent (short) and sample (long) form returns which did not differ substantially. On the average the return rate from sample households was about six percentage points below the short form response.1/

The Collection Systems.--For the mail-out/mail-back areas, some 35 million individual address labels were printed from computer tapes containing city-type residential addresses, up-dated at three different points in time by the Post Office through a "casing" check. These were supplemented by some six million addresses in the non-city delivery portions of the metropolitan areas collected in a special address listing operation by census employees a few months before the census. The operation yielded hand-written address registers for control purposes similar to the computer address register print-outs and hand-addressed mailing pieces comparable to the computer produced materials.

Two variants of the mail-out/mail-back system were used in 1970. Under one approach--designated as "decentralized"--the mail returns for the particular enumeration district were given to the enumerator to be checked in against the address register, reviewed for acceptability, and as necessary, followed-up by telephone, if possible, or by personal visit. The enumerator was also responsible for visiting all addresses from which a mail return was not received.

Under the other approach--designated as "centralized"--the mail returns were checked in, reviewed for acceptability, and as necessary, followed-up by telephone, if possible, by a clerical staff in the local temporary field office. The only work which the enumerator had to do was the personal visits to (a) nonresponse addresses and (b) the households with unacceptable mail returns which could not be resolved by telephone. The centralized procedure had certain potential quality advantages but constituted a difficult office management problem. Therefore, it was used only in the very large metropolitan areas. Furthermore, the centralized-procedure offices were managed by experienced Bureau personnel rather than by new temporary employees as is generally (and necessarily) the case for the other types of decennial field offices.

The "conventional" enumeration system was used in the balance of the country where it was not considered feasible or economical to operate and control the mail system. In the nonmail areas, a collection process much like the 1960 "single-stage" approach was used. A few days prior to Census Day, the mail men left an unaddressed short form (100 percent) questionnaire at every housing unit identical in content with the one used in the mail areas, and also a FOSDIC document to be used for final processing. The purpose of this advance distribution was to obtain the advantages of self-enumeration for the 100-percent items. At every fifth unit, the 1970 enumerator completed the same long-form (15-percent or five-percent) questionnaire as was used in the mail areas. A check was made in selected States by local letter carriers of the conventional enumerators' listings, a coverage review comparable to the check used in mail areas. When the enumerator completed his field collection work he was instructed to prepare cards containing the address of each housing unit he had visited. These cards were then turned over to the post offices for this check of coverage.

Of the 393 temporary district offices established for the 1970 census, 45 were "centralized" mail-out/mail-back offices in large cities where approximately 7,000,000 housing units and their residents were enumerated. The average workload for these offices was 153,000 housing units. There were 167 "decentralized" mail-out/mail-back offices responsible for covering approximately 36,700,000 housing units with an average workload of about 220,000 per office. The balance of the enumeration was handled in 181 "conventional" offices with a total anticipated workload of 26,400,000 housing units and having an average workload of 146,000.

It is of interest to note that despite the successful completion of all of the preparatory operations prior to Census Day, the actual distribution and collection phase was in substantial jeopardy for several days prior to April 1 because of the possibility of a nationwide strike by letter carriers. To lessen the possible impact of such an event, postal authorities moved the mail date from March 28 to March 23 and 24. While this earlier mailing made for some confusion in the radio and TV publicity it had no appreciable effect on the returns. The walkout which occurred in a number of the major mail distribution centers did delay the delivery of the returns to the census offices in several instances and was no doubt responsible for the failure to receive the mailing pieces in the pre-designated arrangements and sequences in many instances. This had the effect of delaying the follow-up of non-responses to some degree.

Costs.--Although the total anticipated cost of a decennial census is developed well in advance of the actual decennial census period as an overall proposed budget for the entire program, funds are provided annually based upon yearly appropriation requests which are subject to review by the Secretary of Commerce, and the Office of Management and Budget before being acted upon by each Congressional body as a part of the total appropriation for the Department of Commerce. While this procedure presents some limitations to the assurance that the entire program can be carried out as planned, it does provide the possibility of revising the budget in the light of unanticipated events as the work progresses to either increase or decrease the funds required.

In the initial consideration of the 1970 census budget the assumption was made that the costs should not exceed those of the 1960 census updated for cost and workload increases. Later, funds were requested and approved for coverage and data improvements.

The currently estimated cost of the Population and Housing census program is slightly more than $205,000,000. This puts the per capita cost of the 1970 Census of Population and Housing at about $1.00 for the entire process from initial planning and pretesting through the final publication of special subject reports at the close of the census period. The comparable per capita cost for the 1960 census was almost $.60 in 1960 dollars, or $.90 when adjusted to 1970 cost level.2/

The 1960 adjusted costs including workload increases, pay and price increases, and other factors were calculated to be $194,000,000 as compared to the actual cost of $105,200,000 for the 1960 Census. Cost changes resulting from new census methods employed in 1970 for the mail-out/mail-back data collection technique represent a reduction in this adjusted 1960 cost of a little over $15,000,000. This yields a net base census program cost of about $179,000,000 for 1970 to which program improvements including special coverage improvement procedures, detailed place-of-work data, additional tabulations and publications, improved availability of unpublished data, and other items were added in the amount of $26,000,000 for a total estimated cost of $205,000,000.

The great bulk of the census cost, of course, is consumed in the data collection phase which was initiated with the opening of the temporary field offices in January 1970, peaking in April and being essentially completed by the end of June. Almost 65 percent of the total cost is chargeable to data collection including the necessary preparatory field work, the preparation and control of mailing lists, and the printing and distribution of questionnaires, field instructions, and operational and adminstrative forms. Data processing including the manual coding of many of the sample items, tabulations, geographic identification, map preparation, evaluation studies, improvements in the availability of unpublished data through summary computer tapes and microfilm and the provisions of public use samples of census returns, and the series of publications of the census results represent 25 percent of the cost of the program. Mail feasibility tests, content pretests, dress rehearsals, and general planning, administration and capital outlay represent the remaining 10 percent of the overall costs.

Work accomplished to date has been within planned costs essentially despite the fact that tabulations and publications are two to three months behind originally established time schedules. It is anticipated that some of the time loss can be recovered in the next calendar year.

One month ago the census met its primary legal requirement when the final State population totals and the number of representatives to be appportioned to each State in the House of Representatives based on these totals were delivered to the President and announced in the press. These State totals including U. S. military and civilian personnel overseas for purposes of Congressional apportionment and the District of Columbia add to a 1970 U. S. population total of 204,765,770.

The Enumeration Experience.--The logistics of establishing 393 temporary census field offices, providing them with the necessary supplies, equipment, and census materials was accomplished successfully before the census date and the planned complements of temporary personnel were initially recruited as required in most offices. Recruiting was very difficult in a number of the offices located in the cores of the large cities.

With the questionnaires in the hands of the respondents a few days in advance of Census Day, April 1, approximately 90 percent of the mail response was made within three days after Census Day. In the conventional areas, about 6 percent of the enumeration was completed by mid-April. By the end of April nearly 85 percent of the conventional assignments had been completed. The enumeration of the conventional areas was virtually complete by May 30, and the last conventional office was closed on June 23.

In the 1960 census, by the end of the first week of enumeration, about one-third of the population had been enumerated. By the end of the second week, almost three-fourths (73.8 percent) had been counted, and by the end of the sixth week 99.1 percent of the population had been enumerated. All of the census district offices were closed by the end of July 1960.

Despite the rapid response overall in the mail areas, the follow-up for non-response fell behind schedule due to difficulties in recruiting and retaining qualified people in many of the centralized offices in the large cities. Due to these difficulties only 50 percent of the follow-up work was completed at the end of May. By the end of June work in the decentralized offices was virtually completed.

At the end of July, some 14 weeks after non-response follow-up was initiated in the centralized offices, 13 of these 45 offices were still striving to complete the enumeration of more than 600,000 housing units. The last centralized office was closed on September 17.

Response Rates.--The higher the response rate by mail and the higher the quality of response from the respondents the less the need for personal follow-up to collect the census questionnaires, the possibility of less cost in the process, and the closer to a true inventory of the population, the characteristics and its housing at a single point in time. Here the evidence of success is clear. The actual response and the quality of response, as shown in the following table, equaled or exceeded our expectations which were based on our pretest experience.

Type of Questionnaire	% of Mail Response		% of Responses Needing Follow-up	
	Exp.	Act.	Exp.	Act.
Short Form (100%)	84	88	11	14
Long Form (100% plus sample)	76	83	71	56
TOTAL	82	87	22	20

The experience in the 1970 census with the mail procedure leads us to hope that it can be extended beyond city delivery areas in the future. We shall soon initiate discussions with the Post Office to explore ways in which suitable address lists linked to the specific geography can be developed for rural areas.[3]

While there was no overall official tally of the extent to which respondents in conventional areas completed their short form questionnaires they received by mail, the census field staff checked a small sample of the 160 enumerator assignments and found that 63-percent of the households had completed their form, another 16 percent had partially filled it out, and only 21-percent had made no attempt to complete it or had mislaid it. In fact the conventional offices received many telephone calls from people who were disappointed when the enumerator did not appear at their door on April 1st.

Census Mail Extension Test.--As a part of the Bureau's philosophy of testing innovations on a small scale in one census to determine their feasibility for incorporation in future censuses, ten census districts essentially rural in character, were selected from the conventional offices for a feasibility test of using the mail technique in non-city delivery areas. The pairing of these ten districts on like population, housing, and geographic characteristics and the random selection of one pair member to a mail area was done to provide a lower estimate of the sample variance, and allow us to make meaningful comparisons of cost, completeness of coverage, timing, coordination and control, and to some extent, data quality, between mail and non-mail techniques when used in highly similar rural type areas. In addition, the mail extension test is expected to provide insight into the operational and logistical problems which might be present in a program to achieve greater use of the mail-out/mail-back procedures. The 5 districts designated as mail areas contained a total population of 2,000,000 in 1960 while the 1960 population of the 5 control districts totaled some 2,350,000.

The analysis of this test will be completed in the early spring of 1971.

Quality Control and Coverage Improvement Efforts in 1970 Census.--In addition to the coverage improvement afforded by the existence of control lists of addresses, a major quality check in the 1970 census was a special audit of housing units reported as "vacant" by census enumerators. This audit was conducted immediately after the field canvass on a sample basis by the regular interviewers on the Bureau staff who are more experienced in interviewing and who have received more training than the census enumerators. This intensive follow-up audit found a certain percentage of the units that, according to proper application of the census residence rules, should have been classified as occupied and the inhabitants enumerated. The improvement achieved by this check had a measurable impact, adding about 1 million persons or one-half of one percent of the total count, well dispersed over all areas.

Another quality effort was the post-enumeration check by the Post Office in areas in which the mail-back procedure was not used. The addresses listed by the enumerators as they canvassed their areas were transcribed to cards and turned over to the Post Office, where they were sorted down to carrier routes. Each carrier was then asked to review the cards and to prepare a blue card for each address on his route for which he was not given a white card. The blue cards then became the basis for follow-up to make sure the household was added to the census rolls if it had actually not been recorded by the enumerator. Because this procedure is both time consuming and relatively expensive it was used in 16 States where earlier experience indicated this procedure would most likely be useful. This procedure added about .8 percent to the total population count for these 16 States.

Another significant innovation adopted for the 1970 census was the employment of full time community educators to work with organized minority groups well in advance of Census Day. Twenty community educators were appointed in our regional offices to call in person on prominent members of minority groups, including religious leaders, community leaders, elected officials and radio personalities. Particular efforts were made to reach the black community, the Mexican-American, Puerto Rican-Americans, Chinese-speaking people, and American Indians. Strong support and cooperation was received from the Urban League, the Cabinet Committee on Opportunity for the Spanish-speaking, and the Puerto Rican Government, as well as many community organizations, members of the clergy, the press, and the entertainment industry. While the impact of this effort cannot be quantified, the belief is that it helped significantly in reducing the under-representation among minority groups identified in our post-analysis of the 1960 census. We plan to expand this activity in future census efforts.

Other measures introduced or greatly expanded in 1970 to improve the quality of census coverage included: (1) a special check in large cities of the coverage of persons moving from one address to another within the period immediately preceding and following the census date--a class of persons more subject than most to potential omission in the census; (2) special foreign language translations of questionnaires and instruction sheets used in appropriate areas; (3) telephone assistance centers; (4) locally organized personal assistance centers; (5) bilingual enumerators to aid the householders in completing their census reports where needed; and (6) smaller enumerator assignments, smaller supervisory assignments, and higher pay rates in difficult-to-enumerate areas together witn a manifold increase in the deployment of experienced census staff to assist temporary workers in preparing for and executing the work.

The Undercount.--Places of all sizes are much more concerned over the 1970 population counts than previous census, since the decade of the '60s has seen the establishment of 22 new Federal funding programs which base their grants or assistance, in whole or in part, on the population of the area involved. Of equal impact is the Supreme Court "one-man-one-vote" decision effecting all levels of districting for the purpose of voting for elective offices.

Most of the concerns over possible undercounts are based on a serious effort to reconcile the census figures with other data related to local growth. A city official who knows that the number of housing units has increased, school enrollment is higher, water meters are more numerous and automobile registrations greater, is naturally concerned when the census shows a lesser increase in total population or no gain at all. The Bureau too, is concerned since we know the importance of the census results. However, the census is a name-by-name accounting of each dwelling unit or other place of abode for the entire population. It is not an estimate. No process that projects a population count over a 10-year period from the last census by means of indirect indicators can be as accurate as the census count. Rising affluence, changes in family size, changes in living arrangements all operate to lessen the reliability of indirect indicators of population change. Often indirect indicators of activity relate to the broad local area, whereas census figures for a city relate to the area within its corporate limits, which often experiences changes that are at variance with those in its suburban areas.

In spite of the improvements in the 1970 field collection procedures, there have been a large number of expressions of concern by local officials concerning the quality of the census in particular areas. More than 500 localities, large and small, registered formal complaints concerning the preliminary counts announced by the census district managers at the conclusion of the field canvass. As the preliminary counts were announced, a "Were You Counted?" campaign was conducted by the census district office by placing a form in the newspapers which anyone who thought he had not been enumerated could fill out and turn in. These were checked against the records and if not found, this person was enumerated. Country-wide this effort added less than 100,000 to the count. Communities who thought they had not been properly counted were given the opportunity to conduct additional campaigns to determine that everyone was enumerated. The Bureau stipulated that specific names and addresses had to be furnished in order that such claims could be individually checked against the census records. Such claims were accepted up to early November and whenever these indicated that some small area had been omitted such as a block or a small neighborhood, such possibilities were immediately field checked by Bureau personnel. Overall these local efforts produced about 75,000 claims of missed households. Checks of these claims resulted in the addition of about seven persons per 10,000 originally enumerated or about five-one hundredths of one percent to the national total.

Three places have brought suit against the Bureau concerning an alleged undercount of their population. Two of these have been dismissed by the court. The third was scheduled for court action on December 9, 1970.

The Census Evaluation Program.--A total of 23 evaluation studies have been established for the 1970 census. Some have been initiated during the field collection phases. Others must await access to the questionnaires before certain comparisons can be made with independent records. It is anticipated that the results and analysis of most of these studies will be issued in the summer and fall of 1971.

Seven of the studies are concerned with population and housing unit coverage, ten others are concerned with content and sample bias. The content studies include a CPS-Census match, a comparison with IRS income data, an employer record check on occupation and industry entries, and several other subject items. A reinterview of a national sample of 10,000 households was conducted during August and September. The reconciliation of these returns with the census reports for these households and the analysis of differences will take several months.

The analysis and publication of the results of a response and enumerator variability study which was embedded in the field collection through the assignment of specific enumerator workloads in 35 district offices will not be completed until early 1972.

Publication Program.--While similar in general format to the 1960 census publication series, the 1970 census publications, with the exception of the first series containing official counts for all places, will be prepared by a computerized process at the Government Printing Office called LINOTRON wherein the table images carrying letter-press type are displayed on a cathrode ray tube and captured on microfilm for ultimate offset plate preparation with much greater speed and greater data capacity per page than printing systems heretofore available. As a result, a substantially greater number of pages will be published in the 1970 program--200,000 as compared to 130,000 for the 1960 census.

All of the preliminary reports for both population and housing showing counts for States, counties, places, SMSA's and for specified areas have already been issued. Most of the Population and Housing Advance Reports containing final data from the first computer tabulations have been sent to the printer. The last of these will be issued next month.

A computerized news release is being prepared for each county in a State as the first tabulation run is completed, and is mailed to the news outlets in the appropriate cities and counties of that State.

Census Data Accessibility.--More effective utilization of census data products has been a long-term goal of the Census Bureau. Current activities directed toward the achievement of this goal have been profoundly affected by three factors: (1) pervasive need for data; (2) widespread use of general-purpose digital computers; and (3) data processing procedures used by the Census Bureau.

Bureau programs and activities with respect to improving utilization of census products stem from its interpretation of a selected number of key critical concepts. These concepts are concerned with access and confidentiality of data, data use, and data delivery.[4]

In processing the 1970 census results, the Bureau creates two sets of basic record tape files as part of the process of arriving at the tabulations for the published reports, one for population and one for housing data. The Bureau develops a set of detailed tabulations from these basic record tapes and transfers the tabulations to intermediate tape files called summary tapes. One set of the summary tape files will be reproduced on industry-compatible tape in standard format and technical language. After being subjected to a confidentiality review procedure in order to prevent disclosure of information identifying individual respondents summary tape files are made available for public use at a cost per tape including documentation.

The Bureau will make available six sets of these public use summary tape files, referred to as "counts." These tapes contain approximately 10 times as much data as it is possible to publish. The first three counts provide information generated from the 100 percent basic record tape. The last three will contain data developed from the basic record tapes which contain the sample information. The complete set of summary files will consist of some 2,050 tape reels which may be purchased from the Census Bureau at a reproduction cost of about $60 per reel. This price includes the accompanying technical documentation and the cost of mailing and handling.

The First Count provides statistical information for areas as small as the enumeration districts and block groups; the Second Count at the census tract and minor civil division level; the Third Count for some 1,500,000 blocks. The Fourth Count provides the sample data results at the level of minor civil divisions and tracts; the Fifth Count provides information by 3- and 5-digit ZIP code areas, and the Sixth Count gives detailed cross-tabulations for relatively large areas.

In addition to summary tapes, the census data user needs the Geographic Area Code Index, which provides the alphabetic equivalents of the numeric codes appearing on the summary tapes. Also, appropriate census maps should be acquired to be able to properly identify the small areas corresponding to the tabulations on the summary tapes.

The Bureau has prepared a display software package which the user might consider purchasing. This program, called DAULLIST, has been designed to display the contents of the First Count and is available in both FORTRAN and COBOL versions on the same tape reel. The cost of this package, including documentation and tape reel, is $60. Plans call for the preparation of additional display programs for the other summary tape counts.

A microfilm version of the First Count has been developed and is available at an equivalent cost of approximately $15 for each First Count reel of magnetic tape. Microfilm versions of other counts will be developed on the basis of user interest and demand.

The Bureau is making available for public use other products which it has developed for use in the 1970 census. These are maps, Address Coding Guides, and Geographic Base Files.

In order to speed up delivery of special tabulations, and to reduce their costs, a special tab generator for processing special tabulation requests is being developed to be operative late in 1971. A general extraction program is also being designed to reduce the cost of providing data for selected geographic areas from summary tapes; and a control system is being installed which would also enable the user under certain circumstances to purchase a tape reel within a file.

Release of Unpublished Summary Data.--Some 130 census summary tape processing centers throughout the country which have been recognized by the Bureau are providing copies of the census summary tapes to users and performing special tabulations from the tapes on request. The Bureau has received orders for more than 13,000 summary tapes to date from these centers as well as others who prefer to purchase direct from the Bureau.

A Central Users' Service has been organized within the Bureau to take summary tape orders and handle requests for special tabulations from the basic record tapes. The Service also provides consultation assistance on other census products to interested users.

Public Use Sample Tape Files.--In general, publicly available data from a census appears as statistical summaries only. However, samples of basic data from the 1970 census, with all identifying information removed, will be made available for large areas. The 1970 Public Use Sample Program is greatly expanded over that provided in 1960. The intent is to make the samples much more appropriate for analysis of subgroups as defined by geographic or demographic considerations.[5]

The basic elements of the Public Use Sample Program are as follows:

1. There will be several basic Public Use Sample Files, each of which contains sample records for 1 percent of the population, or roughly 2 million individuals.
2. No names or addresses will appear in the file. Geographic codes will identify only areas (e.g., States or SMSA's) or at least 250,000 population.
3. All of the characteristics of people and households as recorded in the census basic records will be on the Public Use Sample records, except for name, address, and geographic information as noted above.
4. Records will be organized on a household-by-household basis so that characteristics of the various family members may be interrelated.
5. The user may obtain a Public Use Sample File drawn either from 15-percent sample records or 5-percent sample records (see informational copy of the census questionnaire).
6. The user will have the option of obtaining a Public Use Sample File including "Characteristics of the Neighborhood," a set of social indicators which allows comparison of the individual's characteristics with information about the kind of neighborhood he lives in.

More than one option of geographic identification will be available. Present plans include three: (1) one identifying large SMSA's, large counties, and related groups of counties elsewhere; (2) another identifying each State and its urban and rural, metropolitan and non-metropolitan parts; and (3) one identifying large States and groups of small States with size-of-place reported and with community characteristics for each household included. Combinations of these three identifiers will not be available on a particular Public Use Sample File.

Public Use Sample Files will be available on IBM magnetic tape, 7- or 9-track, at the cost of reproduction. One 1-percent Public Use Sample File for the United States will likely take up 30 reels of tape. Smaller subfiles or subfiles of Negro headed households, will also be available on a special order basis. These samples are expected to be available near the end of 1971.

Early Results.--More than three fourths of the national growth between 1960 and 1970 occurred in metropolitan areas with suburban rings showing rapid and substantial population growth. Suburbanites now outnumber those living in central cities. Population in the suburban rings now stands at 75 million, with 63 million in the central cities, and 65 million outside the SMSA's

Population in many central cities declined sharply. Declines also occurred in the decade of the 1950's, but the number of cities showing population losses in the 1960's is greater than in the previous decade. Only 12 of the 25 largest central cities show an increase in population over 1960, and three of these are due to substantial annexations.

About one-half of the Nation's 3,000 counties lost population between 1960 and 1970. In an additional one-fourth of the counties, there was a low rate of growth. Approximately two-thirds of the counties showing a decrease in the 1960's also had lost population between 1940 and 1960.

The States with the greatest gains in numbers of people are California with 4.2 million; Florida, 1.8 million; Texas, 1.6 million; and New York, 1.4 million. The greatest percentages gainers are Nevada with 71.3 percent; Florida, 37.1 percent; and Arizona, 36.1 percent. Three States and the District of Columbia lost population during the decade. The States are West Virginia, North Dakota, and South Dakota.

Migration patterns continued much as in the past--from mid-country out, from rural to urban; and the coastal areas continued to grow. Overall, the population shift has been to the south and west. The farm population declined by approximately one-third since 1960, dropping from about 15 million to some 10 million.

FOOTNOTES

1/ Kaplan, David L., "Plans for the 1970 Census of Population and Housing," DEMOGRAPHY, Volume 7, Number 1, Feb. 1970, p.8.

2/ U.S. Bureau of the Census, "1960 Censuses of Population and Housing--Procedural History," Mar. 1966, p.344.

3/ Brown, George H., Statement before Subcommittee on Census and Statistics, Committee on Post Office and Civil Service, U. S. House of Representatives, Sep. 1970; Bureau of Census

4/ Gura, Benjamin, "Census Tape Delivery; Dates, Costs, and Contents," Urban Regional Information Systems Association, Jeffersonville, Indiana, Sep. 1970.

5/ Bureau of the Census, "Small Area Data Notes," Vol. 5, No. 1, Jan. 1970.

POPULATION TRENDS IN METROPOLITAN AREAS

Meyer Zitter, U.S. Bureau of the Census

It has already been noted in the popular press that the overall population counts from the recently completed census reveal few real surprises. The picture of population growth and redistribution of the 1960's emerging from the census parallel more or less population patterns reported by the Census Bureau in past years based on data from the Current Population Survey and by its independent population estimates. Yet the extent and intensity of some of these changes are worthy of additional comment and a number of individual situations rate special mention.

Metropolitanization and suburbanization of our population continued at a rapid rate during the 1960's although much below that of the 1950's. More than two-thirds of the population now live in areas defined as metropolitan. Flight from the cities continued, with the suburban areas growing heavily at the expense of the cities. Now for the first time the population of suburbia substantially exceeds that of its cities and of nonmetropolitan America. Close to 75 million of us are now suburbanites, compared with 62 million in the cities and 64 million in the balance of the country. In 1960 our population was roughly evenly distributed with 1/3, or about 60 million, in the cities, 1/3 in the suburbs, and 1/3 in the balance.

Regional Trends

Growth in metropolitan areas was pervasive; all regions of the country except the Northeast experienced more rapid population gains in metropolitan than in nonmetropolitan areas. In the South, which is at the same time both our most populous region and most rural (in terms of percent living in nonmetropolitan areas), about 55 percent now live in metropolitan areas. This proportion is still far behind the rest of the country but the differential has narrowed significantly over the past several decades. Almost 35 million (34.3) Southerners reside in metropolitan areas, not too far different than the 37 million (36.8) in metropolitan areas of our North Central States. Yet, there are still over 27 million (27.3) in the South living outside of metropolitan areas, a far larger number than in any other region. Thus, a substantial reservoir of potential migrants to metropolitan areas still exists in the South.

In the Northeast, on the other hand, it's the nonmetropolitan areas that have been growing faster, as the urbanization process spills over into the nonmetropolitan territory. Although this pattern of metropolitan-nonmetropolitan growth rate is different than that in the rest of the country, it is by no means surprising. The picture was not too different in the 1950's when the metropolitan-nonmetropolitan growth rates in the Northeast were about equal, while at the same time growth rates in metropolitan areas were far outpacing those in nonmetropolitan areas in the rest of the country. The extremely high proportion of population already living in metropolitan areas in the Northeast, it was close to 80 percent as far back as 1940, and the high densities of many of the cities are factors contributing to this abnormal differential in metropolitan-nonmetropolitan growth patterns.

Population growth in the fastest growing region of our country--the West--is also primarily metropolitan in nature, with a 27 percent increase in its metropolitan population since 1960. Better than 3/4's of the 35 million persons in the West live in SMSA's. Its metropolitan percentage--77.6--is close to the 81.2 percent of the very populous Northeast.

Metropolitan growth has been far from uniform with metropolitan areas of the West and South growing far more rapidly than those of the North. Population movement has been toward the pleasant-climate and resort type areas of Florida, Arizona, Nevada, and California, making their metropolitan areas amongst the fastest growing areas in the country. But even at that, metropolitan growth rates in this decade are well below those of the preceding decade. One major factor in the lower growth rates for metropolitan areas has been the substantial drop in the birth rates during the 1960's. Overall national population growth was down to 13.3 percent compared with 18.5 percent in the 1950's. There were 2 million less babies born, and 2½ million more deaths in the 1960-70 period than in the 1950-60 period. Another factor, perhaps even more important, is that the census figures imply a significant reduction in population redistribution through net migration between metropolitan and nonmetropolitan areas. The figures are quite striking in this respect. Roughly speaking, in the 1940's and 1950's net out-migration from nonmetropolitan areas amounted to 5½ to 6 million in each decade; in comparison net out-migration from nonmetropolitan areas amounted to only about 2½ million in the 1960's. The rate in the 1960's, about -4.0 percent, was much less than 1/2 that of the 1950's. At the receiving end, metropolitan areas in the 1960's gained about 5½ million through net migration (and net immigration from abroad), a not inconsiderable amount but much less than the 8 million net migration gain of the 1950's.

Regionally, both the South and the West gained large numbers of net migrants in their metropolitan areas, 2 million and 3 million, respectively, respectable amounts but reduced somewhat from the corresponding net gains in the 1950's. The North Central region metropolitan areas gained virtually zero net migrants in this decade, compared with gains of over one million net migrants in both the 1940's and the 1950's. One of the more remarkable trend shifts in net migration is that which occurred in the South where net out-migration for nonmetropolitan areas in the 1960's amounted to only 1½ million; in the two preceding decades the South's nonmetropolitan areas lost over 4 million in each decade. It is this reduction in net out-migration from nonmetropolitan areas that permitted the South to register an overall net migration gain for the first time in many a decade. A significant fact that emerges from these metropolitan-nonmetropolitan shifts is that it appears that net immigration from abroad (which amounted to almost 4 million persons over the decade) played a very major role in metropolitan population growth in the 1960's. Although immigration from abroad also contributed to metropolitan growth in previous decades, out-migration from nonmetropolitan areas was by far the major component of net migration into the metropolitan areas in the previous decades.

The SMSA definition used here is as of 1970. There are almost 7 million (6.7) persons living in areas in 1970 that were not considered part of the metropolitan population in 1960, thus further reinforcing the trend toward metropolitanization. These areas grew at about the same rate as the population living in the suburban ring. About 60% of this group represent "outlying" counties that became part of the existing SMSA's. The balance represent newly-established SMSA's as the core city reached the 50,000+ class.

Pattern by Size

Rates of metropolitan population growth differ significantly by size, with the very largest and the very smallest areas showing the slowest growth. In fact, growth in SMSA's of 2 million or more and those under 1/4 million was only just about large enough to accommodate their own natural increase, suggesting little population gain (or loss) through net migration. The fastest growing appear to be the more moderate size ones, those in the 1 and 2 million class and those between 1/2 to 1 million. The former grew by about 25 percent and the latter by 17 percent during the past decade. Yet, size class is probably a very imperfect measure of place of residence preference and doesn't really explain whether Americans prefer to live in small or large metropolitan areas. The picture varies depending upon whether places are classified by size in 1960 to 1970--that is, at the beginning of the growth period or at the end. Regional variations also exist by size. Furthermore, population growth itself tends to move areas out of one size class to another. For example, SMSA's of 1/2 to 1 million population in 1960 added 5½ million or 22 percent to their population--a very significant growth level--but this had the effect of increasing the population of a number of these SMSA's to well over the million mark, thus moving them into the next higher size category when viewed from the end of the period. If size is an important factor in attracting migrants to areas, this very factor of growth may lessen its attraction to later migrations.

In any event, the census population figures by size suggest that we have been and continue to be a nation of large area dwellers, and if a significant portion of our population truly yearn to return to "small-town America" they have yet to demonstrate this by their choice of residence. 1/ In 1960, when we were a nation of 180 million, about 1/4 of our population lived in SMSA's of 2 million+ and another 1/4 of our population in SMSA's of 1/2 to 2 million (with 10 percent in SMSA's of 1 and 2 million). Now with a total population of 205 million we still find about 1/4 in SMSA's of 2 million+ and 1/4 in 1/2 to 2 million (with 13 percent in those of 1-2 million). In absolute terms, however, we find 52 million persons now living in the large SMSA's (SMSA's of 2 million+) compared with 43 million in this size category in 1960--a gain of 9 million. There are now 26 million living in SMSA's of 1 and 2 million, whereas there was less than 19 million in this size class back in 1960, an increase of 7 million or almost 40 percent. The number of people living in SMSA's of 1/2 to 1 million and 1/4 to 1/2 million increased very slightly in each instance, but we now have less people living in SMSA's of under 250,000 than was the situation 10 years earlier. Thus, in the face of or perhaps because of an increase of 18 million in metropolitan population (and an overall increase of 24 million in total population) less people are living in small SMSA's than before. I guess we need to adjust our definition of small areas to accommodate population growth which causes SMSA's to shift to larger class sizes.

Change Within Metropolitan Areas

Significant population shifts continue to take place between cities and suburbs within metropolitan areas. Although we have been expecting cities to show up poorly in the way of population growth in this past decade, as indicated by a number of special censuses conducted during the decade and from the Current Population Survey, the intensity of the losses in some cities comes as a surprise. Almost half of the 50 largest cities lost population in this last decade, in many cases intensifying losses that had occurred in the 1950's. One of the steepest declines was experienced by St. Louis City with a 19 percent population loss. This is on top of a 13 percent loss in the 1950's. Other large cities that have had population losses of more than 10 percent include Detroit, Cleveland, Buffalo, Cincinnati, Minneapolis, and Pittsburgh. In general, it's the very large cities of the Northeast and North Central regions that show the largest population declines but even moderate and small-size cities in the SMSA's of these regions lacked growth. Cities in metropolitan areas of the South and West tended to show some growth and in some instances, as in California and Arizona, they show substantial growth, but very frequently city growth in these regions is primarily a function of annexations of territory. In fact, although the figures indicate a small overall population gain for all central cities, it is likely a population loss would have resulted if it were not for annexation. What the figures suggest is that the pattern of suburbanization so well known to our larger cities has also caught up with our smaller ones.

These overall census totals give us only a very general picture of what's happening within metropolitan areas. The more interesting and significant data relate to population distribution by race and age. Unfortunately, the census data reflecting on these changes are not yet available for most of the large cities in the country. However, let me discuss first some overall changes for the 1960-1970 decade based on figures from our Current Population Survey and then cite some figures for a few cities for which data have already become available. The survey figures on metropolitan residence are based on 1960 definitions but the picture would be about the same when converted to a 1970 base.

Although there has been very little change in the overall population in the cities between 1960 and 1970, there have been significant population shifts between cities and suburbs. There was a loss of about 2½ million of the white population of central cities of SMSA's. This was more than offset by a population gain of about 3 million Negro and other races. Stated more dramatically, the 2½ million net loss of the white population implies roughly a net out-migration of about 6 million whites. On the other hand, the majority of the 3 million gain of the Negro population in the cities was as a result of an excess of births over deaths but net in-migration also contributed significantly--about 900,000 to a million. In other words, 2/3's of the increase of Negro population of cities results from its own natural increase, and about 1/3 through net in-migration. About 55 percent of all Negroes live in central cities compared to only 25 percent of the white population. Negroes make up 21-22 percent of the population of cities in 1970 compared with 17 percent in 1960.

The white population gained about 16 million in the suburban ring--about a 30 percent increase during the period--while about 1.1 million Negroes were added to the suburbs. Although the rate of growth of Negroes in the suburbs appears to be fairly high, the overall total is still relatively small. About 3½ million Negroes live in metropolitan areas outside of central cities so that now about 15 percent of all Negroes live in suburban areas (it's almost 40 percent for whites). Negroes now represent 5 percent of the suburban population, about the same as in 1960.

Because of sampling error, the survey data do not provide a clear picture on whether, or the extent to which, Negroes have been moving into suburban areas. Whatever the situation, Negro suburbanization is bound to vary significantly from area to area so that we need to wait for more census data to fill in the story on white-nonwhite metropolitan redistribution and on the nature of Negro suburbanization, if any--i.e., a mere spillover from central cities to heavy concentration in the close-by near-in suburb; or a more diffusive pattern throughout in the suburban area. At this point, there is very little evidence of any appreciable movement of Negroes to suburban areas for the few SMSA's for which data are available, but the evidence is clear on the increasing percentage of Negroes in central cities.

The following are illustrative:

City	Percent Negro in Selected Cities: 1960 and 1970	
	1960	1970
Bridgeport	10	16
Hartford	15	28
New Haven	15	26
Gary 1/	25	33
Indianapolis	14	18
Milwaukee	8	15
Wilmington	26	43

1/ Combined total for cities of Gary, Hammond and East Chicago.

To summarize, the main findings of the census totals on population growth are:

(1) Metropolitan growth continues but at rates below those of the previous decade. This reflects a product of both reduced levels of national population increase and less net out-migration from nonmetropolitan areas.

(2) Suburbanization continues with most large cities still losing population. There were significant population shifts, however, with white out-migration from central cities being offset by in-migration of Negroes and other races. Most large cities will show significant increases in percent Negro and other races.

(3) The extent of Negro suburbanization, if any, cannot be fully determined from the data now available (but such data should be forthcoming very shortly). For the areas already on hand there is hardly any evidence of appreciable movement of Negroes into suburban areas, but there is much evidence of substantial increases of Negroes in central cities.

Footnotes

1/ This paper focuses mainly on metropolitan growth trends. There are, undoubtedly, a number of nonmetropolitan areas that have also grown rapidly in the 1960-70 decade but the nature and extent of nonmetropolitan growth has not as yet been fully examined and will be the subject of a separate report. A brief review of preliminary census data suggests to this writer that many nonmetropolitan growth areas are associated with special situations such as sites of universities or military installations.

Table 1.--POPULATION OF THE UNITED STATES AND PERCENT METROPOLITAN: 1950 TO 1970
(In thousands. Metropolitan areas as defined in 1970)

	U. S. Resident Population	Population in Metropolitan Area	Percent in Metropolitan Area
1950................	151,326	92,913	61.4
1960................	179,323	118,415	66.0
1970 (preliminary)...	200,252	136,261	68.0

Table 2.--PERCENT OF POPULATION IN METROPOLITAN AREAS, BY REGION: 1940 TO 1970
(In thousands)

Region	1970 Resident Population (Preliminary)	Percent metropolitan				
		1970 Definition		1960 Definition		
		1970	1960	1960	1950	1940
United States....	200,252	68.0	66.0	63.0	59.0	55.1
Northeast..........	48,417	81.2	81.6	79.0	79.1	78.8
North Central.......	55,956	65.7	63.8	60.1	56.5	52.8
South..............	61,533	55.7	52.0	48.1	41.2	34.5
West...............	34,347	77.6	74.9	71.8	67.1	62.0

Table 3.--POPULATION CHANGE BY REGION, BY METROPOLITAN RESIDENCE: 1950 TO 1970
(In thousands. 1970 data preliminary. SMSA's as defined in 1970)

Region	Metropolitan			Nonmetropolitan		
	Population 1970	Percent Change		Population 1970	Percent Change	
		1960-70	1950-60		1960-70	1950-60
United States, total ..	137,035	+15.2	26.4	63,214	+4.8	7.5
Regions						
Northeastern	39,302	+7.8	13.2	9,111	+11.0	13.1
North Central	36,774	+11.7	23.7	19,186	+2.6	6.9
The South	34,289	+20.0	36.6	27,239	+3.2	3.4
The West	26,669	+26.9	48.6	7,678	+9.2	2.3

Table 4.--POPULATION CHANGE, BY METROPOLITAN RESIDENCE, BY REGION: 1960-70
(In millions. Based on preliminary data. SMSA's as defined in 1970)

Region	Metropolitan		Central City		Ring	
	Amount of Change	Percent Change	Amount of Change	Percent Change	Amount of Change	Percent Change
United States total..	18.0	+15.2	2.8	4.7	15.3	25.6
Northeast...............	2.8	7.8	-0.7	-3.6	3.5	18.2
North Central...........	3.9	11.7	-0.1	-0.7	4.0	24.5
South...................	5.7	20.0	1.9	12.5	3.8	28.8
West....................	5.7	26.9	1.6	16.1	4.1	36.1

Table 5.--ESTIMATES OF NET MIGRATION, BY REGION, BY RESIDENCE
(In millions. Figures are rough approximations based on preliminary data. Metropolitan definition as of end of each decade.)

Region and Residence	Amount of net migration			Net migration as percent of population a/	
	1960-70	1950-60	1940-50	1960-70	1950-60
Metropolitan, total.....	+5.4	+8.1	+7.2	+4.5	+9.2
Northeast..........	+0.2	+0.3	+0.4	+0.5	+1.0
North Central......	+0.1	+1.3	+1.3	+0.3	+5.3
South..............	+2.1	+2.7	+2.4	+7.4	+14.6
West	+3.0	+3.8	+3.2	+14.5	+28.8
Nonmetropolitan, total..	-2.4	-5.5	-5.9	-4.0	-8.7
Northeast..........	+0.2	+0.0	-0.0	+2.0	+0.4
North Central......	-0.9	-1.4	-1.6	-4.6	-7.0
South	-1.5	-4.1	-4.6	-5.6	-14.4
West...............	-0.2	+0.0	+0.3	-2.5	+0.2

a/ As percent of population at beginning of period.

Table 6.--POPULATION IN SMSA'S, CENTRAL CITIES, AND RING BY SIZE OF SMSA: 1970, 1960, AND 1950

(In thousands. Based on preliminary 1970 data. SMSA's defined as of 1970)

Size Category	Population 1970	Population 1960	Population 1950	Percent Change	
				1960-70	1950-60
United States Total........	200,252	179,323	151,326	+11.7	+18.5
SMSA, Total..................	136,261	118,415	92,913	+15.1	+27.4
SMSA's of 2,000,000 or more......	51,656	46,604	37,920	+10.8	+22.9
SMSA's of 1,000,000 to 1,999,999.	26,143	20,872	14,775	+25.3	+41.3
SMSA's of 500,000 to 999,999.....	24,182	20,598	15,837	+17.4	+30.1
SMSA's of 250,000 to 499,999.....	17,967	15,676	12,332	+14.6	+27.1
SMSA's of under 250,000..........	16,313	14,666	12,048	+11.2	+21.7
Central Cities, Total........	62,161	59,396	52,999	+4.7	+12.1
SMSA's of 2,000,000 or more......	22,837	23,567	23,566	-3.1	+0.0
SMSA's of 1,000,000 to 1,999,999.	10,657	9,615	7,863	+10.8	+22.3
SMSA's of 500,000 to 999,999.....	11,569	10,468	8,854	+10.5	+18.2
SMSA's of 250,000 to 499,999.....	8,180	7,548	6,013	+8.4	+25.5
SMSA's of under 250,000..........	8,918	8,199	6,703	+8.8	+22.3
Ring, Total..................	74,100	59,019	39,913	+25.6	+47.9
SMSA's of 2,000,000 or more......	28,819	23,037	14,354	+25.1	+60.5
SMSA's of 1,000,000 to 1,999,999.	15,486	11,258	6,912	+37.6	+62.9
SMSA's of 500,000 to 999,999.....	12,613	10,130	6,983	+24.5	+45.1
SMSA's of 250,000 to 499,999.....	9,786	8,127	6,319	+20.4	+28.6
SMSA's of under 250,000..........	7,395	6,467	5,344	+14.4	+21.0
Non-metropolitan.............	63,991	60,908	58,413	+5.1	+4.3

PERCENT OF METROPOLITAN POPULATION LIVING IN SMSA'S

Size Category	1970	1960	1950
2,000,000+..................	37.9	39.4	40.8
1,000,000+..................	57.1	57.0	56.7
1 to 2 million	19.2	17.6	15.9
½ million to 1 million	17.7	17.4	17.0

Table 7.--DISTRIBUTION OF METROPOLITAN POPULATION, BY SIZE: 1960 AND 1970

(In millions)

	By size in 1970				By size in 1960			
	1970	1960	Change		1970	1960	Change	
			Number	Percent			Number	Percent
2 million+	52.3	47.1	5.2	11.0	48.2	43.4	4.8	11.0
1 to 2 million	26.1	20.9	5.3	25.3	22.0	18.7	3.3	17.6
1/2 to 1 million	26.7	22.7	3.9	17.3	30.5	25.1	5.5	21.8
1/4 to 1/2 million ...	18.2	15.7	2.4	15.3	18.9	16.8	2.1	12.7
Under 1/4 million	14.8	13.4	1.5	11.1	18.4	15.9	2.6	16.2

	1970 population by size in 1970		1960 population by size in 1960		Percent change in population 1960-70
	Number	Percent of U.S. total	Number	Percent of U.S. total	
2 million+.............	52.3	26.1	43.4	24.2	+20.5
1 to 2 million.........	26.1	13.1	18.7	10.4	+39.6
1/2 to 1 million.......	26.7	13.3	25.1	14.0	+6.4
1/4 to 1/2 million.....	18.2	9.1	16.8	9.4	+8.1
Under 1/4 million......	14.8	7.4	15.9	8.9	-6.5

Table 8.--CHANGES IN METROPOLITAN POPULATION, BY COLOR: 1960-70

(In millions. Based on Current Population Survey. SMSA's as defined in 1960)

	White		Negro	
Absolute change 1960-70	**Number**	**Percent**	**Number**	**Percent**
SMSA............................	+13.6	+13.7	+4.0	+32.7
Central City....................	-2.5	-5.2	+2.8	+29.4
Ring............................	+16.1	+30.8	+1.1	+45.6
Percent living in:	**1970**	**1960**	**1970**	**1960**
SMSA............................	63.8	63.1	70.6	64.7
Central City....................	25.3	30.0	54.8	51.5
Ring............................	38.5	33.1	15.8	13.2
Percent of area that is of specified race:				
SMSA............................	86.4	88.3	12.3	10.8
Central City....................	76.9	82.1	21.4	16.8
Ring............................	94.1	94.9	5.0	4.5

HOUSING CHANGES IN METROPOLITAN AREAS--PRELIMINARY FINDINGS

Aaron Josowitz, Bureau of the Census

It's been close to ten years since the results of the 1960 Census of Housing were published. During the 1960's, very little current information has been available about the housing inventory. Users have often had to resort to estimates based on 1960 data which, as the decade progressed, became more "guestimates" than estimates. A limited amount of housing statistics were, of course, provided by the Census Bureau during the intercensal years. For example, data on vacancy rates are provided in a quarterly series for the United States and the four regions. The Bureau has also provided data on second homes, mobile homes, characteristics of newly constructed units and their occupants, absorption rates for newly constructed multiunit structures, and a large body of statistics on construction. These studies, for the most part, are specialized and national in scope, providing very little information about the thousands of local housing markets that comprise the Nation's total housing inventory.

The ten-year waiting period is about to end. The planning of the census, including the extensive consultations held with users to determine subject content, the development and implementation of the field operations, and the preparation of computer programs and procedures to process the results--all of this has been accomplished. The 1970 data are now beginning to flow to the public.

Compared with the first census of housing in 1940, and with the censuses of 1950 and 1960, the 1970 census promises to meet the needs of users to a far greater extent than previously, with a large assist from advanced computer technology. In response to users, the 1970 printed reports will contain more data than were published in 1960 and earlier censuses. Similarly, the amount of unpublished summary data available on magnetic tape and microfilm in 1970 is substantially greater and more comprehensive than the corresponding tabulations of the preceding censuses. Let me cite a few examples. The block statistics program, essentially limited to cities of 50,000 or more inhabitants in 1960, is extended to cover all of the urbanized area around these cities in 1970. Data on the much needed rent-income ratios--the cost of housing relative to income--are published for the first time by census tract. Unpublished cross-tabulations of housing and household characteristics are available on summary tapes for counties and cities, including census tracts.

As of today, we have printed all of the preliminary housing reports, including the United States Summary. These reports present only preliminary counts of housing units. Selected 100-percent characteristics of the housing inventory are provided in the next series, the Advance Housing Reports, several of which have been released. By the end of January, we expect to release all of the Advance Reports. We have also begun to review data showing 1960 to 1970 demographic and housing changes, to be released soon in a joint population-housing series on General Demographic Trends for Metropolitan Areas.

Now what do the early results tell us about the state of our housing?

1. The rate of housing growth for the Nation declined during the sixties.--The preliminary count of the 1970 Census of Housing indicates that the total for the United States reached 68.7 million housing units, an increase of about 10.4 million units, or 18 percent, since 1960. This compares with an increase of 12.2 million, or 26 percent during the fifties, and 8.7 million or 23 percent during the forties.

2. The greatest growth occurred in the suburbs of the Nation's metropolitan areas.--The trend toward suburban living is indicated by the sharp difference between the rate of growth for the central cities of the standard metropolitan statistical areas (SMSA's), and for the suburban areas outside the central cities. Since 1960, the gain in the suburbs of the metropolitan areas was 5.8 million units, or 32 percent, in contrast to the 10 percent gain in the central cities, and the 13 percent gain in the nonmetropolitan areas.

The distribution of housing in 1970 by metropolitan residence is about the same as the corresponding distribution of the population. Two-thirds of the Nation's housing is in metropolitan areas in 1970--same as 1960. In 1970, there were 46.0 million housing units inside SMSA's, and 22.7 million outside SMSA's. For the first time, however, there were more housing units in the suburbs than in the central cities; there were 23.7 million units in the suburban areas and 22.3 million units in the central cities.

The term "suburban" or "suburbs" is used to define the area within the metropolitan area but outside the central city or cities. It is recognized that many large noncentral cities are, thus, included as "suburbs."

3. All States gained housing units.--All 50 States shared in the increase during the 1960-1970 decade, as was true during the fifties. California with its 7 million housing units replaced New York as the State with the greatest number of housing units. Relative increases ranged from 3 percent in West Virginia and Wyoming, to 70 percent in Nevada. Other fast growing States include: Florida, with a 43 percent increase; Arizona, 42 percent; Alaska, 34 percent; and Maryland, 33 percent.

4. The West continues to lead in rate of housing growth.--Although not as marked as the 43-percent increase during the fifties, the West had a 26-percent increase in housing units during the sixties, closely followed by the South with a 22-percent increase. The North Central and North-

east regions each had increases of 13 percent. In terms of numerical growth, the South experienced the largest gain (3.8 million units) during the decade; the Southern States with substantial increases were Florida (770,000) and Texas (665,000).

The final data in the Advance Housing Reports and in the series on General Demographic Trends for Metropolitan Areas for the States reviewed to date, do not necessarily herald national or regional trends. The observations that follow are based on summaries for a limited number of States. These early reports represent generally the less populated States with relatively smaller metropolitan areas.

The data for these reports indicate the following:

1. The rate of housing growth is greater than the rate of population change.--Housing is a stable commodity and as a rule does not readily disappear with the net out-migration of population from any one locality or area. In such instances, housing might remain vacant and on the market for some time, even though demand has fallen off considerably. Secondly, more and more housing is being used as second homes. In addition, household size is declining, reflecting the need for, and use of, more housing than in former years. The young adults and the elderly are forming separate households to a larger extent than in prior decades. In Maine, for example, one-person households comprised 17 percent of all households in 1970, compared with 13 percent in 1960.

Size of household declined in almost all States reviewed to date. In many cases, the decline occurred in owner as well as renter households. In a few States, the average number of persons per occupied unit remained relatively stable for owner households, but declined among renter households. In Delaware, for example, the ratio of "population per occupied unit" for owner households was 3.5 in 1960 and 3.4 in 1970. For renter households, these ratios were 3.2 in 1960 and 2.8 in 1970, reflecting the tendency of the young singles, young marrieds, and the elderly to live in rental dwellings. The 1970 statistics also indicate the continuing tendency of the smaller households to live in the central cities of the metropolitan areas.

2. The rate of homeownership continues to rise in the large majority of the States.--For many States, the increase occurred in both the metropolitan and nonmetropolitan areas; some showed no change or decrease in homeownership in the central cities. Owned homes continue to be more common in the suburban areas of the SMSA's, although in a few cases the higher rates occur outside the metropolitan areas.

A greater proportion of white households own their homes than Negro households. For example, in the Wilmington, Delaware SMSA, the homeownership rate for white households was 71 percent, compared with 50 percent for Negro households; in the Jackson, Mississippi SMSA, the corresponding figures are 74 percent for whites and 49 percent for Negroes. The increase in homeownership is greater, however, among Negro households than among white households. In each of these SMSA's, the proportion of Negro owner occupied units rose from approximately 40 percent in 1960 to the 50 percent level in 1970. The change for white households was from 72 percent in 1960 to 71 percent in 1970 in the Wilmington SMSA, and from 71 to 74 percent in the Jackson SMSA.

3. The proportion of the inventory accounted for by single-family homes is declining, but is still relatively high.--Compared with 1960, there was a decrease in the proportion of one-family houses, in both metropolitan and nonmetropolitan areas. This may be due, in part, to losses from the inventory of a relatively high proportion of one-unit structures[1] and the construction of more multifamily buildings. To illustrate the decline of one-family structures in metropolitan areas, the proportion of single-family homes declined from 83 percent in 1960 to 73 percent in 1970 in the Phoenix, Arizona SMSA; from 82 to 75 percent in the Wilmington, Delaware SMSA; and from 78 to 74 percent in the Des Moines, Iowa SMSA.

As might be expected, there is a higher proportion of one-unit structures outside SMSA's than inside SMSA's. Within SMSA's, the proportion is higher in suburban areas than in central cities. In South Carolina, for example, housing units in one-unit structures comprise 60 percent of all units in the central cities, 83 percent in the suburban areas of the SMSA's, and 87 percent in the nonmetropolitan areas.

4. The percentage of the total inventory accounted for by mobile homes or trailers increased markedly during the past decade.--In the State of Kansas, for example, mobile homes or trailers represented 3.1 percent of the inventory in 1970 and 1.6 percent in 1960, or about a 100-percent increase over 1960. Increasing use of mobile homes occurred both inside and outside metropolitan areas. There are, however, relatively more mobile homes outside SMSA's than inside SMSA's. Further, there is a greater proportion of mobile homes or trailers in suburban areas than in the central cities.

Of the States reviewed so far, Nevada has the highest proportion of mobile homes, 11 percent, followed closely by Alaska, 10 percent, and North Carolina has the highest number, 86,000.

It should be noted that, by definition, mobile homes and trailers are included in the housing inventory only if occupied.

5. Relatively fewer units lack the basic plumbing facilities,[2] reflecting code enforcement as well as inventory losses and gains since 1960.--In several States, the proportion lacking plumbing facilities inside SMSA's declined to about the 2-percent level--in some cases, under 2 percent. There were also significant declines in

the proportion of units lacking plumbing facilities outside SMSA's. For example, in the nonmetropolitan areas of Delaware, this proportion declined from 32 percent to 14 percent; in New Hampshire, the corresponding percentage dropped from 24 to 9.

In 1970, units lacking the basic plumbing facilities were still more prevalent outside SMSA's than inside SMSA's. In South Dakota, for example, 14.8 percent of all units outside SMSA's lack some or all plumbing facilities; the proportion inside SMSA's is 5.6 percent.

6. The percentage of housing units with more than one person per room--often used as an indicator of crowding--declined during the sixties in both metropolitan and nonmetropolitan areas.--All States reviewed to date showed declines in "crowded" units. In many States, housing with more than one-person per room in 1970, was more prevalent in the nonmetropolitan areas, although in a few, "crowding" was greater in the suburban areas.

7. As may be expected, housing costs in terms of property values and rents have gone up considerably in all geographic areas--in the central cities, in the suburbs, and in the nonmetropolitan areas.--In Delaware, for example, the median value of owner-occupied single-family homes in metropolitan areas rose from $13,200 in 1960 to $18,000 in 1970; the corresponding median contract rent was $66 in 1960, compared with $98 in 1970.

Value and rent data, for 1960 and 1970, are expressed in current dollars (the dollar value at the time of the respective censuses). Any comparison must take into account the inflation which occurred over the past decade, as well as the inventory additions and losses since 1960. Demolitions and other losses tend to remove low-priced units from the housing stock while newly built homes tend to concentrate at the upper levels of the price scale. (Counts and characteristics of additions and losses to the housing inventory will be provided in Volume IV of the 1970 Census of Housing, covering the 1960-1970 period, for the United States and 15 large SMSA's.)

8. The 1970 census data on vacancy rates highlight the fact that housing statistics reflect a series of different local markets.--There does not seem to be a consistent pattern. The data show increases in the vacancy rates for some States, decreases for others, and both increases and decreases within the same State. On a national basis, data on vacancy rates in our Current Housing Reports on Housing Vacancies, Series H-111, show that both the homeowner and rental vacancy rates for the United States have declined, compared with 1960. These reports reveal very clearly, however, that any change since 1960 is not necessarily a "straight line" change over the 10-year period. Reflecting many factors affecting the housing market, such as the volume of construction and the availability of funds, the vacancy reports show that homeowner and rental vacancy rates were fairly stable until about the mid-decade, and declined since then to the current levels.

This presentation is only a brief preview of the findings from the 1970 Census of Housing, available at this time for a limited number of States. The release of all of the Advance Reports and the series on General Demographic Trends for Metropolitan Areas will provide an early detailed documentation, based on the 100-percent items, of the changing characteristics of metropolitan and nonmetropolitan housing for each State and for each SMSA and its central cities. The complete 1970 census results, 100-percent data as well as those based on a sample, will be published shortly in the regular census volumes.

1/ 1960 Census of Housing, Volume IV, Part 2-1, Components of Inventory Change, 1957-1959 components.

2/ Toilet and bathing facilities for exclusive use and piped hot water.

COVERAGE OF POPULATION IN THE 1970 CENSUS: PRELIMINARY FINDINGS AND RESEARCH PLANS

Jacob S. Siegel, U. S. Bureau of the Census*

National Population

Introduction[1/]

At various times during the 1960's the Census Bureau or staff members of the Bureau issued various estimates of underenumeration of the population in the 1960 Census of the United States. The latest of these reports, published in 1967, indicated a net underenumeration of 5.7 million, or 3.1 percent of the estimated total population. This estimate was derived by the method of demographic analysis as compared with the use of reinterview surveys or record-check studies. Demographic analysis involves the manipulation of various types of demographic data such as census data, birth, death, immigration, and emigration statistics, etc. for purposes of analysis, estimation, or projection.

This estimate was viewed as a rough one although it seemed to be the best possible at the time. Just after publishing this estimate, new data on the aged population became available from the Social Security Administration's Medicare program. Data on the number of persons 65 and over enrolled for Medicare in 1967 permitted us to derive alternative estimates of the population 60 and over by age, sex, and color, on April 1, 1960. These estimates are believed to be more accurate than those previously used to represent the "true" population 60 and over in 1960. The coverage of the aged population by Medicare is substantially complete and proof of age is required. On the other hand, some persons may have been omitted by Medicare, and the tabulations must be supplemented for groups excluded by law from the program. Furthermore, the adjustment for seven years of population change between 1960 and 1967 is considerable and, hence, the accuracy of the estimates for 1960 depends heavily on the accuracy of the data on deaths and net migration. The use of the Medicare data makes a substantial difference in the overall estimate of underenumeration of the population in 1960. The estimates based on Medicare data suggest a reduction of 457,000 in the net undercount for ages 60 and over in 1960 and a corresponding reduction in the overall amount of net underenumeration, which then drops from 5,702,000 to 5,245,000. The overall rate in 1960 drops from 3.1 percent to 2.8 percent.

These revised estimates of the overall level of net underenumeration in 1960 remain preliminary, however, since a reexamination of the estimates of coverage in 1960, including a review of the basic data and assumptions, is under way and will be concluded only after we have had an opportunity to review this research with outside experts and after the evaluation of the 1970 Census has proceeded to a point where the age-sex-color patterns of net undercounts in the two censuses can be compared. In fact, we recently met with a group of specialists to review the procedures we used to estimate underenumeration in the 1960 Census. Principal areas for our reconsideration are the adjustments for the underregistration of births, the quality of the data on deaths and immigration, and the possibility of employing the Coale method of historical birth reconstruction to estimate the underenumeration of the nonwhite population.

In view of the limitations of the reenumerative and record-checking procedures (i.e., case-by-case matching studies) in establishing the level of underenumeration in 1960 and 1950, we have decided to employ demographic analysis as the principal basis for estimating the level of underenumeration in the 1970 Census. Some case-by-case matching studies will also be conducted and these will be employed in conjunction with the studies using demographic analysis in making the final evaluation. Because of their secondary role in measuring the extent of census coverage, I will describe the matching studies after I have discussed the use of demographic analysis for 1970.

Coverage of total population

Relative net underenumeration.--We now know the 1970 Census count of the total resident population of the United States and, hence, we can begin to develop estimates of the completeness of enumeration of the population in this census. The "true" total population in 1970 may be represented by the sum of the "true" figure in 1960 and the estimated change due to births, deaths, and net migration during the 1960-70 decade. Even assuming that we do not know the "true" population in 1960, a comparison of the estimated population in 1970, based on the 1960 Census count and the change during the decade, with the 1970 Census count, gives an indication of the amount by which net underenumeration in the 1970 Census exceeds or falls below net underenumeration in the 1960 Census, provided that the estimate of population change for the 1960-70 decade is free of error. The Census Bureau's published postcensal estimate of the resident population for April 1, 1970 (Series P-25, No. 445) was 203,657,000 and the census count was 203,185,000. If we accept these figures at face value, the implied net underenumeration in 1970 is 490,000 greater than in 1960. Since the above estimate of the population in 1970 was published, the estimate of intercensal change has been revised downward by 250,000 as a result of a corresponding downward revision of the estimate of net immigration. No changes have been made in the estimates of births and deaths. On this basis, the implied net underenumeration in 1970 is 240,000 greater than in 1960.

*The author wishes to acknowledge the technical assistance of Sylvia D. Quick in the preparation of this paper. Andrea E. Word assisted in the computations.

Inasmuch as the estimate of intercensal change is important for this calculation, I want to make some comments regarding the quality of the data entering into this estimate.

1. The provisional data on births and deaths used for the most recent years of the decade in the present estimate are expected to differ very little from the final estimates of births and deaths for these years. The present estimate of intercensal population change is based on final statistics of deaths through 1967 and final statistics of births through 1968, and provisional data thereafter. This kind of revision has had little effect on the estimates in the past.

2. Registered births for the 1960-70 decade were adjusted for underregistration on the basis of factors obtained by extrapolation of the results of national tests conducted in 1940 and 1950. A new birth registration test conducted in 1969 and 1970 gives results which are quite comparable with those from the two previous tests and lends a measure of support to them. The 1969-70 Test provided estimates of the percent completeness of birth registration in 1964-68, by calendar year and by color. This test indicated that 98.9 percent of all births (99.2 percent white and 97.6 percent nonwhite) were registered in these years taken as a whole. These figures differ insignificantly from the figures used for this period to prepare the postcensal population estimates for 1970 published by the Census Bureau (99.0 percent total, 99.4 percent white, and 97.2 percent nonwhite). If we use the new measures of underregistration for these years to adjust the registered births for the 1960-70 decade, we arrive at an estimate of total births for this period which hardly differs from the previous estimate.

3. The number of deaths as registered for this decade has been accepted, without modification, even though we may reasonably assume that there is some underregistration of deaths. If we assume specifically, for illustrative purposes, that the rate of underregistration of deaths was one quarter as great as that for births during 1960-70 (an extreme assumption), we would increase the estimate of deaths, and reduce the estimate of intercensal change, by only 45,000. No national test of the completeness of death registration has been conducted.

4. As is reported by Irwin and Warren in another session of this conference, a reexamination of the estimate of net civilian immigration for the 1960-70 decade suggests that we ought to reduce the estimate of net civilian immigration implicit in the published population estimate for April 1, 1970 by about 250,000 (from 4,051,000 to 3,801,000).[2] This correction represents an allowance for both alien emigration and net departures of private citizens (i.e., those who have no affiliation with the Federal Government as workers or dependents of workers) to foreign countries, groups previously not allowed for in our estimates. The allowance for overseas movement of former residents may still be too low but we are unable to establish the figure more closely at this time. On the other hand, no allowance has been made for illegal immigration, and there are indications that it was sizeable during the 1960's. Because of the considerable uncertainty regarding the exact amount of migration, we plan to continue our reexamination of these data.

<u>Absolute amount and rate of net underenumeration in 1970.</u>--As I have said, our procedure for estimating the amount and rate of net underenumeration in 1970 depends on the amount or rate of net underenumeration in 1960 and the estimates of population change between 1960 and 1970. We can suggest the possible amount and rate of error in 1970, and particularly the comparative level of the percent error for 1960 and 1970, therefore, by positing various amounts of error in 1960 and various estimates of intercensal change, 1960-70 (or, alternatively, various assumptions regarding the change in coverage between 1960 and 1970).

Under nearly all the combinations of assumptions shown in table 1, the percent of net underenumeration declined between 1960 and 1970. If we accept a 1960 rate of net underenumeration of 2.8 percent (representing our latest published estimate of net underenumeration adjusted to take account of Medicare data for 1967), and an intercensal population change of 24,102,000, or 250,000 less than the published figure of 24,352,000 (corresponding to a coverage decrease of 240,000 between 1960 and 1970), then the 1970 rate is 2.6 percent. If the intercensal population change and net civilian immigration as originally measured for the 1960-70 decade are accepted (24,352,000 and 4,051,000, respectively), this implies that the additional ¼ million net "emigration" is offset by illegal immigration of about the same amount and that there was a coverage decrease of 490,000 between 1960 and 1970. Under these conditions given the rate of underenumeration of 2.8 percent in 1960, we would have a rate of 2.7 percent in 1970. Even if the rate of underenumeration in 1960 corresponded to the results of a composite of demographic analysis and the reinterview studies (2.6 percent)--viewed as a minimum reasonable estimate--the above two assumptions regarding the 1960-70 change in coverage would imply a decline in the rate of net underenumeration to 2.4 or 2.5 percent in 1970. We plan to conduct a thorough study of the components of these estimates, and the Bureau plans to make its official position known sometime in the next year or two.

A historical series of estimates of rates of net underenumeration, principally for white males and white females for 1880-1970, which I have computed partly on the basis of the data developed by Coale and Zelnik,[3] supports the view that the coverage of censuses has been improving, albeit irregularly, and that the 1970 Census had the lowest rate of net underenumeration over this period.

Demographic factors affecting coverage in the 1970 Census.--Although demographic analysis cannot shed much light on the socio-economic causes and correlates of census underenumeration, limited explanatory information is provided by the estimates of net undercount for age, sex, and color groups which we have developed for 1960. These estimates show that certain age-sex-race groups are more difficult to enumerate than others. We should, therefore, "expect" a higher overall rate of net underenumeration in 1970 than in 1960 if a larger proportion of the population falls in the more-difficult-to-enumerate categories in 1970 than in 1960. We have measured the potential effect of the shifts in the size and age-sex-color composition of the population between 1960 and 1970 by applying the rates of net undercounts for each age-sex-color group, as we have estimated them for 1960,4/ to the corresponding estimates of the "true" population in 1970.

The rate of net underenumeration might then be "expected" to rise from 2.8 percent in 1960 to 3.0 percent in 1970. The value for 1970 is purely hypothetical, of course. Since an intermediate "expected" overall rate (2.9) is secured in 1970 when the 1960 rates for all whites and all nonwhites (not by age and sex) are assumed for 1970, it is evident that the rise in the "expected" value between 1960 and 1970 is in part due to the shift in race composition and in part due to the shift in age and sex composition. In fact, however, the "actual" rate of net underenumeration in 1970 corresponding to these figures is 2.7 percent. Since the "actual" rate of net underenumeration apparently declined slightly between 1960 and 1970, it is clear that the 1970 Census succeeded in overcoming the "demographic" tendency for the rate to rise and even achieved some further improvement. (The effects of the geographic redistribution of the population between 1960 and 1970 on the "expected" overall rate of underenumeration are not known.)

Age, sex, and race distribution

In order to measure the extent of the net undercount for age-sex-race groups in 1970, we need to carry our estimates of the "true" population distributed by age, sex, and race in 1960 forward to 1970 on the basis of estimates of the population change for these specific categories between 1960 and 1970. Assuming that we have satisfactory estimates of "true" population by age, sex, and color for 1960, we now become concerned with the accuracy of our estimates of the age-sex-race distribution of "net immigrants" and deaths for the 1960-70 decade and, particularly, with any biases in the reporting of age of decedents. The age distribution of "net migrants" may be affected particularly by the accuracy of the estimates of the characteristics of migrants for which reported data are not available or by the omission of certain migrants from the totals (i.e., aliens entering illegally).

In general, however, the component of deaths is a far more important one than the component of net immigration; net migration is numerically dominant only in the ages of very low mortality. A comparison of age as reported on death certificates and age as reported in the census for persons who died in a 4-month period following the 1960 Census indicated some marked differences and suggested some substantial errors in age reporting.5/ If the percent differences between deaths at each age observed in these four months were assumed to apply to the whole decade 1950-60, and census age of decedents were substituted for death-certificate age, the percents of net undercount in 1960 for nonwhites would be increased by about 2 percentage points for the age group 55-59, nearly 4 percentage points for the age groups 60-64, and about 3 percentage points for the age group 65 and over. At present there are no plans for a more current study designed to evaluate age on death certificates, whether by comparison with age as reported in the census or with age as reported for beneficiaries in social security records.

Research plans.--We have already alluded to some studies we are carrying out to ascertain the "true" population in 1970 and its characteristics in terms of age, sex, and race. Several additional studies will be undertaken; some of these will provide actual estimates of net undercounts and others will provide definite indications of "weak" spots in the census data. The studies employing demographic analysis include the following:

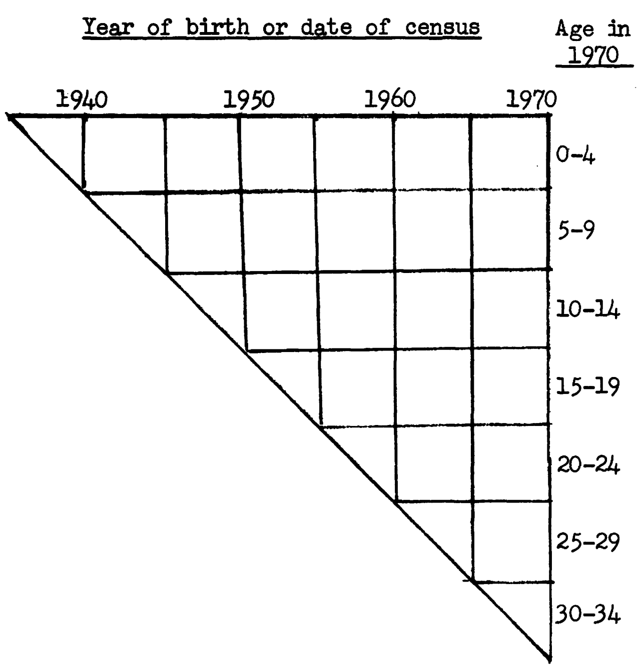
1. We want to reexamine the procedure previously used to estimate expected sex ratios (ratios of males to females) for 1960 and then extend these expected sex ratios, revised as necessary, to 1970. We will reexamine particularly the mortality component employed in estimating the expected sex ratios at the older ages, and seek to determine the effect of using alternative levels of mortality in general and life tables based on population figures corrected for net census undercounts in particular. It should be recognized that the available life tables are based essentially on death statistics and population census data as reported without adjustment for undercounting or misreporting with respect to age, sex, or race. We need to reexamine the survival rates again to note whether these errors have any substantial effect on the expected sex ratios.

2. Second, we will compare estimates of the "true" population 65 and over on April 1, 1970, by age, sex, and color, based on enrollments for Medicare for 1969 and 1970, with the census counts for these categories. Again, the Medicare data will require supplementation for groups excluded from the program. These data can also be used to estimate the population 55 and over by age, sex, and color in 1960. Accordingly, recalling the estimates for 1960 based on Medicare data for 1967, we will have two estimates of the population 60 and over in 1960 and two estimates of the population 70 and over in 1970, which can be

compared. As was suggested earlier, because of the considerable changes which occurred in the aged population over these years, adjustment of the Medicare figures to represent the population at earlier or later dates necessarily introduces some additional error into the estimates. For example, the population cohort 65 and over in 1960 declined by 39 percent by 1967 (when it was aged 72 and over) and the population cohort 55 to 59 in 1960 declined by 18 percent between 1960 and 1970 (when it was aged 65 to 69).

It is possible that Medicare data for 1970 are relatively more complete than the data for 1967, simply because of the historical development of Medicare as a registration system. If this is true, the estimates based on Medicare data for 1970 would tend to state the population more completely than the estimates based on Medicare data for 1967. The difference between the estimates of "true" population 65 and over, by age, sex, and color, for 1967 and 1970 can be compared with the estimates of population change for 1967 to 1970 based on death and migration statistics, in order to evaluate the accuracy of the two bases of estimating change in the aged population.

3. Third, we plan to compare the pattern of errors in the 1960 and 1970 Censuses in order to improve the estimates of the errors in both censuses. Specifically, we will be able initially to compare the pattern of errors for young persons in four censuses covering all cohorts born since 1935, i.e., children under 5 in 1940, children under 15 in 1950, children and youth under 25 in 1960, and persons under 35 in 1970. We may sketch these cohort relationships as follows:



These are the cohorts for which we believe we have quite satisfactory estimates of the number of births (births since 1935), and, hence, for which we believe we can make rather accurate estimates of the population at several successive census dates. We could attempt to apply the Coale iterative procedure (used previously to estimate the "true" nonwhite female population for 1950 aged 15 and over on the basis of the censuses of 1930, 1940, and 1950), to estimate the net undercount rates at ages 35 and over in 1970, assuming some relationship between the patterns of net undercount by age in 1960 and 1970. Note also that we will have estimates of the net undercounts for the population 55 and over by age, sex, and color in 1960 and 65 and over in 1970 from the Medicare data.

4. We plan once again to compare tabulations of Social Security account holders with census counts tabulated by age, sex, and color. In spite of the limitations of this comparison (e.g., incomplete coverage of the population by Social Security, particularly certain age-sex-color groups; incomplete elimination of deaths and duplicate cardholders, etc.), we should be able to detect certain areas of serious omission.

5. Because of the difficulties in establishing the "true" size of the aged population in past censuses, we plan to apply the "method of extinct generations."6/ This method involves reconstructing the aged population for an earlier date by cumulating deaths from the oldest ages and most recent years backward to younger ages and earlier years. The population 65 and over for about 1940, and the population 75 and over for about 1950, can now be estimated by this method since nearly all members of this group have died. The precise accuracy of reporting age for these decedents is not important since a cumulative age group of deaths is used for each year (e.g., 72 and over, 73 and over, etc.). The method would give estimates only for the oldest population groups and with a considerable lag in time.

6. We want to investigate the possibility of employing the Coale method of historical birth reconstruction7/ to derive improved estimates of net undercounts by age and sex for the nonwhite population 25 to 54 or 64 in 1960. In this procedure data from a number of censuses are used to generate several estimates of births for each year or period from, say, 1855 to 1934 (after which date the estimates are based on birth registration data). A "best" estimate of births is derived for each year or period and then carried forward to each later census date by survival rates.

7. We want to examine the effect of assuming an alternative annual trend in the improvement in birth registration from 1935 on, including the effect of incorporating the results of the 1969-70 test.

8. As soon as the final census counts of the age, sex, and race distribution of the population of the United States for April 1, 1970 are

available from the computers--and this should occur early in 1971--we shall be able to make our first very preliminary estimates of the net census undercounts in terms of (abridged) age groups, sex, and race. We may employ for this first analysis estimates of the "true" population for 1970 10 to 64 years of age based on the estimates of the "true" population for 1960 which we published in 1967, estimates of the population 65 and over based on the Medicare data for 1970, and estimates of children under 10 based on births between 1960 and 1970. Since the estimates of the population under 35 years of age in 1970 are based directly on birth, death, and migration statistics, and since the estimates of the population 65 and over are based directly on registration data, we will have a rather firm basis for measuring the undercounts for these groups, covering about two-thirds of the population.

Special Match Studies

Two special studies in the 1970 Census Evaluation Program, the Census-CPS Match Study and the Census-Medicare Match Study, may contribute to our knowledge of the extent of undercoverage of the population in the 1970 Census even though they are not expected to provide acceptable over-all measures of the completeness of coverage.

1. The Census-CPS Match. A sample of households enumerated in the March 1970 Current Population Survey will be matched with the census returns. The study will provide national estimates of missed housing units and of persons missed in those housing units on the basis of a match of all CPS units with the Census, and estimates of missed persons in enumerated housing units on the basis of a sample of about 10,000 occupied CPS units. The matching operation is being conducted now and there will be a reconciliation of differences in the field in February and April, with a determination of the reasons for the differences. The CPS coverage of persons in enumerated units will be assumed to be correct and a two-way reconciliation with the census will not be carried out. Although estimates of the number and proportion of missed persons (as shown by CPS) will be secured in this match study, in view of previous experience with reinterviews and record checks, it is not expected that this study will provide an adequate estimate of underenumeration in the census. The Census-CPS Match will also provide information on the accuracy of reporting of various characteristics including age, sex, and color or race.

2. Census-Medicare Match. A systematic sample of 8,000 individuals from the Medicare records will be matched with the census schedules to measure the coverage of persons 65 and over in the census and the accuracy of census reporting of age by sex and race in this age range. Again, assuming that coverage of the population by Medicare is correct, there will be only a one-way reconciliation between the census and Medicare records. This match study will directly provide estimates of the number and proportion of persons included in the Medicare rolls who were missed in the census; but, with an appropriate assumption, it will also be possible to estimate the total number missed and the "true" number of persons 65 and over.

The 1970 Census reenumerative study ("Content Reinterview Study") will not attempt to determine the completeness of population coverage in the census as did the 1960 Census reenumerative study, but will try only to measure the accuracy of reporting of various characteristics (not including age, sex, color or race).

Regional Estimates of Coverage

We are acutely aware of the interest on the part of many users of census data in estimates of coverage for various geographic units within the United States. In the absence of such estimates for 1960, a number of users have inquired about the propriety of applying the available national estimates by age, sex, and color for 1960 to particular areas in 1960, and some have actually done this. In addition to the fact that the estimates are rough, even as national estimates, the level of the net undercount for a particular age-sex-color group must be assumed to vary widely from area to area and we have no way of measuring this variation satisfactorily.

We know little in formal quantitative terms about geographic variations in coverage. We learned from the Census Evaluation Study of 1960 that the rates of omission of persons in missed housing units were greater for the open country and the very large cities than for suburbs, smaller cities, and towns. The record-check studies of 1960 suggested that, among the four regions, the rate of gross omissions was greatest in the South. Procedural difficulties in taking censuses have been greatest in the inner zones of the very large cities of the North East and North Central Regions, but perhaps the coverage problem is only most visible in these areas. For example, could it be that underenumeration is also relatively great in the urban and rural slums of the South?

I do not know whether it is possible to measure underenumeration for geographic units within the United States satisfactorily. The problem of deriving accurate estimates of the "true" resident population of geographic units at any level may be insurmountable. Use of the conventional component method requires data on internal migration, for which the census itself is the only source. (On the other hand, the national population is a relatively closed population.) We do, however, plan to conduct research to determine the feasibility of measuring underenumeration at the subnational level, particularly for geographic Divisions and possibly States. The measurement problem for areas within States is even more difficult and would involve very different methods. Although we plan to investigate this matter too, it seems very likely that the measurement error here is much greater than the error we are trying to determine.

Two component procedures may be considered for geographic Regions, Divisions, and States. We may try to develop estimates of the expected native resident population of each area under 35 years of age in 1970, on the basis of birth statistics since 1935, life table survival rates, and data on State of birth of the native population from the 1970 Census itself, and compare these estimates with the 1970 Census figures for the resident native population of the area. Or, we may try to develop estimates of the expected native population in 1970 under 35 years of age born in each area on the basis of birth statistics and life table survival rates, and compare these estimates with the population born in the area as indicated by the census. The former comparison is more appropriate to our needs--providing estimates of net underenumeration of the 1970 resident population of each area although limited to ages under 35--but the expected population is based on census data which are themselves subject to underenumeration. Any calculation of this kind must, therefore, allow for underenumeration of migrants, recognizing differential underenumeration by migration status and by State of birth or State of residence. The second comparison will provide indications of geographic variations in coverage but for populations which sometimes bear little relation to the present resident population of States or Divisions. (In 1960 about one-quarter of the population was not living in its State of birth.)

In these experimental calculations we would use as working units States, by color, sex, and 5-year age groups under 35 years of age. Given the limitations of the data and the method, we would hope to secure estimates of underenumeration for the following categories at most: each State, total population under 35; each geographic Division, population by color and 5-year age groups under 35; each Region, population by color, sex, and 5-year age groups under 35. We also want to try to estimate the difference in the rate of coverage for urban and rural areas, for Regions and the United States as a whole. The method in the latter case would not be a component one but would assign part of the underenumeration for each State (age, sex, and color group) to the urban and rural sectors or would involve use of regression analysis. Given the kind of data required and the scope of the calculations, results cannot be expected before early 1973. Whether these will be sufficiently accurate to justify publication remains to be seen.

List of Notes

1. This paper uses as a point of departure the following papers: (1) Jacob S. Siegel, "Completeness of Coverage of the Nonwhite Population in the 1960 Census and Current Estimates, and Some Implications," Social Statistics and the City, David M. Heer, Editor, Report of a Conference Held in Washington, D. C., June 22-23, 1967, Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, 1968; (2) Jacob S. Siegel and Melvin Zelnik, "An Evaluation of Coverage in the 1960 Census of Population by Techniques of Demographic Analysis and by Composite Methods," Proceedings of the Social Statistics Section, 1966, American Statistical Association; and (3) Eli S. Marks and Joseph Waksberg, "Evaluation of Coverage in the 1960 Census of Population Through Case-by-Case Checking," Proceedings of the Social Statistics Section, 1966, American Statistical Association.

2. Richard Irwin and Robert Warren, "American Immigration in the Sixties," paper presented at the annual meeting of the American Statistical Association, Detroit, Michigan, Dec. 29, 1970.

3. Ansley J. Coale, "The Population of the United States in 1950 Classified by Age, Sex, and Color--A Revision of Census Figures," Journal of the American Statistical Association, 50(269):16-54, March 1955; and Ansley J. Coale and Melvin Zelnik, New Estimates of Fertility and Population in the United States, Princeton, N. J., Princeton University Press, 1963.

4. Siegel, op. cit. The rates used for 1960 for ages 60 to 64 years and 65 years and over are based on Medicare data.

5. M. G. Sirken, J. S. Siegel, and R. S. Murphy, "Errors in Postcensal Population Estimates Due to Age Reporting Errors on Death Certificates," paper presented at 1969 Annual Meeting of the Population Association of America, Atlantic City, New Jersey, April 12, 1969.

6. Paul Vincent, "La Mortalité des Vieillards" (The mortality of the aged), Population (Paris), 6(2):181-204, April-June 1951; and Ira Rosenwaike, "On Measuring the Extreme Aged in the Population," Journal of the American Statistical Association, 63(321):29-40, March 1968.

7. Siegel and Zelnik, op. cit., p. 72; and Coale and Zelnik, op. cit., pp. 5-14.

Table 1.--Estimated Amount and Percent of Net Underenumeration in 1970, for Various Amounts and Percents of Net Underenumeration in 1960 and Various Amounts of Change in Coverage between 1960 and 1970

(Numbers in thousands. Base of percent is corresponding estimate of corrected population)

Underenumeration, 1960	1970 net underenumeration according to the 1960-70 change in amount of coverage		
	No change in coverage[1]	Coverage decrease of 240,000[2]	Coverage decrease of 490,000[3]
Amount			
5,702[4]	5,702	5,942	6,192
5,245[5]	5,245	5,485	5,735
4,744[6]	4,744	4,984	5,234
3,328[7]	3,328	3,568	3,818
Percent			
3.1[4]	2.7	2.8	3.0
2.8[5]	2.5	2.6	2.7
2.6[6]	2.3	2.4	2.5
1.8[7]	1.6	1.7	1.8

[1]Assumes a postcensal estimate of 203,185,000 for the resident population on April 1, 1970, implying 23,862,000 population increase and 3,561,000 net civilian immigration for 1960-70 (490,000 less than "published" in _Current Population Reports_, Series P-25, No. 445).

[2]Assumes a postcensal estimate of 203,425,000 for the resident population on April 1, 1970, implying 24,102,000 population increase and 3,801,000 net civilian immigration for 1960-70 (250,000 less than "published" in Series P-25).

[3]Assumes a postcensal estimate of 203,675,000 for April 1, 1970, implying 24,352,000 population increase and 4,051,000 net civilian immigration (as "published" in Series P-25).

[4]Estimate based on demographic analysis. See Siegel, _op. cit._

[5]Estimate based on demographic analysis and Medicare data.

[6]Estimate based on composite of results from demographic analysis and reinterview surveys. See Siegel and Zelnik, _op. cit._

[7]Estimate from reinterview surveys. See Marks and Waksberg, _op. cit._

A SIMULTANEOUS EQUATIONS MODEL OF SPLIT TICKET VOTING IN ISRAELI ELECTIONS[1]

Martin S. Geisel, Carnegie-Mellon University
Naomi E. Kies, Hebrew University
Timothy W. McGuire, Carnegie-Mellon University
Howard Rosenthal, Carnegie-Mellon University

1. Introduction

An election in which citizens vote simultaneously for more than one office is a complex social process where two or more decisions are interrelated. Where more than one office is at stake and where candidates are labeled as to political party voters can vote a "straight" party ticket or they can "split" their ballot, voting for candidates for two or more parties. Politicos have long felt that the two choices were mutually dependent. In this paper we present models of one such election, that of Israel in 1965. Although the choice of the Israeli 1965 data was largely one of circumstance, it was fortunate in that the Israelis voted for only two offices, the national legislature (Knesset) and local municipal councils. We thus avoid the contaminating effects of additional contests. Furthermore, for the city of Jerusalem, much split-ticket voting was suggested by the presence of Teddy Kollek, an attractive candidate at the local level representing the otherwise unsuccessful Rafi party. Our data base consists of voting and census statistics for the sixty-four census tracts of (pre-1967) Jerusalem.

To describe these interrelated phenomena a system of equations is required since a party's municipal vote depends on, among other things, its national vote, and vice versa. The statistical analysis of systems of simultaneous linear equations has received much attention from econometricians. The models with which we shall deal are somewhat more complicated since they involve both non-linearities and inequality constraints on the parameters of the system.

We can introduce our estimation problem through an example involving only a two equation model. Let R_{Ki} denote the proportion of Rafi voters in the i-th district for the Knesset election in 1965; R_{Mi} the corresponding municipal proportion; p's, q's, β's, and γ's constants to be estimated; u's and v's stochastic terms in the model; and let $i = 1,2,\ldots,n$ index the census tracts. Let us first examine a single-equation model originally suggested by Goodman [4] and frequently used in political science applications in the study of two temporally separated elections.[2] For each census tract, assume that the Rafi municipal list maintained a proportion $p_1 + u_{1i}$ of its Knesset vote while winning a proportion $p_2 + u_{2i}$ of the Knesset vote for all other parties (including abstentions). Then

$$\begin{aligned} R_{Mi} &= (p_1+u_{1i})R_{Ki} + (p_2+u_{2i})\ (1-R_{Ki}) \\ &= p_2 + (p_1-p_2)R_{Ki} + u_{2i} + (u_{1i}-u_{2i})R_{Ki}. \end{aligned} \tag{1.1}$$

Let $e_i = u_{2i} + (u_{1i}-u_{2i})R_{Ki}$ and assume the R_{Ki}'s are given numbers. Then, if $E(e_i) = 0$ and $E(e_ie_j) = \sigma^2$ if $i=j$ and zero otherwise, ordinary least squares applied to

$$R_{Mi} = \beta_0 + \beta_1R_{Ki} + e_i \tag{1.2}$$

yields minimum variance linear unbiased estimates of p_1 and p_2 via the relations $\hat{p}_o = \hat{\beta}_0 + \hat{\beta}_1$, $\hat{p}_2 = \hat{\beta}_0$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimates from equation (1.2). For the constant variance assumption to hold, either $u_{2i} = u_{1i}$ or $R_{Ki}{}^2\ \mathrm{Var}(u_{1i}) + (1-R_{Ki})^2\ \mathrm{Var}(u_{2i}) + 2R_{Ki}(1-R_{Ki})\ \mathrm{Cov}(u_{1i},u_{2i})$ is constant for all i. Otherwise, procedures which allow for heteroscedastic errors will be required in order to obtain best linear unbiased estimates.

A variant of the Goodman model was suggested by Rosenthal [12]. Instead of letting the proportion won from the remaining parties be a constant, he suggested letting this proportion be a function of the Knesset vote proportion, for example, $p_2/(1-R_{Ki})$. Then,

$$R_{Mi} = (p_1+u_{1i})R_{Ki} + \left(\frac{p_2}{(1-R_{Ki})} + u_{2i}\right)(1-R_{Ki}). \tag{1.3}$$

Simple algebra will show that (1.3) also leads to (1.2) except that the estimates are $\hat{p}_1 = \hat{\beta}_1$, $\hat{p}_2 = \hat{\beta}_0$. The reader can conceive of still further models that could be associated with (1.2). Although the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ may enable one to eliminate some of these models on logical grounds (see Rosenthal [12] and Goodman [4]), any regression poses difficult problems of interpretation.

If we regard R_{Ki} and R_{Mi} both as endogenous (mutually determined by the electoral system) variables, then the least squares estimates of the parameters of (1.2) will be biased and inconsistent since e_i and R_{Ki} will be correlated in general (see, e.g., Johnston [6,pp.231-4]).

To analyze the simultaneous problem another relationship involving R_{Ki} is required. An equation analogous to (1.2) with the roles of R_{Mi} and R_{Ki} interchanged will not do, for the parameters of this two equation system are not estimable or identifiable. The parameters can be identified if we complicate the equations by adding exogenous (determined outside the system) variables to the model. To illustrate this, let

X and Y be two sociological variables (e.g. educational level and proportion of Oriental birth). Modifying (1.1) so as to make the proportions functions of the sociological variables leads, for example, to:

$$R_{Mi} = (p_1+p_3X_i+u_{1i})R_{Ki} + (p_2+p_4X_i+u_{2i})(1-R_{Ki}), \tag{1.4}$$

$$R_{Ki} = (q_1+q_3Y_i+v_{1i})R_{Mi} + (q_2+q_4Y_i+v_{2i})(1-R_{Ki}). \tag{1.5}$$

This leads to the structural equations:

$$R_{Mi} = \beta_0 + \beta_1 R_{Ki} + \beta_2 X_i + \beta_3 R_{Ki}X_i + e_i, \tag{1.6}$$

$$R_{Ki} = \gamma_0 + \gamma_1 R_{Mi} + \gamma_2 Y_i + \gamma_3 R_{Mi}Y_i + g_i, \tag{1.7}$$

where the relations between the parameters of (1.4)-(1.5) and (1.6)-(1.7) are immediate.

An interesting feature of (1.6) is that even in this simple case there are variables in the system which are products of endogenous and exogenous variables. This does not typically occur in models considered by econometricians.

In the next section of the paper we give a brief discussion of identification and estimation of systems of linear equations. In section 3 we present methods for dealing with the "adding up" constraint (that is, for each census district the sum of the Municipal proportions is unity, and similarly for the Knesset proportions). In section 4 we discuss the methods employed for dealing with the product variables and parameter constraints. The results of the analysis for the 1965 Jerusalem elections are given in section 5. Concluding remarks are given in section 6.

2. Identification and Estimation of Systems of Linear Equations.[3]

In this section we give a brief exposition of methods of estimation of the parameters of systems of equations of the form:

$$\sum_{\ell=1}^{G} y_{t\ell}\gamma_{\ell g} + \sum_{m=1}^{K} x_{tm}\beta_{mg} = u_{tg}, \qquad t=1,\dots,T,\ g=1,\dots,G. \tag{2.1}$$

The y's are jointly dependent or endogenous variables and the x's are predetermined variables which are stochastically independent of the errors, u_{tg}. The equations (2.1) may be expressed conveniently in matrix notation as

$$Y\Gamma + XB = U, \tag{2.2}$$

where Y is a $T \times G$ matrix of values of the endogenous variables, X a $T \times K$ matrix of values of predetermined variables, U a $T \times G$ matrix of unobservable stochastic disturbances, and Γ and B are $G \times G$ and $K \times G$ matrices of structural parameters.

We assume that the columns of X are linearly independent so that the rank of X is K.[4] Further, we assume that $|\Gamma| \neq 0$ so we can solve for the reduced form

$$Y = -XB\Gamma^{-1} + U\Gamma^{-1} = X\Pi + V. \tag{2.3}$$

Regarding the u's we assume

$$E(u_{tg}) = 0, \quad t=1,\dots,T, g=1,\dots,G,$$

and

$$E(u_{sg}u_{th}) = \begin{cases} \sigma_{gh} & \text{if } s=t, \\ 0 & \text{if } s\neq t. \end{cases}$$

If we let $u'(t) = [u_{t1},\dots,u_{tg}]$ we may express the above as

$$E(u'(t)) = [0,\dots,0],$$

$$E(u(s)u'(t)) = \begin{cases} \Sigma = \{\sigma_{gh}\} & \text{if } s=t, \\ \underset{\sim}{0}_{GG} = \{0\} & \text{if } s\neq t, \end{cases}$$

where Σ is a positive definite symmetric $G \times G$ matrix[5] and $\underset{\sim}{0}_{GG}$ is the null matrix of order G.

The reduced form (2.3) is similar to a multivariate linear regression model. This suggests estimation of the reduced form by ordinary least squares and then estimation of structural parameters using the relation

$$\Pi = -B\Gamma^{-1}.$$

To see that this is impossible in general let us post-multiply the structure (2.2) by an arbitrary non-singular matrix L to give

$$Y\Gamma L + XBL = UL$$

or

$$Y\Gamma^* + XB^* = U^* \tag{2.4}$$

with reduced form

$$Y = X\Pi^* + V^*,$$

where

$$\Pi^* = -B^*\Gamma^{*-1} = -BL(\Gamma L)^{-1} = -B\Gamma^{-1} = \Pi,$$

so that the two structures have the same reduced form.[6] The two structures (2.2) and (2.4) are observationally equivalent in the sense that the likelihood function for one structure is the same as that for the other. It is thus impossible in general to estimate structural parameters from the reduced form without the imposition of some additional prior information about the structural parameters. This is called the identification problem.[7]

It should be emphasized that identification is not a problem unique to simultaneous equation systems nor is it related to sample size. For example, suppose we are interested in estimating the weights of an apple and an orange. If we always weigh them together we can by repetitive weighing obtain a very precise estimate of the sum of the weights, but regardless of sample size we never can gain any information about the individual weights from this experiment.

We shall discuss only one kind of prior information in relation to the identification problem; namely, knowledge that certain elements of Γ and B are zero. We shall discuss conditions for identification of the parameters of the gth equation.

Let us write the gth equation of the system as

$$y_g = Y_g\gamma_g + X_g\beta_g + u_g, \tag{2.5}$$

where Y_g is a $T \times G_g$ matrix of values of endogenous variables which are included in the gth equation, X_g is a $T \times K_g$ matrix of included exogenous variables, y_g is a $T \times 1$ vector of observations on one endogenous variable which we put on the left hand side, u_g is a $T \times 1$ vector of disturbances, and γ_g and β_g are $G_g \times 1$ and $K_g \times 1$ vectors of parameters, whose elements are the appropriate elements of the gth columns of Γ and B of (2.2), respectively, all divided by $-\gamma_{gg}$.

Now write the part of the reduced form corresponding to y_g and Y_g as

$$[y_g : Y_g] = [X_g : X_g^*] \begin{bmatrix} \pi_g & \Pi_g \\ \pi_g^* & \Pi_g^* \end{bmatrix} + \text{reduced form disturbances},$$

where X_g^* is the $T \times (K-K_g)$ matrix of exogenous variables excluded from (2.5), π_g is $K_g \times 1$, π_g^* is $(K-K_g) \times 1$, Π_g is $K_g \times G_g$, and Π_g^* is $(K-K_g) \times G_g$. If we post-multiply by $[1: -\gamma_g']'$ we have on the left hand side

$$y_g - Y_g\gamma_g = X_g\beta_g + u_g ,$$

where the equality is by (2.5). Now, neglecting disturbances,[8] we must have on the right

$$[\pi_g : \Pi_g]\,[1: -\gamma_g']' = \beta_g, \tag{2.6}$$

$$[\pi_g^* : \Pi_g^*]\,[1: -\gamma_g']' = 0; \tag{2.7}$$

(2.6) gives β_g in terms of γ_g given π_g and Π_g. It does not imply restrictions on π_g and Π_g. Relation (2.7) says

$$\pi_g^* = \Pi_g^*\,\gamma_g .$$

This system of equations has a unique solution for γ_g if and only if the ranks of Π_g^* and $[\pi_g^* : \Pi_g^*]$ are both G_g. This is the <u>rank condition</u> for identification. This rank can be G_g only if $K - K_g \geq G_g$; that is, only if the number of exogenous variables excluded from the gth relation is greater than or equal to the number of endogenous variables in the equation minus one. This necessary condition is called the <u>order condition</u>. Both conditions may easily be extended to the case of homogeneous linear restrictions on the parameters of the gth equation. The rank condition may also be stated in terms of the structural coefficient matrices (see Johnston [6, p.251]).

Having established criteria for the possibility of estimation, we now proceed to discuss estimation methods. If ordinary least squares were applied directly to (2.5) the estimators would be inconsistent because of the correlation between Y_g and u_g. Theil [14,15] developed the following approach which yields consistent estimators and is known as <u>two stage least squares</u> (2SLS).[9]

Now Y_g is a submatrix of $Y = -XB\Gamma^{-1} + U\Gamma^{-1}$, so in (2.5) let us replace Y_g with its equivalent expression from (2.3); that is,

$$y_g = [(-XB\Gamma^{-1})_g : X_g] \begin{bmatrix} \gamma_g \\ \beta_g \end{bmatrix} + u_g + (U\Gamma^{-1})_g\gamma_g .$$

However, $B\Gamma^{-1}$ is unknown. Nevertheless, we may estimate Y_g by ordinary least squares from the reduced form:

$$\hat{Y}_g = X(X'X)^{-1}X'Y_g = X\hat{\Pi}_g .$$

Then

$$Y_g = X\hat{\Pi}_g + V_g ,$$

where

$$V_g = [I - X(X'X)^{-1}X']Y_g .$$

This is the first stage. The second stage consists of regressing y_g on $X\hat{\Pi}_g$ and X_g; that is,

$$y_g = [Y_g - V_g : X_g]\,[\gamma_g' : \beta_g']' + u_g + V_g\gamma_g .$$

Then

$$(\hat{\gamma}_g' : \hat{\beta}_g')' = M_{\hat{z}\hat{z}}^{-1}\, m_{\hat{z}y_g} ,$$

where

$$M_{\hat{z}\hat{z}} = \begin{bmatrix} (Y_g-V_g)'(Y_g-V_g) & (Y_g-V_g)'X_g \\ X_g'(Y_g-V_g) & X_g'X_g \end{bmatrix}$$

and

$$m_{\hat{z}y_g} = [(Y_g - V_g):X_g]'y_g \ .$$

For the inverse of the matrix $M_{\hat{z}\hat{z}}$ to exist it is necessary that the order condition be satisfied. Note that $X_g'V_g = \underset{\sim}{0}_{K_g G_g}$ (the $K_g \times G_g$ null matrix) and $(Y_g - V_g)'(Y_g - V_g) = Y_g'Y_g - V_g'V_g$. The The 2SLS estimator has a normal limiting distribution with mean $[\gamma_g':\beta_g']'$ and covariance matrix $\sigma_{gg}M^{-1}_{\hat{z}\hat{z}}$ [however, finite sample moments need not exist (see, e.g., Dhrymes [2,pp. 180, 204])]; σ_{gg} may be estimated consistently by

$$s_{gg} = \frac{1}{T}(y_g - Y_g\hat{\gamma}_g - X_g\hat{\beta}_g)'(y_g - Y_g\hat{\gamma}_g - X_g\hat{\beta}_g).$$

The method of <u>three stage least squares</u> (3SLS) is a method for simultaneous estimation of the entire system. If we let $Z_g = [Y_g:X_g]$ and $\delta_g = [\gamma_g':\beta_g']'$ and premultiply each equation by X' we may write

$$\begin{bmatrix} X'y_1 \\ \cdot \\ \cdot \\ \cdot \\ X'y_G \end{bmatrix} = \begin{bmatrix} X'Z_1 & 0 \ \dots \ 0 \\ \cdot & \cdot \ \cdot \ \ \cdot \\ \cdot & \cdot \ \ \cdot \ \ \cdot \\ \cdot & \cdot \ \ \cdot \ \cdot \\ 0 & 0 \ \dots \ X'Z_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \cdot \\ \cdot \\ \cdot \\ \delta_G \end{bmatrix} + \begin{bmatrix} X'u_1 \\ \cdot \\ \cdot \\ \cdot \\ X'u_G \end{bmatrix}$$

or

$$w = W\delta + \epsilon \ ,$$

with

$$E\ \epsilon\ \epsilon' = \Sigma \otimes X'X.$$

Then the 3SLS estimator of δ is

$$\hat{\delta} = \{W'[\Sigma^{-1} \otimes (X'X)^{-1}]W\}^{-1}\ W'[\Sigma^{-1} \otimes (X'X)^{-1}]w, \quad (2.9)$$

with the unknown matrix Σ replaced by an estimate of itself constructed from the 2SLS residuals. The estimator $\hat{\delta}$ is consistent, has covariance matrix approximately equal to $\{W'[\Sigma^{-1} \otimes (X'X)^{-1}]W\}^{-1}$, and is asymptotically efficient unless some of the elements of Σ are known (see Rothenberg and Leenders [13] and Madansky [10]).

3. The Adding-Up Conditions

<u>Implications for the reduced form</u>. Denote the typical reduced form observation by

$$\underset{(1\times G)}{y'(t)} = \underset{(1\times K)}{x'(t)}\ \underset{(K\times G)}{\Pi} + \underset{(1\times G)}{v'(t)}, \quad t=1,\dots,T, \quad (3.1)$$

with

$$y'(t)\ \underset{\sim}{1}_G = 1 \quad \left(\sum_{i=1}^{G} y_i(t) = 1\right), \quad t=1,\dots,T, \quad (3.2)$$

where $\underset{\sim}{1}_G$ is a G-component (column) vector with each element equal to unity. Clearly

$$y'(t)\ \underset{\sim}{1}_G = x'(t)\Pi\underset{\sim}{1}_G + v'(t)\underset{\sim}{1}_G\ , \quad t=1,\dots,T,$$

so from (3.2),

$$1 = x'(t)\Pi\underset{\sim}{1}_G + v'(t)\underset{\sim}{1}_G, \quad t=1,\dots,T; \quad (3.3)$$

taking expectations,

$$x'(t)\Pi\underset{\sim}{1}_G = 1, \quad t=1,\dots,T. \quad (3.4)$$

Substituting this result back into (3.3),

$$v'(t)\underset{\sim}{1}_G = 0, \quad t=1,\dots,T; \quad (3.5)$$

thus

$$\Omega\ \underset{\sim}{1}_G = \underset{\sim}{0}_G\ ,$$

where

$$\Omega = E[v(t)\ v'(t)], \quad t=1,\dots,T.$$

Let us examine the restrictions on Π more closely. First, it follows immediately from (3.2) and the fact that

$$\Pi_{\cdot} \equiv \Pi\ \underset{\sim}{1}_G = (\pi_{1\cdot}, \pi_{2\cdot}, \dots, \pi_{K\cdot})'$$

(where $\pi_{i\cdot} = \sum_j \pi_{ij}$) is a constant vector that there exists at least one linear combination of the columns of X that sums to a given non-zero number (say unity), and the condition that X be of full column rank (K) implies that there can be at most one such linear combination; consequently there is exactly one such combination, and hereinafter it is assumed that $x_{t1} = 1$, $t=1,\dots,T$.

Next it is shown that $\pi_{i\cdot} = 0$, $i=2,\dots,K$. Condition (3.4) may be written equivalently as

$$\sum_{i=1}^{G} x_{ti}\pi_{i\cdot} = 1, \quad t=1,\dots,T.$$

Suppose this restriction holds for a given X; then it cannot hold if we perturb any x_{ti} for which $\pi_{i\cdot} \neq 0$. Since the x_{ti}'s are unconstrained for $i=2,\dots,G$, it follows that $\pi_{i\cdot} = 0$, $i=2,\dots,G$. Finally, it follows that $\pi_{1\cdot} = 1$. Although stated differently, these conditions are essentially equivalent to those stated in McGuire <u>et al</u>. [11].

It should be noted that if $y'(t)$ is subject to inequality constraints of the form

$$\underset{\sim}{0}'_G \leq y'(t) \leq \underset{\sim}{1}'_G, \quad t=1,\dots,T,$$

then

$$\underset{\sim}{0}'_G \leq x'(t)\ \Pi \leq \underset{\sim}{1}'_G\ , \quad t=1,\dots,T. \quad (3.6)$$

One obvious implication of this condition is that the $x'(t)$'s must be bounded, for if one or more components of $x'(t)$ is unbounded, then it is always possible to choose a value great enough that (3.6) is violated.

Implications for the structural equations. Define

$$\Gamma^{\cdot} = \Gamma^{-1}\underset{\sim}{1}_G = (\gamma^{1\cdot}, \gamma^{2\cdot}, \ldots, \gamma^{G\cdot})' ,$$

where $\gamma^{i} = \sum_j \gamma^{ij}$ is the i,j-th element of Γ^{-1}.

It was shown above that $[1:\underset{\sim}{0}'_{K-1}]' = \Pi\, \underset{\sim}{1}_G$; in terms of restrictions on the coefficient matrix B in the structural equations this condition becomes

$$[1:\underset{\sim}{0}'_{K-1}]' = \Pi\underset{\sim}{1}_G = (\Pi\Gamma)(\Gamma^{-1}\underset{\sim}{1}_G) = -B\Gamma^{\cdot}.$$

Similarly, (3.5) implies

$$0 = v'(t)\underset{\sim}{1}_G = (v'(t)\Gamma)(\Gamma^{-1}\underset{\sim}{1}_G)$$

$$= u'(t)\Gamma^{\cdot} , \qquad t=1,\ldots,T;$$

thus

$$\underset{\sim}{0}_G = \Omega\, \underset{\sim}{1}_G = (\Gamma')^{-1}\Gamma'\Omega\Gamma\Gamma^{-1}\underset{\sim}{1}_G$$

$$= (\Gamma')^{-1}\Sigma\, \Gamma^{\cdot} = \Sigma\, \Gamma^{\cdot};$$

hence, the rank of Σ, which is equal to the rank of Ω since Γ has full rank (G), is less than G.

The linear dependence of the elements of the disturbance vector impinges on the identification of the elements of Γ. Define

$$z^{(i)}(t) = (z_{t,1},\ldots,z_{t,i-1}, z_{t,i+1},\ldots,z_{t,m})',$$

where $z(t)$ is an m-component (column) vector, and partition the structural model

$$y'(t)\Gamma + x'(t)B = u'(t), \qquad t=1,\ldots,T,$$

$$(y^{(G)'} : y_G)\begin{bmatrix} \Gamma_{11} & \Gamma_{21} \\ \Gamma_{12} & \Gamma_{22} \end{bmatrix} + (B_{11} : B_{21}) + x^{(1)'}[B_{12} : B_{22}] = (u^{(G)'} : u_G),$$

where the nature of the partitioning is obvious. Using (3.2), this relationship may be written equivalently as

$$y^{(G)'}[\Gamma_{11} - \underset{\sim}{1}\,\Gamma_{12} : \Gamma_{21} - \underset{\sim}{1}\,\Gamma_{22}] + (B_{11} + \Gamma_{12} : B_{21} + \Gamma_{22}) + x^{(1)'}[B_{12} : B_{22}] = (u^{(G)'} : u_G).$$

Define

$$\bar{\Gamma}_{11} = \Gamma_{11} - \underset{\sim}{1}\,\Gamma_{12}, \quad \bar{\Gamma}_{21} = \Gamma_{21} - \underset{\sim}{1}\,\Gamma_{22} ,$$

$$\bar{B}_{11} = B_{11} + \Gamma_{12}, \quad \bar{B}_{21} = B_{21} + \Gamma_{22};$$

then this system is described completely by

$$y^{(G)}\bar{\Gamma}_{11} + \bar{B}_{11} + x^{(1)}B_{12} = u^{(G)} \tag{3.7}$$

since $\bar{\Gamma}_{21}$ is a linear combination of the columns of $\bar{\Gamma}_{11}$. This set of structural relations is in the form of standard classical econometric models and can be analyzed as such (i.e., the identification conditions and estimation procedures described in the previous section apply directly to this specification). It should be obvious that identification of $\bar{\Gamma}_{11}$, $\bar{B}_{11}$ and B_{12} is the most that possibly can be achieved in this model, for model (3.7) assumes that form for all Γ_{12}, Γ_{21}, and Γ_{22}.

4. Non-Linear Estimation and Identification

In the previous sections we treated identification, estimation, and the adding-up condition in linear simultaneous equations systems. We now briefly discuss these areas with respect to the non-linear problem to be treated here.

The problem of identification of certain non-linear systems has been treated by Fisher [3] and Kelejian [7]. Consider a system of simultaneous equations

$$y'(t)\Gamma + F'(t)A + x'(t)\, B = u'(t),$$

where $F'(t)$ is a vector of non-linear functions of contemporaneous endogenous and exogenous variables, and let $H(t) = E[F(t)]$. Kelejian proves the following result. If the columns of X and $F = [F(1),\ldots,F(T)]'$ are linearly independent, "each additional endogenous function may be considered, for identification purposes, as just another linearly independent predetermined variable" [Kelejian, 7, p.7]. By this rule, all the equations in non-linear systems we shall consider in this study are over-identified.

Estimation poses more of a problem. The technique used here is to approximate the reduced form with a second order Taylor Series expansion, giving the i-th reduced form equation[10]

$$y_{ti} = \sum_{k=1}^{K} \sum_{j=1}^{K} x_{tk}x_{tj}\beta_{ij} + v_{ti}, \qquad \begin{array}{l} i=1,\ldots,G, \\ t=1,\ldots,T. \end{array} \tag{4.1}$$

The predicted values of the endogenous variables from (4.1) are then used in the second and third stages.

There are two sets of adding-up conditions for each observation (census tract) in our problem: the proportions of the Knesset vote received by each party (including abstentions) sum to unity, as do the corresponding Municipal proportions. Accordingly, we eliminate one Knesset equation and one Municipal equation prior to estimation.

5. Results for the 1965 Jerusalem Elections

In this section we analyze a model based on a ten equation model. The endogenous variables in these structural relations are the proportions of the registered voters casting ballots for the (1) Rafi, (2) Gahal, (3) All Religious, (4) All Other Secular [about three-fourths of which consists of the Alignment (of Mapai and Ah'dut A'avoda) vote] parties and (5) abstaining in the Knesset and Municipal races. Since this model is primarily illustrative and is not what we regard as the best specification of the election, we aggregated the All Religious, All Other Secular, and Abstention votes into an All Other category to prevent the analysis from becoming too cumbersome. This aggregation is not likely to conceal any interesting switching among religious parties, since Kies and Rosenthal [9] have shown that the squared correlations between the Knesset and Municipal votes received by the three religious parties, Poalei Aguda, NRP, and Aguda, are 0.92, 0.95, and 0.98, respectively; similar results were obtained for abstentions. If there are any explanatory variables influencing the All Religious, All Other Secular, or Abstention votes which do not drop out when these variables are aggregated into the All Other category and which are not included in at least the Rafi or Gahal pair of structural relations, then the absence of such variables is a specification error.

We chose to retain the pair of Rafi equations for this illustrative example because we hypothesized that the feedback effects were greatest for this party due to (1) Kollek's presence and (2) Rafi being a new party. Gahal was retained because it was shown in Kies and Rosenthal [9] that, at least with OLS estimation, the Gahal municipal equation benefitted substantially from the inclusion of "product" variables as against ordinary linear variables.

In any simultaneous equations problem, the choice of exogenous variables is obviously limited by available data. We had the results of the 1961 elections, when only the Knesset was elected, and detailed census data (occupational structure, ethnic origin broken down by period of immigration, sex, age, housing conditions, etc.). Unfortunately, we did not have data for religious practice, one of the sociological variables which may play an important role in political analysis.

It seems important to use the 1961 Knesset vote variables since voter behavior will reflect organizational and historical as well as socioeconomic influences on preferences. In fact, analysis presented in Kies [8] indicates that in this case past vote variables are generally better predictors than socioeconomic variables and that socioeconomic variables frequently add little to explained variance over what is "explained" by the past vote variables.

In fact, we used only two socioeconomic variables in the analysis chosen as markers for the two major dimensions found in a Guttman-Lingoes SSA-I analysis for the socioeconomic variables (see Kies [8]). The two variables were the proportion of the total population immigrant from "Oriental" countries (chiefly Sephardic Jews from North Africa and the Middle East) before 1948, which we denote by Seph. 48-, and the proportion of the adult population with less than one complete year of formal schooling, denoted by Low Ed.; both variables come from the 1961 census.

Since the original ten-equation model has been reduced to a six-equation model by aggregation and since one Knesset equation and Municipal equation are estimated indirectly by using the "adding up" conditions, we estimated four equations. We chose to eliminate the pair of All Other relations, so the structural equations we estimated are for the (1) Rafi Knesset, (2) Rafi Municipal, (3) Gahal Knesset, and (4) Gahal Municipal proportions, respectively. The OLS and 3SLS estimates of the model parameters are presented in Table 1. We also estimated a linearized version of the model; the 3SLS estimates of the parameters and the implied reduced form (RF) paramter estimates (i.e., $\hat{\Pi} = \hat{B}\hat{\Gamma}^{-1}$ in the notation of sections 2 and 3) are presented in Table 2.

Comparison of the estimation methods. The OLS estimates would seem to be reasonably satisfactory insofar as the R^2 values, which range from .80 to .93 for the nonlinear model (.76 to .92 for the linear model, although these estimates are not shown due to space limitations) are evidence of good cross-sectional specification. However, since the equations contain endogenous variables as regressors, the estimates are inconsistent. The 3SLS estimates, which are consistent, also indicate good explanatory power. The importance of using 3SLS is evidenced by the differences in the estimates obtained using the two procedures, which in most cases exceed one standard error (using either the OLS or the 3SLS estimate of the standard error of the parameter estimates).

The reduced form estimates. The reduced form estimates for the linear model are consistent with what we know about the structures of the parties and the relationships between the 1961 and 1965 parties. The Rafi party was formed by Mapai dissidents; in the 1965 elections Rafi captured about 55 percent of the Mapai 1961 vote in the Municipal election and 25 percent in the Knesset election. Gahal, a coalition of the Herut and part of the Liberal organizations, picked up 42 percent of the Herut 1961 vote in the Municipal election and 71 percent in the Knesset election while retaining 25 percent of the Liberal 1961 vote in the Municipal election and 42 percent in the Knesset election. It is

TABLE 1. ORDINARY AND THREE STAGE LEAST SQUARES ESTIMATES OF THE NON-LINEAR MODEL

Regression Variable	Rafi-Knesset OLS	Rafi-Knesset 3SLS	Rafi-Municipal OLS	Rafi-Municipal 3SLS	Gahal-Knesset OLS	Gahal-Knesset 3SLS	Gahal-Municipal OLS	Gahal-Municipal 3SLS
Constant	0.0114 (0.0049)	0.0072 (0.0053)	0.0053 (0.0115)	-0.0014 (0.0112)	0.0352 (0.0112)	0.0417 (0.0103)	0.0053 (0.0088)	-0.0043 (0.0097)
Herut, 1961					0.3054 (0.0581)	0.2126 (0.0579)		
Liberal, 1961	-0.1296 (0.0429)	-0.1972 (0.0443)	0.4051 (0.0702)	0.4443 (0.0694)	0.1947 (0.0650)	0.1091 (0.0584)		
Mapai, 1961			0.2200 (0.0604)	0.2651 (0.0639)				
Low Ed.					-0.0586 (0.0277)	-0.0920 (0.0292)		
Seph. 48-	-0.1650 (0.0427)	-0.1267 (0.0381)			-0.0721 (0.1053)	-0.2102 (0.1172)		
Rafi-Knesset			1.3281 (0.1500)	1.1179 (0.1936)				
Rafi-Municipal	0.2625 (0.0440)	0.3600 (0.0495)						
Gahal-Knesset							0.4618 (0.0524)	0.5245 (0.0578)
Gahal-Municipal					0.9429 (0.1058)	1.1438 (0.1312)		
Herut, 1961 x Rafi-M	1.0499 (0.1935)	0.7685 (0.1817)						
Low Ed. x Gahal-K			-0.2534 (0.0904)	-0.1006 (0.0827)			0.4286 (0.0559)	0.3757 (0.0566)
Seph. 48- x Gahal-K							1.1975 (0.2172)	1.0381 (0.2203)
Adjusted variance x 10000[a]	1.3881	1.4222	5.7244	6.5218	3.6629	3.7132	2.1404	2.1412
Var. of dep. var. x 10000	6.5825	6.5825	35.760	35.760	38.507	38.507	30.351	30.351

Note a. The adjusted variances for the 3SLS estimates are based on the 2SLS estimates.

interesting to note that Kollek's strategy for cultivating the vote of Sephardic Jews (he gave *Sefaradim* two of the first five places on his list) apparently failed; he captured none of the Herut 1961 vote and, *ceteris paribus*, lost about three-eighths of a vote for every voting Sephardic Jew. (There appears to be little difference between the attractiveness of the Rafi Municipal and Knesset lists to Sephardic Jews; the strong ethnic appeal of Gahal is evidenced by the large positive coefficients of Seph. 48- in the Gahal equations). However, his appeal to high status voters obviously paid off; he captured 36 percent of the Liberal 1961 vote while this variable actually exerted a slight negative effect on the Rafi Knesset vote. Although the educational achievement of the constituency appears not to have been terribly important, its effect is in the predicted direction. It often is argued that independents are more likely to appeal to better educated, "thinking" voters [9]. The negative coefficients of "Low Ed." (proportion of voters completing less than one year of formal education), with the more negative occurring in the Rafi Municipal equation, and the positive coefficients in the Gahal equations are consistent with this hypothesis.

Estimates of a proportions model. In this section we examine the implications of the parameter estimates in the context of a proportions model such as that discussed in the Introduction.

TABLE 2. REDUCED FORM AND THREE STAGE LEAST SQUARES ESTIMATES OF THE LINEAR MODEL

	Equation							
	Rafi-Knesset		Rafi-Municipal		Gahal-Knesset		Gahal-Municipal	
Regression Variable	RF	3SLS	RF	3SLS	RF	3SLS	RF	3SLS
Constant	-0.0033	-0.0033 (0.0077)	0.0001	0.0045 (0.0162)	0.0324	0.0384 (0.0103)	-0.0049	-0.0243 (0.0085)
Herut, 1961	0.0089	0.0095 (0.0286)	-0.0014		0.7102	0.1947 (0.0559)	0.4244	
Liberal, 1961	-0.0585	-0.2251 (0.0469)	0.3606	0.4374 (0.0701)	0.4231	0.1160 (0.0558)	0.2529	
Mapai, 1961	0.2532		0.5481	0.2464 (0.0631)				
Low Ed.	-0.0386		-0.0835	-0.0365 (0.0254)	0.0613	-0.1010 (0.0292)	0.1336	0.0970 (0.0151)
Seph. 48-	-0.3094	-0.1350 (0.0445)	-0.3775		0.5254	-0.2652 (0.1176)	0.6511	0.3370 (0.0710)
Rafi-Knesset				1.1915 (0.1935)				
Rafi-Municipal		0.4620 (0.0433)						
Gahal-Knesset				-0.0168 (0.0623)				0.5977 (0.0531)
Gahal-Municipal						1.2144 (0.1282)		
Adjusted variance x 10000 (see Table 1, Note a)		1.6559		6.1791		3.7132		2.3374
Variance of dependent variable x 10000	6.5825	6.5825	35.760	35.760	38.507	38.507	30.351	30.351

In constructing proportions interpretations we assume: (a) all voters who voted for the Rafi Knesset list also voted for the Rafi Municipal list; (b) all voters who voted for the Gahal Municipal list also voted for the Gahal Knesset list. These severe assumptions obviously are simplifications, allowing for switching in only one direction for each party. Nevertheless, they may not be grossly incorrect in the context of this election, and the parameter estimates of the implied proportions model do not violate any theoretical constraints (the terms in brackets are the proportions).

$$\text{Rafi}_K = [.3600 + .7685(\text{Herut}_{61}) + \{.0724 - .1972(\text{Lib}_{61}) - .1267(\text{Seph. 48-})\}/\text{Rafi}_M]\ \text{Rafi}_M$$

$$\begin{aligned}\text{Gahal}_K = {} & [1.0]\ \text{Gahal}_M \\ & + [\{.1855 + .2126(\text{Herut}_{61}) + .1091(\text{Lib}_{61}) - .2102(\text{Seph. 48-}) \\ & - .0920(\text{Low Ed.}) - .1438(\text{Gahal}_M)\}/(\text{Other}_M + \text{Rafi}_M)] \\ & \times (\text{Other}_M + \text{Rafi}_M)\end{aligned}$$

$$\begin{aligned}\text{Rafi}_M = {} & [1.0]\ \text{Rafi}_K \\ & + [\{.2651(\text{Mapai}_{61}) + .4443(\text{Lib}_{61}) + .1179(\text{Rafi}_K) \\ & - .1006(\text{Low Ed. x Gahal}_K)\}/(\text{Other}_K + \text{Gahal}_K)](\text{Other}_K + \text{Gahal}_K)\end{aligned}$$

$$\text{Gahal}_M = [.0043/(\text{Gahal}_K) + .5245 + .3757(\text{Low Ed.}) + 1.0381(\text{Seph. 48-})]\ \text{Gahal}_K$$

A couple of remarks are in order.

(i) Although we need to undertake an

investigation as to whether the proportions are bounded between zero and one for all sample values, investigation of these expressions for various trial values suggests that the functions for the proportions behave appropriately.

(ii) That the various proportions are the complex expressions in brackets reflects the finding that simple proportions models lead to estimates far outside the [0, 1] interval (see Kies and Rosenthal [9]). That we reject the simple proportions model is hardly a reason for accepting the present estimates. Given that the present model is intended as an illustrative example, great caution should be placed on the use of these proportions models. On the other hand, this complexity in "proportions" interpretations of the results would seem to lend further support to the use of simultaneous models.

6. Directions for Further Research

One of the present difficulties of econometric analyses of split ticket voting is the absence of well-developed mathematical theory of voting behavior in multiple-office elections. Progress in empirical analysis is not likely to be made without prior improvements in theory. Accordingly, we now outline the directions which we believe theoretical advances might take and the difficulties likely to be encountered. First we discuss an alternative specification of coattails. We then discuss the problems of modeling cognitive dissonance. Finally, we consider the relationship between the coattails and the cognitive dissonance specifications.

<u>Coattails revisited</u>. For expositional clarity let us consider the pair of equations for Rafi only, which in the absence of coattails may be written

$$R_K = x_1\beta_1 + x_2\beta_2 + \epsilon_K, \tag{6.1}$$

$$R_M = x_1\alpha_1 + x_3\alpha_3 + \epsilon_M. \tag{6.2}$$

Noticeably absent from these equations is a variable measuring the special appeal of a Teddy Kollek. Thus, if we knew the true values of the vectors α_1 and α_2 (or had independent estimates of them) we surely would underpredict R_M. A reasonable estimate of the personal appeal of Kollek is the difference between the vote he received and the vote he was expected to receive; this difference is simply ϵ_M [or, equivalently, $(R_M - x_1\alpha_1 - x_3\alpha_3)$]. One way to model Kollek's coattails is to add the term $\beta_4\epsilon_M$ to (6.1), giving

$$R_K = x_1\beta_1 + x_2\beta_2 + \beta_4\epsilon_M + \epsilon_K. \tag{6.3}$$

We currently are attempting to gather relevant data for one or more other Israeli cities so that we might obtain independent estimates of the α's which can be used in estimating (6.3).

<u>Cognitive dissonance</u>. Consider a two-party [say Rafi (1) and Gahal (2)], two-office [say Knesset (1) and Municipal (2)] election. One way to model the simultaneous choice problem is to assume that each individual first decides how he would vote for each office if that were the only decision and then evaluates the set of independent choices later to check for cognitive balance. Let p^*_{ij} be the proportion of individuals who would vote for party i in election 1 and party j in election 2 if these decisions were treated as independent and let p_{ij} be the corresponding proportion in the actual simultaneous choice situation; also, let R^*_K and R^*_M be the vote proportions which would have been obtained in the hypothetical independent choice situation. Then

$$R^*_K = p^*_{11} + p^*_{12}\ , \quad R^*_M = p^*_{11} + p^*_{21}\ ,$$

$$R_K = p_{11} + p_{12}\ , \quad R_M = p_{11} + p_{21}\ .$$

Now assume that because of political realties, party loyalty, or cognitive dissonance (these concepts are not identifiable at this level of analysis), individuals voting for different parties in the two elections (that is, those voters comprising the p^*_{12} and p^*_{21} terms) reconsider their decisions. It seems reasonable to assume

$$p_{11} = p^*_{11} + \alpha_1 p^*_{12} + \alpha_2 p^*_{21}\ , \quad p_{12} = (1-\alpha_1-\beta_1)p^*_{12},$$

$$p_{22} = p^*_{22} + \beta_1 p^*_{12} + \beta_2 p^*_{21}\ , \quad p_{21} = (1-\alpha_2-\beta_2)p^*_{21},$$

$$0 \le \alpha_1,\alpha_2,\beta_1,\beta_2 \le 1, \quad \alpha_1+\beta_1 \le 1, \quad \alpha_2+\beta_2 \le 1\ .$$

Then it is easily shown that

$$R_K - R^*_K = \alpha_2 p^*_{21} - \beta_1 p^*_{12} \gtreqless 0\ ,$$

$$R_M - R^*_M = \alpha_1 p^*_{12} - \beta_2 p^*_{21} \gtreqless 0\ .$$

Thus at this level theory provides no clues about the magnitude or direction of adjustments due to cognitive dissonance; indeed, it even provides no clues about the variables that should be in the model.

When we move on to two elections each involving more than two parties the difficulties are compounded. For example, consider the individual whose independent choices are Gahal in the Knesset election and Independent Liberal in the Municipal election. One reasonable way (in addition to the previously discussed ways) that the individual might achieve cognitive balance is to vote for Rafi in both elections.

The prospects for modeling cognitive dissonance adjustments in an aggregate model are dim. Apparently the first two sentences in chapter 11 of Brown[11] should not be overlooked: "The general experimental design for discovering determinants of attitude change is a simple one. Some sort of an attitude must be measured <u>before and after</u> [italics ours] the interpolation of persuasive communications which differ from one another in some known respect."

Coattails and cognitive dissonance. It would appear that the coattails phenomenon is a subset of cognitive dissonance. What we mean by an unusually attractive candidate is one who can attract votes which his party ordinarily would not have received. His ability to carry some portion of these votes for the party in elections for other offices would seem to result from the cognitive balancing on the part of the extra individuals who were attracted by the man (as opposed to the party). The practical difference between coattails in particular and cognitive dissonance in general is that in the former case we believe that we have a method for dealing with the situation empirically at an aggregate level while in the latter case we do not.

FOOTNOTES

[1]This research was supported in part by National Science Foundation grant GS-2945 to Howard Rosenthal and NSF grant GS-2751 to Timothy McGuire (and others).

[2]We emphasize that this approach abstracts from the simultaneity problem which is of central interest here. We include a discussion of it for the sake of continuity with earlier literature and to motivate the simultaneous problem.

[3]This subject is discussed in detail in any econometrics text. A good introductory exposition is given in Johnston [6].

[4]We also assume $\underset{T \to \infty}{\text{plim}} \frac{1}{T} X'X = Q$ with $|Q| \neq 0$. This assumption guarantees that the estimation procedures discussed below have desirable asymptotic properties.

[5]In many econometric models, some equations are identities with no disturbance terms. In this case Σ is non-negative definite.

[6]A simple counting argument will also indicate this. Π contains KG elements whereas the total number of elements in B and Γ is $KG + G^2$.

[7]For a general treatment of identification see Fisher [3]. For a Bayesian approach see Zellner [16].

[8]It is straightforward to show that this function of the reduced form disturbances is exactly u_g.

[9]The method was independently developed by Basmann [1]. Also, a substantial amount of work has been done on maximum likelihood methods (see Hood & Koopmans [5]). For a more recent exposition of estimation methods see Dhrymes [2].

[10]In actual fact we omitted three variables, (Seph. 48-x Liberal, 1961), (Low Ed. x Mapai, 1961), and (Herut, 1961 x Liberal, 1961), due to the high degree of collinearity in the set of first and second order exogenous variables. Since the number of such variables is (K+1)K/2, the multi-collinearity problem is likely to be serious in this situation. One possibility might be to use principal components analysis to eliminate unimportant variables.

[11]Roger Brown, Social Psychology (New York: The Free Press, 1965), p. 549.

REFERENCES

[1] Basmann, R. L., "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," Econometrica, 25 (1957), 77-83.
[2] Dhrymes, P. J., Econometrics. New York: Harper & Row, 1970.
[3] Fisher, F. M., The Identification Problem in Econometrics. New York: McGraw-Hill, 1966.
[4] Goodman, L. A., "Some Alternatives to Ecological Correlation," American Journal of Sociology, 64 (1959), 610-25.
[5] Hood, W. C. and Koopmans, T. C., eds., Studies in Econometric Method. New York: John Wiley & Sons, Inc., 1953.
[6] Johnston, J., Econometric Methods. New York: McGraw-Hill Book Company, Inc., 1963.
[7] Kelejian, H. H., "Identification of Nonlinear Systems: An Interpretation of Fisher (also, the Multiple Solution Problem)," presented at the Second World Congress of the Econometric Society, Cambridge, England, September, 1970.
[8] Kies, N. E., Constituency Support and the Israeli Party System: An Analysis of Elections in Jerusalem, 1959-65. Unpublished Ph.D. dissertation, MIT, 1969.
[9] Kies, N. E. and Rosenthal, H., "Split Tickets and Aggregate Analysis: Jerusalem, 1965, unpublished Carnegie-Mellon University Working Paper WP-28-69-7, 1969.
[10] Madansky, A., "On the Efficiency of Three-Stage Least-Squares Estimation," Econometrica, 32 (1964), 51-56.
[11] McGuire, T. W., Farley, J. U., Lucas, R. E. and Ring, L. W., "Estimation and Inference for Linear Models in Which Subsets of the Dependent Variable are Constrained," Journal of the American Statistical Association, 63 (1968), 1201-13.
[12] Rosenthal, H., "Aggregate Data," in I. Pool et al., eds., Handbook of Communications, Chicago: Rand McNally, forthcoming
[13] Rothenberg, T. J. and Leenders, C. T., "Efficient Estimation of Simultaneous Equation Systems," Econometrica, 32 (1964), 57-76.
[14] Theil, H., Economic Forecasts and Policy, Second Revised Edition. Amsterdam: North-Holland Publishing Company, 1961.
[15] ________, "Estimation and Simultaneous Correlation in Complete Equation Systems," The Hague: Central Planning Bureau, 1953.
[16] Zellner, A., "Bayesian Inference in Econometrics," unpublished University of Chicago manuscript, 1970.

CAUSAL RELATIONS AND STRUCTURAL MODELS*

Richard Juster, Survey Research Center, University of California, Berkeley

1. Introduction

Following early work by Simon (1954), a method of testing the adequacy of certain a priori assumptions of non-experimental causal models has gained considerable acceptance among social investigators.[1] This approach to empirical theory construction has been developed further by Blalock (1962, 1964), Alker (1965, 1966), and Boudon (1968), among others, and is generally referred to as "Simon-Blalock causal modeling."

Simon and others who have contributed to this approach to theory building have limited their interest almost exclusively to recursive-form linear structural models and have been concerned with implications of such models in so far as they are representations of "one-way" or "non-reciprocal" causal relations among sets of variables.[2] In his classic paper, Simon (1954) derives necessary and sufficient conditions (which can be empirically approximated) for specified structural coefficients in the three-equation recursive-form linear model to be zero. Assuming that the dependent variable in each structural equation is subject to change according to the causal laws postulated by the equation, he interprets his statistical results to be a test of causal relations represented in a recursive-form linear model.

The purpose of the present paper is two-fold. First, after a review of Simon's work, we shall elucidate some of the inherent weaknesses of the recursive-form linear structural model for the representation of causal processes and shall argue against its use, even when it seems an appropriate choice for such a representation. A case will be made for the employment of linear structural systems which are less restrictive than the recursive-form in two important respects. Specifically, we are interested in linear structural systems which (1) have the ability to represent reciprocal causal relations and (2) require weaker restrictions on the covariances of the disturbance terms than is required for the recursive-form system. Second, we shall derive correlational conditions which are both necessary and sufficient for specified structural coefficients to be zero in a type of linear structural model which is more general than the recursive-form in the two respects just mentioned. As does Simon for the recursive-form system, we shall take our statistical results to be a test of causal relations in this more general structural model.

Greatly simplifying Simon's paper (1954), we shall try to capture the most important implications of his efforts. In an earlier piece (1953), Simon formally develops the concepts of causality and the causal relation.[3] Offering an implicit definition of the causal relation as an asymmetrical relation between two variables, Simon points out that the temporal sequence of the variables is not the basis of the asymmetry which defines the causal relation. In fact, his implicit definition admits relations where no temporal sequence even appears. He argues that such a definition corresponds more closely to the consensual scientific usage of the concept than does a definition which employs time sequence as the basis of the asymmetry between two variables. At the heart of Simon's notion of the causal relation is his concern with a "production" or "influence" operation. A mechanistic relation exists between two variables whereby impulses from one variable influence behavior in the other. It is this influence operation which forms the basis of the asymmetry that defines the causal relation. For example: "a thrown rock produces a broken window" is a relationship which exhibits an operational relation providing the basis of an asymmetry between two variables. While a temporal sequence is certainly involved in such a relationship, it is not the sequence which forms the basis of the asymmetry.

Yet, Hume argues (and Simon agrees) that the only relationship we can observe between two variables is a "constant conjunction" in the past (an association). Since an association is all we are able to observe, it is impossible for us to establish a necessary connection or ontological relationship between a cause and an effect. Only in a tentative sense are we able to establish a causal relationship. Hence, it makes little sense to affirm that some variable is the "true" cause of some other variable. In our example, for instance, we could never demonstrate that a thrown rock really causes a broken window. We can never dismiss the possibility that other variables exist which would "explain away" our causal relationship. Nor can we avoid the possibility that the analytic level chosen by another investigator to explain a given empirical process is different than the level we have chosen -- we may choose a "microscopic" level of analysis, while someone else may choose a "macroscopic" level. Even a controlled experiment cannot demonstrate conclusively a "real" causal relationship between two variables.

However, it is possible to view the causal relation in a manner which does not violate Hume's argument. We may take a "subjective" view of the causal relation and argue that whether or not an empirical relationship is causal depends on the context in which we make our description. While it makes little sense to argue that some process is unconditionally causal, it is nevertheless consistent with the Humean view that a relationship may be called causal in a conditional sense. When we argue that a relationship is causal, we are exhibiting a perception of (or hypothesis about) the empirical world where we have made, either explicitly or implicitly, certain ceteris paribus assertions about other empirical relations. We are making causal statements about an abstraction of the real world, not about the real world itself. For example, it is necessary to consider the empirical process we are interested in as practically isolated from the rest of the world. However, a truly isolated process is almost never encountered in the real world nor are many of the other ceteris paribus conditions which yield the context in which we call a relationship causal. But, if we are willing to assume that these ceteris paribus conditions are satisfied and confine our causal descriptions to abstractions of the empirical world, then Hume's critique becomes irrelevant.

It is for the above reasons that Simon restricts his formal definition of the causal relation to refer only to models of empirical processes rather than to the empirical processes themselves. Inspired by his implicit notion of the causal relation as an asymmetrical production operation, he formally defines the causal relation within the context of a non-stochastic "segmentable" system of linear non-homogeneous equations -- segmentable in the sense that the current predetermined variables are determined by equations which do not contain any of the current endogenous variables.[4] A paraphrase of Simon's formal definition of the direct causal relation, X causes Y, is: X directly causes Y if X appears as a predetermined variable in the equation for Y, when Y is currently endogenous in a segmentable system of linear non-homogeneous equations.[5] This formal definition provides the basis for Simon's later paper (1954), to which we now turn our attention. However, before proceeding to our discussion, we shall digress to introduce an important preliminary to the remainder of this paper.

The recursive-form linear structural model is that model in which the structural disturbances are independent of each other, and the matrix of coefficients of the endogenous variables has only zeros to one side of the main diagonal such that:

$$
\begin{array}{llcl}
 & Y_1 & = & \sum_{k=1}^{K} \gamma_{1k} X_k + u_1 \\
 & \beta_{21} Y_1 + Y_2 & = & \sum_{k=1}^{K} \gamma_{2k} X_k + u_2 \\
(1.1) & \vdots & & \vdots \qquad \vdots \\
 & \beta_{G1} Y_1 + \beta_{G2} Y_2 + \cdots Y_G & = & \sum_{k=1}^{K} \gamma_{Gk} X_k + u_G
\end{array}
$$

where $Y_1,\ldots,Y_G$ are variables endogenous to the whole system, $X_1,\ldots,X_K$ are predetermined variables, and $u_1,\ldots,u_G$ are random disturbances with

$$E(u_g) = 0, \quad E(u_g u_{g'}) = \begin{cases} \phi_{gg}, & g' = g \\ 0, & g' \neq g, \end{cases}$$

and u_g is independent of $Y_1,\ldots,Y_{g-1},X_1,\ldots,X_K$, for all g, $1 \leq g \leq G$, so that in the gth equation there is only one endogenous variable, Y_g, all the variables $Y_1,\ldots,Y_{g-1}$ being predetermined by equations preceding in the hierarchy, and $X_1,\ldots,X_K$ being predetermined for the entire model.[6] Hence, an obvious property of such a system is that influences move only in one direction: that is, from equations earlier in the hierarchy to those appearing later. Moreover, it can be shown that every equation in such a system is identified.[7] More will be said about the recursive-form structural system after we complete our discussion of Simon's contribution to theory building.

While Simon has made an important contribution to the methodology of scientific inquiry simply on the basis of his formal definition of the causal relation, by no means is this the extent of his contribution. We shall now consider Simon's concern with the derivation of necessity and sufficiency conditions for the adequacy of specific three-variable causal models.

In order to carry out his derivation, Simon employs a three-equation recursive-form linear structural model to represent the causal relations among a set of three variables. From a close examination of the properties of system (1.1), it should be obvious that the recursive-form is segmentable and an appropriate choice for the representation of causal relations as defined by Simon in his earlier piece (1953). The recursive-form linear structural model is a stochastic analogue of the non-stochastic segmentable system employed by Simon in his development of the formal definition of the causal relation. Moreover, the direct causal relation, Y_i causes Y_j, is represented in the recursive-form linear model by the non-zero structural coefficient β_{ji} (i.e. $\beta_{ji} \neq 0$).

Assuming a system of three endogenous variables, each variable being determined by a linear mechanism so that the model has three equations, Simon represents this situation as:

$$
\begin{aligned}
Y_1 + \beta_{12}Y_2 + \beta_{13}Y_3 &= u_1 \\
\beta_{21}Y_1 + \quad Y_2 + \beta_{23}Y_3 &= u_2 \\
\beta_{31}Y_1 + \beta_{32}Y_2 + \quad Y_3 &= u_3
\end{aligned}
\tag{1.2}
$$

where the normalization rule $\beta_{ii} = 1$ is used and where the u's are random disturbances with $E(u_1) = E(u_2) = E(u_3) = 0$. However, (1.2) is not segmentable (or recursive) unless additional a priori assumptions are made. As in (1.1), the matrix of coefficients of the endogenous variables (in this case the variables Y_1, Y_2, Y_3) must be triangular and the covariances of the disturbances must be zero for the system to be recursive-form (this insures the identifiability of each equation in lieu of other information).[8] Simon assumes that Y_1 is not causally dependent on Y_2 or Y_3, and Y_2 is not causally dependent on Y_3. This assertion amounts to an a priori substantive ordering of the three variables. Additionally, restricting the covariances of the disturbances to be zero so that $E(u_1u_2) = E(u_1u_3) = E(u_2u_3) = 0$, yields the following recursive-form system:

$$
\begin{aligned}
Y_1 \qquad\qquad &= u_1 \\
\beta_{21}Y_1 + \quad Y_2 \qquad &= u_2 \\
\beta_{31}Y_1 + \beta_{32}Y_2 + Y_3 &= u_3
\end{aligned}
\tag{1.3}
$$

Given the assumed substantive ordering of the three variables, the assumed independence of the disturbances, and assuming that the variables are measured from their respective means, Simon proves the following: if certain restrictions are placed on the set $\{\beta_{21}, \beta_{31}, \beta_{32}\}$ (that is, certain combinations of the elements of this set are zero), then certain conditions in terms of zero-order correlations between various pairs of variables in the model hold, and conversely. Simon then deduces that if none of these correlational conditions holds, then all three structural coefficients β_{21}, β_{31}, β_{32} are non-zero.

Alternatively, we may look at the matter from a different perspective. It can be shown that the two restrictions placed on (1.3) -- that the matrix of coefficients is triangular and that the covariances of the disturbances are zero -- are jointly necessary and sufficient to identify every equation in the model without having further information on the model. Moreover, each equation of (1.3) is just identified. Now, with additional restrictions on the set $\{\beta_{21}, \beta_{31}, \beta_{32}\}$ (that some elements of this set are zero), either the second or the third equation (or both of these equations) becomes overidentified. Thus, upon observation of the population data, Simon's correlational conditions allow us to determine which one of the possible situations of overidentification (there are $C_1^3 + C_2^3 + C_3^3 = 7$ possible situations) is consistent with the a priori just identifying restrictions on (1.3). If none of the overidentified conditions holds, then we deduce that every equation is just identified.

Perhaps the most important interpretation we can give to Simon's statistical results is that he provides a test of the existence of direct causal relations as formally defined in his earlier paper (1953). Since his correlational conditions tell us which specific structural coefficients are non-zero or zero, and since a non-zero structural coefficient defines a causal relation, we can determine which specific causal relations hold among the variables in the model (assuming a priori, of course, that some of them do not hold). With reference to (1.3), we can summarize Simon's results as follows, where ρ_{ij} is the zero-order correlation coefficient between variables Y_i and Y_j:

Given the a priori restrictions on (1.3), then

1) $\beta_{21} = 0$, $\beta_{31} \neq 0$, $\beta_{32} \neq 0$ if and only if
$\rho_{21} = 0$, $\rho_{31} \neq 0$, $\rho_{32} \neq 0$.

2) $\beta_{31} = 0$, $\beta_{21} \neq 0$, $\beta_{32} \neq 0$ if and only if
$\rho_{21} \neq 0$, $\rho_{31} \neq 0$, $\rho_{32} \neq 0$, and $\rho_{31} = \rho_{21}\rho_{32}$
(or, equivalently, $\rho_{31.2} = 0$).

3) $\beta_{32} = 0$, $\beta_{21} \neq 0$, $\beta_{31} \neq 0$ if and only if
$\rho_{21} \neq 0$, $\rho_{31} \neq 0$, $\rho_{32} \neq 0$, and $\rho_{32} = \rho_{21}\rho_{31}$
(or, equivalently, $\rho_{32.1} = 0$).

4) $\beta_{21} = \beta_{31} = 0$, $\beta_{32} \neq 0$ if and only if
$\rho_{21} = \rho_{31} = 0$, $\rho_{32} \neq 0$.

5) $\beta_{21} = \beta_{32} = 0$, $\beta_{31} \neq 0$ if and only if
$\rho_{21} = \rho_{32} = 0$, $\rho_{31} \neq 0$.

6) $\beta_{31} = \beta_{32} = 0$, $\beta_{21} \neq 0$ if and only if
$\rho_{31} = \rho_{32} = 0$, $\rho_{21} \neq 0$.

7) $\beta_{21} = \beta_{31} = \beta_{32} = 0$ if and only if
$\rho_{21} = \rho_{31} = \rho_{32} = 0$.

8) $\beta_{21} \neq 0$, $\beta_{31} \neq 0$, $\beta_{32} \neq 0$ if and only if none of the correlational conditions (1)-(7) holds.[9]

While Simon's conditions refer to population correlation coefficients, we can nevertheless approximate his conditions by employing sample correlation coefficients. However, it should be kept in mind that his results have utility only with respect to the a priori simplyfying assumptions of the model and the initial just identifying restrictions on its equations. Consequently, any practical application of Simon's results to sample data must be done with a sensitivity to the possibility that some of the a priori assertions may not actually hold (of course, this warning could be given to the use of any statistical model). However, even though Simon's results refer only to formal statistical models of the real world, the application of his procedures can lead to tentative real world inferences.

On heuristic grounds, Simon's work has appeal since it may be viewed as an attempt to explicate the assumptions and logical processes that are usually involved in making causal inferences from correlational data. When we observe a non-zero correlation coefficient between two variables and we wish to make some causal inference from this association, we ordinarily introduce a third variable if we have doubts whether the observed correlation is "genuine."[10] This third variable may account for the observed zero-order correlation and render it "spurious." In order to investigate this possibility, we compute the partial correlation coefficient between the original two variables, holding the third variable constant as a control. Comparing the computed partial correlation coefficient with the initial zero-order correlation coefficient, we attempt to make causal inferences. If the computed partial correlation coefficient is approximately zero, then we are apt to conclude either (1) that the third variable is an intervening variable, implying that the causal relationship between the original two variables is mediated by the third variable; or (2) that the third variable causes both of the original variables and thus accounts for the initially observed zero-order correlation. However, this procedure cannot tell us the causal direction in (1), let alone distinguish between (1) and (2).

We have seen above how, by moving to a formal representation of causal processes within the confines of the recursive-form, Simon makes clear what assumptions are necessary to distinguish between (1) and (2) and to determine the causal direction in (1). The correlational data alone do not allow us to decide among these possible conclusions. Not only does he point out the necessary assumptions, but his analysis also provides us with a set of correlational predictions that preserve the flavor of less rigorous causal investigations. The heuristic appeal of Simon's work, then, rests not so much on his ability to provide a test for causal theories embedded in a formal statistical model, but rather on his ability to use the algebraic relationships of the formal model to deduce a correlational test which maintains highly intuitive properties.

II. The Present Problem

We have seen that Simon's attempts to define formally the causal relation takes place exclusively within the confines of a recursive-form linear structural system of equations.[11] While we find his approach to be useful and important, it has several significant limitations. In this section of the paper, we shall be concerned with the nature of these limitations.

A. Reciprocal Causal Relations

Since the basis of Simon's implicit definition of the causal relation is the notion of an asymmetrical production relation between two variables, and not the temporal sequence of the pair, his formal definition of the concept is not entirely adequate. The choice of the recursive-form linear structural model to play a role in the formal definition of the causal relation prevents, in general, our consideration of reciprocal asymmetrical production relations. If we employ Simon's formal definition of the causal relation, it is clear that the recursive-form structural model does not permit a variable Y_i to both cause and be caused by some other variable Y_j, without the explicit introduction of temporal sequence into the system by "lagging" one or the other of the two variables.[12] This can easily be seen by noting that in the recursive-form linear structural model the matrix of coefficients of the endogenous variables must be triangular. (See (1.1)) Hence, if a structural coefficient $\beta_{ij} \neq 0$, then β_{ji} must be zero (or conversely).

Although we would argue that temporal sequences are present and ought to be explicitly included in the formal representations of such processes as symbiosis in biological systems, interpersonal attitude influence in the husband-wife dyad, or the influence relationship between constituency and legislative representative in the formation of public policy, it is not these sequences which lead us to consider the processes as causal.[13] If we are inclined to consider a process like symbiosis as a reciprocal causal process, in which we believe temporal sequence to be no more the basis for either of the two asymmetries than it is for the single asymmetry of our rock-window example, then Simon's approach to formalizing the causal relation is inadequate for our purposes. When we think of a symbiotic relationship between two biological organisms, we are most concerned with the fact that each organism asymmetrically provides products necessary for the existence and development of the other. The gist of such a relation is one of reciprocal production or influence. The temporal sequences involved do not form the bases of the mutual asymmetries we are interested in; it is the existence of a situation in which benefit is reciprocally given and received by both organisms which performs this function.

Of course, it can be argued that every causal relation, reciprocal or not, involves temporal sequences to some extent, or otherwise no change could ever take place as a consequence of the relation. Thus, it might be inferred from this argument that temporal sequences should be explicitly included in the formal representation of any process which we regard as causal. Yet, no matter how closely the concept of the causal relation may hang together in empirical situations with the concept of temporal sequence, these two principles are logically independent.[14] Simon does not argue that time is not involved in causal relations, but rather that the basis of the asymmetry of a causal relation is provided by the mechanistic or operational notion of production. Knowledge of the temporal sequence of two variables does not imply that a mechanistic relation holds between the pair. Hence, the concept of time should play no role whatsoever in either the development or the statement of a satisfactory formal definition of the causal relation.[15]

However, as already mentioned, if we admit the possibility of reciprocal causation, it is impossible to define formally the causal relation within the context of the recursive-form structural model without the explicit representation of temporal sequence. But, by the above argument, such an explicit representation would violate the logical independence of the causal relation and temporal sequence. Therefore, Simon's formal definition actually precludes the notion of reciprocal causation, even though this notion is perfectly consistent with his implicit definition of the causal relation.

In other words, we regard a reciprocal causal process as a process exhibiting two asymmetries, each of which satisfies Simon's conceptual notion of the causal relation. There is nothing contained in his implicit definition of the causal relation which excludes the possibility that two asymmetrical relations can hold between a pair of variables. Indeed, the concept of reciprocal causation does not in any way change or distort Simon's implicit definition of the causal relation. Furthermore, we feel quite comfortable with his implicit definition. The notion of reciprocal causation merely implies that two causal relations, rather than only one, are present between a pair of variables. We intend no additional meaning, either implicit or explicit, to be attached to this notion.

Like Simon, we shall restrict our formal definition of the causal relation to refer only to a representational model. However, as we have seen in the preceding discussion, the recursive-form linear structural system is inappropriate for our purposes (as is any segmentable system) if we desire to consider reciprocal causation. Since we desire to confine our interest in causal relations to the domain of a structural model, it is clear that we must have at our disposal a structural model which has the ability to represent reciprocal causal relations. After a discussion of some of the additional problems of the recursive-form, we shall consider a type of linear structural model which meets our requirements.

B. Restrictions on the Covariances of Disturbance Terms

In addition to its inability to represent reciprocal causal relations, the recursive-form linear structural system has another disadvantage: its restriction that the covariances of the disturbance terms are zero is a rather stringent a priori condition to impose upon models of many empirical processes. In practice, this restriction is usually not a consequence of the theory which gives rise to the particular structural model being considered. It is more likely to be either a consequence of the investigator's implicitly held ideas about some causal process being contained in the disturbance terms or a consequence of the desire for computational convenience. We generally regard the disturbances to be the net effects of many independently operating tiny influences that have been excluded from the structural parts of our equations. Therefore, when we assume that the disturbances of two equations are independent, we are implying that the omitted variables influencing the two equations have few elements in common. Whether or not such an assumption is plausible may depend largely on the theoretical proximity of the mechanisms described by the two equations. For example, even if we were to consider attitude influence in the husband-wife dyad to be non-reciprocal, with the husband influencing his wife, but not conversely, it would probably be rather unreasonable for us to expect that the variables omitted from a stochastic two-equation formal description of this process have no large number of elements in common. On the other hand, with respect to a stochastic two-equation economic model, where each equation describes widely separated, yet operationally linked, segments of some market, there is probably a greater likelihood that the excluded variables affecting the two equations have few common elements.[16] However, even in the latter example, we would probably be at a loss to give a theoretically meaningful defense of the assertion that the disturbances are independent.

Unfortunately, the restriction that the covariances of the disturbances are zero is often made simply because it offers a gain in the ease of estimation of the structural parameters of an equation system with a triangular matrix of coefficients. This restriction is used to aid in the identification of individual equations in such a system, and helps to guarantee the consistency of ordinary least squares estimates.[17] However, as indicated above, there is a good deal of risk associated with this restriction. Invoking this restriction becomes even more problematical if the only reason for doing so is the desire for computational convenience. This is not only due to the reasons already mentioned, but also due to the critical relationship between this restriction and the identification of equations in systems with triangular coefficient matrices.

The concept of an identified equation is generally defined in the econometric literature in terms equivalent to the following: a particular structure equation of a model is identified if that equation is the only equation, among the entire set of structure equations compatible with the data, which is also compatible with the a priori restrictions imposed by the model on that particular equation. (Essentially, an equation is identified if we have enough information to distinguish that equation from other equations in a simultaneous system.) Using this definition, it can be shown that in a linear model only those structure equations which are linear combinations of the true structure equations are the equations which satisfy the data. Moreover, an equation in such a model is identified if and only if no more than one of this set of linear combinations is consistent with the restrictions placed by the model on that particular equation.[18] When the only a priori information employed is that which specifies that certain variables are excluded from certain equations (i.e. that certain structural coefficients are zero), a necessary (but not sufficient) condition (order condition) for the identifiability of a particular equation in a linear model of G equations is that at least $G-1$ of the variables that appear in the entire model be excluded a priori from that equation. Using the same type of a priori exclusion information, a necessary and sufficient condition (rank condition) for the identifiability of a particular equation in a model of G linear equations is that it be possible to form at least one nonzero determinant of order $G-1$ from the matrix of coefficients constructed as follows: from the matrix of coefficients of the entire system, delete every column which corresponds to a variable not excluded a priori from that equation, and delete the row of coefficients of that equation.[19] (It is presumed, however, that the reader is already familiar with the concept of identification and with the order and rank conditions for the identifiability of equations in linear models, since the discussion we have just offered is certainly not sufficient to prepare the reader for all of what follows in this paper. We intend the above discussion to be more of a reminder than anything else. If the reader is not familiar with the identification problem, he should consult one of the standard works: for example, Christ (1966), Johnston (1963), or Fisher (1966).)

Even though it can be shown that every equation in a recursive-form linear structural system is identified, it is not uncommon for investigators to hold the belief that it is the triangularity of the coefficient matrix that is sufficient to identify all equations of the system. However, without additional a priori information, the triangularity of the coefficient matrix is insufficient to identify any equation of the system except the first. To see this, consider the following two-equation model, where the normalization rule $\beta_{ii} = -1$ has been used:

$$\begin{aligned} -Y_1 \qquad &= u_1 \\ \beta_{21}Y_1 - Y_2 &= u_2 \end{aligned} \tag{2.1}$$

where we assume $\beta_{12} = 0$, and all that we assume about the disturbances, u_1, u_2, is that their variance-covariance matrix, call it Σ, is positive definite. While the coefficient matrix of (2.1), call it B, is clearly triangular (since $\beta_{12} = 0$), nevertheless the second equation is not identified, since it does not satisfy the order condition for identifiability (a necessary condition). However, the first equation is identified (trivially), since it satisfies the rank condition (a necessary and sufficient condition).

Let us examine the two equations of (2.1) more closely. The matrix of coefficients of the entire system is

$$\underset{\sim}{B} = \begin{pmatrix} -1 & 0 \\ \beta_{21} & -1 \end{pmatrix} \tag{2.2}$$

For the first equation of (2.1), the matrix of coefficients relevant to the rank condition is found to be

$$(-1) \tag{2.3}$$

which has been formed by deleting from B: (a) every column not containing an a priori assumed zero in the first equation, and (b) the row of coefficients of the first equation. Now, for the first equation to satisfy the rank condition for identifiability, which is both necessary and sufficient, we must be able to form at least one non-zero determinant of order $G-1$, where G is the number of equations in the system, from matrix (2.3). Since the determinant of (2.3) is non-zero and of order $G-1 = 1$, the first equation of (2.1) is identified.

However, by the order condition for identifiability, which is necessary, the second equation of (2.1) must a priori exclude at least $G-1$ of the variables that appear in the system. For system (2.1), $G-1 = 1$. But, the second equation excludes zero of the variables appearing in the system. Hence, by the order condition, the second equation of (2.1) is not identified.

It should now be clear that the triangularity of the coefficient matrix is insufficient to identify every equation of a linear model. Moreover, in order to identify the entire system, having only the knowledge that the coefficient matrix is triangular, it is both necessary and sufficient to make the additional restriction that the covariances of the disturbances are zero. (See Fisher (1966) for a proof of this.)

Nor does the triangularity of the coefficient matrix give license to readily assume zero covariances of the disturbances, even though the hierarchical appearance of the system for some reason seems to tempt many investigators to make this assumption. The triangularity of the coefficient matrix only guarantees that perturbations in a disturbance directly influence only one of the endogeneous variables which appear in the same equation as that disturbance. The triangularity does not imply that movements in one disturbance are not associated with movements in other disturbances. Indeed, if the triangularity did imply such independence among the disturbances, then the triangularity alone would suffice to identify every equation in the model. But, we already have seen that the triangularity alone is not sufficient to identify the entire system. Therefore, it is clear that the triangularity does not imply that the disturbances are independent of each other.

Now, let us refer to our earlier example of attitude influence in the husband-wife dyad. If we regard influence in this dyad as non-reciprocal, then we could certainly represent the process in a structural system with a triangular matrix of coefficients. Yet, we already know that this triangularity does not alone suffice to identify the entire model. Let us suppose that we do not impose any a priori restrictions on the model other than those exclusion restrictions which led to the triangularity of the coefficient matrix. Since we have no other information, it is clear that we must additionally assume that the covariances of the disturbances are zero, if the entire system is to be identified. However, as indicated above, we cannot rely on the triangularity alone to permit us to assume that the disturbances are independent.

Nor, are we really convinced that it is even desirable to make such an assertion about the disturbances. We have a strong feeling that such an assertion would have rather implausible implications. If we were to make this assertion, we would be implying that the influences omitted from the structural parts of our equations, yet whose net effects are represented by the disturbances, are essentially different for each equation in the model. But, we are inclined to think that such an implication is not very plausible for the husband-wife dyad. The only alternative left to us is to make the assertion about the disturbances out of a desire for computational convenience. However, given that we already believe that the assertion is probably inappropriate, it certainly would be foolish for us to choose this last alternative.

Let us now examine how a situation of associated disturbances may affect the application of Simon's correlational testing procedure. We shall use a set of computer simulated data to illustrate such a situation. We are interested in assessing whether Simon's technique is sensitive to the "true" structural model, given a situation of associated disturbances. This illustration may shed some light on the problems associated with making the assumption that the disturbances in a triangular system are independent when, in fact, they are not.

The following structural system was used to generate a sample of 1000 cases of computer simulated data.[20]

$$\begin{aligned} -Y_1 \qquad\qquad + u_1 &= 0 \\ 5Y_1 - Y_2 \qquad + u_2 &= 0 \\ 6Y_1 \qquad - Y_3 + u_3 &= 0 \end{aligned} \tag{2.4}$$

where u_1 is a normally distributed variable with $E(u_1) = 0$ and $\sigma u_1 = 1$; and where $u_2 = u_1^2$, $u_3 = u_1^3$. While the coefficient matrix of (2.4) is triangular, we have a situation where the disturbances of this system are not independent. If we were to assume the disturbances to be independent (although, in fact, they are not) and since $\beta_{32} = 0$, Simon's procedure would predict that (2.4) is correct if and only if $\rho_{21} \neq 0$, $\rho_{31} \neq 0$, $\rho_{32} \neq 0$, and $\rho_{32.1} = 0$.

Upon observation of the correlation coefficients computed from the simulated data generated by (2.4), we find that $\rho_{32.1} \neq 0$ but rather is equal to 0.20. Therefore, Simon's method fails to detect the "true" structural model (2.4) which generated the data. This analysis suggests that to naively assume independence of structural disturbances may lead to serious consequences. The application of Simon's procedure to such a misspecified model may produce quite misleading conclusions about the configuration of causal relations within the model.

It should now be clear that we should be very cautious in choosing the recursive-form linear model to represent our theories, even when a triangular matrix of structural coefficients seems appropriate (i.e. when a complete a priori substantive ordering of the variables is theoretically justified). For unless we are willing to make the additional restriction (having no other identifiability information at hand) that the covariances of the disturbances are zero -- a restriction that is most likely not a consequence of our theory, but more likely a consequence of implicit notions of isolability (ceteris paribus) or of a desire for computational ease -- the recursive-form is not suitable. Indeed, even when we do think it suitable, its advantages usually just do not seem to outweigh its disadvantages.

If, as a consequence of our abstraction of some empirical situation, we feel that our theory can be best represented in a structural system with a triangular matrix of coefficients, it might be preferable to employ some structural system which requires weaker restrictions on the disturbance covariances, and avoid the recursive-form altogether, along with its high risks. More often than not, the theory which we intend to represent has implications only for the structural parts of our equations, and does not give rise to restrictions about the behavior of the disturbances. Fortunately, there do exist alternatives to the recursive-form for the representation of theories which inspire a triangular matrix of structural coefficients. One of these alternatives will be discussed in the following pages. Although the structural system which we shall discuss has some drawbacks, it will be seen to be superior to the recursive-form in a number of important respects.

III. An Alternative Formal Definition of the Causal Relation

A. Introductory Remarks

In the preceding sections of this paper, we have focused on a number of issues associated with the concept of causality. A large part of our discussion was devoted to a critical examination of Simon's approach to defining the causal relation. We saw that his approach is both useful and important, but that it is not entirely satisfactory. While we found his implicit notion of the causal relation (an asymmetrical influence or production relationship between two variables) to have desirable properties, we found his formal definition of this concept (confined to the recursive-form) to be inadequate for our purposes. In particular, we saw that if we admit the possibility of reciprocal causation, it is impossible to define formally the causal relation within the context of the recursive-form linear structural model without the explicit representation of temporal sequence. We argued, however, that such an explicit representation would violate the logical independence of the causal relation and temporal sequence.

In addition, we found the recursive-form to be unsatisfactory due to its rather stringent restriction on the covariances of the disturbances. We saw that the recursive-form requires the assumption that the disturbance covariances are zero. We examined this assumption in some detail and pointed out its critical role in the identification of systems with triangular coefficient matrices. While it was seen that this assumption leads to a computationally convenient situation, it was also seen that high risks are involved when it is invoked. We saw that even when a system has a triangular coefficient matrix, to assume that the disturbances are independent often has implausible implications and can lead to misleading causal conclusions (employing Simon's procedure).

Therefore, since the recursive-form is unsatisfactory in a number of important respects, we must offer some alternative structural system in which we may adequately define the causal relation and in which we may adequately represent causal theories. Such an alternative structural system must (1) have the ability to represent reciprocal causal relations, without requiring the explicit representation of temporal sequence <u>and</u> (2) allow the possibility that disturbance terms are associated. We shall now proceed to the description of one such structural system which meets our requirements. After this description, we shall then be ready to offer our formal definition of the causal relation.

B. The Simultaneous Linear Structural Model

Before we begin our description of the structural model in which we shall eventually imbed our notion of the causal relation, let us make a distinction between two types of variables. Let us call a variable which is determined in a particular model by other variables in that model an <u>endogeneous</u> variable. A variable which is not determined in that particular model will be called a <u>predetermined</u> variable.[21] While we have occasionally used this distinction in the preceding pages of this paper, we were not very precise about its meaning. However, for the model we are about to describe, this distinction plays a very crucial role. This role will become apparent in the course of our description.

Let us assume that our model refers to characteristics of a population, and not to characteristics of some sample of that population. Let us further assume that we have knowledge of the values of the variables in that population, but not of the structual coefficients. We shall treat these structural coefficients as unknowns. We shall also assume that the equations of our model are linear in the variables and unknowns. Thus, we consider the following model of G equations:[22]

$$(3.1)\qquad \begin{array}{l} y_1\beta_{11}+\cdots+y_G\beta_{1G}+z_1\gamma_{11}+\cdots+z_k\gamma_{1k}+u_1 \;\cdots \\ y_1\beta_{G1}+\cdots+y_G\beta_{GG}+z_1\gamma_{G1}+\cdots+z_k\gamma_{Gk}+u_G = 0 \end{array}$$

which may be written in matrix form as:

$$(3.2)\qquad YB + Z\Gamma + U = 0$$

where

(3.3) $Y = (y_1,\ldots,y_G)$ is the $1\times G$ row vector of endogenous variables in the model;

(3.4) $B = \begin{pmatrix} \beta_{11} & \cdots & \beta_{G1} \\ \vdots & & \vdots \\ \beta_{1G} & \cdots & \beta_{GG} \end{pmatrix}$ is the $G\times G$ matrix of coefficients (constant but unknown population parameters) of the endogenous variables;

(3.5) $Z = (z_1,\ldots,z_K)$ is the $1\times K$ row vector of predetermined variables in the model;

(3.6) $\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{G1} \\ \vdots & & \vdots \\ \gamma_{1K} & \cdots & \gamma_{GK} \end{pmatrix}$ is the $K\times G$ matrix of coefficients (constant but unknown population parameters) of the predetermined variables;

and

(3.7) $U = (u_1,\ldots,u_G)$ is the $1\times G$ row vector of the random structural disturbances.

Clearly, (3.2) contains as many equations as there are endogenous variables in the model.

<u>Assumption 3.1</u>: B is nonsingular.

Given that B is non-singular, postmultiplication of (3.2) by B^{-1} yields a solution for the values of the G endogenous variables of the model in terms of the predetermined variables and the structural disturbances. Thus, we have:

$$(3.8)\qquad Y = Z\Pi + V$$

which is the <u>reduced form</u> of (3.2) where

(3.9) $\Pi = -\Gamma B^{-1}$ is the $K\times G$ matrix of reduced form coefficients

and

(3.10) $V = -UB^{-1}$ is the $1\times G$ matrix of reduced form disturbances.

It is obvious that each equation of the reduced form (3.8) has only one endogenous variable.

<u>Assumption 3.2</u>: $E(U) = 0$

That is, each disturbance has zero expectation.

Now, denote the variance-covariance matrix of the disturbances by Σ, so that: $E(U'U) = \Sigma$, where the prime mark stands for transposition.

<u>Assumption 3.3</u>: Σ is positive definite.

<u>Assumption 3.4</u>: The predetermined variables are linearly independent.

<u>Assumption 3.5</u>: $E(Z'U) = 0$

In other words, the predetermined variables are assumed to be independent of the disturbances.

Let us now discuss some properties of the reduced form (3.8). Since the reduced form disturbances are linear combinations of the disturbances in (3.2), it can easily be shown that:

$$(3.11)\qquad E(V) = 0;$$

(3.12) $E(V'V)$, call it Ω, is positive definite;

and

$$(3.13)\qquad E(Z'V) = 0$$

Let us assume that the only kind of identifying restrictions on model (3.2) are exclusion restrictions. Suppose that extraneous <u>a priori</u> information indicates to us that certain of the y's, H in number, are permitted to appear in the <u>gth</u> equation of (3.2), and the remainder of them, $G-H$ in number, are excluded. Suppose also that certain of the z's, J in number, are permitted to appear, and the remainder of them, $K-J$, are excluded. It is always possible for the y's to be numbered in such a way that $y_1,\ldots,y_H$ are the ones that appear and $y_{H+1},\ldots,y_G$ are excluded, and for the z's to be so numbered that $z_1,\ldots,z_J$ appear and $z_{J+1},\ldots,z_K$ are excluded. Of course, this <u>a priori</u> information amounts to exact linear restrictions on the coefficients of (3.2) such that $\beta_{g,H+1} = \cdots = \beta_{gG} = \gamma_{g,J+1} = \cdots = \gamma_{g,K} = 0$. Let us further assume that each equation in (3.2) satisfies the rank condition for identifiability, which implies that the order condition must also be satisfied, i.e. $(G-H) + (K-J) \geq G-1$ or $K-J \geq H-1$ for all g, $1 \leq g \leq G$. Moreover, let us employ the normalization rule $\beta_{gg} = -1$ for all g, $1 \leq g \leq G$. Hence, the g^{th} equation of (3.2) may be written, after solving for y_g, as:

$$(3.14)\qquad y_g = y_1\beta + z_1\gamma + u_g$$

where y_1 is a $1\times H-1$ row vector of the $H-1$ endogenous variables other than y_g included in the equation, so that:

$$(3.15)\qquad y_1 = (y_1,\ldots,y_H);$$

β is an $H-1\times 1$ column vector of the coefficients of y_1, so that:

$$(3.16)\qquad \beta = \begin{pmatrix} \beta_{g1} \\ \vdots \\ \beta_{gH} \end{pmatrix};$$

z_1 is a $1\times J$ row vector of the J predetermined variables included in the equation:

$$(3.17)\qquad z_1 = (z_1,\ldots,z_J);$$

γ is a $J\times 1$ column vector of the coefficients of z_1:

$$(3.18)\qquad \gamma = \begin{pmatrix} \gamma_{g1} \\ \vdots \\ \gamma_{gJ} \end{pmatrix};$$

and u_g is the structural disturbance for the equation.

C. Defining the Causal Relation

Wold argues that a causal interpretation should not be given to the behavior relations represented by the structural coefficients of model (3.2), i.e., the elements of $\underset{\sim}{B}$ and $\underset{\sim}{\Gamma}$ (See Wold and Jureen (1953), Wold (1959, 1960), and Strotz and Wold (1960).). His argument is that such coefficients do not lend themselves to direct operative use in the sense of permitting a stimulus-response interpretation of the relations which they represent. In other words, Wold uses the concept of causality to correspond with the usual laboratory meaning of the term. The basis for this use of the concept rests largely on the notion of control. Therefore, control in a system like (3.2) is a result of the operative significance of the behavior relations of the reduced form of (3.2), i.e., (3.8). This is so because an analogy to the direct control of the stimulus in the laboratory setting can only be gained by the assumed manipulability of the predetermined variables in (3.2), whose behavioral relations are represented in the reduced form. Wold argues that the structural relations do not lend themselves to this interpretation since they contain interdependencies among the various endogenous variables in the model -- variables which are not subject to direct operational control.

However, Wold's notion of causality is much narrower than ours. While we agree that direct operative significance in the experimental sense cannot be attributed to the structural relations of (3.2), our view of causality does give causal meaning to these structural relations. Our concept of causality as an asymmetrical influence process includes more than just the notion of operational control. This should be apparent from earlier sections of this paper. We are also interested in giving causal significance to the asymmetric relations among the endogenous variables themselves. Relying only on the reduced form, as Wold would have it, tells us nothing about such structural relationships. When a structural model is constructed, it is not done so arbitrarily. It is formulated in an attempt to represent meaningful relations derived from theory and astute perception. The structural model is not just a set of arbitrary linear combinations of variables (See Goldberger (1964)). In fact, the structural model, not the reduced form, is the formulation which is inspired by the underlying theory of the model. Therefore, we consider giving causal significance to the structural relations of (3.2) entirely appropriate.

We are now ready to offer a formal definition of the causal relation. We shall do this in terms of the $g^{\underline{th}}$ equation (3.14) of model (3.2).

<u>Definition 3.1</u>: A variable is a <u>direct cause</u> of y_g if that variable is an element of either $\underset{\sim}{y}_1$ or $\underset{\sim}{z}_1$.

Therefore, a variable which is a direct cause of y_g has a (non-zero) structural coefficient which is an element of either $\underset{\sim}{\beta}$ or $\underset{\sim}{\gamma}$.

Thus, we have defined the direct causal relation within a model which admits the possibility of reciprocal causal relations, without the explicit introduction of temporal sequence. However, it should be apparent that not all variables may be involved in reciprocal causal relationships; this clearly applies to the predetermined variables in the model. While a predetermined variable may be the cause of some other variable(s) in the model, it cannot itself be caused by <u>any</u> other variable in the model. That certain variables are restricted from engaging in reciprocal causal relationships in a particular model should not, however, detract from the usefulness of Definition 3.1. It is certainly reasonable to suppose that, as a consequence of the theory which gives rise to a particular structural model, certain variables in that model are given as predetermined. Our critique of Simon's formal definition of the causal relation should not be misconstrued. We did <u>not</u> argue that all variables in a given model should enjoy the possibility of taking part in reciprocal relationships. We merely pointed out that Simon's formal definition is inadequate because it does not permit the possibility of <u>any</u> variable taking part in such a relationship.

We also noted that to assume the independence of the structural disturbances in a given model, as is required by the recursive-form, is an arbitrary decision usually not a consequence of the underlying theory and often having rather implausible implications. A similar charge might be leveled at the assumption in (3.2) that the predetermined variables are independent of all disturbances in the model. Since we can never actually observe the disturbances, we cannot obtain more observational information about the relationships between the disturbances and the variables we consider predetermined than we can about the relationships among the disturbances themselves. Yet, while both assumptions are arbitrary, one may be more arbitrary than the other. It is considerably more difficult to give theoretical justification for the assertion that variables determined in the <u>same</u> model have independent disturbances than for the assertion that variables determined <u>outside</u> of the model are independent of the disturbances of variables determined within the model. Therefore, we feel somewhat more comfortable assuming that predetermined variables are independent of the model's disturbances than we feel assuming that the model's disturbances themselves are independent.

IV. A Test of Causal Theories

A. Causal Theories

Let us suppose that (3.2) represents some particular theory. In addition, let us require that all previous assumptions made about this system still hold. Suppose further that the structural relations in (3.2) are causal relations, as defined in Definition (3.1). We have already specified that certain variables are excluded from certain equations in (3.2) -- that is, that certain elements of $\underset{\sim}{B}$ and/or $\underset{\sim}{\Gamma}$ are <u>a priori</u> assumed to be zero. Let us assume that such exclusion restrictions on the model are consequences of our theory. In other words, our theory indicates to us which variables are, and which variables are not, direct causes of the $g^{\underline{th}}$ endogeneous variable, for all g, $1 \leq g \leq G$. Moreover, we have already assumed that the exclusion restrictions placed on (3.2) are sufficient to identify each equation of the model.

However, we may not be certain that our <u>a priori</u> exclusion restrictions are correct with respect to the population. Our causal theory, represented by (3.2), may not be appropriate. Therefore, it is desirable to have some statistical test of the exclusion restrictions on the equations of (3.2). Alternatively, we may interpret such a test to be a test of the causal relations postulated in (3.2). It should be obvious that we shall be unable to provide a separate test for <u>each</u> <u>a priori</u> assumption of (3.2). We shall have to consider at least <u>some</u> of the assumptions of (3.2) to be correct without their being tested. Therefore, our statistical test of the <u>a priori</u> exclusion restrictions of (3.2) will be useful only if we are willing to <u>assume</u> that certain other assumptions of (3.2) are valid. In fact, it will be seen that not even every exclusion restriction on (3.2) will be tested. At least some of these restrictions will have to be assumed to be correct. Hence, our statistical test of the causal relations of (3.2) will allow us to make only <u>conditional</u> statements about the adequacy of the model. But, this should not bother us. We have already argued, much earlier in this paper, that it makes little sense ever to state that a relationship is unconditionally causal.

Even though a number of tests of identifying restrictions may be found in the econometric literature (for example, tests developed by Anderson and Rubin (1950) and by Basmann (1960); also see Christ (1966), Chapter X.), none of these tests has the heuristic appeal of the correlational test developed for the recursive-form by Simon. While it is clear that Simon's testing procedure is inappropriate for non-recursive systems, it would be desirable to develop a test which has some of the intuitive characteristics associated with his procedure. Particularly, it would be desirable to develop a correlational test which in some sense reflects the logical process associated with introducing control variables into a zero-order correlational relationship and examining the partial correlations as a test for "spuriousness." We earlier argued that Simon's procedure is appealing precisely because it reflects this logical process, even though the procedure was deduced from a formal statistical model. Therefore, it is hoped that the results of the derivation that follows will also reflect this process.

B. Derivation of the Statistical Test

Without loss of generality, let us assume that all variables in (3.2) have zero means. We still assume that we are referring to the population and not to a sample of that population. Employing our normalization rule, we again consider the $g^{\underline{th}}$ equation of (3.2):

(4.1) $$y_g = \underset{\sim}{y}_1\underset{\sim}{\beta} + \underset{\sim}{z}_1\underset{\sim}{\gamma} + u_g$$

Let

(4.2) $\underset{\sim}{z}_2 = (z_{J+1},\ldots,z_K)$ be the $1\times K\text{-}J$ row vector of the predetermined variables excluded from the $g^{\underline{th}}$ equation.

Since $\underset{\sim}{z}_1$ is the $1\times J$ row vector of predetermined variables included in the $g^{\underline{th}}$ equation, we may write the reduced form of $\underset{\sim}{y}_1$ as:

(4.3) $$\underset{\sim}{y}_1 = \underset{\sim}{z}_1\underset{\sim}{\pi}^{11} + \underset{\sim}{z}_2\underset{\sim}{\pi}^{12} + \underset{\sim}{v}_1$$

where

(4.4) $\underset{\sim}{\pi}^{11}$ is the $J\times H\text{-}1$ matrix of reduced form coefficients of $\underset{\sim}{z}_1$,

(4.5) $\underset{\sim}{\pi}^{12}$ is the $K\text{-}J\times H\text{-}1$ matrix of reduced form coefficients of $\underset{\sim}{z}_2$,

and

(4.6) $\underset{\sim}{v}_1$ is the $1\times H\text{-}1$ row vector of reduced form disturbances for $\underset{\sim}{y}_1$.

Now, let us define:

(4.7) $$y_1^* = y_1 - v_1 = z_1\pi^{11} + z_2\pi^{12}$$

Solving (4.7) for y_1 we obtain:

(4.8) $$y_1 = y_1^* + v_1$$

Substituting (4.8) into (4.1) for y_1 we have:

(4.9) $$y_g = y_1^*\beta + z_1\gamma + (u_g + v_1\beta)$$

Since the elements of y_1^* are linear combinations of the original predetermined variables in (3.2), we have, using Assumption (3.5) and (3.13):

(4.10) $$E(y_1^{*\prime}U) = 0$$

and

(4.11) $$E(y_1^{*\prime}V) = 0$$

Therefore, we may consider the elements of y_1^* to be predetermined variables. Let

(4.12) $y_2 = (y_{H+1},\ldots,y_G)$ be the $1 \times G\text{-}H$ row vector of the $G\text{-}H$ endogenous variables excluded from the g^{th} equation.

We may write the reduced form of y_2 as:

(4.13) $$y_2 = z_1\pi^{21} + z_2\pi^{22} + v_2$$

As with y_1^*, we define:

(4.14) $$y_2^* = y_2 - v_2 = z_1\pi^{21} + z_2\pi^{22}$$

and we obtain:

(4.15) $$E(y_2^{*\prime}U) = 0$$

and

(4.16) $$E(y_2^{*\prime}V) = 0$$

Hence, we may consider the elements of y_2^* to be predetermined variables. Therefore, we have as predetermined variables the elements of

(4.17) $$z_1,\ z_2,\ y_1^*,\ \text{and}\ y_2^*.$$

So, let us define:

(4.18) $$I = \begin{pmatrix} y_1^{*\prime} \\ z_1{}' \\ y_2^{*\prime} \\ z_2{}' \end{pmatrix}$$ to be the $G\text{-}1\text{+}K \times 1$ column vector of instrumental variables formed so that the predetermined variables appearing in the right-hand side of (4.9) come first, and the predetermined variables not appearing in (4.9) come last.

Now, let us partition I so that

(4.19) $$I = \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}$$

where

(4.20) $$I_1 = \begin{pmatrix} y_1^{*\prime} \\ z_1{}' \end{pmatrix} \quad \text{and} \quad I_2 = \begin{pmatrix} y_2^{*\prime} \\ z_2{}' \end{pmatrix}$$

Rewriting (4.9), we obtain:

(4.21) $$y_g = (y_1^* \ \ z_1)\begin{pmatrix}\beta \\ \gamma\end{pmatrix} + (u_g + v_1\beta)$$

Premultiplying (4.21) by (4.19), we have:

(4.22) $$\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} y_g = \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}(y_1^* \ \ z_1)\begin{pmatrix}\beta \\ \gamma\end{pmatrix} + \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} u_g + \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} v_1\beta$$

which may be written as:

(4.23) $$\begin{pmatrix} I_1 y_g \\ I_2 y_g \end{pmatrix} = \begin{pmatrix} I_1 y_1^* & I_1 z_1 \\ I_2 y_1^* & I_2 z_1 \end{pmatrix}\begin{pmatrix}\beta \\ \gamma\end{pmatrix} + \begin{pmatrix} I_1 u_g \\ I_2 u_g \end{pmatrix} + \begin{pmatrix} I_1 v_1\beta \\ I_2 v_1\beta \end{pmatrix}$$

Taking the expected value of (4.23), we have:

(4.24) $$E\begin{pmatrix} I_1 y_g \\ I_2 y_g \end{pmatrix} = E\left[\begin{pmatrix} I_1 y_1^* & I_1 z_1 \\ I_2 y_1^* & I_2 z_1 \end{pmatrix}\begin{pmatrix}\beta \\ \gamma\end{pmatrix}\right] + E\begin{pmatrix} I_1 u_g \\ I_2 u_g \end{pmatrix} + E\begin{pmatrix} I_1 v_1\beta \\ I_2 v_1\beta \end{pmatrix}$$

and this gives us:

(4.25) $$E\begin{pmatrix} I_1 y_g \\ I_2 y_g \end{pmatrix} = E\left[\begin{pmatrix} I_1 y_1^* & I_1 z_1 \\ I_2 y_1^* & I_2 z_1 \end{pmatrix}\begin{pmatrix}\beta \\ \gamma\end{pmatrix}\right],$$

since $E\begin{pmatrix} I_1 u_g \\ I_2 u_g \end{pmatrix} = E\begin{pmatrix} I_1 v_1\beta \\ I_2 v_1\beta \end{pmatrix} = 0$, by the fact that the elements of I_1 and I_2 are predetermined variables. Rewriting (4.25) we get:

(4.26) $$E\begin{pmatrix} I_1 y_g \\ I_2 y_g \end{pmatrix} = E\begin{pmatrix} I_1 y_1^* & I_1 z_1 \\ I_2 y_1^* & I_2 z_1 \end{pmatrix}\begin{pmatrix}\beta \\ \gamma\end{pmatrix},$$

which is a system of $G\text{-}1\text{+}K$ equations in $H\text{-}1\text{+}J$ unknowns (unknown structural coefficients).

Assume Equation (4.1) is correct. Thus, all $G\text{-}1\text{+}K$ equations of (4.26) are satisfied by the true population parameter vector $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$, since all of these equations are consequences of (4.1). However, $G\text{-}1\text{+}K > H\text{-}1\text{+}J$; that is, (4.26) has more equations than unknown structural parameters. Therefore, there is a possibility of extracting additional information if we substitute the true vector $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$ into system (4.26). Indeed, we might find necessary conditions for (4.1) to be correct. Yet, the true parameter vector $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$ is unknown to us, by (3.4). Hence, in order to investigate the possibility of extracting such additional information, we must find some way of obtaining the true parameter vector $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$ from the information we have at hand.

Since the true parameter vector $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$ satisfies the entire system (4.26), it is a solution to any subsystem of (4.26). In particular, if we choose any subsystem of $H\text{-}1\text{+}J$ equations from (4.26), we get at least one solution (the true parameter vector). For various reasons, it is possible to extract a particular subsystem of $H\text{-}1\text{+}J$ equations from (4.26) such that one of the solutions has as its expected value the true parameter vector and such that the matrix of coefficients (of this solution) is nonsingular; that is, invertible. Therefore, if we substitute this obtained solution into system (4.26), we should be able to extract the same (in an expectational sense) additional information as if we were to substitute the true parameter vector $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$ into (4.26).

Now, let us examine this solution to (4.26) which has as its expected value the true parameter vector $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$.

Let us partition (4.26) into two subsystems so that one subsystem is

(4.27) $$E(I_1 y_g) = E(I_1 y_1^* \ \ I_1 z_1)\begin{pmatrix}\beta \\ \gamma\end{pmatrix},$$

a system of $H\text{-}1\text{+}J$ equations in $H\text{-}1\text{+}J$ unknowns, and the other subsystem is

(4.28) $$E(I_2 y_g) = E(I_2 y_1^* \ \ I_2 z_1)\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$$

a system of $G\text{-}H\text{+}K\text{-}J$ equations in $H\text{-}1\text{+}J$ unknowns. It is clear that this partitioning corresponds to the distinction initially made in (4.19) and (4.20) when we partitioned I. That is, system (4.27) is the system of equations derived from (4.9) by using as instruments those predetermined variables appearing in (4.9) (i.e., the elements of I_1, $I_1 = \begin{pmatrix} y_1^{*\prime} \\ z_1{}' \end{pmatrix}$) and system (4.28) is the system derived from (4.9) by using as instruments those predetermined variables not appearing in (4.9) (i.e., the elements of I_2, $I_2 = \begin{pmatrix} y_2^{*\prime} \\ z_2{}' \end{pmatrix}$).

If we solve (4.27) for $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$, we obtain a solution which has as its expected value the true parameter vector.[23] $E(I_1 y_1^* \ \ I_1 z_1)$ is a square matrix of order $H\text{-}1\text{+}J$. By Assumption (3.4), $E(I_1 y_1^* \ \ I_1 z_1)$ exists. By the assumption that (4.1) is identified, $E(I_1 y_1^* \ \ I_1 z_1)$ is nonsingular. Therefore, the inverse of $E(I_1 y_1^* \ \ I_1 z_1)$ exists. Hence, by Cramer's rule, we solve (4.27) for $\begin{pmatrix}\beta \\ \gamma\end{pmatrix}$ and we get:

(4.29) $$\begin{pmatrix}\beta \\ \gamma\end{pmatrix} = [E(I_1 y_1^* \ \ I_1 z_1)]^{-1} E(I_1 y_g),$$

a unique solution to the subsystem (4.27).

To show that (4.29) has as its expected value the true parameter vector $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$, let us denote (4.29) by $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}''$. We define the error of (4.29) to be the difference between (4.29) and the true parameter vector $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$, so that:

$$e = \begin{pmatrix}\beta\\ \gamma\end{pmatrix}'' - \begin{pmatrix}\beta\\ \gamma\end{pmatrix}, \tag{4.30}$$

where e is the error of $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}''$. Our task is to show that $E(e) = 0$.

Substitution of (4.9) into (4.29) for y_g yields:

$$\begin{pmatrix}\beta\\ \gamma\end{pmatrix}'' = [E(I_1y_1^* \;\; I_1z_1)]^{-1}E\left\{I_1[(y_1^* \;\; z_1)\begin{pmatrix}\beta\\ \gamma\end{pmatrix} + u_g + v_1\beta]\right\} \tag{4.31}$$

$$\begin{aligned} &= [E(I_1y_1^* \;\; I_1z_1)]^{-1}E(I_1I_1')\begin{pmatrix}\beta\\ \gamma\end{pmatrix} \\ &\quad + [E(I_1y_1^* \;\; I_1z_1)]^{-1}E(I_1u_g) \\ &\quad + [E(I_1y_1^* \;\; I_1z_1)]^{-1}E(I_1v_1\beta) \end{aligned} \tag{4.32}$$

$$\begin{aligned} &= [E(I_1y_1^* \;\; I_1z_1)]^{-1}E(I_1y_1^* \;\; I_1z_1)\begin{pmatrix}\beta\\ \gamma\end{pmatrix} \\ &\quad + [E(I_1y_1^* \;\; I_1z_1)]^{-1}E(I_1u_g) \\ &\quad + [E(I_1y_1^* \;\; I_1z_1)]^{-1}E(I_1v_1\beta) \end{aligned} \tag{4.33}$$

$$\begin{aligned} &= \begin{pmatrix}\beta\\ \gamma\end{pmatrix} + [E(I_1y_1^* \;\; I_1z_1)]^{-1}E(I_1u_g) \\ &\quad + [E(I_1y_1^* \;\; I_1z_1)]^{-1}E(I_1v_1\beta) \end{aligned} \tag{4.34}$$

$$= \begin{pmatrix}\beta\\ \gamma\end{pmatrix} + 0 + 0; \tag{4.35}$$

by Assumption (3.5) and Equations (3.13), (4.10), and (4.11). Hence $E(e) = 0$. Therefore, $E\begin{pmatrix}\beta\\ \gamma\end{pmatrix}'' = \begin{pmatrix}\beta\\ \gamma\end{pmatrix}$.

Earlier we pointed out that, by substituting an appropriate solution into (4.26), we should be able to extract some additional information. Instead of substituting $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}''$ directly into (4.26), we will substitute it into another system, each of which can be derived from the other.

Let us define

$$\sigma = \begin{pmatrix} \sigma_{y_1^*} & 0 & \cdots & & & & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & \sigma_{y_H^*} & & & & \vdots \\ \vdots & & & \sigma_{z_1} & & & \vdots \\ \vdots & & & & \ddots \; \sigma_{z_J} & & \vdots \\ \vdots & & & & & \sigma_{y_{H+1}^*} & 0 \\ 0 & \cdots & & & 0 & \ddots & \sigma_{z_K} \end{pmatrix} \tag{4.36}$$

to be the $G\text{-}1\text{+}K \times G\text{-}1\text{+}K$ diagonal matrix of population standard deviations of the elements of I. We see that

$$\sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}, \tag{4.37}$$

where σ_1 is the $H\text{-}1\text{+}J \times H\text{-}1\text{+}J$ diagonal matrix of population standard deviations of the elements of I_1 and σ_2 is the corresponding $G\text{-}H\text{+}K\text{-}J \times G\text{-}H\text{+}K\text{-}J$ matrix for the elements of I_2. Similarly, let σ_g be the population standard deviation of y_g.

Since none of the diagonal elements of σ is zero (assuming none of our variables degenerates to a constant), σ is invertible. Therefore, σ_1 is also invertible, as well as the scalar σ_g. Premultiplying system (4.26) by σ^{-1}, postmultiplying by σ_g^{-1}, and noting that $\sigma_1^{-1}\sigma_1$ is the identity matrix, we have:

$$\sigma^{-1}E\begin{pmatrix} I_1y_g \\ I_2y_g \end{pmatrix}\sigma_g^{-1} = \sigma^{-1}E\begin{pmatrix} I_1y_1^* & I_1z_1 \\ I_2y_1^* & I_2z_1 \end{pmatrix}\sigma_1^{-1}\sigma_1\begin{pmatrix}\beta\\ \gamma\end{pmatrix}\sigma_g^{-1} \tag{4.38}$$

Since a diagonal matrix is symmetric, and since we assume that all variables in (3.2) have zero means, and since the elements of y_1^* and y_2^* have zero means (since they are linear combinations of variables with zero means), (4.38) becomes:

$$\begin{pmatrix} P_{I_1y_g} \\ P_{I_2y_g} \end{pmatrix} = \begin{pmatrix} P_{I_1y_1^*} & P_{I_1z_1} \\ P_{I_2y_1^*} & P_{I_2z_1} \end{pmatrix}\sigma_1\begin{pmatrix}\beta\\ \gamma\end{pmatrix}\sigma_g^{-1} \tag{4.39}$$

where

$$P_{I_1y_g} = \begin{pmatrix} \rho_{y_1^*y_g} \\ \vdots \\ \rho_{y_H^*y_g} \\ \rho_{z_1y_g} \\ \vdots \\ \rho_{z_Jy_g} \end{pmatrix} \tag{4.40}$$

is the $H\text{-}1\text{+}J \times 1$ column vector of population correlation coefficients between the elements of I_1 and y_g;

$$P_{I_2y_g} = \begin{pmatrix} \rho_{y_{H+1}^*y_g} \\ \vdots \\ \rho_{y_G^*y_g} \\ \rho_{z_{J+1}y_g} \\ \vdots \\ \rho_{z_Ky_g} \end{pmatrix} \tag{4.41}$$

is the $G\text{-}H\text{+}K\text{-}J \times 1$ column vector of population correlation coefficients between the elements of I_2 and y_g;

$$P_{I_1y_1^*} = \begin{pmatrix} \rho_{y_1^*y_1^*} & \cdots & \rho_{y_1^*y_H^*} \\ \vdots & & \vdots \\ \rho_{y_H^*y_1^*} & \cdots & \rho_{y_H^*y_H^*} \\ \rho_{z_1y_1^*} & \cdots & \rho_{z_1y_H^*} \\ \vdots & & \vdots \\ \rho_{z_Jy_1^*} & \cdots & \rho_{z_Jy_H^*} \end{pmatrix} \tag{4.42}$$

is the $H\text{-}1\text{+}J \times H\text{-}1$ matrix of population correlation coefficients between the elements of I_1 and the elements of y_1^*; similarly,

(4.43) $P_{I_1z_1}$ is the $H\text{-}1\text{+}J \times J$ matrix of population correlation coefficients between the elements of I_1 and z_1;

(4.44) $P_{I_2y_1^*}$ is the $G\text{-}H\text{+}K\text{-}J \times H\text{-}1$ matrix of population correlation coefficients between the elements of I_2 and y_1^*;

and

(4.45) $P_{I_2z_1}$ is the $G\text{-}H\text{+}K\text{-}J$ matrix of population correlation coefficients between the elements of I_2 and z_1.

System (4.39) is the above mentioned system into which we will substitute $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}''$.

We will now examine the necessary and sufficient conditions for (4.1) to be correct. We have assumed (4.1) to be identified; it may be just identified (i.e., $K\text{-}J = H\text{-}1$) or overidentified (i.e., $K\text{-}J > H\text{-}1$). If (4.1) is just identified, certain (necessary) conditions will hold. If (4.1) is overidentified, then we will extract a suitable amount of information such that we can treat (4.1) to be just identified, on the basis of only this information. We will then find necessary and sufficient conditions for the "left-over" information to hold. It is this latter case in which we are most interested.

Case I. Assume (4.1) is just identified (i.e., $K\text{-}J = H\text{-}1$). Let us define

(4.46) $\beta^2 = \begin{pmatrix} \beta_{g,H+1} \\ \vdots \\ \beta_{gG} \end{pmatrix}$ to be the $G\text{-}H \times 1$ column vector of coefficients of the elements of y_2 (the endogeneous variables excluded from (4.1))

and

$$(4.47)\qquad \underset{\sim}{\gamma}^2 = \begin{pmatrix}\gamma_{g,J+1}\\ \vdots \\ \gamma_{gK}\end{pmatrix} \quad \text{to be the } K\text{-}J \times 1 \text{ column vector of}$$

coefficients of the elements of $\underset{\sim}{z}_2$ (the predetermined variables excluded from (4.1)).

Furthermore, let

$$(4.48)\qquad \underset{\sim}{Q} = \begin{pmatrix}\underset{\sim}{\beta}^2\\ \underset{\sim}{\gamma}^2\end{pmatrix}$$

Let us also define

$$(4.49)\qquad \underset{\sim}{P}_{\underset{\sim}{I}_2 y_g.\underset{\sim}{I}_1} = \begin{pmatrix}\rho_{y_{H+1}y_g.y_1^*\cdots y_H^* z_1\cdots z_J}\\ \vdots \\ \rho_{y_G y_g.y_1^*\cdots y_H^* z_1\cdots z_J}\\ \rho_{z_{J+1}y_g.y_1^*\cdots y_H^* z_1\cdots z_J}\\ \vdots \\ \rho_{z_K y_g.y_1^*\cdots y_H^* z_1\cdots z_J}\end{pmatrix}$$

to be the $G\text{-}H\text{+}K\text{-}J \times 1$ vector of partial correlation coefficients between the elements of $\underset{\sim}{I}_2$ and y_g, controlling for every element of $\underset{\sim}{I}_1$. Let us state the following theorem:

Theorem 4.1: If $\underset{\sim}{Q} = \underset{\sim}{0}$, then $\underset{\sim}{P}_{\underset{\sim}{I}_2 y_g.\underset{\sim}{I}_1} = \underset{\sim}{0}$.

Essential to the proof of Theorem 4.1 is the following lemma about partial correlation coefficients. Let us state, without proof, this easily verified lemma.

Lemma 4.1: $\rho_{12.34\ldots k} = 0$ if and only if $\rho_{12} = B_{23}\rho_{13} + B_{24}\rho_{14} + \cdots + B_{2k}\rho_{1k}$ where $B_{ij} = \beta_{ij}\frac{\sigma_j}{\sigma_i}$.

Proof of Theorem 4.1: By the assumption that $\underset{\sim}{Q} = \underset{\sim}{0}$, (4.1) is identified. Therefore, $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$" exists. Substitution of $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$" into (4.39) for $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$ yields a set of relations which, by Lemma 4.1, implies that $\underset{\sim}{P}_{\underset{\sim}{I}_2 y_g.\underset{\sim}{I}_1} = \underset{\sim}{0}$. Q.E.D.

We see that when (4.1) is just identified $\underset{\sim}{Q} = \underset{\sim}{0}$, since both $\underset{\sim}{\beta}^2$ and $\underset{\sim}{\gamma}^2$ are equal to zero, by the assumption that the elements of $\underset{\sim}{y}_2$ and $\underset{\sim}{z}_2$ are excluded from (4.1). Therefore, by Theorem 4.1, for the just identified case, $\underset{\sim}{P}_{\underset{\sim}{I}_2 y_g.\underset{\sim}{I}_1} = \underset{\sim}{0}$.

However, Theorem 4.1 does not really provide us with a test (i.e., necessary and sufficient conditions) of the correctness of the identifying restrictions on equation (4.1). We have merely provided a necessary condition for these restrictions to be correct. This is not a sufficient condition for the correctness of these restrictions. Indeed, if the restrictions on (4.1) were incorrect (i.e., if some elements of $\underset{\sim}{Q}$ were nonzero), then (4.1) would be underidentified. Therefore, $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$" would not exist, since $E(\underset{\sim}{I}_1\underset{\sim}{y}_1^* \;\; \underset{\sim}{I}_1\underset{\sim}{z}_1)$ would be singular and, hence, not invertible. Thus, using our method, we cannot deduce any sufficient conditions for the identifying restrictions to be correct in the just identified case.

Case II. Assume (4.1) is overidentified (i.e., K-J > H-1).

We have seen that our method cannot provide a test of the correctness of the a priori exclusion restrictions on (4.1) if (4.1) is just identified. However, if (4.1) is overidentified, then our method does provide a test for some of the a priori exclusion restrictions on (4.1) to be correct. If (4.1) is overidentified, then there are more exclusion restrictions on (4.1) than are needed to insure its identifiability. Therefore, if we choose a subset of these a priori exclusion restrictions -- just enough to guarantee the identifiability (4.1) -- and consider the restrictions contained in this subset to be untestable, then we may consider the remainder of the exclusion restrictions on (4.1) to be testable. That is, by our method, we are able to find necessary and sufficient conditions for the correctness of these overidentifying (extra) restrictions, given that our chosen subset is assumed to be sufficient to identify (4.1).[24]

Let us assume that, for theoretical reasons, we feel that the restrictions contained in our chosen subset have a stronger basis than the remainder of the restrictions on (4.1). Let us additionally assume that even if our test rejects the validity of all of the overidentifying restrictions on (4.1), the identifiability of all other equations in (3.2) is unaffected. Renumbering if necessary, we can order the variables excluded from (4.1) (i.e., the elements of $\underset{\sim}{y}_2$ and $\underset{\sim}{z}_2$) so that:

$$(4.57)\qquad \underset{\sim}{y}_2 = (y_{H+1},\ldots,y_{H'},y_{H'+1},\ldots,y_G)$$

and

$$(4.58)\qquad \underset{\sim}{z}_2 = (z_{J+1},\ldots,z_{J'},z_{J'+1},\ldots,z_K),$$

where the exclusion from (4.1) of the set of variables $\{y_{H'+1},\ldots,y_G,z_{J'+1},\ldots,z_K\}$ suffices to identify (4.1). The number of elements in this set is $G\text{-}H'\text{+}K\text{-}J' = G\text{-}1$ (since an equation in a linear system of G equations must exclude at least $G\text{-}1$ of the variables that appear in the entire system in order to be identified and exactly $G\text{-}1$ to be just identified). Let us assume that the exclusion of this set of variables is taken as unquestioned (for theoretical reasons). Let the exclusion of the set of variables $\{y_{H+1},\ldots,y_{H'},z_{J+1},\ldots,z_{J'}\}$ from (4.1) be the overidentifying restrictions on (4.1). The number of elements in this set is $H'\text{-}H\text{+}J'\text{-}J$. Since the exclusion of the variables contained in the set $\{y_{H'+1},\ldots,y_G,z_{J'+1},\ldots,z_K\}$ suffices to identify that equation, the exclusion of the variables of the set $\{y_{H+1},\ldots,y_{H'},z_{J+1},\ldots,z_{J'}\}$ is a testable assumption.

Now, partition $\underset{\sim}{y}_2$ and $\underset{\sim}{z}_2$ so that

$$(4.59)\qquad \underset{\sim}{y}_2 = (\underset{\sim}{y}_2^1 \;\; \underset{\sim}{y}_2^2)$$

where $\underset{\sim}{y}_2^1 = (y_{H+1},\ldots,y_{H'})$

$\underset{\sim}{y}_2^2 = (y_{H'+1},\ldots,y_G)$

and

$$(4.60)\qquad \underset{\sim}{z}_2 = (\underset{\sim}{z}_2^1 \;\; \underset{\sim}{z}_2^2)$$

where $\underset{\sim}{z}_2^1 = (z_{J+1},\ldots,z_{J'})$

$\underset{\sim}{z}_2^2 = (z_{J'+1},\ldots,z_K)$

Corresponding to the distinction just made and recalling that $\underset{\sim}{I} = \begin{pmatrix}\underset{\sim}{I}_1\\ \underset{\sim}{I}_2\end{pmatrix}$, let us partition $\underset{\sim}{I}_2$ so that

$$(4.61)\qquad \underset{\sim}{I}_2 = \begin{pmatrix}\underset{\sim}{I}_2^1\\ \underset{\sim}{I}_2^3\\ \underset{\sim}{I}_2^2\\ \underset{\sim}{I}_2^4\end{pmatrix}.$$

where $\underset{\sim}{I}_2^1 = (\underset{\sim}{y}_2^{1*})$, $\underset{\sim}{I}_2^3 = (\underset{\sim}{y}_2^{2*})$, $\underset{\sim}{I}_2^2 = (\underset{\sim}{z}_2^1)$, and $\underset{\sim}{I}_2^4 = (\underset{\sim}{z}_2^2)$.

Likewise, we partition (4.39), so that:

$$(4.62)\qquad \begin{pmatrix}\underset{\sim}{P}_{\underset{\sim}{I}_1 y_g}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^1 y_g}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^3 y_g}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^2 y_g}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^4 y_g}\end{pmatrix} = \begin{pmatrix}\underset{\sim}{P}_{\underset{\sim}{I}_1\underset{\sim}{y}_1^*} & \underset{\sim}{P}_{\underset{\sim}{I}_1\underset{\sim}{z}_1}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^1\underset{\sim}{y}_1^*} & \underset{\sim}{P}_{\underset{\sim}{I}_2^1\underset{\sim}{z}_1}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^3\underset{\sim}{y}_1^*} & \underset{\sim}{P}_{\underset{\sim}{I}_2^3\underset{\sim}{z}_1}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^2\underset{\sim}{y}_1^*} & \underset{\sim}{P}_{\underset{\sim}{I}_2^2\underset{\sim}{z}_1}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^4\underset{\sim}{y}_1^*} & \underset{\sim}{P}_{\underset{\sim}{I}_2^4\underset{\sim}{z}_1}\end{pmatrix} \underset{\sim}{\sigma}_1 \begin{pmatrix}\beta\\ \gamma\end{pmatrix} \sigma_g^{-1}$$

We can rearrange (4.62) into two subsystems so that one subsystem is:

$$(4.63)\qquad \begin{pmatrix}\underset{\sim}{P}_{\underset{\sim}{I}_1 y_g}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^1 y_g}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^2 y_g}\end{pmatrix} = \begin{pmatrix}\underset{\sim}{P}_{\underset{\sim}{I}_1\underset{\sim}{y}_1^*} & \underset{\sim}{P}_{\underset{\sim}{I}_1\underset{\sim}{z}_1}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^1\underset{\sim}{y}_1^*} & \underset{\sim}{P}_{\underset{\sim}{I}_2^1\underset{\sim}{z}_1}\\ \underset{\sim}{P}_{\underset{\sim}{I}_2^2\underset{\sim}{y}_1^*} & \underset{\sim}{P}_{\underset{\sim}{I}_2^2\underset{\sim}{z}_1}\end{pmatrix} \underset{\sim}{\sigma}_1 \begin{pmatrix}\beta\\ \gamma\end{pmatrix} \sigma_g^{-1},$$

a system of $H'\text{-}1\text{+}J'$ equations in $H\text{-}1\text{+}J$ unknowns and the other subsystem is

$$(4.64)\qquad \begin{pmatrix} P_{\tilde{I}_2^3 y_g} \\ P_{\tilde{I}_2^4 y_g} \end{pmatrix} = \begin{pmatrix} P_{\tilde{I}_2^3 \tilde{y}_1^*} & P_{\tilde{I}_2^3 \tilde{z}_1} \\ P_{\tilde{I}_2^4 \tilde{y}_1^*} & P_{\tilde{I}_2^4 \tilde{z}_1} \end{pmatrix} \tilde{\sigma}_1 \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \sigma_g^{-1},$$

a system of $G-H'+K-J'$ equations in $H-1+J$ unknowns, where (4.63) is the subsystem of (4.62) which corresponds to those equations derived from (4.9) by using as instruments the predetermined variables appearing in (4.9) (i.e., the elements of $\tilde{I}_1$) and the predetermined variables which are elements of $\tilde{I}_2^1$ and $\tilde{I}_2^2$, and where (4.64) is the subsystem of (4.62) which corresponds to those equations derived from (4.9) by using as instruments the predetermined variables of $\tilde{I}_2^3$ and $\tilde{I}_2^4$. Since the elements of $\tilde{I}_2^3$ and $\tilde{I}_2^4$ correspond to the set of variables $\{y_{H'+1},\ldots,y_G, z_{J'+1},\ldots,z_K\}$ whose exclusion from (4.1) is taken as unquestioned, we shall ignore (4.64) and concern ourselves only with (4.63).

Define

$$(4.65)\qquad Q = \begin{pmatrix} \tilde{\beta}_2^1 \\ \tilde{\gamma}_2^1 \end{pmatrix}$$

to be a $H'-H+J'-J \times 1$ column vector, where

$$(4.66)\qquad \tilde{\beta}_2^1 = \begin{pmatrix} \beta_{g,H+1} \\ \vdots \\ \beta_{gH'} \end{pmatrix}$$

is the $H'-H \times 1$ column vector of coefficients of the elements of $\tilde{y}_2^1$ and

$$(4.67)\qquad \tilde{\gamma}_2^1 = \begin{pmatrix} \gamma_{g,J+1} \\ \vdots \\ \gamma_{gJ'} \end{pmatrix},$$

is the $J'-J \times 1$ column vector of coefficients of the elements of $\tilde{z}_2^1$. Let

$$(4.68)\qquad \begin{pmatrix} P_{\tilde{I}_2^1 y_g.\tilde{I}_1} \\ P_{\tilde{I}_2^2 y_g.\tilde{I}_1} \end{pmatrix} = \begin{pmatrix} \rho_{y_{H+1}^* y_g.y_1^*\ldots y_H^* z_1\ldots z_J} \\ \vdots \\ \rho_{y_{H'}^* y_g.y_1^*\ldots y_H^* z_1\ldots z_J} \\ \rho_{z_{J+1} y_g.y_1^*\ldots y_H^* z_1\ldots z_J} \\ \vdots \\ \rho_{z_{J'} y_g.y_1^*\ldots y_H^* z_1\ldots z_J} \end{pmatrix}$$

be the $H'-H+J'-J \times 1$ vector of partial correlation coefficients between the elements of $\begin{pmatrix} \tilde{I}_2^1 \\ \tilde{I}_2^2 \end{pmatrix}$ and y_g, controlling for every element of $\tilde{I}_1$.

Consider the following theorem:

Theorem 4.2: $\tilde{Q} = \tilde{0}$ if and only if

$$\begin{pmatrix} P_{\tilde{I}_2^1 y_g.\tilde{I}_1} \\ P_{\tilde{I}_2^2 y_g.\tilde{I}_1} \end{pmatrix} = \tilde{0}.$$

Proof:

Part I: If $\tilde{Q} = \tilde{0}$ then $\begin{pmatrix} P_{\tilde{I}_2^1 y_g.\tilde{I}_1} \\ P_{\tilde{I}_2^2 y_g.\tilde{I}_1} \end{pmatrix} = \tilde{0}.$

Proof of Part I: By assumption, $\tilde{Q} = \tilde{0}$. Therefore, $\tilde{\beta}_2^1 = \tilde{0}$ and $\tilde{\gamma}_2^1 = \tilde{0}$, which implies that the elements of $\tilde{y}_2^1$ and $\tilde{z}_2^1$ do not appear in equation (4.1). Since (4.1) is identified, $\begin{pmatrix} \beta \\ \gamma \end{pmatrix}^*$ exists. Therefore, substituting $\begin{pmatrix} \beta \\ \gamma \end{pmatrix}^*$ into (4.63), we obtain, by Lemma 4.1, our result:

$$\begin{pmatrix} P_{\tilde{I}_2^1 y_g.\tilde{I}_1} \\ P_{\tilde{I}_2^2 y_g.\tilde{I}_1} \end{pmatrix} = \tilde{0}.$$

Part II: $\begin{pmatrix} P_{\tilde{I}_2^1 y_g.\tilde{I}_1} \\ P_{\tilde{I}_2^2 y_g.\tilde{I}_1} \end{pmatrix} = \tilde{0}$, then $\tilde{Q} = \tilde{0}$.

Proof of Part II: Assume $\tilde{Q} \neq \tilde{0}$; then either $\tilde{\beta}_2^1 \neq 0$ or $\tilde{\gamma}_2^1 \neq 0$ (or both). Therefore there exists some j such that $H+1 \leq j \leq H'$ (or $J+1 \leq j \leq J'$) and $\beta_{gj} \neq 0$ (or $\gamma_{gj} \neq 0$). We now have k non-zero parameters in equation (4.1), where $k \geq H+J$. Therefore $\begin{pmatrix} \beta \\ \gamma \end{pmatrix}$ has k rows. Since, by assumption, the elements of $\tilde{y}_2^2$ and $\tilde{z}_2^2$ are excluded from (4.1), then (4.1) is identified and $\begin{pmatrix} \beta \\ \gamma \end{pmatrix}^*$ exists. Therefore substitution of $\begin{pmatrix} \beta \\ \gamma \end{pmatrix}^*$ into (4.63) for $\begin{pmatrix} \beta \\ \gamma \end{pmatrix}$ yields an expression for $\rho_{y_j^* y_g}$ (or $\rho_{z_j y_g}$) which, by Lemma 4.1, implies that $\rho_{y_j^* y_g.y_1^*\ldots y_H^* z_1\ldots z_J} \neq 0$ (or $\rho_{z_j y_g.y_1^*\ldots y_H^* z_1\ldots z_J} \neq 0$). Therefore,

$$\begin{pmatrix} P_{\tilde{I}_2^1 y_g.\tilde{I}_1} \\ P_{\tilde{I}_2^2 y_g.\tilde{I}_1} \end{pmatrix} \neq \tilde{0}.$$

This is a contradiction to our hypothesis; therefore we must have $\tilde{Q} = \tilde{0}$. Q.E.D.

C. Interpretation of Results

By Definition (3.1), we can see why the results of Theorem 4.2 can be interpreted to be a statistical test of the adequacy of causal theories represented in the general linear structural model. If we are willing to make certain untestable *a priori* assumptions, then, by our statistical results, we are able to determine whether or not certain other restrictions on the equations of our model are correct. Since those restrictions which we test are exclusion restrictions, we have by Definition (3.1), that they are, additionally, restrictions about causal relations.

We have also accomplished our heuristic objective in deriving a statistical test which reflects the logical process associated with introducing control variables into a zero-order correlational relationship and examining the partial correlations as a test for "spuriousness." Our results give us precisely which partial correlations (if any) ought to be equal to zero in any given causal situation (as we have defined such a situation).

Furthermore, it should be easily seen that, as a consequence of Theorem (4.2), we can determine which one of the $C_0^{(K-1)} + C_1^{(K-1)} + \ldots + C_{(K-1)}^{(K-1)} = 2^{(K-1)}$ possible causal mechanisms represented by (4.1) is the correct one, given the *a priori* just identifying restrictions on (4.1). We can do this by observing which set of partial correlational conditions holds, since it is obvious that only one set of these conditions holds for each one of the $2^{(K-1)}$ possibilities. Also, since (3.2) is a system of G equations, we can determine which of the $[2^{(K-1)}]^G$ possible causal theories represented by (3.2) is correct, given the just identifying restrictions on all g, $g = 1,\ldots,G$.

While our test was developed only in an expectational sense, employing population parameters and expected values, we conjecture that our test can be approximated with information on samples of data. In fact, there is a close relationship between much of the derivation of our test and the parameter estimating method of two-stage least squares (See Footnote 23). A good part of our thinking was inspired by this method. Combining such an estimating procedure with the use of sample statistics rather than the population parameters used in this piece, one should be able, at least in principle, to approximate our test on sample data. However, as mentioned earlier when we discussed Simon's test for the recursive-form, the possibility of violation of untestable *a priori* assumptions is a serious problem in empirical work and, therefore, any application of our test to sample data should be done with a good deal of caution.

FOOTNOTES

[1]For example, see the following works: Alker (1966); Cnudde and McCrone (1966); Forbes and Tufte (1968); and Goldberg (1966).

[2]See Wold (1953), Wold (1960), and Strotz and Wold (1960), for discussion of the causal implications of the recursive-form.

[3]Throughout this paper, we shall use "causal relation" interchangeably with "direct causal relation."

[4]A good discussion of segmentable systems may be found in Christ (1966), pp. 61-62.

[5]See Simon (1953), p. 18, for the exact statement of his definition.

[6]The basic distinction between an endogeneous variable and a predetermined variable is the following: a predetermined variable is a variable which is independent of all disturbances in the model at time <u>t</u>. All other variables at time <u>t</u> are endogeneous variables. The distinction between these two types of variables will be made more explicit later in this paper.

[7]The notion of identifiability is not essential for an understanding of this section. We will discuss this notion in a later section.

[8]We shall consider the identification problems associated with triangular systems in a later section.

[9]See Simon (1954), p. 47.

[10]See Lazarsfeld (1959).

[11]We will use "recursive-form" interchangeably with "segmentable" in the following pages.

[12]See Bentzel and Hansen (1954-1955), pp. 153-168, for a good discussion of the problems associated with the use of the recursive-form for dealing with extended periods of time. Also see Samuelson (1965), pp. 139-40.

[13]Miller and Stokes (1963) initially desired to consider some reciprocal links in their model of constituency influence in Congress, but they never really carried out such an analysis. See Alker (1969) for a discussion of alternatives to the Miller and Stokes formulation.

[14]For a somewhat different approach to the relationship between time and causality see Fisher (1970). In this piece Fisher takes a different epistomological position than do we and admits a <u>logical</u> relationship between time and causality. He then examines the implications of the stance that simultaneous equation models are limiting approximations to nonsimultaneous ones as time lags go to zero. See also Granger (1969).

[15]See Bunge (1963), pp. 188-190 and J. Simon (1969), pp. 458-460.

[16]This paragraph is essentially a paraphrase of Fisher (1966), p. 92.

[17]Basically, an estimator is consistent if it converges in the probability limit to the "true" parameter being estimated.

[18]See Christ (1966), p. 317, for a proof of this.

[19]See Fisher (1966), pp. 36-39, for a proof.

[20]The data was generated by a program called DATSIM on the Berkeley Computer Center CDC 6400.

[21]We will include in this definition both variables which are independent of the disturbances at all times <u>t</u> (exogenous) and variables which are independent of the disturbances at a single time <u>t</u> (predetermined).

[22]See any standard econometric text for discussion of the general linear structural model. For example, Christ (1966) or Goldberger (1964).

[23]This is essentially the two-stage least squares solution. See Christ (1966), pp. 432-46; Goldberger (1964), pp. 329-36, or Johnston (1963), pp. 258-60 for discussions of the method of two-stage least squares.

[24]See Christ (1966), pp. 539-40.

*I wish to thank J. Merrill Shanks of the University of California, Berkeley and Hayward R. Alker, Jr. of the Massachusetts Institute of Technology for inspiration and helpful criticisms of an earlier draft of this paper. I also wish to thank Ruth Suzuki for her many hours on the typewriter.

REFERENCES

Alker, Hayward R., Jr. <u>Mathematics and Politics</u>. New York: Macmillan Co., 1965.

Alker, Hayward R., Jr. "Causal Inference and Political Analysis." in <u>Mathematical Applications in Political Science, II</u>. Edited by Joseph A. Bernd. Dallas: Southern Methodist University Press, 1966, pp. 3-43.

Alker, Hayward R., Jr. "Statistics and Politics: The Need for Causal Data Analysis." in <u>Politics and the Social Sciences</u>. Edited by Seymour Martin Lipset. New York: Oxford University Press, 1969. pp. 244-313.

Anderson, Theodore Wilbur, and Herman Rubin. "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations, <u>Annals of Mathematical Statistics</u>, 21 (December, 1950), pp. 570-582.

Basmann, Robert L. "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics, <u>Journal of the American Statistical Association</u>, 55 (December, 1960), pp. 650-659.

Bentzel, R. and Hansen, B. "On Recursiveness and Interdependency in Economic Models." <u>Review of Economic Studies</u>, XXII (1954-1955), pp. 153-68.

Blalock, Hubert M., Jr. "Four-Variable Causal Models and Partial Correlations." <u>American Journal of Sociology</u>, LXVIII (September, 1962), pp. 182-94.

Blalock, Hubert M., Jr. <u>Causal Inferences in Nonexperimental Research</u>. Chapel Hill: University of North Carolina Press, 1964.

Boudon, Raymond. "A New Look at Correlation Analysis." in <u>Methodology in Social Research</u>. Edited by Hubert M. Blalock, Jr. and Ann B. Blalock. New York: McGraw-Hill, Inc., 1968, pp. 199-235.

Bunge, Mario. <u>The Myth of Simplicity</u>. Englewood Cliffs: Prentice-Hall, Inc., 1963.

Christ, Carl F. <u>Econometric Models and Methods</u>. New York: John Wiley and Sons, Inc., 1966.

Cnudde, Charles F., and McCrone, Donald J. "Constituency Attitudes and Congressional Voting: A Causal Model." <u>American Political Science Review</u>, LX (March, 1966), pp. 66-72.

Fisher, Franklin M. <u>The Identification Problem in Econometrics</u>. New York: McGraw-Hill, 1966.

Fisher, Franklin M. "A Correspondence Principle for Simultaneous Equation Models," <u>Econometrica</u>, 38 (January, 1970), pp. 73-92.

Forbes, Donald, and Tufte, Edward. "A Note of Caution in Causal Modelling." <u>American Political Science Review</u>, LXII (December, 1968), pp. 1258-64.

Granger, C.W.J. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," <u>Econometrica</u>, 37 (July, 1969), pp. 424-38.

Goldberg, Arthur S. "Discerning a Causal Pattern among Data on Voting Behavior." <u>American Political Science Review</u>, LX (December, 1966), pp. 913-22.

Goldberger, Arthur S. <u>Econometric Theory</u>. New York: John Wiley and Sons, Inc., 1964.

Johnston, J. <u>Econometric Methods</u>. New York: John Wiley and Sons, Inc., 1963.

Kadane, Joseph B. "Testing a Subset of Overidentifying Restrictions," Unpublished Manuscript, 1970.

Kaplan, Abraham. <u>The Conduct of Inquiry</u>. San Francisco: Chandler Publishing Co., 1964.

Koopmans, Tjalling C. "Identification Problems in Economic Model Construction," in <u>Studies in Econometric Method</u>. Edited by William C. Hood and Tjalling C. Koopmans. Cowles Commission Monograph 14, New York: John Wiley and Sons, Inc., 1953, pp. 27-48.

Lazarsfeld, Paul F. "The Algebra of Dichotomous Systems," in <u>Item Analysis and Prediction</u>. Edited by H. Solomon. Stanford: Stanford University Press, 1959.

Liu, Ta-Chung and William J. Breen. "The Covariance Matrix of the Limited Information Estimator and the Identification Test," <u>Econometrica</u>, 37 (April, 1969), pp. 222-227.

Miller, Warren E., and Stokes, Donald E. "Constituency Influence in Congress." <u>American Political Science Review</u>, LVII (March, 1963), pp. 45-56.

Samuelson, Paul A. "Some Notions on Causality and Teleology in Economics," in <u>Cause and Effect</u>. Edited by Daniel Lerner. New York: The Free Press, 1965, pp. 99-143.

Simon, Herbert A. "Causal Ordering and Identifiability." in <u>Studies in Econometric Method</u>. Edited by William C. Hood and Tjalling C. Koopmans. Cowles Commission Monograph 14, New York: John Wiley and Sons, Inc., 1953, pp. 49-73. Reprinted in <u>Models of Man</u>. By Herbert A. Simon. New York: John Wiley and Sons, Inc., 1957, pp. 10-36.

Simon, Herbert A. "Spurious Correlation: A Causal Interpretation." <u>Journal of the American Statistical Association</u>, XLIX (September, 1954). Reprinted in <u>Models of Man</u>. By Herbert A. Simon. New York: John Wiley and Sons, Inc., 1957, pp. 37-49.

Simon, Julian L. <u>Basic Research Methods in Social Science: The Art of Empirical Investigation</u>. New York: Random House, 1969.

Strotz, Robert H. "Interdependence as a Specification Error," <u>Econometrica</u>, 28 (April, 1960), pp. 428-442.

Strotz, Robert H. and H.O.A. Wold. "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis," <u>Econometrica</u>, 28 (April, 1960), pp. 417-427.

Wold, H.O.A., in association with Lars Jureen. <u>Demand Analysis: A Study in Econometrics</u>. New York: John Wiley and Sons, Inc., 1953.

Wold, H.O.A. "Ends and Means in Econometric Model Building. Basic Considerations Reviewed," in <u>Probability and Statistics</u>, Edited by U. Grenander. Stockholm: Almqvist & Wiksell, 1959.

Wold, H.O.A. "A Generalization of Causal Chain Models," <u>Econometrica</u>, 28 (April, 1960), pp. 443-463.

DISCUSSION

Gudmund R. Iversen, The University of Michigan

The Richardson model on arms race discussed in the Ferejohn paper and Herbert Simon's [2] mathematical treatment of the verbal propositions in The Human Group by George C. Homans [1] are classical items for anyone interested in the use of mathematics and statistics in the social sciences. A new paper proposing a more extended arms race model is therefore received with considerable interest. However, my own experience in this area has been rather limited, and these comments can only be of a general nature.

One may say that to empirically identify an arms race there is no need for mathematical models or differential equations. Tense international relations and large arms budgets are all too obvious just from reading newspapers, but that is an unfair comment on the Ferejohn model. His model seems very promising for the understanding of the complicated relationships that exist between nations.

Any model of social phenomena is only as good as the assumptions that go into the model. Such assumptions are abstracted from the substantive theory developed to account for the phenomena. International relations has become a special topic of study within political science, and one could have expected that the assumptions in the Ferejohn model would reflect some of the work done in international relations. With the assumptions more deeply anchored in substantive theory the consequences of the model would have been more relevant for current political research.

The Ferejohn model deals with an extremely difficult problem. Many variables enter into determining the arms budgets in different nations. That would lead one to believe that in order to explain these phenomena more fully, larger and more complicated models are needed. Such work undoubtedly goes on under classified cover, which means that it is difficult to see the Ferejohn model in full perspective. But in spite of this the Ferejohn model is a welcome addition. The model shows imagination in its dealing with a difficult problem, and one can only hope that this work will be pursued and made better known.

The paper by Geisel, McGuire, Rosenthal and Kies deals with a problem that is guaranteed not to have a solution. This is no criticism of the paper, however, quite on the contrary. With problems of this kind one is more free to formulate models and investigate the properties of the models. The basic underlying problem dealt with in this paper and which does not have any solution, is the problem of estimating the cell entries in a set of contingency tables where the margins are known and the cell entries are unknown. One can look at such tables until one is blue in the face without the tables divulging their secrets. But information about the cell entries may be obtained from the observed margins by insightful modelling of the relationships between the cell entries and the margins. The social science world is full of unidimensional distributions from census statistics, voting records and other sources. With substantive theory as a collection of statements on how variables are related one wants to relate the variables from the distributions above. But with unidimensional variables the relationships can only be measured on the group level, even though for most purposes one would want the relationship measured on the level of the individual. This, however, is only possible if the cell entries are known.

The strength of the paper lies in its statistical parts, which is appropriate for this meeting. The main contribution consists of the estimation procedures that are developed. But I am not certain that the methods are sufficiently justified. It may be possible to show that ordinary least squares estimators have undesirable statistical properties and thereby conclude that one should move on to two and three stage least squares. But at the same time one can show examples with known cell entries that the estimates obtained from the margins using ordinary least squares are very close to the true cell entries. This is the type of problem where it may be more profitable to invest more heavily in the model and spend less on the estimation methods. I am not coming out against complicated estimators with good statistical properties; I am arguing that simple estimators should not necessarily be excluded, and we need in this case to know more about why ordinary least squares should not be used. One point that can be noted in this connection is that the correlation between e_i and R_{ki} is not as important as the magnitudes of the e_i's in deciding whether ordinary least squares is appropriate or not.

The vitality of the statistical profession is dependent on the input and challenges from those who are using statistical methods in their substantive area to obtain results otherwise not obtainable. This paper is a good example of such a challenge. But joint papers also run the risk of being disjointed. Sections 2, 3 and 4 are highly technical and cannot be read or understood by most social scientists. These sections have no references to the substantive problem, even most of the notation is changed.

In some ways it seems as if Sections 5 and 6 lost sight of Section 1. For example, in the choice of variables it is not clear why EI 48- is selected for Rafi-K and LOW.ED is selected for Rafi-M. Perhaps a more serious question has to do with the phenomenon of coattails. The concept does not seem to be clearly defined in terms of the model that is analyzed. This may have to do with a lack of distinction between aggregate and individual level data. In Section 1 we get set up to investigate the individual level data, but in the later sections we do not make it back again. After analyzing the structural regression equations we only make it back in the end to the proportion model but not to the original equations in Section 1.

The substantive model ought to do justice to the estimation procedures that are developed. With that the authors will have made a distinct contribution to the whole topic of cross level analysis where data are available on one level and we want to do analysis on another level.

The paper by Richard Juster was not received in time to be included in this discussion.

REFERENCES

[1] Homans, George C., The Human Group, New York: Harcourt, Brace and Company, 1950.

[2] Simon, Herbert, A., "A Formal Theory of Interaction in Social Groups," American Sociological Review, 17(1952).

THE NEED FOR REFORM OF NATIONAL INCOME STATISTICS

Kenneth E. Boulding, University of Colorado

There is no doubt that the body of economic indicators known as National Income Statistics has been a fantastic success. From being a somewhat erudite and even rather secondary interest of the National Bureau for Economic Research in the 1920's, its concepts have become standard categories not only of economic argument but also of political discourse. It is hard to believe that magic letters GNP were virtually unknown before 1930. However, one of my favorite anti-proverbs is that nothing fails like success, and the very success of National Income Statistics perhaps has induced a certain complacency among economists, who sit a little too secure in the high prestige of which National Income Statistics have been a considerable part of the purchase price.

This complacency is now being challenged from a number of sources. The ecologists and conservationists have been challenging us quite rightly because the categories of national income accounting do not usually take account of depreciation of the environment and the production of "bads." The worship of the Gross National Product has been attacked as a source of pollution, environmental degradation, and even of future environmental catastrophe. The Gross National Product can be increased by destroying irreplacable natural resources, and a good deal of the activity which it represents may consist of activity necessary to overcome the bads which are produced as a by-product of goods which are otherwise counted in the Gross National Product. We can perhaps defend the Gross National Product on the grounds that it is intended only as a measure of economic activity and says nothing about what that activity is for. This however is not quite adequate as an answer, because the GNP purports to measure the product of economic activity, not merely the activity itself. Hence, if it does not include the production of bads as negative items, it is seriously defective, simply because it is too gross. The failure to include bads, indeed, is a technical defect quite similar to the problem of avoiding double counting in the production of goods, a problem which was worked out in the very early days.

While the theoretical point is easy to make, its practical application is quite difficult, mainly because of severe difficulties which we encounter in the evaluation of bads. In order to get a figure for the GNP we have to add up an enormous heterogeneous list of items of goods, weighting each by some shadow price or valuation coefficient. We would have to do the same thing for the negative commodities, which should appear as negative items in the addition. The price system, however, does not easily create negative prices for negative commodities, so that it is often hard to tell what these negative prices ought to be. We can visualize this by imagining that we go to a complete system of effluent taxes in which everybody who produces a bad is taxed accordingly. The controversies which would arise over the simple word "accordingly" can well be imagined. The difficulty is that there are very poor markets for bads, mainly because we do not cherish them and exchange them, but rather try to push them off onto somebody else without him noticing it.

We may try to wriggle out of this bind by saying that what we are really looking for is a <u>net</u> national product. Even this, however, is not too satisfactory. The Net National Product in the standard accounts is largely an accounting convention, which is largely a function of the accounting habits of last year modified slightly by the impact of the tax system. The Capital Consumption Allowance which is subtracted from the GNP to produce the NNP is about as much an exercise of the fertile human imagination as any statistic can be. If we want to satisfy the ecologists and conservationsists, we should certainly add depreciation of the environment to this Capital Consumption Allowance to get a true NNP, although, unless this were also registered in private accounts and in private net worths, this would cause trouble when it came to evaluate the accounts on the income side. Degredation of the environment, unfortunately, frequently turns up as somebody's income, whereas if it were accounted properly it would appear as a deduction from somebody's income.

Exhaustion of resources should also be deducted from GNP to get a true GNP. When we ask ourselves how much to deduct, however, this seems like an even more difficult problem than the problem of evaluating bads. Natural resources are a function of human knowledge. Because of the spectacular rise in the "quantity of science," as Adam Smith calls it, we have probably been increasing natural resources faster than we have been using them up in the last two hundred years. But whether we can continue to do this is a serious question. We may be reaching the point of diminishing returns in the application of increased knowledge to the discovery of new resources. This, however, is a matter for the future which is extremely difficult to estimate now. If, for instance, we solve the problem of fusion power in the next fifty years, we may not have to worry very much about the exhaustion of oil, coal, and gas. If we do not solve it, we may have to worry a great deal. Who is to say, therefore, whether we should be deducting from the Gross National Product all the oil, coal, and gas that we use up, or whether we should be adding the capitalized value of the unknown future knowledge that we are going to have in fifty years? These are problems which can daunt the most skilled account of any kind and we can hardly blame national income accountants for simply pretending that they are not there.

Another problem of "netting" involves the evaluation of public goods. We could take almost any view here between two extremes. One of these would be to assume that public goods, that is, the government sector of the economy, represents simply something that is extorted by superior force from an unwilling citizenry and adds nothing whatever to their welfare. All taxes then are a simple deduction from real income. There are indeed some kleptocracies which may not fall altogether short of this model. At the other extreme, we have the exchange theory of public finance which supposes that people get something for their taxes equivalent to what they pay. Under these circumstances, the public sector of the economy is simply added to the private sector as individual welfare. The truth certainly lies somewhere between these two extremes, but where it is not easy to say. Even on the exchange theory, it is clear that public goods are much more like Christmas presents than they are like purchases, that is, they involve reciprocity rather than exchange, and just as one seldom gets what one wants for Christmas, one seldom gets what one wants from taxes, so that obviously in this sense the public sector has to be deflated in order to get a welfare index, but by how much it has to be deflated is almost anybody's guess. These problems revolving around the relation of National Income Accounts to welfare, either public or private, have been dealt with with great expertise by A. W. Sametz[1] and there is no need to beat this well-beaten bush much further. One or two other complaints, however, are not so often heard and as long as this is a complaining session we might as well try to get all the complaints out in the open.

One important set of complaints relates to the possible use or lack of use of National Income Statistics for the understanding of the problems of distribution in society. The standard National Income Accounts do have a breakdown of national income into distributional components such as Wages and Salaries, Business and Professional Income, Farm Income, Rental Income of Persons and Corporate Profits. How this fantastic hodge podge ever got into the national accounts is something which would be well worth a Ph.D. thesis. In the first place, the national income concept itself is an absurdity. It is surely one of the least defensible of all the aggregates. It is net of indirect taxes and gross of direct taxes, and there really seems to be no case for it at all, except perhaps on the assumption that only direct taxes measure the real value of the public economy and that indirect taxes represent that part of the public economy which is sheer waste. I doubt, however, if this theory was behind the construction of the accounts in the first place.

We start off, therefore, by distributing a quantity which is absurd to distribute in the first place, and we then proceed to distribute it in the oddest way imaginable. Farm Income, for instance, should surely belong to distribution by the industrial sector of the economy, something which, incidentally, is quite hard to get in any very satisfactory form. Then we have this preposterous aggregate called "Business and Professional" which includes presumably the income of doctors, lawyers, and unincorporated business. I have not been able to get a breakdown of this into business and professional, and one gets an uneasy feeling sometimes that these numbers are simply made up by a little man with a green eyeshade in the attics of the Department of Commerce. It is very peculiar, for instance, that this segment of national income has been declining, relatively, in the last forty years, whereas one would surely have thought that we have been witnessing a considerable expansion of the professions. Another very odd segment is "Rental Income of Persons," which again simply reflects a certain lack of incorporation in the real estate market with, I suppose, a certain amount of literary income. Corporate Profits here are gross of direct taxes and hence are probably overestimated. It would be hard to imagine, indeed, a more preposterous breakdown of a more preposterous total.

What then do we want? In the first place, we want distributional breakdowns of a number of different totals. For some purposes, the distribution of the Gross National Product itself would be quite interesting--for instance, by industries, by regions, and by various segments of the labor force, such as union versus non-union, corporate versus non-corporate, and so on. For other purposes, a breakdown of the Net National Product or something rather like it--perhaps it might be called the "gross national income," that is, gross income before taxes of all kinds. For other purposes, a breakdown of gross private <u>disposable</u> income, which would roughly be equal to the Gross National Income minus the total government sector, would also be of great interest. There are one or two rather tricky technical problems here of minor items such as the statistical discrepancy, certain items of income which are hard to allocate to individuals and the total government surplus or deficit, which are a little too technical to go into here. However, they are all fairly small items in normal times.

For both of these aggregates it would be extremely interesting to be able to break them down into labor income and non-labor income, which we cannot do in the present accounts, and it would be useful to be able to divide non-labor income into interest, profit, and rent in the good old classical style. These categories, unfortunately, are not as simple as they look and some compromises would have to be made, but it would certainly not be difficult to achieve something more significant than the present breakdowns. On the labor income side, it would be extremely nice to have a breakdown, even if somewhat arbitrary, into, say, unskilled, semi-skilled and highly professional, or something of this sort, which again we do not have now. It surely ought to be easy to answer from National Income Statistics questions such as "Is the economy getting more and more corporatized?" or "Are the service industries really increasing?" or "Is distribution going more towards labor or towards capital?" The plain fact is we simply cannot answer any of these important questions from the information as it is now presented.

Other distributions of national income would also be of enormous interest, but would be harder to get. We have certain distributions by race, although the aggregates here can be extremely misleading, as they are likely to hide certain poverty sectors within all the different racial groups. All blacks are certainly not poor and all whites are not rich. Breakdowns of income by religious preference would be of enormous interest, but very difficult to obtain. A breakdown by years of education would also be of great interest. The ideal here, of course, would be a system of computerized information from individuals which would enable us to get any kind of breakdown that we wanted, but this perhaps is too much to expect, and also raises all the specters of the national data bank, in terms of privacy, manipulation and so on.

Another very fundamental complaint against the national accounts is that they are extremely deficient on the side of capital accounting. There is no annual accounting of the national capital and its distribution. And yet, as I have been arguing, apparently without anybody listening to me for years, when it comes to developing an index of human welfare, capital accounts are much more important than income accounts. Welfare is a condition or state, that is, a stock rather than a flow, although it is not unrelated to certain flow elements. With the recent interest in the environmental deterioration the notion of welfare as a stock variable has suddenly become very fashionable. Yet National Income Accounts pay no attention to this at all. It is very difficult to find out from national accounts anything about the distribution of equity in the total capital stock of the society. It is equally difficult to find out much about the significance of the financial variables. What is the real significance, for instance, from the point of view of the distribution either of capital or income, of an increase in debt, both public and private, or of the significance of a change in the rate of interest? The fact that we cannot tell the distributional impact of almost any act of public policy, either in the distribution of capital or in the distribution of income, is one of the gravest defects in our economic information system. I am not suggesting the national accounts are necessarily the answer to this problem, but they are certainly part of the answer.

One of the great difficulties in interpreting the national accounts as a consumer, that is, as an outsider, is that, in spite of the publications on the subject, one does not really know how these things are actually put together in the shop. Anyone who has had any experience at all with the production of statistics by public bodies knows how much finagling, estimation, compromise, and argumentation has to go on inside the establishment before the beautiful tables appear in print. Any statistics-producing enterprise develops in the course of time a subculture of its own, along with a good deal of conventional wisdom and what we might call "private learning from experience," which is seldom subject to any outside test, and very rarely subjected to any formal outside valuation. A thorough, outside valuation of these "statistical subcultures" is something that should surely be done at least once every generation.

May I raise the question, therefore, as to whether the time is not ripe for a substantial reexamination of the whole system of aggregate statistics? I do not intend in any way to deprecate the magnificent work which has been done in the last forty years. Indeed, if one were to apply the overworked term "revolutionary" to any social change of this century, it is surely the development of National Income Statistics and the related improvements in the whole apparatus of social indicators. This has probably produced a larger change in social policy than any other single change in the institutional structure we can name, and the influence on the whole has been markedly beneficial. I have often pointed to the contrast between the twenty years after the First World War and the twenty years after the Second as an example of a fantastic transformation in the economic life of the world, a considerable part of which is a result of the change in political and social images which has come about through the use of National Income Statistics, coupled with the development of a macro-economic theoretical system which fitted in with the cumulation of statistical information. The interest which has developed in recent years in the extension of the system of aggregative indicators from National Income Statistics in the field of social indicators[2] is suggestive both of the tremendous impact which National Income Statistics have had and also of a certain sense of inadequacy in regard to them. The corollary of the anti-proverb that I quoted at the beginning--that nothing fails like success--is that we only learn from failure, which is why it is so important to identify the right failures and to learn the right things from them. I suggest, therefore, that the time has come when a major research inquiry, not only into National Income Statistics, but also into social indicators in general, is clearly desirable. How this should be financed and organized is of course a question. It should clearly be in some sense independent enough of government to be capable of making sharp criticisms where these are required. On the other hand, it has to be close to government, which is the major provider of the statistical enterprise, and to have the confidence of those who are presently engaged in the business of producing statistics. Perhaps the natural body to sponsor such an enterprise would be the National Academy of Sciences, were it not for the fact that the relations of the National Academy with the social sciences are so marginal and unsatisfactory. Whatever the machinery, however, the need is very clear. Perhaps the preliminary work might be done by a joint committee of the American Economic Association and the American Statistical Association, though this might not have the essential entrée into government that the enterprise would really require. Whatever the machinery, it is important that the enterprise should be well financed and taken very seriously and that it should involve the participation of large numbers of consumers of statistics. It goes without saying that it should be an interdisciplinary enterprise, that it should involve all the social sciences, and it should probably have an element of collective bargaining in it between the users and the producers of statistics, so that it might almost be conceived as a semi-permanent institution after the initial work has been done.

REFERENCES

(1) Sametz, A. W., "Production of Goods and Services: The Measurement of Economic Growth," in Indicators of Social Change, edited by E. B. Sheldon and W. Moore. New York: Russell Sage Foundation, 1968.

(2) Toward A Social Report. Washington, D.C.: U.S. Department of Health, Education and Welfare.

ALTERNATIVE FUTURES: CONTEXTS IN WHICH SOCIAL INDICATORS MUST WORK

O. W. Markley, Stanford Research Institute

"Social indicators," "social reports," and "social accounts" are three terms which are increasingly used to describe a comprehensive but empirical approach for describing, understanding, and managing society.[1] The road ahead for the rapid development and early widespread use of these tools does not appear to be an easy one. They are being proposed at a time when there is an increasing distrust of excessively rationalistic macro-policy management methods.[2] Also, there seems to be a real possibility that they are being oversold with exaggerated claims made as to their utility.[3] And finally, there are a host of politically related value problems which are attendant upon their use which have no presently viable way to be resolved.[4]

One thing seems clear however. Although these tools may never become useful for such far-reaching purposes as the development of a master social accounting scheme (with a balance sheet for comprehensive national, social, and economic accounting), they may well prove useful for more limited purposes such as an improved descriptive reporting and conventional understanding of society. Either way, any set of social indicators which are devised today and used as a bases for present social policy decisions for future decades, must be designed with that future in mind. Since social indicators are specifically intended to provide "bench marks" of various qualitative aspects of society that may be compared over time, they must be of sufficient range and diversity to include the spectrum of circumstances which seem of highest plausibility in the future. Thus, social indicators need to be designed in terms of what tomorrow may bring, and not be limited to what today has already brought.

The purpose of this paper is to briefly describe how projections of the future can usefully be made, to present a synopsis of a selected set of alternative future histories for the United States, and to illustrate their relevance to the design and use of social indicators. Understanding of the method, however, is not essential for appreciation of the results.

Precis of the Methods Used to Project Alternative Futures

We seek to describe alternative futures because it is impossible to predict a single most probable course of future evolution for the world, the nation, or for any significant aspect therein. However, even brief reflection reveals that the number of different but plausible future histories that can be written exceeds the number that can usefully be used. Therefore, we construct a "planning cone" which contains a reduced set of alternative lines of societal development which we hope will "bracket" the one future that comes to pass. This reduced set of future histories is made up of those future possibilities which (1) seem most plausible, (2) differ significantly from each other, and (3) have important characteristics with respect to policy analysis and planning.

Each projected alternative must be schematically commensurate with what actually will emerge. That is, each must be (insofar as imagination and analytical skill can manage) an internally consistent whole; each must merge aspirations and the more mundane considerations of feasibility; each must evolve partly because of purposive efforts and partly because of forces beyond the reach of conscious desires.

Two different, but complementary types of analysis are followed to realize the above constraints. One is quite formal and methodical, the second somewhat more holistic and intuitive. The first and more methodical type of analysis is based on a newly developed approach called Field Anomaly Relaxation.[5] It employs a "morphological" expansion of basic societal descriptors in ways that use the principles of relaxation often used in the modeling of complex dynamic systems as in thermodynamics. The procedures constitute a method for qualitative analysis of complex fields of partly or wholly non-quantifiable information. They allow one to describe both the state and the dynamics of a complex society in initially simple and imprecise terms. Then, the method is recycled to eliminate internal inconsistencies and errors in approximation as well as to add new constructs and new input data. Thus, the analyst continuously applies his "common sense" as well as relevant theoretical and empirical data--with the confidence that through such recycling, serious faults will gradually be corrected.

The first step of the initial development of the method involved choosing a set of six sectors as a minimal descriptive framework for the United States (see the outer ring of the "map" in Figure 1). Each of these six sectors were then elaborated with five or six alternative states for each (shown in Table 1), thus covering the likely range of variation.

In the first re-cycling of the method, which is now being completed, the first six sectors were somewhat revised, while the six inner sectors shown in Figure 1 (and the roster of factor "states" shown in Table 2) were added to more adequately model society.

If full alternation of all combinations (one factor in each of the 12 sectors) were undertaken, it would lead to more than a billion possible descriptions of society. To bring it within realistic constraints, this number was reduced first by identifying sets of factors which were internally self-consistent. (For example, "economic depression" is not consistent with "rapidly expanding technology" and hence could not co-exist in a plausible societal configuration). This process limited the number to several hundred patterns from each of the two types of sectors. The two sets were then merged, and the number of societal configurations reduced still further to less than 50 by including only those which might conceivably occur through specifiable and plausible scenarios.

In the second and more informal type of analysis the constraining limitations of present-time realities were added, reducing the number of plausible configurations still further. The approach used here is hard to describe because it partakes of the methods used in intelligence work as well as those of conventional social science. Suffice to say that competing views regarding events and trends of the recent past and present are examined, and compared with both long-standing national goals and the more short-term desires and expectations of various stakeholders in society. Thus this task involves continued monitoring of contemporary events and literature, as well as contact with persons who hold divergent viewpoints concerning society. It often involves examination of social issues from the standpoint of the various basic value premises and "organizing images" which are prevalent, and earlier involved a major research effort devoted to analysis of forces which might lead to discontinuous or revolutionary change in society.[6]

Both methods involve numerous judgments by the investigators as to what constitutes "plausible" sequences of states in society. Criticism of these judgments is possible, however, as the important steps in both underlying analyses and the resulting scenarios are explicitly stated.

Tentative Results

By use of these methods we have projected the "tree" of alternative future histories which is shown on Figures 2 and 3. The five main lines to the year 1985 and secondary branches represent a distillation of some 40 highly plausible histories, and are being used as a working "planning cone." Again, by planning cone is meant simply the bracketing set of alternative future histories that should be useful as societal contexts which can be assumed in long-range planning.

A "Tree" of Alternative Future Histories

The five main lines to the year 1985 and secondary branching lines shown in Figures 2 and 3 represent a distillation of some 40 highly plausible future histories, and are being used as the basis for a working "planning cone." While not clear from the shorthand labels, given to the various "year 2000" states they tend to differ in two especially significant dimensions. One dimension concerns the degree to which society is adept in the Faustian sense (i.e., both competent and motivated to attempt control of its own destiny). The other dimension relates to the degree of social "openness," or "civility"--both terms which imply flexibility, the social coherence which flows from trust, tolerance for diversity, and the ability to sustain decentralization decision-making without undue internal violence.

A "planning cone" type representation of the "year 2000" slice of the future tree, with the alternative states arrayed in these dimensions, is shown in Figure 4.

A World Macroproblem

Before giving a brief description for each of the five alternative primary future histories, it is useful to describe the central set of societal problems which have had to be considered throughout this research.

When we initiated our studies in early 1969, we accepted the essential plausibility of the relatively optimistic forecasts which dominated most of the "futurist" literature.[7] It was clear that there were societal problems which had to be solved, and many would need social as well as technological innovations, but there seemed to be no reason to believe that any given difficulties were in principle insurmountable by conventional means.

As our work progressed, however, we came to recognize that while most contemporary problems are interrelated, their import can more readily be grasped if they are viewed--not as individual problems--but as a network of social forces that have been brought about by a combination of proliferating knowledge, industrial development unmoderated by a larger sense of social responsibility, rising population levels (which in turn are a consequence of technology-produced mortality rate), and an expanding have-have not gap.

These forces are mutually exacerbating and systemic in nature and therefore are not likely to be "solved" by special programs aimed at one or more component parts. Also, they appear in all the plausible futures and hence will be encountered in one fashion or another.

With Peccei,[8] we have come to view the composite of these social forces as a world macroproblem. One aspect of this world macroproblem is the host of familiar problems of the ecosystem: ecological imbalances, fouling of the environment, resource depletion, overpopulation with consequent famine and plague.

A second is the area of technological threats; weapons of mass destruction; vulnerability of a complex society to sabotage or breakdown; misused capabilities to "engineer" the human body, mind, foetus, and genetic transmission; threats to privacy and individual rights; mental stress of complex living; etc.

A third is the persistent and increasing "have-have not" gap with the resulting internal and external dissension, intensified by the belief that the world agricultural-industrial system could easily produce enough of the necessities to meet the needs of all.

A fourth is the incipient crisis of specialization and rapid growth, in which increasing "bits" of knowledge are created, used, transmitted and stored without adequate "overall" perspectives with which to satisfactorily relate the pieces.

It further became apparent that the expectation that there would be a shift from industrial to post-industrial development[9] and that either technological or governmental interventions would be adequate to ameliorate the world macro-problem was no longer very credible. One block to such a shift is the difficulty which Garrett Hardin has described as the "tragedy of the commons,"[10] in which collectively held resources (such as clean air, or low population density) are depleted by individual behaviors which, though personally profitable, are in the long-run self-defeating for society. Another major problem is that of rapid technological and cultural change which provides a sense of "future shock"[11] and makes conventional management techniques obsolescent.[12] A third problem is what Mendel has termed the "great refusal" of youth to go along with social institutions as presently operative.[13]

For a variety of reasons, technological solutions are impossible for a significant subset of the world macroproblem.[14] Political solutions, without pervasive changes in the underlying culture and political institutions, are similarly not feasible,[15] nor are increasing extensions of basic "golden rule" morality.[16]

From the perspectives of our various analyses as well as those of others, the various aspects of the world macroproblem have gradually appeared more as surface manifestations of a fundamental cultural condition rather than as difficulties which are open to conventional solutions. This was revealed when we looked for plausible alternative future histories, where desirable future histories appeared hard to come by and given the problems just mentioned, requires significant changes in operative values and cultural morality. It appeared again as we attempted to analyze the roots of our present problems and began to see that these problems of the commons were implicit in both the premises and successes of our present form of Western technological-industrial culture, awaiting only increased levels of population and technological application to become intolerable. It showed up again as we grappled with the significance of contemporary revolutionary forces and found that the crucial gap is not between generations,[17] nor between liberals and conservatives, but between those who anticipate a continuation of present trends, and those who insist that a drastic change is inevitable and possibly desirable.

In short, the results of these various analyses raise the question as to whether the operative values which have served to bring us to the present point of development in the "great ascent" (Heilbroner) of civilization will continue to serve well in dealing with the problems created by that development.

While the logic of this analysis has seemed persuasive to many analysts, it is not possible to empirically demonstrate either the present severity of what we have termed the "world macroproblem" or the degree to which our cultural premises are undergoing transition; nor is it possible to predict the outcome of these forces. The various alternative future histories which are projected as a minimum set for long-range planning therefore reflect a plausible range of variation for these realities as well as variation in the other types of societal descriptors listed on Figure 1. Similarly they reflect a variation in the degree of success that is assumed for remedial attempts, again a hedge to cover our inability to predict the outcome of major attempts at social change.

Of the many plausible alternative lines of future history for the United States, the following five have been selected to provide the widest and most balanced coverage of alternatives was possible, yet small enough to be useable for the majority of policy analytic uses. Of course these results must be considered tentative and preliminary given this stage of the research and the rapidly developing state of the art. Hence one should be hesitant about drawing hard and fast conclusions from either the "tree" or from the brief descriptions of the five primary scenarios given next unless adequate analysis accompanies such inferences. Nevertheless these materials are useful as a framework from which to derive useful implications for long-range plans.

Five Alternative Mid-Range Futures[18]

1. A "War" on Eco-System Imbalance. This scenario differs from the others in that a national effort (a "moral equivalent to war") is undertaken to re-establish an ecological balance and to re-distribute the flow of material wealth so as to eliminate extreme domestic poverty. This effort is undertaken during the 1970s and early 80s, and is pervasive, judicious, self-sacrificing, and ultimately relatively "victorious." While some easily seen calamity would trigger such a "war," both a national consensus supporting it and a favorable combination of education and leadership must also be assumed if characteristic American impulses toward one-shot solutions, bureaucratic competition and scapegoating (e.g., young hoodlums or private industry) are to be superceded by a continuing national effort which is supported by a wide consensus.

As ecological sensitivity in society increases, the "war" on eco-system imbalance comes to be seen, not as a "war" to be "won" but as a set of cultural lessons to be learned.

The outcome of this scenario is uncertain, but seems to include substantial changes in operative cultural premises, hence the title "new" society.

2. "Surprise-Free" High Growth. This relatively optimistic line of development seems to best describe the future imagined by most "futurists." If it turns out to closely resemble the actual future, the various elements which combine to form the "world macroproblem" will prove in retrospect to have been grossly exaggerated. Both the economic and political patterns during the next 15 years prove to be quite similar to those of preceding decades, except that the continued increasing rate of both technological and cultural change slows down due to limitations of retraining and of management. The current trends toward growth and urban problems continue although the more severe problems of pollution are brought under control. Except for an assumed re-emergence of the international cold-war (a plausible "binder" for an otherwise

marginally coherent future), this might be thought of as a "good-luck" version of scenario 4 below.

3. Imprudent Optimism, Leading to a Left-Centrist Recession and Bureaucratic Stultification. This scenario explores a sequence of events in which efforts which are too hurried, too many, and too fragmented are made through governmental channels to correct presently perceived environmental and social ills. Hence it can be thought of a "bad-luck" version of the first scenario in which the "war" was "victorious."

Although initially optimism regarding the domestic reforms is high, the remedial programs prove inept and the commitments to the numerous competing stakeholder groups turn out to have exceeded the national productivity. Although a number of very plausible lines of evolution flow from this beginning (some of which are reflected on Figure 2 and 3), this scenario follows a persistent pursuit of welfare policies under bureaucratic control, which "lock in" to a slow drift toward recession. Social dissatisfaction becomes more and more general as the level of capitalization decreases, with concern for stability and economic growth then taking precedence over the "world macroproblem," which continues to worsen, but the pattern is relatively stable as each individual sees retention of existing conditions a least disadvantageous choice in the short-run.

4. Excessive Reprivatization, Leading to a Right Centerist Recession and Garrison State. This alternative future exemplifies one of the kinds of recessional developments that might find its roots in present conditions, if the events of the early 1970s indicate clearly the inadequacy of bureaucratic intervention as a strategy to deal with social problems and control of the economy. Here extensive reprivatization is undertaken as a major reform movement. "Funding of the people" instead of centrally administered programs is followed, attempting to stimulate "individual" initiative and to obtain the flexibility and efficiency that the profit motive often provides. An initial optimism continues as long as most stakeholder groups have some chance of realizing their objectives. Gradually, however, recession threatens as the government fails to successfully tune the economy, and stakeholder coalitions pre-emtively try to "get theirs." Scape-goats are easier to blame than failures of the socio-economic system and progressively more severe forms of repression are brought against those who protest violently. The domestic "garrison state" is paralleled late in the century by an international one, as recession imposes politico-military disengagement and then economic isolationism and the North Atlantic Community finds itself in continual change from the inward seeping of politicized violence from the chaotic Third World.

5. Escalating Violence. The character of this alternative future flows from an escalation of present trends in the use of confrontation politics as a means of accomplishing pervasive societal reform and premature "nourishment" of sub-cultural differences. As the thrust and confidence throughout the society breaks down, societal authorities increasingly come to rely on force as a means of maintaining control, and power soon replaces consensual authority. The outcome of this line of development depends to a large extent on the type of authoritarian form that gains power. However, both the paralyzing effects of violent terrorism and the repressive inflexibility inherent in an authoritarian response make generally recessive trends seem most plausible. A Caesarist take-over (analagous to that of Hitler) would be one alternative, leading toward supernationalization and extremely Faustian domestic and foreign politics.

Implications for Social Indicator Development

Two implications of this work for the development of social indicators stand out in importance. One is substantive, the other methodological.

First--We do not yet know the severity of what we have termed the world macroproblem. Nor do we know which of the several alternative futures is most probable. Nevertheless, it appears highly plausible that the various aspects of the macroproblem are intrinsic in the basic operative premises of present industralized culture. If this is correct, they may in the short-term be ameliorated or postponed by appropriate technological advances but will in the long run get more intense as the problems associated with cultural change also rise. If the experience of the past is any guide, numerous "one-shot" programmatic solutions will be attempted--efforts that will surely aggravate the situation unless they stem from an adequate understanding of the larger situation. Systems of well-selected, well-designed, and well-executed social indicators can help provide that understanding but only if they are designed with the overall societal context in mind. Thus, it seems important that any comprehensive set of social indicators should reflect the status of what we have termed the "macroproblem" and should monitor changes in social values as well.

Second--It is no new insight that normative social indicators should be used with caution because what is "good" for society at one time may not be so good for society sometime else.[19] The alternative future histories provide a convenient way to illustrate the need for this precaution, and suggest a methodological corrective as well.

Figure 4 illustrates how the five primary alternatives differ along the dimension of openness or civility. Note that a future with an efficient authoritarian government would likely be very high in Faustian competence, but low in civility; that a successful war on ecosystem imbalance could be expected to produce a society which limited its Faustian propensities, but attained a relatively high degree of civility; and that the other three futures suffer in both dimensions.

To illustrate the import of these differences consider one component of the dimension of civility--that of tolerance for diversity or pluralism. Obviously, it is not in society's best interest to be highly tolerant of diversity in times of social crisis such as war. In such an instance the avoidance of diversity would be sought, not the reverse. So it is with other values. A realistic priority of values must reflect the state of the system at the time and place they are to be operative, hence social indicators should not have a necessarily fixed direction of evaluative scoring. Consequently, if a system of normative social indicators are to be used to help guide the setting of national policy, their direction of evaluative scoring should not be fixed, but should have alternative directions specified in advance of use according to what embracing societal context was assumed. The five primary alternative future histories presented here may prove helpful in this regard.

By way of conclusion it seems worthwhile to assert that the way to a desirable future and avoidance of catastrophe will not be found exclusively through "top-down" control of such issues as population, technology, or "law and order." Control will be useful and acceptable only if it is in harmony with the basic cultural changes which seem to be taking place. If deeply help premises and values are to be re-examined and perhaps altered, it is to be expected that social goals, and hence some of the aims of social policy will also change. During the continuing transition, as with any adaptive organism, there will be "error signals" which document various discrepancies between the state of the present system and what is required. Social and policy scientists must help practicing politicians and the populace as well to see these discrepancies as necessary data for social management and not as evidence of failure (hence to be hidden from view). Social indicators, societal reporting and social accounting can help in this task, but only if they are sufficiently flexible and realistic that they adequately describe future possibilities as well as the present.

FOOTNOTES

1. R. A. Bauer (Ed.), Social Indicators (Cambridge: M.I.T. Press, 1966);

   B. M. Gross, and M. Springer "A New Orientation in American Government", (The Annals of the American Academy of Political and Social Science, 371 and 372,

   U. S. Department of Health, Education, and Welfare, Toward a Social Report. Washington, D.C.: U.S. Government Printing Office, 1969).

2. H. S. Rowan, "Some Futures of Operations Research" (Santa Monica, Calif.: Rand Corporation, Report No. P-4001, 1968); Aaron Wildavsky, "Rescuing Policy Analysis from PPBS," The Public Administration Review, (Vol. XXIX (2), 1969, 189-202).

3. E. B. Sheldon and H. E. Freeman "Notes on Social Indicators: Promises and Potential," Policy Sciences, (1, 1970, 97-111).

4. P. J. Henriot "Political Questions about Social Indicators, "Western Political Quarterly (23, 1970, 235-255).

5. The methodological details of this part of the analysis are reported in R. F. Rhyne, "Projecting Whole-Body Future Patterns--The Field-Anamoly Relaxation (FAR) Method. (Santa Barbara, California: Johnson Research Associates; or Menlo Park, California: Educational Policy Research Center, Stanford Research Institute, 1971).

6. Reports which have resulted from analyses of this type include: W. W. Harman, "The Nature of our Changing Society: Implications for Schools," in P. K. Piele, et al (Eds.) Social and Technological Change: Implications for Education in Temporary Society (Eugene, Oregon: The Center for the Advanced Study of Educational Administration, University of Oregon, 1970, 1-67); W. W. Harman "The New Copernican Revolution," Journal of Humanistic Psychology (9 1969, 127-134). N. McEachron, and C. Persico, "Transitional Forces and Social Change in the U.S. 1950 - 1990" (Menlo Park, Calif.: Educational Policy Research Center, Stanford Research Institute, 1970).

7. See for example, H. Kahn and A. Wiener, The Year 2000: A Framework for Speculation on the Next Thirty-Three Years (New York: MacMillan, 1967); S. Chase, The Most Probable World (New York: Harper & Row, 1968); Also various issues of The Futurist.

8. A. Peccei, The Chasm Ahead (New York: Harper & Row, 1968).

9. Kahn and Weiner, op cit.

10. G. Hardin "The Tragedy of the Commons," In G. Hardin, (Ed.) Population, Evolution, and Birth Control (San Francisco: Freeman, 1964). (Also published in Science, 1968, 162, 243-1248).

11. A Toffler, Future Shock (New York: Random House, 1970)

12. W. G. Bennis and P. E. Stater, The Temporary Society (New York: Harper Colophon, 1969)

13. A.Mendel, "Robots and Rebels," The New Republic (Jan. 11, 1969). See also: T. Roszak, The Making of a Counter Culture (New York: Doubleday, 1969)

14. G. Hardin, op cit., Wiesner, J. and York, H. "National Security and the Nuclear Test-Ban" Scientific American, 211(4),(1964).

15. B. Crowe, "The Tragedy of the Commons Revisited," Science (3909) 1969, 166 (1103-1107); D. Michael "On the Social Psychology of Organizational Resistances to Long-Range Social Planning," (Ann Arbor, Michigan: Institute for Social Research, 1970); V. Ferkiss, Technological Man: The Myth and the Reality, (New York: Braziller, 1969).

16. G. Hardin, op cit.

17. S. Lipset, and E. Raab "The Non-Generation Gap," Commentary 50 (1970).

18. Scenarios describing these future histories, together with supporting materials are presented in a forthcoming report, "Alternate

"Future Histories for the United States," which may be obtained by writing the author.

19. By normative social indicator is meant "A statistic of direct normative interest which facilitates concise, comprehensive and balanced judgments about the condition of major aspects of a 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 could not be (normative) social indicators, whereas figures on health or crime rates could." (Department of Health, Education and Welfare, op cit, p. 97)

TABLE 1

FACTOR ROSTER: "EXTERNAL" SECTORS

U.S. Economics

E_1:	Prosperous, expanding-free enterprise
E_2:	Slow growth, stagnant-free enterprise
E_3:	Depression--start under free enterprise
E_4:	Prosperous, expanding--strong government control
E_5:	Unsuccessful government control
E_6:	A non-expanding, successful economy
E_7:	Communalism

U.S. Internal Politics

I_1:	Status quo
I_2:	Increased federal power
I_3:	Shift of power locus to state/local
I_4:	Single-party government
I_5:	Direct democracy, multi-party
I_6:	Cybernetic bureaucracy

Science & Technology

S_1:	Rapidly expanding technology
S_2:	Stasis; elite security
S_3:	Stasis; little advance, much application
S_4:	Active science; shift to behavioral science
S_5:	Active science/technology; anti-pollution focus

U.S. Demographic Patterns

D_1:	Status quo, 300 million by 2000
D_2:	Extreme urbanization
D_3:	Population dispersion, pastoral
D_4:	Like D_3, but technological and connective

World Population/Subsistence

H_1:	Optimistic, "Green Revolution" a success
H_2:	Like H_1, but technical failure of G. R.
H_3:	Like H_1, but G. R. negated by violence
H_4:	Reductions in help to developing nations
H_5:	Population Stabilization without G. R. success

U.S. Foreign Relations

F_1:	Status quo
F_2:	Only AID involvement in underdeveloped world
F_3:	Selective AID/Military involvement in underdeveloped world
F_4:	Isolation re underdeveloped world; involvement with developed nations
F_5:	General isolationism
F_6:	"Manifest Destiny"

TABLE 2

FACTOR ROSTER: "INTERNAL" SECTORS

P: Cultural Pluralism

- P_1: Uniformity
- P_2: Non-Plural Diversity
- P_3: Unified Pluralism
- P_4: Non-Hostile Pluralism
- P_5: Hostile Pluralism

V: Internal Violence

- V_1: Sporadic Crime
- V_2: Pervasive Apolitical Violence
- V_3: Visible, Low-Intensity Insurgency
- V_4: Covertly Supported, Low-Intensity Insurgency
- V_5: Higher Intensity Insurgency
- V_6: Private Armies

A: Profiles of Personal Concerns ("A"ims)

- A_1: Anxiety, Individual Solution
- A_2: Anxiety, Collective
- A_3: USA 1965
- A_4: Achievement Orientation, Individual
- A_5: Achievement Orientation, Collectivity
- A_6: Apollonian Calm
- A_7: Person-Centered Unfolding

M: Personal Morality

- M_1: Punishment Oriented
- M_2: Opportunistic Pragmatism
- M_3: Approval Oriented
- M_4: Fixed Social Order Orientation
- M_5: Contractual Social Order Orientation
- M_6: Transpersonal Orientation

C: Cultural Transmission and Change (E = Enculturation, A = Acculturation)

- C_1: Uninhibited Change (E-,A+)
- C_2: Moderated Change (EO,A+)
- C_3: Neutral (EO,AO)
- C_4: Assimilated Change (E+,A+)
- C_5: Conventional (E+,AO)
- C_6: Tradition-Controlled (E+,A-)

O: Organizations and Institutions (P = Pervasiveness, C = Control external vs internal, S = Strength

- O_1: Strong Mandatory ($P_1C_1S_1$)
- O_2: Weak Mandatory Institutional Orientation ($P_1C_1S_2$)
- O_3: Strong Homeostatic Institutional Orientation ($P_1C_2S_1$)
- O_4: Weak Homeostatic Institution Orientation (Relaxed Norms) ($P_1C_2S_2$)
- O_5: Non-silent Minorities ($P_2C_2S_1$)
- O_6: "Laissez Faire" ($P_2C_2S_2$)

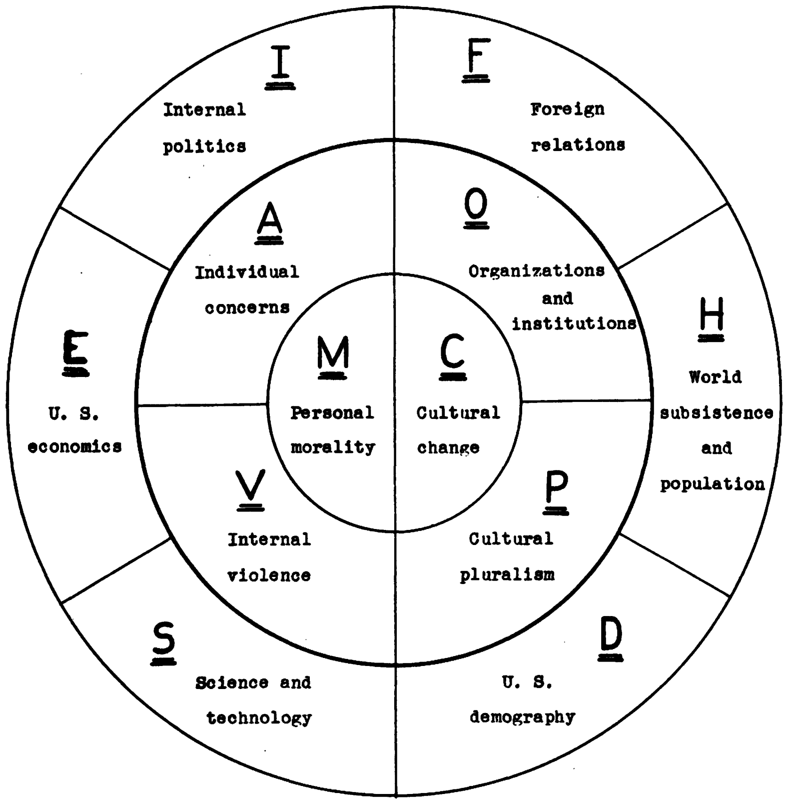


FIGURE 1. ASPECTS OF U.S. SOCIETY FOR GENERATING ALTERNATIVE FUTURE HISTORIES.

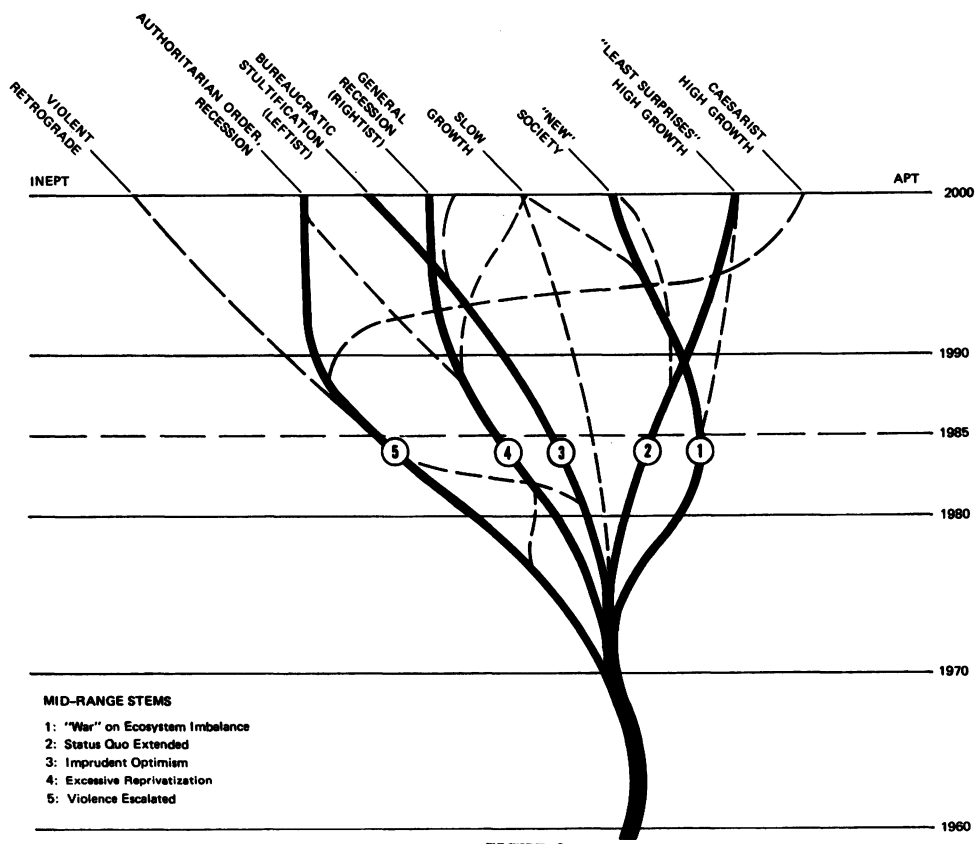


FIGURE 2

"Tree" of alternative futures
(Apt-Inept dimension)

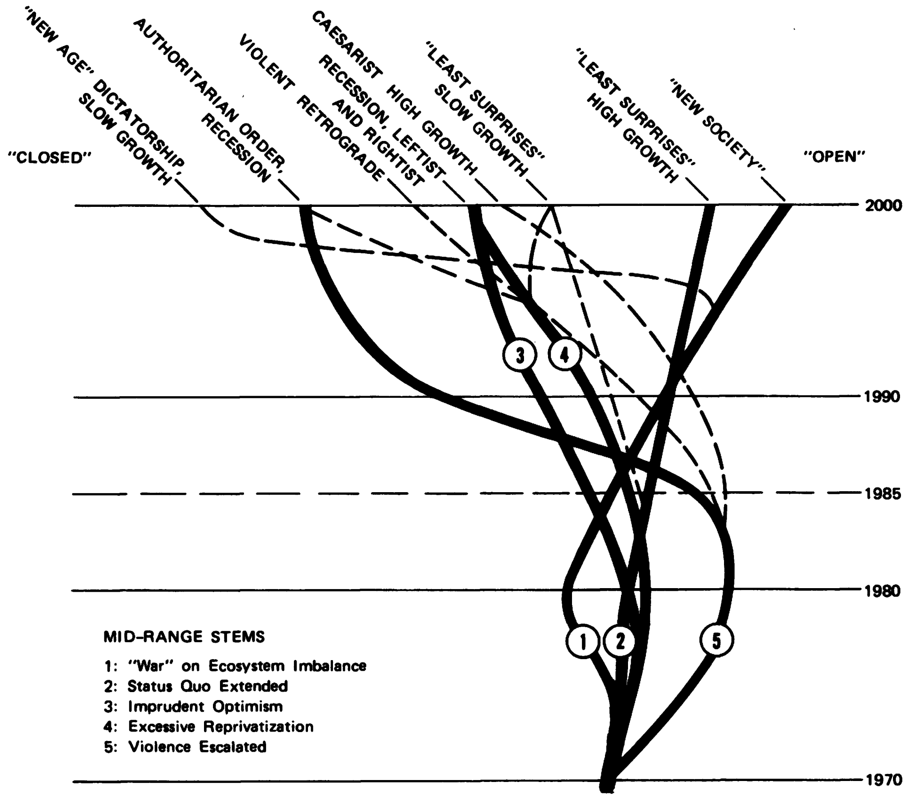


FIGURE 3
"Tree" of alternative futures
(Open-Closed dimension)

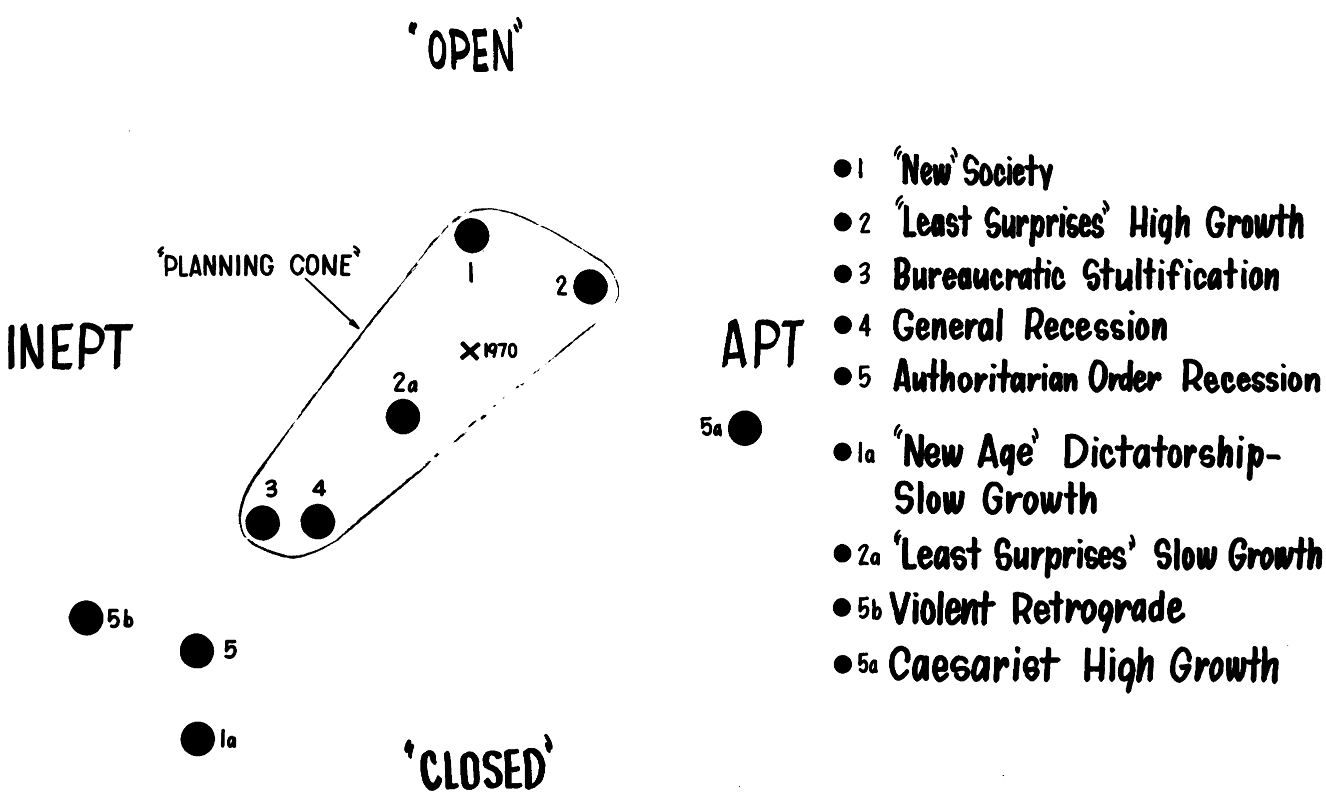


FIGURE 4
Year 2000 slice of "tree"

DEVELOPING A SOCIAL STATISTICS PUBLICATION

Daniel B. Tunstall, U.S. Office of Management and Budget

In the past few years, there has been a heightened interest in using quantified measures to explain social change. Profound changes have occurred within society that no one has been able to identify clearly nor to explain, much less devise means to deal with effectively. Economic progress of the past decade has not solved all social problems. No one ever expected it to, but neither did we expect that so many difficult and controversial problems would be left unsolved in a modern society if the rate of real economic growth was sustained at levels of five percent or more.

A further anomaly is that, although progress has been made with the assistance of government in many social areas -- delivery of medical care to the aged, payments to the needy, access to higher education, school desegregation, transportation, leisure time, and general mobility of the American people -- these advances have not kept us from experiencing major upheavals and discontent. The improvements themselves may even have helped bring on new and even more difficult problems. Not only are people concerned with the availability of social goods and services but they are now concerned about the quality of life of which those goods and services are a part. And they are beginning to ask not only how well existing systems meet social needs, but also whether entirely new ways of doing things might not be better.

Average life expectancy in the United States has remained relatively constant during the past 10 to 15 years, but with females having considerably higher levels than males, enough so to make us wonder why there is such a gap. Could male life expectancy be raised to the level of females or higher with better medical care and more preventive medicine, or will a change in male life expectancy come about only with a change in the life style of males? The environmental crisis is a problem of economics and values. People understand the importance of preserving open spaces and wilderness but to do so means not only great expense but changes in attitudes and life styles. To clean up what is now being dirtied can be done once, maybe twice, but it is too expensive and too dangerous to continue to pollute and clean up afterwards. What is a satisfactory balance between the use of resources and the disposal of wastes and the need to preserve and protect natural resources? Similar kinds of questions are being raised about education, employment, recreation, and almost all other areas that impinge on individual well-being.

If rapid economic growth and technological progress are not synonomous with quality, as it was long assumed they were, then we need to develop information that will permit us to examine the quality of life and its component parts -- information that will add to productive national debate and permit decision-makers to make their decisions on more than economics, pragmatic politics, and intuition. The quality-of-life issue may be divided into two dimensions: the objective, which includes conditions of the environment (housing, roads, recreational resources, ecological balances, etc.) and attributes of persons (health, educational achievement, family stability, etc.) and the subjective, which includes aspects of personal experience such as frustration, satisfaction, aspirations, and perceptions. The distinction between objective and subjective is drawn from ideas expressed by Eleanor Sheldon.

Three works have been extremely influential in suggesting new approaches to information. They are Social Indicators, edited by Raymond A. Bauer (Cambridge, Mass. and London: MIT Press, 1966); Indicators of Social Change: Concepts and Measurements, edited by Eleanor Bernert Sheldon and Wilbert E. Moore (New York: Russell Sage Foundation, 1968); and HEW's Toward a Social Report (Washington: Government Printing Office, 1969). These studies mark the beginning of a movement to identify social indicators that may not only help us monitor social systems but also develop entirely new means for analyzing social problems.

Underlying any new approach to social analysis, however, is the need for a foundation of well-organized and carefully selected social statistics that will appear periodically and be open to revision and change. The social statistics publication now being prepared by the Statistical Policy staff in the Office of Management and Budget is designed to be just that. It draws on the three studies mentioned above as well as the work of Mancur Olson, Stafford Beer, Nestor Terleckyj, Allen Schick, Milton Moss, and other innovative thinkers in the field.

Our publication will be concerned only with the objective dimensions of individual well-being. It will be composed almost entirely of statistical series presented in chart form with accompanying tables. The information will be drawn from existing statistical series, mainly those produced by government agencies, and important single-time series. It will be post-World War II national data, with totals broken down by race, sex, age, and, in some cases, region of the country or SMSA/non-SMSA. The degree of disaggregation will be based on our estimation of the usefulness of the components. Projections will be included only for relatively known quantities, such as number of pupils enrolling in primary school five years from now. There will be almost no information on peoples' preferences or their attitudes toward national social problems. The text will be limited to chapter introductions and short notes explaining the technical problems and limitations of the data; there will be no interpretation of the statistics included.

The Statistical Policy staff began work on this publication in mid-1969. By February of 1970, we had a working outline, and, by September, three draft chapters were ready for review by our advisory groups and consultants. We hope to distribute a limited-edition, preliminary report during the coming summer.

We have organized statistics into nine broad categories -- health, public safety and legal justice, education, employment, income, housing, transportation, environment, and recreation. Basically, these categories, or social goals areas as we called them in earlier outline drafts, comprise a rough listing of basic human needs or wants that are intertwined with public policy.

Obviously, not every social area of national concern is treated separately. It has been suggested that we examine social opportunity or equal opportunity as an independent area. (HEW's Toward a Social Report included chapters on social mobility and participation and alienation.) We have not done so because it seemed that equal opportunity may be understood as opportunity within a specific area, such as employment or education, and so should be included under the respective category for continuity. Within each chapter, we are including detailed information on race, sex, and age. If others want to rearrange this information to construct a comprehensive picture of whether equality of opportunity in the United States is growing, they will have some of the tools to do so.

Essentially, though, these broad social areas are an organizational device. It is likely in the future that we will add new categories or combine some of those already identified, especially if we find that we can not develop them to our satisfaction. Of far more importance to the publication are the statements of what we are measuring within each category, of what we have labeled, for lack of a better name, social concerns.

Figure I is a summarized outline of the social concerns we are examining in five areas. All are stated positively and all reflect the state of affairs at the level of the individual.

Figure I

Social Concerns

Health

1. Long life
2. Physical and mental well-being

Public Safety and Legal Justice

1. Safety of life and property from crime
2. Adequate legal safeguards for the accused

Education

1. Basic skills for everyone
2. Educational opportunity for those able and interested in pursuing advanced learning

Employment

1. Employment opportunities for all who want work
2. Adequate earnings
3. Job satisfaction

Income

1. Rising general level of real income
2. Equitable distribution of income
3. Adequate income

These concerns have been broadly defined to reveal the general well-being of the entire population, not to represent immediately pressing social problems nor the problems of any particular group within society; to depict concerns that can be dealt with by national policies; and to encompass most of the important national social issues. More practically, they serve as a mechanism for sorting the tremendous amount of information available on society and social change and for clearly spelling out where meaningful information is lacking.

Obviously, the concerns within a category overlap. Yet it seems both logical and important to distinguish between, for example, health conditions and long life, for only by isolating important concerns and measuring them as best as we can will we be able to determine the interrelationships among them and to understand how the pursuit of one affects the outputs of others.

Examples from the employment, income, and education chapters will help make clear what we have identified as social concerns. The statistics selected to measure these social concerns in three draft chapters are listed in the Appendix.

Employment opportunities for all who want work, the first employment concern, was explicitly stated in the 1946 Employment Act in which the government was given the task of "creating and maintaining . . . conditions under which there will be afforded useful employment opportunities, including self-employment, for those able, willing, and seeking work." We broadened that goal to include all who want work, whether able or not. Although many individuals do not want employment -- youths in school, adult women who choose to devote their time to household and volunteer duties, retired persons -- there ought to be a range of opportunities to permit all who seek work, full-time or part-time, to find it. This concern as stated is so broad that we have included a breakdown of information to measure labor force participation (Employment Status of Non-Institutional Population 16 Years and Over, Total Labor Force Participation Rates by Selected Age-Sex Groups, etc.), unemployment (Unemployment Rates for All Workers and for Married Men, Unemployment Rates by Race, Unemployment Rates for Sex and Age Groups by Race, etc.), underemployment (Persons on Part-time for Economic Reasons, Men 20-64 Working Part-Year by Major Reason, etc.), special social problems limiting an individual's opportunity for employment (Number of Migratory Workers and Days Worked at Wage Work per Year by Farm-NonFarm, Unemployment Rates of Male High School Dropouts and Graduates for Selected Age Groups, etc.), and those wanting work but not out looking for it, the so-called hidden unemployed.

Two other social concerns within the employment area are adequate earnings for those employed -- a broad measure of the extent to which the economy and society provide jobs with earnings sufficient to maintain various levels of living -- and job satisfaction. Adequacy of earnings is measured for our purposes by estimating the percent of full-time year-round workers with annual earnings, in constant dollars, below the minimum wage, the poverty level, and the Bureau of Labor Statistics low budget. Job satisfaction is more difficult to measure. We have included statistics on working conditions -- level of earnings, risk of unemployment, racial composition, work injury rates, and holiday and vacation time of various occupational groups -- as proxy measures of satisfaction.

The first income concern is also direct and specific: rising general level of real income. This means that income should rise over time, that it should rise for everyone, and that it rise in real terms. The measure is per capita personal income before and after taxes, in constant dollars, for the years 1947 to 1969.

A second concern in the income category is that there be equitable distribution of income. Although not as widely accepted as other income concerns, there is little question that most people would like to see those at the bottom receive more relative to others. Our income tax policies as well as certain government transfer payments rest on a similar assumption. Series included are: Mean Family Income by Quintiles of Families and Top Five Percent; Ratio of Mean Income of Each Quintile of Families to the Mean of All Families; Median Family Income by Race of Head and Ratio of Negro and Other Races to White by Median Family Income, etc.

A third concern is with adequacy of income. How many people have income insufficient to permit them to maintain a decent standard of living? There is consensus that those at the bottom of the income ladder should be helped out, but not over how high the "bottom" goes. In this case, the choice of a statistical series is critical. We have included statistics on the number and percent of persons in poverty (as defined by the federal government) and those with annual income lower than the BLS low budget estimate for the years 1959 to 1969. Characteristics of the poor are given by race, sex, type of family unit, age, and location. Details are also included on the changing composition of the poor and on the size of income deficit for families in poverty.

Not surprisingly, in both the employment and income areas, the social indicators we have selected are important economic indicators as well. Clearly, having the opportunity to work at a meaningful job and the resources with which to buy goods and services are important measures of well-being in the United States.

In identifying social concerns, we have drawn heavily on the unpublished Materials for a Preliminary Draft of the Social Report, prepared for the Panel on Social Indicators of HEW (April, 1968) and the resulting publication Toward a Social Report, and on the work of Nestor E. Terleckyj, some of whose most important ideas appeared in the recent article "Measurement Possibilities of Social Change" in Looking Ahead (the monthly report of the National Planning Association), August, 1970.

In our first outline, we distinguished between performance indicators, those statistics that would best measure the concerns statements, and what we labelled analytical information, statistics less useful for measuring the national picture but important for policy analysis because they pinpointed the groups most affected or identified causes of change in the performance indicators. However, when we put together a draft chapter, the distinction broke down and we found it much easier simply to arrange the information by social concern.

The choice of statistics has been based on two criteria. The first is that the data measure individual well-being. For each series, national totals are given in terms of persons or families and then broken down to reveal the age, sex, race, and other characteristics of those involved.

It has been relatively easy to meet this first criterion in health (Percent of Population with Chronic Ailments, Rate of Infant Mortality, for example), income (Per Capita and Family Income), education (Percent of Age Group Graduating from High School and Going on to College), and employment (Percent Unemployed or Underemployed). Within public safety, we can measure how many people are affected by violent crime (Rate of Victimization), but we do not have information on how violence done to one person affects others in the community, nor do we know how to describe the effect property crime has on those victimized.

It is simpler, although less satisfying and less meaningful, to produce statistics on the changing dollar value of crime than on the emotional costs of violence.

As we develop chapters on housing, transportation, and the environment, it would be tempting to rely on statistics that measure things: number of housing units, miles of road, number of automobiles, major pollutants in the air and water, acres of wilderness. But this information does not measure the objective dimension of well-being. For that, we need information even more difficult to find -- the percent of persons without adequate housing; the number living in poor neighborhoods; the percent having access to rapid, local transportation; or the percent of the population subjected to hazardous pollution. Nor do we have a quantifiable measure of an individual's need for wilderness and open space.

But we are not simply selecting information on people; we are also collecting information that reflects outputs. In the case of transportation, we need a measurable concept of access. We need standards for neighborhood quality and, in environmental matters, we need to know the levels at which major pollutants become harmful before we can determine the percent of the population subjected to hazardous or bothersome pollution. It may turn out that the information included in these chapters initially is less important than the fact that we will have identified some very important conceptual and statistical gaps.

I have already mentioned the second criterion -- that the statistics measure outputs. In the past, we were content to look at the processes by which social goods and services were produced. It was expected that the goals of organizations and bureaucracies were synonymous with the needs of individuals, and that, if competent managers ran those institutions, the products would then meet the needs of the people. These expectations carried over to the kinds of information considered important to collect and publish. In education, for example, there exists a great deal of data on enrollments, expenditures, teachers, buildings, and related input indicators, such as pupil-teacher ratios and per pupil expenditures, but little information on the important variables associated with learning, on the relative value of different learning experiences, or on just how well our education system is working toward its primary purpose of instruction.

So far, finding measures of outputs has been most difficult for the education chapter. We have chosen statistics from the Survey of Educational Opportunity (the Coleman Report) to measure the relative educational achievement between race and age groups. We may also be able to include statistics from the first report on school achievement of noninstitutionalized children 6 - 11 years of age, based on data from the Health Examination Survey of 1963-65. We lack adequate information on student achievement in universities and on changing values in higher education. However, we do have statistics from a number of one-time studies (Project Talent, SCOPE, the Preliminary Scholastic Aptitude Test and Academic Interest Measures, known as PSAT-AIM, conducted by the Educational Testing Service) which, when combined with statistics from Census' Current Population Survey, provide information on opportunities for students to attend institutions of higher education.

In many cases, though, the information we have included in the education chapter is not really adequate. We have compromised by including more statistics on school attendance and high school graduation than on student achievement, and we have no information on college achievement other than guaduation rates.

How can this collection of statistical information be put to work? We envision two main uses for the publication: as a tool in decision-making and as a guide in developing social statistics.

Policy-makers must now deal with technological changes in transportation, energy production and transmission, communications, and many other systems that transform scientific invention into mass-produced technological change for millions of people. Assessing the effects of future technological developments in light of the present social system is not easy. Social indicators may provide information on which planners can build a model of society in order to estimate its requirements for transportation, energy, communication, etc., and from which they can begin to understand the implications for individuals of massive new technological systems. In short, our publication may help them to adapt technology to man's expressed needs and values.

To assess technology, planners need broadly based social information for making decisions with pay-offs far in the future. Government, with its almost daily development and analysis of policy, faces decisions with more immediate consequences. In the past few years, the federal government and state governments have developed programs to lighten the burden of health expenditures for the poor and the elderly. One result has been an increased demand for health services, with rising costs for everyone. However, with only a brief examination of major health indicators currently available -- life expectancy at birth, rate of chronic disability, rate of mental illnes, etc. -- it is difficult to believe that health has improved commensurate with the rapid rise in cost. Soon there will be another round of national and state decisions on medical-care systems and the delivery of medical care, although this time the focus will probably be on paying some of the bills, at least the most expensive ones, for everyone.

The problem here is not simply one of efficiency of distribution, although it is obvious that, if medical care is delivered as it has been in the past, costs will not go down. To meet the increased demand for more services at reasonable costs, there will have to be new systems. And as these systems are developed, basic questions will be raised: What is good health? How much more health output is technically feasible? What kind of care is necessary and desirable to provide greater health output? And, of course, what is the most efficient way of dispensing health care?

Social indicators can provide some of the information needed to develop answers to these questions. Alternative policies will need to rest on a common information base, which is likely to be composed of measures of effectiveness, that is, measures of health output.

Average years of life expectancy, at birth and at other ages, rate of infant mortality, and age-adjusted death rates, along with statistics on the major causes of death, provide information needed to measure the most serious level of health achievement. The percent of population with chronic mental and physical disability measures another important health output, as does rate of acute illnes and injury and levels of nutrition and fitness.

As the major causes of serious illness and death become scientifically linked with human behavior and less with communicable disease, it may be important for health organizations, including the government, to make the public aware of the health hazards of alcohol, drugs, pollutants, improper diet, obesity, smoking, and other physically and mentally debilitating activities. There may be significant trade-offs between expenditures on hospital facilities to care for heart disease and cancer patients and expenditures for preventive care facilities that would permit early detection of these diseases, or offer exercise and weight-reducing programs, or deal effectively with drug addiction or provide any one of a number of services that would reduce the need for medical care. Social indicators of health, if developed sufficiently to meet a wide variety of information needs, can provide a set of measures against which the government and private citizens can evaluate their changing system of medical and health care. Indicators of education, employment, housing, and other areas of social concern can be used to serve the same purpose.

In addition to its usefulness in decision-making, the information gathered in this publication can serve as a guide for the development of new social statistics. At present, the federal statistics community has few means with which to evaluate its social statistics program. It is particularly lacking in ways of incorporating suggestions from policy-makers and researchers to improve information relevant to their needs. One result of this lack of communication is that policy-makers have become satisfied with relatively little quantified social information. What is available is often administrative data produced from ongoing programs. This data is often insufficient to allow one to comprehend the magnitude of important social needs.

It may appear rather circuitous to say that the development of a social statistics publication or a set of social indicators will help to improve the main body of social statistics. However, as I mentioned earlier, we find in many instances that there are no adequate statistics to illuminate the social concerns we have identified and that we are confronted with inconsistent or conflicting information. One of the principal reasons the Statistical Policy staff is actively engaged in this project is to identify statistical gaps and shortcomings of this kind. We believe the approach and structure developed for this publication will provide an excellent framework for a thorough examination of the government's social statistics program and serve as a basis for directing additional resources or reallocating present ones to meet the problem-oriented demands for information that are heard every day.

APPENDIX

EDUCATION OUTLINE

A. Basic Skills for Everyone

Enrollment and Graduation

1. Number Enrolled in Elementary and Secondary Schools, Public and Private, 1950-1974 (projected)
2. Percent of Persons Enrolled in School by Age and Race, 1962-1968
3. Number of Students Graduating from High School, 1948-1968

4a. Percent of 17 Year Olds Graduating from High School, 1946-1968

4b. Percent of 17 Year Olds Graduating from High School by Race, 1960-1968

5. School Retention Rates from Fifth Grade, for Selected Years (percentages)
6. Percent of Persons 25-29 and 25 and Over with Less Than Five Years of School by Race, 1947-1967
7. Percent of Persons 25-29 and 25 and Over with Four Years of High School by Race, 1947-1967

Measures of Achievement

8. Test Scores from the National Assessment of Educational Progress, 1970 (Average raw score of ten Science exercises)
9. Test Scores from the Survey of Educational Opportunity (Coleman Report) 1966, by Grade Level Equivalent for Pupils in Sixth, Ninth, and Twelfth Grades by Race and by Socio-Economic Status of Parent

B. Opportunities for Advanced Learning

Enrollment

10. Number Attending Institutions of Higher Education by Degree-Credit Status and Level, 1957-1969
11. Number Enrolled in 4-Year and 2-Year Degree-Credit Programs by Sex, 1957-1970
12. Percent of 18 to 21 Year Olds Enrolled in Undergraduate Degree-Credit Programs by Sex, 1957-1969

13a. Number Enrolled for the First Time in Degree-Credit Programs by Age and Sex, 1947-1968

13b. Percent Enrolled for the First Time in Degree-Credit Programs by Age and Sex, 1947-1968

14. Percent of Twelfth Grade Students Going on to College, 1947-1968
15. Number and Percent of High School Graduates Enrolling in College the Same Year by Sex, 1960-1967
16. Percent of High School Graduates Attending College by Educational Ability and by Socio-Economic Status of Parents, 1960 (Project Talent)
17. Entrance Rates to 2-Year and 4-Year Institutions by Ability and by Socio-Economic Status of Parents, 1967 (ETS Growth Study)

Graduation

18. Number of Earned Degrees by Level and by Sex of Student, 1948-1968
19. Bachelor Degrees as Percent of 23 Year Olds and as Percent of High School Graduates Four Years Earlier, 1948-1967
20. Percent of Students Receiving Bachelor's Degrees Who Were Enrolled in Degree-Credit Programs Four Years Earlier by Sex
21. Percent of Persons 25-29 and 25 and Over with Four or More Years of College by Race, 1948-1967

EMPLOYMENT OUTLINE

A. Employment Opportunities for All Who Want Work

1. Employment Status of Noninstitutional Population 16 Years and Over, 1950, 1960, 1969 and Projected for 1980 (millions)
2. Major Changes in the Labor Force by Age, 1960's (estimated) and the 1970's (projected) (millions)
3. Total Labor Force Participation Rates, Selected Age-Sex Groups (annual averages; percent), 1947-1968
4. Labor Force Participation Rate of Married Women under 35, by Presence and Age of Children, March 1960, 1964, and 1969 (by labor force as percent of population)

Unemployment

5. Unemployment Rates for All Workers and for Married Men (annual averages; percent) 1947-1968
6. Unemployment Rates by Race and Ratio of Negro and Other Races to White, 1948-1968
7. Unemployment Rates for Sex and Age Groups by Race (annual averages), 1954-1968
8. Number Unemployed by Sex and Age Group by Race 1960 and 1969 (annual average; thousands)
9. Duration of Unemployment and Long-Term Unemployed as a Percent of Total Unemployed, 1947-1968

Underemployment

10. Persons on Part-time for Economic Reasons, 1957-1968, by Sex (thousands), 1959-1968
11. Men 20-64 Working Part-Year by Major Reason, 1960, 1965, and 1968 (percent of total)

Persons on Fringe of Labor Market

12. Persons Not in the Labor Force Who Want Jobs

Special Employment Problems

13. Number of Migratory Workers and Days Worked at Wage Work per Year by Farm - Non-Farm, 1960-1968
14. Unemployment Rates of Male High School Dropouts and Graduates for Selected Age Groups, 1962-1969
15. Number of Persons 16-64 Years Not Working During Year Because of Illness or Disability by Race and Age, 1968 (thousands)
16. Number of Mothers in the Labor Force with Husband Present and with Children under Six Years (millions), 1950-1968
17. Child Care Arrangements of Full-time Working Mothers, 1964 (percent of children of full-time working mothers)

B. Adequacy of Earnings

18. Estimated Percent of Full-time, Year Round, Workers with Earnings Below Certain Levels (1968 dollars), 1968
19. Percent of Full-time, Year Round Workers with Earnings Below Minimum Wage and BLS Low Budget, by Sex and Race
20. Percent of Full-time, Year Round Workers with Earnings Below Minimum Wage and BLS Low Budget, by Age and Family Size
21. Percent of Full-time, Year Round Workers with Earnings Below Minimum Wage and BLS Low Budget, Selected Occupational Groups and Educational Levels
22. Percent of Full-time,Year Round Workers, Who Were Heads of 4-Person Families with No Other Earners, Below BLS Low Budget

C. Job Satisfaction

23. Employment Trends Among Major Occupational Categories 1947-1968 (actual) and 1980 (projected for a services economy with 3 percent unemployment)
24. Median Earnings by Occupation: Year-Round Full-time Workers in 1968 by Longest Job and by Sex (dollars)
25. Percent of Workers Experiencing Some Unemployment in 1968 by Longest Job and by Sex
26. Proportion of Negroes and Other Races by Occupation Groups, 1960 and 1969 (percent of total for each occupation group)
27. Work Injury Rates in Selected Industries, 1950-1968
28. Average Days of Disability per Injury in Selected Industries, 1955 and 1968 (number of days)
29. Average Number of Paid Holidays Provided Plant and Office Workers, 1960 and 1968 (days)
30. Proportion of Private Non-Farm Office and Plant Workers Receiving Paid Vacations by Length of Vacation, 1968 (percent of workers in each category)
31. Life and Work-life Expectancy at Birth, by Sex, 1940, 1950, and 1960 (years)

INCOME OUTLINE

A. Rising General Level of Real Income

1. Per Capita Personal Income Before and After Taxes (in 1968 dollars), 1947-1969

B. Equitable Distribution of Income

2. Mean Family Income by Quintiles of Families and Top 5%(thousands of current dollars), 1947-1970
3. Ratio of Mean Income of Each Quintile of Families to the Mean of All Families, 1947-1969
4. Ratio of Mean Income of the Top 5% of Families to the Mean Income of the Lowest Fifth, 1948-1968
5. Median Family Income by Race of Head (thousands of current dollars) and Ratio of Negro and Other Races to White by Median Family Income, 1948-1968
6. Median Family Income by Race and Age of Head (thousands of current dollars), 1959 and 1968
7. Median Family Income by Sex of Head (thousands of current dollars) and Ratio of Male to Female Family Heads, 1948-1968
8. Median Family Income by Age of Head (thousands of current dollars), 1948-1968

C. Adequacy of Income

9. Number (millions) and Percent of Persons in Poverty and with Annual Income Lower than BLS Low Budget Estimate, 1959-1968
10. Number of Persons in Poverty by Race (millions), 1959-1968
11. Number of Persons in Poverty by Living Arrangement (millions), 1959-1968
12. Number of Persons in Poverty by Age (millions), 1959-1968
13. Number of Persons in Poverty by Residence (millions) by Rural-Urban; Farm-Nonfarm; Inside SMSA-Outside SMSA; Central City-Suburban Ring; 1959-1968
14. Number of Families in Poverty by Sex and Race of Head (millions), 1959-1968
15. Percent of White Persons and Negro and Other Persons in Poverty, 1959-1968
16. Percent of Males and Females in Poverty, 1959-1968
17. Percent of Each Age Group in Poverty, 1959-1968
18. Percent of Farm Population in Poverty, 1959-1968
19. Percent of Persons in Poverty by Region and Race, 1959-1968
20. Changing Composition of the Poor, Selected Groups - Race, Age, Region, Family Relationship (percent of total poor persons), 1959 and 1968
21. Changing Composition of Poor Families, Selected Groups (Sex of Family Head), (percent of total Poor families), 1959 and 1968
22. Size of Income Deficit for Families in Poverty (percent of total poor families), 1959 and 1968

LOCAL SOCIETAL INDICATORS: A PROGRESS REPORT

Eugene D. Perle*, Wayne State University

In the last several years a growing body of literature usually referred to as social indicators has appeared on the academic scene. On occasion it carries alternative labels, such as social accounts, social reports, or monitoring social change.[1] This literature focuses on the measurement and quantification of societal issues in an effort to utilize social science knowledge for public policy purposes. Those who have written about indicators explicitly are few in number, primarily representing the fields of political science, public administration, and sociology. Thus, the literature tends to emphasize a few major themes, almost all directed at a macro scale of analysis, covering the entire nation.

More recently there has emerged a growing interest in urban indicators, which entails the application of social indicator approaches to urban areas and metropolitan regions. This development is taking place before many of the issues associated with social indicator approaches have been resolved. The field is still young, the number of seminal ideas contained in the literature is small, and the number of suggestive uses of indicators is increasing rapidly. These recent developments are intimately related to what we consider urban problems requiring 'solutions,' but even the list of problems is changing rapidly.

Only a few years ago what were considered urban problems focused upon issues related to growth and development. Approximately fifteen years of experience has resulted in the development and testing of reasonably sophisticated models dealing with urban structure, form and growth.[2] Presently we consider issues relating to lawful behavior, political efficacy, environmental quality, segregation, health, ethnicity, education, employment-unemployment, housing, transportation, aspirations, expectations, perceptions, values, objectives and goals, as all constituting urban problems. Of course, there is nothing inherently urban about this list. These issues relate to ethical questions about the structure of society and the distribution of societal benefits and costs. Insofar as we have become an urbanized nation and insofar as certain groups in society are heavily urbanized, many of these ethical issues have taken on the character of urban problems. The issues are real, the need for 'solutions' is evident, and the application of urban indicators for policy purposes is being explored.

*The author acknowledges the contribution of Mr. Zvi Maimon in the preparation of this paper.

INDICATOR PERSPECTIVES

The literature on indicators is so recent that there is little consensus at present about what indicators are, what they should be, and how they are to be utilized. Some of the topics appearing in the literature which have provoked disagreements, include:[3] the necessity for and utility of indicators of direct normative interest, the utility of indicators for the setting of goals and priorities, the utility of indicators for the evaluation of social programs,[4] and indicators contained within a system of social accounts.

Underlying the dialogue concerning the usefulness of indicators are two groups of persons: (1) those who are extremely optimistic about their immediate utility for a wide class of societal issues, both in the public and private sectors, and (2) those who are cautiously optimistic about their eventual use for societal issues, subsequent to an intensive period of fundamental social science research. The first group is composed of public officials as well as academics who are closely related to the political system; the second group is primarily composed of academics who are interested in understanding the structure and functioning of social systems, but are not intimately related to the political system in a professional sense. This categorization reflects an expected distinction between social science activists and research oriented scholars. The distinctions between these two groups is also reflected in their research strategies and writings. By and large, the first group concentrates upon macro social indicators, for the entire nation, and some very general concepts about their utility. On the other hand, the second group tends to be more empirically oriented in terms of operationalizing indicators and, therefore, micro indicator research is more common within this set.

There are at least several major themes which permeate indicators literature in terms of their usefulness.[5] They include: (1) improved descriptive reporting on the state of society, (2) the analysis of social trends and social change, (3) assessing the performance of society, (4) anticipating alternative social futures, and (5) social knowledge for societal control. Obviously, these five themes have significant interdependencies. Unless we have good descriptive reporting for the data base it becomes extremely difficult to analyze trends and social change, performance assessments yield questionable validity, it becomes almost impossible to anticipate societal futures, and the ability to exercise some measure of control over social processes becomes hopeless since there is little understanding of where to intervene in the system.

Although we have not been terribly effective in the utilization of available information for social policy purposes, there is a need for better descriptive reporting on the state of society. Unless the descriptive data base is comprehensive, valid, meaningful, and forthcoming on a regular basis, it becomes difficult to proceed further. Perhaps the most serious problem facing indicator research and utilization, however, centers upon conceptual requirements -- what should be measured and why.

The use of indicators for the analysis of social trends and social change was originally politically motivated; more recently it has become research oriented. In its scientific orientation it represents a basic intellectual concern for understanding how social systems operate and developing and testing hypotheses about the determinants of social change. In the long run, this activity will hopefully yield some knowledge about social systems behavior, the determinants of social change, and the design of appropriate policy for social problems. In the short run it probably has little applied utility.

It has been suggested by some writers that indicators in a social accounting framework can be utilized to assess the performance of society. However, in order to use indicators for normative assessment purposes it is necessary to have the specification of goals against which actual performance can be compared as well as the specification of desirable and undesirable system states. Writers expressing this point of view usually avoid suggesting how normative criteria are to be developed, by whom, and for whom. Given the diversity of the ever changing goals in our society, it is not at all clear that meaningful normative criteria can be developed for complex social systems or that their value will hold for more than very brief periods.

Anticipation of alternative social futures has been suggested as the logical outcome of a process based upon a broad array of indicators. Clearly, forecasting alternative futures for differing sets of policies is directly linked with detailed and accurate understanding of existing system states and their determinants. Otherwise, futurism becomes an activity for self-appointed sages with unique powers. Thus, for futurism to be successful it is necessary for the analysis of social trends and social change to become better developed and more comprehensive in scope.

The use of social knowledge for societal control is an expression of managerial thinking. This view suggests that we can develop assessments of the social order, predict futures, and place social processes in a control theory setting. Though it may be possible to control some major activities of business firms (large or small) and even governmental agencies, it is far less likely that we are or soon will be able to control complex social systems. This entire concept is completely elitist in attitude, and frightening. Who will do the controlling, for what ends, using what means, and who are to be controlled and why are issues usually avoided. The answers to these problems are assumed to be self-evident truths representing an undefined, but broad based popular consensus.

APPLICATIONS

Some of the themes regarding the use of indicators, previously mentioned, may appear to be brilliant conceptions or ludicrous suggestions. It is much too premature for us to settle on one or two approaches while discarding others. Their validity and feasibility are yet to be determined. An intensive period of empirical testing is required for all the approaches suggested in the literature as well as new ones that may be forthcoming. Even those approaches which have received the most favorable comment for their conceptual formulations have yet to be verified.

In order to justifiably realize the promise of indicators, not only is it necessary to suggest apparently brilliant conceptual models but it is also necessary to empirically verify them. Without an active process of empirical testing for model specification, validity and reliability, most of our conceptual models will continue to lie on the shelf for conversation and intellectual purposes alone. Clearly, the search for knowledge can be initiated either deductively or inductively. At some point in the process, however, it is necessary both to empirically verify deductive propositions and theorize about empirical findings.

One of the major difficulties associated with the types of indicator approaches suggested in the literature is their concentration upon highly aggregative forms of analysis spanning the nation. Such an emphasis, which has some theoretical appeal for certain classes of problems, has relatively little utility for application in any specific geographical area. Information designed to represent national trends usually is not constructed by aggregating local or small area totals. Thus, the intra-urban and inter-urban diversity present in our cities and urban regions is completely masked by the national totals and it is impossible to disaggregate back to the local scene. However, the formulation of effective public urban policy at both national and local levels is largely dependent upon understanding the nature of these diversities.

Since we have become a nation of urban regions and since this process probably will continue in an accelerating fashion, it is likely that large payoffs can be derived from the development of urban indicators, both theoretically and empirically. At this level of specificity indicators begin to reflect the issues confronting the real world, where their ultimate utility holds much promise. Every one of the indicator approaches mentioned earlier has an urban analogue. Thus, it is possible for analytical purposes to define a national system of urban regions in terms of their inter-connectedness; a set of urban regions in terms of their individual similarities and differences; and small areas within a particular city region. This shift in emphasis represents a movement from vague abstractions to reality-oriented concepts of planning and design.[6]

A DEVELOPING PROGRAM

Over the past two years the Center for Urban Studies (CUS) at Wayne State University has been attempting to develop a wide-ranging program of research, education and service relative to societal indicators. The program includes work in both social and economic indicators. Simultaneously, there has been a shift in emphasis away from macro societal indicators representing national conditions to local societal indicators, focusing upon neighborhoods, communities, and the metropolitan region. The activities and interest pursued vary from theoretical frameworks and concept formation to empirically based testing of hypotheses and concepts. Thus, we have become more problem oriented in our search for understanding social systems behavior and the role of indicators rather than developing holistic designs for a general system to ascertain metropolitan well-being. It is our hope, however, that this problem oriented focus, over time, will result in a collection of research and understanding that may be the basis for a more general system of local societal indicators.[7]

The development of this indicators program at CUS began with the work of Gross and Springer (1969, 1970). In their writings they are essentially concerned with macro societal indicators useful for national policy planning. The conceptual utility of indicators which they advocate for planning and policy analysis is intellectually appealing; however, the argument is dependent upon the existence and specification of a general theory of social system behavior. To my knowledge no such system is sufficiently well developed yet to permit the implementation of their ideas.

More recently Porter (1970) has developed a model of resource mobilization, in which he analyzes the development of federal aid programs and the resource flows to their intended beneficiaries. He has been able to isolate three different patterns of resource flows: (1) symbolic allocation, (2) catalytic allocation, and (3) perfect allocation. The first pattern occurs when funds released by an earmarked grant are used for purposes other than those stipulated by the donor. The second case, catalytic allocation, occurs when a grant serves to attract additional funds into the aided program. Perfect allocation occurs when the full amount of aid (no more or no less) is added to the "normal growth" of a given program. Porter's model was developed through an analysis of federal aid to education. While the model framework has a macro-orientation, the concepts and criteria for identification and measurement of indicators of resource mobilization and resource flows are quite applicable to the local level.

The second area in which progress has been achieved is in Mattila's (1970) work on metropolitan income estimation. Income estimates are the starting points for any serious analysis of inequality, poverty, wealth, and social stratification. Personal income estimates by metropolitan area do exist, and they represent an internally consistent set of accounts. They have been heavily used as an all purpose measure of wealth. On the other hand, if one were concerned about the structure of the metropolitan economy and policies for industrial concentration, diversification, and development and employment, personal income estimates have numerous defects since they cannot be disaggregated by sector. They represent income received, irrespective of sector of origin and irrespective of whether locally, nationally, or internationally produced. For structural analysis what is required is metropolitan income produced by sector of origin. For this purpose Mattila has developed a recursive model with ten structural equations of the metropolitan Detroit economy with which he is able to estimate income produced by each of twenty-one non-agricultural sectors for the 1956-1968 period.

Were models of the Mattila type available for a number of metropolitan areas, it would be far easier to evaluate the impacts of federal and state policies upon metropolitan areas than is presently the case. In Detroit it is readily seen that there has been a major structural shift in employment during the 50's and 60's out of motor vehicles, even though income produced by the motor vehicle sector has continued to increase. For an industry that is markedly becoming more capital intensive, with an absolute decrease in employment, there are serious implications for the employment structure of the Detroit area. Other manufacturing sectors have maintained a rather even balance between employment and income shares for the region. Although the Detroit economy continues to be heavily dominated by the automobile industry and its suppliers, and is thus cyclically unstable, there is some evidence that the economic base of the region is becoming more diversified.

One of our areas of interest involves information systems for urban and regional planning and policy analysis. In the past few years there has developed a growing collection of literature dealing with urban information systems, and this growing interest has been supported by federal programs. Unfortunately, one of the most fundamental sources of confusion regarding automated information systems centers upon the distinction between the management <u>of</u> information and management <u>by</u> information. In the former case, the objective is to obtain utility from the on-going processes of acquiring, storing and distributing information within or between organizations. The use of information in this fashion is an attempt to rationalize existing archaic and inefficient information channels. There are significant opportunities for achieving these objectives through automation, especially where information usage follows pre-programmed patterns. Management by information involves the utilization of information systems to help decision-makers make decisions or help managers manage.

Although there are scores of publications which suggest urban information system designs, under careful examination most of these system designs are patently absurd. Not because the hardware and software packages will not process data in the manner advertised, but rather because there has been little attention given to what data should be processed and why. This state of affairs exists because information systems in the public sector have been treated as technical computer problems, without the realization that public decision-making takes place in an environment of contesting claims. At the present time, those interested in information systems represent a separate grouping from those concerned with the utility of indicators for public policy purposes. It has been suggested by Perle (1970b) that a merging of interests between these groups is likely to develop, so that descriptive measures of social conditions affecting individuals and families will be merged with more conventional measures of physically observable phenomena. Then, perhaps, urban planning and policy analysis will be better able to cope with the social issues of the urban environment rather than their physical manifestations.

In a related project, we were asked to review the status of societal information for the state of Michigan and the utility of indicators for state-wide planning (Perle et al., 1970). In that report we reviewed the present status of quality of life reporting at the state level and dealt with issues involving the organization of societal information. In addition, for several classes of indicators we reviewed the existing data sources, indicated inadequacies in existing measures, and suggested alternative measures for a number of areas. This review and analysis covered the following types of indicators: demographic, health, economic, lawful behavior, education, and environmental. The study terminated with a set of recommendations to the state for the initiation of a regular system of social reporting, to be initiated immediately with the results forthcoming in 1972 and covering the areas of lawful behavior, environmental quality, health, and education. Throughout the report, special attention was given to urban oriented issues.

More local in nature have been a number of efforts dealing with a variety of subjects, including: citizen calls for help to police departments (Bercal, 1969), consensus and disparity in the public's perceptions of educational goals (Nowikowski, 1970), and employment as a determinant of Detroit's future (Warner, 1971). The paper by Bercal argues that metropolitan police departments should be viewed as service agencies which are involved in dispensing a wide variety of services, both to the individual and society. This thesis is presented in opposition to the traditional view that police departments should be studied as quasi-military organizations which "enforce the law."

After analyzing citizen calls for help in Detroit, New York and St. Louis, Bercal concludes that police departments dispense a wide array of personal and societal services that have little, if anything, to do with law enforcement. In Detroit in 1968 only 38.7% of dispatches of police patrols had to do with crime, prowlers, alarms, and recovery of property; 34.8% of the patrol dispatches dealt with public disorder, like family trouble, missing persons, neighbor trouble, or rubbish complaints; 12% dealt with crimes of negligence (accidents - vehicles); 10% with health service (sick person, animal bites, etc.) and 4.5% with matters of safety (directing of traffic, hazards, etc.). Bercal suggests, therefore, that by viewing metropolitan police departments as service agencies, the needs of the community as well as the services offered will be identified. In so doing the "real" nature of these needs may be determined and the limitations of effective police service in "satisfying" these needs recognized. As a result it is hoped that such an analysis will lead to a more rational restructuring of the responsibilities of metropolitan police departments within the communities they serve and that it will lead to innovative solutions to what are now defined as "police problems."

One of our experimental projects aimed to discover whether different socio-economic groups in the urban environment have similar or different perceptions concerning the goals of public education. Nowikowski describes a two-stage research design, where open-ended discussion groups (5-8 persons) were used to elicit their expectations which subsequently were to be used for the construction of a large sample questionnaire. The first stage included groups of parents of students, teachers, and high school students drawn from three differing socio-economic environments. Thus, nine groups were interviewed. The findings of stage one were somewhat interesting for the education we received, but they did not indicate as wide a spread of views as we had anticipated; however, the resource commitments necessary to conduct the second stage were not available. Therefore, the project has been terminated. Another experimental endeavor, reported by Warner (1971) involves the use of several existing employment forecasts of the Detroit region as the basis for tracing through a variety of economic impacts, including the occupational structure of the population, the propensity to attract in-migrants, and residential locations for differing income groups.

One study attempted to utilize traditional socio-economic variables to explain some attitudes and patterns of behavior of the residents within the Model Neighborhood Area (MNA) of Detroit. The data source was a 1968 survey of the MNA, conducted on behalf of the City of Detroit. Maimon (1970a) provides some evidence that the common use of income, education, race and sex as explanatory variables may not be of great assistance in the Detroit inner city. Moreover, preliminary analysis of satisfactions and dissatisfactions among inner city residents regarding their physical and social environments in terms of some personality traits (internal versus external control) yielded similar results. It appears that a very heterogeneous area of Detroit, in terms of income, education, race, and sex, has been found to display a high level of homogeneity in so far as some attitudes and some modes of behavior are concerned. Maimon suggests that there is a certain phenomenon associated with inner city life in Detroit which tends to elicit this similarity and homogeneity, and this phenomenon appears to offset some of the usual social differences. These preliminary findings require further analysis since they cast some doubt upon "conventional wisdom."

Recently increasing attention has been given to the problems of undesirable, unintended, and unanticipated secondary consequences of various kinds of activities. The damage caused to the natural environment by our present modes of living and the social and political implications of the war in Southeast Asia are, perhaps, the most dramatic examples of the day. Our present ability to understand, anticipate, and control some of these consequences is quite limited, yet their relative importance in our lives is increasing so rapidly that they soon may become of primary importance. Among social scientists, economists traditionally have paid the most systematic attention to these problems, utilizing the concepts of social costs and benefits and externalities. The insistence of economists to quantify consequences in monetary terms and the desire of the proponents of "social indicators" to detach themselves from economic constraints have brought increasing attention to the questions.

It appears that a more general framework of analysis is required to deal with secondary consequences, one which can encompass the considerations of both economics and the other social sciences. Maimon (1970b) has been attempting to develop such a framework and it is hoped that some empirical testing will be forthcoming soon. The major components of this analytical framework include the distinction between primary and secondary consequences, the cutoff point in a time horizon, decision makers and strategies of decision, identification of the affected population, trade-offs between performance characteristics and future capability, and reviews of some techniques for measurement and anticipation.

Located as we are in Detroit it is possible to implement some of our ideas about indicators. In this regard we perform a service function by lending our abilities to local agencies. We have given advice to one local agency for the evaluation of a job training program for teenagers in the inner city; one member of our staff is presently engaged in setting up an information system for a comprehensive medical care program and clinic for inner city residents; and we have been assisting in the creation of a social planning unit in the city of Detroit in addition to assisting in the design of an evaluation strategy for the Model Neighborhood Agency (Musial, 1969).

A number of the substantive areas in which we have already made some explorations will be pursued, such as the development of societal indicators for urban information systems, understanding the behavior and attitudes of differing socio-economic groups, and developing a framework for the analysis of secondary consequences. In addition, several other areas are being developed, including the analysis of mass behavior leading to varieties of urban violence and the analysis of alternative urban transportation systems and their locational and societal impacts.

NOTES

[1] A good selective bibliography of the indicators literature through the fall of 1969 is contained in Agocs (1970a). Since the preparation of that bibliography other materials on indicators have appeared, including: Ferriss (1969a, 1969b), Duncan (1969), Cooper (1970), and Henriot (1970).

[2] Good reviews of urban development models are contained in Irwin (1965) and Hemmens (1968).

[3] These issues are critically reviewed by Sheldon and Freeman (1970). Another critical review of the indicators literature, from a very different point of view, can be found in Hoos (1970).

[4] An excellent presentation dealing with requirements for effective evaluation has appeared in Campbell (1969).

[5] This section has benefited by the comments of Springer (1970).

[6] This perspective is forcefully presented by Gross (1969a). In addition, it is exemplified by the appearance of a journal volume dealing with urban indicators (Perle, 1970a).

[7] From Sept. 1969 through July 1971, this program has been supported by an unrestricted grant from the Bank of the Commonwealth (Detroit). That support is gratefully acknowledged.

REFERENCES

[1] Agocs, C., "Social Indicators: Selected Readings," The Annals, (March 1970a), 127-132.

[2] ________, "Toward Ethnic Studies in Detroit: Research Problems, Data Sources and Indicators," Discussion Paper, Center for Urban Studies, Wayne State University, (June 1970b).

[3] Bercal, T., "Metropolitan Police Departments: Indicators of 'Consumer Demands for Governmental Service,'" Discussion Paper, Center for Urban Studies, Wayne State University, (November 1969).

[4] Campbell, D. T., "Reforms as Experiments," American Psychologist, 24, 4 (April 1969), 409-429.

[5] Cooper, W. W., ed., "Urban Issues," Management Science, 16, 12 (August 1970).

[6] Duncan, O. D., "Toward Social Reporting: Next Steps," Social Science Frontiers, Russell Sage Foundation, (1969).

[7] Ferriss, A. L., Indicators of Trends in American Education, Russell Sage Foundation, (1969a).

[8] ____________, Indicators of Change in the American Family, Russell Sage Foundation, (1969b).

[9] Gross, B. M., "Urban Mapping for 1976 and 2000," Urban Affairs Quarterly, (December 1969a), 121-142.

[10] ___________, Social Intelligence for America's Future, Boston: Allyn and Bacon, (1969b).

[11] Gross, B. M. and Springer, M., eds., "Political Intelligence for America's Future," The Annals, (March 1970).

[12] Hemmens, G. C., ed., Urban Development Models, Highway Research Board Special Report No. 97, (1968).

[13] Henriot, P. J., "Political Questions about Social Indicators," The Western Political Quarterly, 23, 2 (June 1970), 235-255.

[14] Hoos, I. R., "Information Systems and Public Planning," Working Paper No. 112, Space Sciences Laboratory, Berkeley: University of California, (January 1970).

[15] Irwin, N. A., "Review of Existing Land Use Forecasting Techniques," Highway Research Record 88, (1965), 182-213.

[16] Maimon, Z., "The Inner City Impact," Urban Affairs Quarterly, (December 1970a).

[17] _________, "The Concept of Second-Order Consequences," Discussion Paper, Center for Urban Studies, Wayne State University, (December 1970b).

[18] Mattila, J., "Estimating Metropolitan Income: Detroit 1950-1968," Center for Urban Studies, Wayne State University, (January 1970).

[19] Musial, J. J., "Guidelines for Evaluation," The Model Cities Program in Detroit, Center for Urban Studies, Wayne State University, (June 1969).

[20] Nowikowski, E., "Consensus and Disparity in the Public's Perception of Educational Goals," Discussion Paper, Center for Urban Studies, Wayne State University, (Sept. 1970).

[21] Perle, E. D., ed., "Urban Indicators," Urban Affairs Quarterly, (December 1970a).

[22] ___________, "Notes on Urban Information Systems and Urban Indicators," Proceedings: Association for Computing Machinery, (forthcoming, 1970b).

[23] Perle, E. D. et al., "Social Reporting in Michigan: Problems and Issues," Technical Report A-37, Office of Planning Coordination, Bureau of Policies and Programs, State of Michigan, (1970a).

[24] Porter, D. O., "Who Slices the Pie? A Study of Resource Mobilization and the Politics of Budgeting. The Case of Federal Aid to Education," Center for Urban Studies, Wayne State University, (1970).

[25] Sheldon, E. B. and Freeman, H. E., "Notes on Social Indicators: Promises and Potential," Policy Sciences, 1, 1 (Spring 1970), 97-111.

[26] Springer, M., "Social Indicators, Reports, and Accounts: Toward the Management of Society," The Annals, (March 1970), 1-14.

[27] Warner, D., "Employment as a Determinant of Detroit's Future," Center for Urban Studies, Wayne State University, (forthcoming 1971).

Discussion

Kenneth C. Land, Russell Sage Foundation

I find the Perle and Tunstall papers interesting for somewhat different reasons. In order to elucidate my reactions, I would like to characterize the development of the interest in social indicators. In a recent survey of the literature on social indicators over the years 1964-1969, I found three recurring themes or claims for social indictors. That is to say, there occurs repeatedly the claims that social indicators can help (1) set goals and priorities, (2) evaluate specific programs, and (3) establish a system of social accounts. Suffice it to say that these are rather grandiose claims for social indicators and that they are rather easy to criticize. In fact, they have come under attack recently as noted in the Perle paper and we now realize that social indicators are more likely to contribute to improved descriptive reporting and social analysis.

Against this weak historical foundation, one particular problem seems to have emerged and come to the fore: how can we stimulate a commitment of professional social scientists to the creation and maintenance of a system of social indicators? For, without such a commitment, social indicators are certain to be unbearably subject to changing political pressures and circumstances.

One possible approach to the creation of a professional social science commitment to social indicators is to appeal to their ongoing interests in the theoretical and empirical analysis of social processes. Of these, two model building interests of social scientists seem to be particularly relevant to social indicators. First, social scientists are interested in models which relate various aggregate indices of social activity. For example, a social scientist might construct a model relating measures of the aggregate social mobility and the aggregate divorce rates of marriage cohorts. A second class of models which are of interest to social indicators focuses on the individual rather than the aggregate level. These seek to determine the distribution of various social activities among individuals. In brief, these models could be called models of social inequality. As an example, consider the various sociological models of the relation of achieved status characteristics to ascribed statuses.

From this model building perspective, the contributions of these two papers may be specified. First, with respect to Perle's paper, I must concur with the comment that not only is it necessary to suggest apparently brilliant conceptual models but is is also necessary to verify them empirically. Moreover, I am pleased with the brief reports on specific projects at the Center for Urban Studies which are oriented towards the specification of theoretical models for various social processes and their estimation from data for a specific metropolitan area. However, without more detail, it is impossible to comment further on these efforts except to say that this is indeed the direction for future work if we are to achieve a commitment of professional social scientists to social indicators.

On the other hand, the Tunstall paper is a revealing report on the progress of the Office of Statistical Policy of the Federal Government on the development of a publication of social statistics. It appears to me that the development of the kinds of models (aggregate and distributional) described above is crucial to a clear understanding of what is happening in each of the nine areas of social concern which the publication will cover. Consider, for example, the topic of health care. If we had possessed well-specified models of the distribution of health care facilities, and if we had possessed the data requisite to the estimation of such models, then I am sure that we would have observed some very interesting changes in the parameters of the models during the last decade. As it is now, we can do little more than speculate about the changes which have created a crisis in health care. Thus, I find the work reported in the Tunstall paper useful because it will facilitate the development of models of social processes by social scientists. Perhaps, in the process, we will be able to specify other indicators which should be made available. At this point, I can only suggest that these may include (1) indicators on institutions, and (2) social psychological indicators.

MICROANALYTIC SIMULATION: A DEMOGRAPHIC MODEL*

Guy H. Orcutt, Gerald E. Peabody, Steven B. Caldwell
The Urban Institute

Work is under way at the Urban Institute on a microanalytic model of the income distribution.[1] In this model, an extension of earlier work by Orcutt, Greenberger, Korbel and Rivlin [6], the behavior of family members that affects the family's income status over time is simulated. To describe dynamically population growth and the formation and dissolution of families, birth, death, marriage, separation, and divorce are simulated, where appropriate, for each person in the sample population. To simulate income activity, the number of weeks worked in the year and the wage rate are imputed for each working member of the family. Transfer payments and income from wealth are added to the total family wage income. Taxes are then calculated and subtracted from the total family income to give disposable income, from which savings are imputed to generate the accumulation of assets. Migration and the educational attainment of children are also simulated. The demographic and education sections of the model are essentially complete, and these operating characteristics, along with the auxiliary macroanalytic model, are described in this paper.

The model is still in the process of development and the description in this paper should not be taken as the final product. In fact, the model is being designed so that it can be easily modified whenever new results can be brought to bear.

This model is designed to provide a dynamic representation of the United States population of individuals and families which will be useful in tracing the effects of various public policies, singly or in combination, on the behavior and well-being of individuals and families over time. However, since this model is focused entirely on individuals and families it cannot present an essentially closed representation of the entire economic system. By itself it cannot be expected to generate unemployment, price level changes, GNP or fractions of GNP going to earned income and to wealth holders. Thus, an auxiliary macroanalytic model has been designed to provide a serviceable interim closure.

Work on the demographic sector of the model has been completed for a simplified family structure model--a persons model. In the persons model marital status is imputed for each adult in the sample but the files of a man and woman who are to be married are not actually merged to form a new family file. This procedure vastly simplifies the computer implementation but at the cost of the loss of most information about the husband in the family. A family model is under development in which the files of the marrying couple are merged to form a complete data file for all members of the family.

The family model will be used to implement the complete model of the income distribution. The persons model, which will be described in more detail below, is being used to test the various demographic components of the model and constitutes a reasonable demographic model in its own right. By simulating marriage, separation and divorce the marital status of each adult in the population is determined. The marital status of women, along with their parity, race and socioeconomic status, provide the necessary input for the birth simulation. With the simulation of deaths we have a complete vehicle for providing population growth predictions. The detailed information about race, social class, etc., of persons in the model gives estimates of the future size distribution by various subgroups of the population.

The persons model contains six operating characteristics: death, marriage, separation, divorce, birth and education. The first five constitute the demographic core of the model while the sixth operating characteristic begins our explicit concern with the intergenerational distribution of status and income. For each pass of the simulation the operating characteristic will generate a probability of the corresponding event occurring to that person during the year. For example, each person will first be subjected to the death routine in which the probability that the person dies during the year will be determined. Via the Monte Carlo technique of drawing a random number and comparing it with the probability of the event, the occurrence, or not, of death will be imputed (if a random number r, $000 \le r < 999$, is drawn and the probability of death is .05, then death will be imputed if $r < 50$).

For the initial population we are using a self-weighting random sample drawn from the 1967 Survey of Economic Opportunity (SEO) [9]. For

*This research is being conducted under grants from the Ford Foundation and the Social Security Administration. Opinions expressed are those of the authors and do not necessarily represent the views of The Urban Institute or its sponsors.

[1] Our colleagues in this project are John B. Edwards, Harold Guthrie, Sara Kelly, George Sadowsky and James Smith. Edwards achieved the computer implementation of the persons model reported in this paper. Guthrie is co-director of the project. Orcutt, Peabody and Caldwell are listed as authors of this paper because they had responsibility for developing the operating characteristics reported here--Peabody for birth and education, Caldwell for death, marriage, separation and divorce, and Orcutt with Kelly for the auxiliary macroanalytic model.

the persons model most of the family and household structure of the SEO file has been eliminated. A mother and her children constitute the "familv" unit, and males are isolated individuals in the file.

In the following sections we will describe the structure of each operating characteristic for the persons model and then discuss the macro model that is being developed to close the model.[2]

Deaths

In the previous work [6] the annual probability of death for an individual was taken to be a function of his race, sex, age and the calendar year being simulated. Corrected central death rates covering the period 1933-54 were fitted to a logarithmic function; i.e., the death probability within each race-sex-age class was assumed to be decreasing by a fixed percent each year toward a limiting value of zero. However, mortality levels since 1954 have proven to be consistently higher than predicted by this two-parameter model.

Rather than re-estimate the same two-parameter model and incorporate subsequent (1955-1968) data, we have respecified the model by relaxing the assumption that mortality is declining toward zero in all classes and included an asymptote as a third parameter. Within each of 76 race-sex-age class we have estimated the following function using uncorrected death rates:

$$P(\text{death}_{it}) = a_o + a_1 e^{a_2(t-t_o)} \qquad (1)$$

where

$P(\text{death}_{it})$ = the central death rate in year t for an individual in age-race-sex category i

and $t_o = 1950$

is an arbitrary base year. The asymptote (a_o) could be interpreted as a measure of whatever mortality risk has remained constant over the past 36 years, a "floor"toward which mortality rates are tending.

Data from census years [3] reveal that within age-sex-race categories, substantial mortality differentials exist for the four marital states: single (never married), married, divorced and widowed. Using these data to calculate ratios of marital status-specific mortality within age-race-sex classes, and taking the mean of the ratios for the three points[3] for which we have data, enables us to generate a parameter with which we make the death probability a function of marital status.

The excellent research of Kitagawa and Hauser [5] provides estimates for national socioeconomic differentials in mortality in May-August, 1960.[4] We know of no subsequent re-estimates of these parameters which could serve to give some idea of their temporal stability so we have assumed that these parameters are constant.

Kitagawa-Hauser carried out separate estimates within race-sex classes. They also standardized for age, but presented the estimates only in terms of two age classes (25-64, 65 and over); consequently we do not know if there are significant age differences in the parameters which were not captured by their results. For whites, Kitagawa gives separate estimates for family members and for unrelated individuals, but since the differences are slight [5, p.10], we have chosen to apply the combined estimates across marital status categories. For all persons aged 65 and over, except white females, no clear and significant pattern of education differentials was detected. For all others a strong inverse relationship between years of school completed and mortality was noted and these differentials provide the basis for adding a fifth parameter to the death probability.

Finally, parity differentials were also examined in the above study [5, p.23]. These differentials were calculated for ever married white women 45 years of age and over. No consistent pattern was evident for those 65 and over. However, for those aged 45-64 there was a J-shaped relationship in which women with a parity of three had the lowest death probability. No separate tabulations for non-whites were presented. However, we have decided for the model that it would be less reasonable to leave out parity differentials for non-whites altogether than to assume that the same parameters hold for whites and non-whites, so we have done the latter. The parity parameters were standardized for age and education, to account for the known negative association between fertility and education, but they were not presented in education-specific form.

Thus, in the present version of our mortality operating characteristic the death probability is a function of the person's age, race, sex,

---

[2]More detailed descriptions of the operating characteristics and a discussion of simulation results are in preparation and will be available, upon request, from the authors.

[3]It is difficult to impute time trends to these ratios using only three points; so far we have not done so, though it is a trivial matter to make the parameter a function of time if evidence so suggests.

[4]We have not attempted to correct for possible seasonal biases. For a discussion of this and other possible biases see [4].

education (for all persons aged 25-64, and also for white females aged 65 and over), marital status, parity (for ever-married females aged 45-64), and the calendar year. Full interaction between age, race, sex and marital status is allowed. Education interacts with race, sex and (very roughly) age. Parity is sex-specific, marital status-specific (roughly) and age-specific (very roughly). The time trend is age, race and sex specific.

With respect to all independent variables, but particularly those that are in some degree manipulable by individuals or policymakers (e.g., education, parity, or marital status), the question of causality must be faced. For an individual in the model a change in marital status, years of school completed or parity gives rise to a sudden change in his or her death probability. Only in the case of marital status is this situation likely to generate unrealistic histories, since we do not apply the other differentials until after the age when most people will have completed their education and child-bearing. Undoubtedly the effect due to marital status overstates its causal importance and understates the extent to which the differentials are due to selection. But more important for our purposes is the aggregate question: does a change in socioeconomic composition (by marital status, education, or parity) of a race-sex-age class have an effect on the level of mortality rate for that class, or does it affect only the distribution within the class of the predicted rate. Largely because we have estimated the time trend within age-race-sex classes, we are reluctant to attribute additional temporal variations to changes in any of the other independent variables. Thus a tracking routine uniformly scales the education and marital status parameters up or down each year to arrive at an expected value of deaths consistent with predictions in each age-race-sex class.[5] However, since the parity differentials are relatively minor and we are inclined to attribute some causal importance to changes in parity, we have not scaled the parity parameters. Thus, changing parity composition is allowed to affect aggregate mortality rates, which is not the case for changing marital status and education composition.

Marriage

All unmarried persons over fourteen are assigned a probability of marriage for each year of the simulation. For never-married persons the probability of marriage is a function of race, sex, age, (by single years to age 50) and a trend factor. For divorced persons the probability of remarriage is generated as a function of sex, number of years divorced, and a trend factor. For widowed persons remarriage is a function of age and sex. No matching of marriage partners is carried out in the persons model, but there will be matching, according to various relevant characteristics, in the family model.

The most important missing variable at present in the marriage probability functions discussed above is education. Analyses in progress of SEO and other data should enable us to fill this gap. Such analyses might also yield parameters on significant differences in remarriage rates by number of living children.

Marriage Dissolution

Marriages can be dissolved in our model by divorce or death. Separation, a step prior to divorce that dissolves the family unit, is also simulated. Persons are widowed by assuming that each married person has a spouse of the same race, education, and marital status, of opposite sex and with a two year age difference (male older) and subjecting this shadow spouse to the death routine. In the family model the spouse will be present so we will have the advantage of knowing his or her particular characteristics.

Data concerning separation are scarce (especially for rates) and imperfect; e.g., half again as many women as men generally report themselves as separated [1, p.14] though the two numbers ought to be equal. However, substantial numbers of persons report themselves as separated (over 2 million in 1960) and sharp differentials exist by race. Five or six times as many non-whites, in percentage terms, report themselves as separated. Leaving out separation leads to an over-estimate of the number of intact families, particularly non-white families, and consequently, introduces error when, in the family model, we determine total family income. Thus, it seems preferable to deal with separation, even if ad hoc assumptions must be made, rather than ignore it completely.

Given the lack of data on the rates at which married persons separate and separated persons divorce (i.e., existing data generally skips separation; divorce rates are calculated by using married persons as the base) we are forced to make certain assumptions. We have chosen to use evidence on divorce rate differentials in constructing separation rate differentials. Accordingly, separation rates become a function of sex, duration of marriage, race, number of children, and a trend factor. To account for the fact that a much larger fraction of non-whites report themselves as separated, we assume that non-whites have a longer duration of separation. Thus, the probability of divorce is made a function of race and the length of separation. We do not at present allow reconciliations in the model. Having different functions for

---

[5] A mixed strategy could easily be implemented by adjusting the scaling factor. For example, moving it one-half the distance toward unity would have the affect of attributing some of the effect to selection and letting the rest reflect a causal influence.

separation and divorce gives us the capability of having some interaction between the two rates. We can use the traditional divorce rates to constrain the separation and divorce probabilities in the model.

Birth

In the previous work [6] the birth probability for a woman was taken to be a function of her age, parity and the interval from her last birth (or marriage). This procedure allocates the proper number of children and distribution of family sizes, but it does not capture the variation of family size with social class. Since our focus is on the distribution of inequality, and since the number of children that a couple has can have a large impact on their financial situation, we have devised a birth simulation that more closely represents the individual couple's approach to, and success with, family planning. We can then more adequately deal with the effects, for example, that unwanted or too closely spaced children can have on the financial status of low income families.

The simulation draws heavily on the various fertility surveys that have been conducted in the past fifteen years, including the Growth of American Family Studies [2], the Princeton Study [8], and the 1965 National Fertility Survey [7]. These studies, along with other demographic research, have provided a wealth of information about the distribution of desired number of children, the circumstances under which (married) women use contraception and how effectively they use it, the intervals between successive births, and fecundity impairments that may limit a couple's fertility. The model incorporates all these features of fertility.

The birth probability in year t for each woman, i, who is not using contraception is simply the couple's fecundability,[6]

$$P(\text{birth}_{it}) = f_{it}\, F(\text{Age}_{it}), \qquad (2)$$

which is a function of the age of the woman and her fecundity, f_{it}. For a fecund couple $f_{it} = 1$, and f_{it} assumes values less than one for sub-fecund couples down to zero for a sterile couple. The monthly probability of conception is about 0.2 for women in their early to mid-twenties and then declines with age to zero at the onset of menopause. Most couples are fecund, but about 30% of American couples are subfecund to some degree, including slightly over 10% who are sterile.

[6] Demographers usually define fecundability as the probability that a fecund couple not in a period of post-partum amenorrhea will conceive in a month; since we will simulate in intervals of a year, we will stretch the definition to be the probability that the couple will have a live birth in a year.

Most couples try to control their fertility through the use of contraceptives. When birth control is being practiced the birth probability of Eq.(2) is reduced by a factor that measures the effectiveness of the contraceptive usage,

$$P(\text{birth}_{it}) = f_{it}F(\text{Age}_{it})[1-\text{Eff}_{it}]. \qquad (3)$$

A couple that is using a perfect contraceptive device will have $\text{Eff}_{it}=1$, while a couple that practices birth control so ineptly that their fertility is not reduced by its use will have zero Effectiveness. Effectiveness includes both the effects of the imperfections of the particular method used and the skill with which the couples use their method. In this country effectiveness is a function of race, education, and purpose (to limit fertility or to control timing of births) and length of use.

For the model we make the assumption that all non-married women use contraception. Illegitimate births will be allowed in the model by assigning an Effectiveness less than one to some unmarried women. The only married couples who will be assumed not to use contraceptives are those whose desired number of children is larger than the number of children they now have and whose interval from their last birth (or marriage) exceeds their desired minimum interval (if they want to control the timing of births). All other married women will be assumed to practice birth control and will have a value for Effectiveness imputed to them.

Thus the couple's desires about family size and spacing of births are used to control the woman's birth probability. To each woman we will impute a desired number of children and the length of the interval after a birth during which the couple uses contraception. Most families in America want two, three or four children. Religion accounts for more of the variance of the distribution of desired family size among couples than any other social variable, but it does not account for much, with Catholics desiring larger families than other religious groups. Socioeconomic status has little impact on desired fertility--except that women who work have much lower fertility than those who don't--and race a slight effect with Negroes wanting fewer children than other racial groups. Social class has more effect on desired spacing with upper class couples spacing their children further apart than lower class couples.

The birth simulation proceeds, then, by assigning to each married woman a desired family size, a desired minimum interval, a fecundity, and an effectiveness of birth control usage. At each pass a decision to use or not use contraception is imputed to each married woman by comparing her desired fertility with her actual fertility and comparing her elapsed interval with her desired minimum interval. Then a birth probability for the year is determined from Eq.(2) or (3) depending upon the use, or non-use, of contraceptives. If a birth occurs it is determined whether or not it is a multiple birth and a sex is assigned to each child born.

Education

The educational attainment of children provides the major mechanism in the model for examining intergenerational distribution of income and status. We deal only with educational attainment, the number of years completed or degree obtained, rather than with educational achievement, the grades obtained or what the student learned, for two reasons. First, it would be very difficult to simulate achievement in a national model, and the necessary data probably do not exist. Second, there is no evidence in the literature that achievement has much effect upon future earnings or status, which is our primary interest. (Achievement is, of course, a very important intermediate variable in determining attainment, but we feel we can capture the important socioeconomic differentials in attainment without including achievement.)

The data that we have accumulated indicates that our present provisions for educating children also turn out to constitute a fairly effective mechanism for maintaining class positions over successive generations. Children of upper class parents are much more likely to graduate from high school, to enter college, and then to graduate from college than are the children of the lower class. The cumulative effect of these inequities leads to a stratification system with far less inter-generational mobility than is commonly supposed.

Our simulation will concentrate on these effects of the parent's status upon the educational attainment of their children. We will also include other family effects such as number of children in the family, race, sex and age of the child. School and community variables have also been shown to be of importance in explaining attainment, but we do not have enough locational detail in the model to be able to incorporate these variables very effectively. We will, however, try to capture the differences between the four census regions and urban-rural differences.

To simplify the simulation we do not pay attention to each grade level that a person passes through. Instead we consider the following levels of schooling:

n	school level
0	no schooling
1	grade school
2	8th grade graduation
3	high school
4	high school graduation
5	college
6	college graduation

We plan to include two year as well as four year schools in the college sector, and we hope to be able to include vocational schools as another alternative for high school graduates. We may also add graduate school as levels seven and eight.

Children will be entered into the first grade at ages five, six or seven. For children in school, at each level (n) there is a function that gives the probability of being retained at that level for a year, $P^R(n)$, of dropping out of school at that level, $P^D(n)$, or of advancing to the next level, $P^A(n+1|n)$. For children in grade school the retention function will be tested first to determine whether or not the child remains at that level for the next year. If he is retained, we exit from the education simulation. To further simplify the simulation all children will be automatically retained at the grade school level for eight years. At the beginning of the child's ninth year, $P^R(1)$ will be tested to see if he was retained an extra year in grade school. If he is not retained, then $P^D(1)$ and $P^A(2|1)$ are tested (with $P^D(1)+P^A(2|1)=1$). If the person advances to level 2, a test is made in the same pass to see if he enter, high school ($P^D(2)+P^A(3|2)=1$).

The procedure for high school and college is slightly different in that we allow a person to drop out of school each year; hence $P^D(n)$ is tested on each pass. For the first four years of high school or college $P^D+P^R = 1$, so that if leaving school is not imputed, the person is automatically retained for one more year. After four years of high school we will test $P^R(3)$ first to see if the person remained in high school a fifth year. If not, we test $P^D(3)$ to determine whether he dropped out before graduating or he graduated, $P^D(3)+P^A(4|3)=1$. Next we determine whether the person leaves school after graduation from high school or enters college.

If it is determined that the person leaves school, the amount of schooling obtained by the individual to that time becomes a permanent attribute and he never re-enters the schooling simulation. At the present time we do not deal with people who leave school for some period of time and then return to school.

Macroanalytic Model

The value of providing closure between the microanalytic model of individuals and families and the macro economy is two-fold. In the first place the microanalytic models under development need an environment in which to operate. In the second place economists think they know something about the control of some macro variables such as the percent unemployed and the rate of price change. It would be useful to trace out the impact of fiscal, monetary and other policies operated at the macro level on the behavior and well-being of individuals and families. The macroanalytic model being developed is intended to provide both an environment for the micro-analytic models and a useful link to variables which can be controlled, or at least influenced, by available monetary and fiscal tools.

The simplest expedient for providing a needed environment for the model of the population of individuals and families would be to treat unemployment, real GNP, price level changes and fractions of GNP going to earned income and wealth holders as direct exogenous inputs. The disadvantage of this approach is that no explicit account is taken of the extensive multicollinearity of these variables. By leaving such variables entirely unconnected, the user of the microanalytic models would be given an entirely unrealistic view of the extent to which outcomes could be influenced by manipulation of policy tools at the macrolevel. The objective behind the auxiliary model was to take a useful step towards capturing the close inter-connectedness of household inputs from the macrolevel and still leave points at which policy assumptions could be entered either by alteration of target unemployment or by alteration of parameter values in appropriate operating characteristics.

In developing an auxiliary macroanalytic model extensive simplification has been achieved by assuming that the federal government can and will cause aggregate demand to vary so as approximately to control the fraction of the labor force which is unemployed. The advantage of this assumption is that if total aggregate demand actually is controlled by the federal government it becomes less critical and probably unnecessary to account for the role of non-household sectors in generating aggregate demand. Their behavior in this area is simply supplemented or offset as necessary to achieve a desired unemployment rate given past price movements. Of course this approach would not do for a model intended to be useful in guiding short-run stabilization efforts. It is hoped and expected to be useful for consideration of long-run consequences of policy measures.

Given the auxiliary macroanalytic model the capability exists for constructing links between micro events and the macro economy. For example, one might argue that age at marriage or the desired number of children are related to the over-all condition of the economy, in excess of the specific monetary effects at the individual level. Given evidence for such connections we could link, for example, the race-sex-age-specific first marriage probabilities to changes in total unemployment, the price index, etc. No such links have been incorporated into the operating characteristics developed to date. However, such questions will be explored as the family model develops.

The auxiliary model should be regarded as a first serious step in establishing useful links from monetary and fiscal policies into micro-analytic models of the population of individuals and families. It also is of interest in that it provides for and makes important uses of outputs of such microanalytic models as inputs into a macroanalytic model. This model has several deficiencies which hopefully can be reduced with additional effort. Perhaps the most serious of these is that the gap between what policy makers might do at the macrolevel and appropriate alteration of parameter values in the auxiliary model is still uncomfortably large.

Implementation

The persons model has been implemented on a PDP-10 in an interactive mode. We are currently operating with a random, self-weighting initial population drawn from the 1967 SEO that consists of 500 interview units (families or single individuals) and 1553 persons. The sample was drawn in such a way as to control for the distribution of family types (female-headed families, single persons, etc.), and the age, race, and sex distribution of the sample was compared with the total SEO file to check the drawing procedure.

The birth rates, death rates, etc. obtained from the simulations are reasonable, but the initial population size is far too small to give meaningful population projections. We are using the small sample to check the computer logic of the model and the design of the operating characteristics. After completing this first rough stage of the model validation we will simulate with an initial population of approximately 10,000 households. We expect that simulations with this much larger population will give reasonable population projections under different assumptions about the future time trends of the parameters in the model.

References

1. Hugh Carter and Paul C. Glick, Marriage and Divorce: A Social and Economic Study, Cambridge, 1970.

2. R. Freedman, P.K. Whelpton and A.A. Campbell, Family Planning, Sterility, and Population Growth, New York: McGraw-Hill, 1959.

   P.K. Whelpton, A.A. Campbell and J.E. Patterson, Fertility and Family Planning in the United States, Princeton, 1966.

3. Robert D. Grove and Alice M. Hetzel, Vital Statistics Rates in the United States, 1940-1960. Public Health Service Publication No. 1677, U.S. Government Printing Office, 1968.

4. Evelyn M. Kitagawa and P.M. Hauser, "Educational Differentials in Mortality by Cause of Death: United States, 1960," Demography, Vol. 5, No. 1, 1968, pp.318-353.

5. Evelyn M. Kitagawa, "Social and Economic Differentials in Mortality in the United States, 1960," paper delivered at Conference of International Union for Scientific Study of Population, London, 1969, to be published in the Proceedings.

6. G.H. Orcutt, M. Greenberger, J. Korbel, and A.M. Rivlin, Microanalysis of Socioeconomic Systems: A Simulation Study, Harper, New York, 1961.

7. Norman Ryder and Charles Westoff, to be published.

8. C.F. Westoff, R.G. Potter, Jr., P.C. Sagi, and E.G. Mishler, Family Growth in Metropolitan America, Princeton, 1961.

C. F. Westoff, R.G. Potter, Jr., and P.C. Sagi, The Third Child, Princeton, 1963.

9. U.S. Office of Economic Opportunity, "Surveys of Economic Opportunity," Codebook and other technical documentation.

10. U.S. Public Health Service, Vital Statistics of the United States, U.S. Government Printing Office, Washington, published annually since 1945.

SIMULATION OF HOSPITAL UTILIZATION*

D. G. Horvitz, T. D. Hartwell and J. R. Batts, Research Triangle Institute

I. INTRODUCTION

The purpose of this paper is to present a dynamic microsimulation model for projecting utilization of general hospitals by the U.S. civilian non-institutional population. The model has been developed to provide essential inputs to a model designed to project hospital manpower requirements [1], but may be used independently. It attempts to meet the need for improved methods of analyzing the effects of changes in the composition of the population and of changing utilization patterns on hospital admission rates and length of stay, as well as for more refined projections of utilization. Further, the model may be used effectively in simulation experiments to provide prior measures of the consequences of alternative programs designed to improve personal health care in the U.S.

The model has two components. The first is POPSIM, a demographic simulation model [2] which creates a population of individuals in the computer and then projects this computer population forward in time, generating births, deaths and changes in marital status for each individual by the Monte Carlo method and on a competing risks basis. The output of POPSIM for a given simulation period (e.g. a year) provides the input to HOSPEP, the second component of the hospital utilization simulation model. HOSPEP is also a stochastic model, generating hospital episode histories for each individual for a selected simulation period.

The principal variables and characteristics generated by POPSIM for each individual in the population file and used by HOSPEP are age (birth date), sex, race, family income, residence (metropolitan or non-metropolitan) and date of death. Using the POPSIM input, HOSPEP classifies each individual by hospital insurance and generates hospital admission and discharge dates. Further, it specifies the diagnosis (18 classes), whether surgery was performed or not, and the bedsize of the hospital.

The hospital utilization model discussed in this paper is essentially descriptive in nature. By varying those parameters (conditional probability distributions) of the model which determine admission rates, diagnosis, surgery status, length of stay and hospital bedsize, the model projects the resultant changes in utilization. The model is not self-adaptive in the sense of having parameters which adjust automatically through interaction with resource constraints in the system. For example, the model assumes there is a hospital bed available for every episode generated by HOSPEP. However, the model does have a definite potential for linkage with resource models of components of the personal health care system and hence for introducing parameter adjustments through feedback. This implies a distinct possibility that the current model could be adapted to project <u>economic demand</u> for hospital services and not just <u>utilization</u>.

The model through appropriate choice of parameters could be used to project <u>need</u> for hospital services rather than utilization. Currently, statistics on need, and particularly unmet need, are available to a considerably lesser degree than utilization statistics. Although the initial interest and emphasis of the model is in projections of utilization at the national level, it can be adapted for local, state or regional projections provided appropriate data are available.

Most models concerned with projecting utilization of health services are aggregate or macro-models. The microsimulation models discussed in this paper appear to offer greater versatility and flexibility in terms of assumptions and output than might be possible with macro-models. Microsimulation models, such as POPSIM and HOSPEP, by their very nature include the inherent variability in the system and can easily generate distributions for the output variables as well as averages. This is a particularly useful feature when assessing alternative health care programs. A more detailed comparison of microsimulation and macrosimulation demographic models suggests there is a definite trade-off in choice of model based on the number of variables and levels of each variable and on the size of the simulated population required to arrive at reasonable conclusions [4].

II. THE DEMOGRAPHIC MODEL (POPSIM)

A. General Description of POPSIM

POPSIM is a dynamic demographic model designed for computer simulation of the principal demographic processes occurring in human populations. It is classed as a microsimulation model because it generates a vital event history, including the dates of birth, marriage, divorce, widowhood, re-marriage and death for each individual in the computer population. Although POPSIM is a two-sex model, it may be used for simulating cohort as well as period data. POPSIM is a stochastic model in the sense that random sampling from probability distributions is used to determine which events occur to an individual and when they occur. It is a dynamic continuous time model, permitting the probabilities to change with time. The model can be made self-adjusting

* Research sponsored by the Bureau of Health Professions Education and Manpower Training, National Institutes of Health under contract number NIH 70-4085.

or recursive through the use of feedback mechanisms. For example, feedback models appropriate for POPSIM to account for the effect of changing marriage patterns on marriage rates are discussed in [5].

The demand in recent years for new techniques for studying complex demographic phenomena with fewer restrictive assumptions than has been possible with analytic models has led to the development of a variety of microsimulation models [6, 7, 8, 9, 10, 11]. POPSIM is a laboratory research tool designed in response to that demand. It can be used to considerable advantage in a wide variety of simulated longitudinal studies. Further, it can be linked easily to other models (e.g. as the demographic component of models of economic, education or health systems), whether they be micro- or macro-models.

1. The Initial Population (Phase I)

The POPSIM computer program consists of two distinct parts or phases. The first is used to create an initial population. The initial population is conceptualized as a random sample of individuals (stratified by age, sex and marital status) from a hypothetical population. The program, through the use of control cards, offers the user considerable flexibility in specifying the nature of the hypothetical population. For example, the user can choose to confine his population to a cohort of women x years of age in 1970, or to all persons 65 years of age and older in 1966, or to the civilian non-institutionalized U.S. population as of 1960.

The initial population can be considered usefully as a random sample of individuals selected from a population register, without regard to familial relationships. Thus, for example, a married female may be selected for the initial population, while her husband and children may not be chosen. Individuals in the computer population (initial sample plus births) are referred to as primary individuals, and marriage partners and children as secondary individuals. Since secondary individuals are not members of the computer population, information concerning them must be carried by the primary individual. In the simulation of vital event histories (second phase of POPSIM), all events which take place are considered as events to primary individuals. All tabulations produced are counts of primary individuals, or of events which happened to primary individuals. Secondary individuals enter the model only in the sense that they influence the vital event risks to which primary individuals are subject. For example, the risk of a married woman (primary individual) becoming a widow is a function of the age of her husband, a secondary individual.

The list of characteristics describing the primary individuals and the associated secondary individuals is shown in Figure 1. The nature of the entries in the list is dictated in part by the fact that POPSIM is a time-oriented simulation model, with continuous time. Consequently, rather than carry age as a characteristic and update periodically, date of birth is used.

Figure 1

List of Characteristics of Each Primary Individual in the Sample Population*

RECØRD(1)	=	sex of individual
RECØRD(2)	=	date of birth
RECØRD(3)	=	current marital status (single, married, widowed, divorced)
RECØRD(4)	=	parity
RECØRD(5)	=	number of living children
RECØRD(6)	=	current contraceptive method**
RECØRD(7)	=	number of marriages (0, 1, 2+)
RECØRD(8)	=	date of current marital status
RECØRD(9)	=	date of next event
RECØRD(10)	=	next event
RECØRD(11)	=	number of children (secondary individuals) for whom information on age, sex and date of death (if appropriate) is available in a separate array for each mother
RECØRD(12)	=	date of birth of spouse
RECØRD(13)	=	previous contraceptive method**
RECØRD(14)	=	date of last birth
RECØRD(15)	=	number of events which happen to the individual during a simulation
RECØRD(16)	=	race (white or non-white)
RECØRD(17)	=	residence (SMSA or non-SMSA)
RECØRD(18)	=	family income deviate
RECØRD(19)	=	family income (dollars)
RECØRD(20)	=	date family income was computed

*RECØRDS(16) through (20) were added for use with hospital utilization simulation model.

**Not used for hospital utilization simulation model.

The computer program to generate initial populations is written in Fortran for use on IBM 360 System/Model 50 computers. The program is designed to generate a specified number of initial sample populations of a designated size in a single run and to write each population on a disk or on magnetic tape. The input to the program includes (in addition to parameters which control various options in the program and class limits used in the output tables) the following variables:

1. The size of the initial population.
2. The proportion of individuals in each of the eight sex-marital status groups.
3. The proportion of individuals in specific (but arbitrary) age groups for each sex and marital status class.
4. The probabilities with which married females are classified as remarried or not.
5. Parameters to assign age of husband given age of wife.
6. Parameters to assign age of wife given age of husband.
7. Parameters to assign the date of current marital status.
8. Monthly birth probabilities by age, parity and marital status (married or single), appropriate to three time points in the 30 year period prior to the date of the initial population. These values are used to generate a birth history, and hence the parity and date of last birth for each married or single female.
9. Monthly death probabilities by age group, sex and marital status to determine number of living children for married or single females.
10. Parameters to assign parity to widowed or divorced females.

The characteristics corresponding to REC∅RD(6), REC∅RD(9), REC∅RD(10), REC∅RD(13) and REC∅RD(15) are not assigned when the initial population is generated.

POPSIM creates each initial sample population in the computer by means of a series of subroutines which use random sampling of inverse probability distribution functions, for the most part, to assign a consistent set of characteristics to each individual. For example, a stratified random sampling procedure is used to assign age (date of birth), sex and marital status. Using the first three input variables listed above, a density function is fitted by the computer for each age-sex-marital status group. The distribution function and its inverse are then computed for each of these groups. The age assignment routine then sets up the records for the individuals in each age-sex-marital status group and assigns their ages by stratified random sampling of the associated inverse probability function. This is accomplished by first dividing the (0, 1) interval into n sub-intervals or strata of length $1/n$, where n is the number of persons in the particular age-sex-marital status group, and then generating a uniformly distributed random number for each sub-interval to sample the appropriate inverse. This stratified sampling procedure distributes very effectively the ages of individuals over the entire age interval for each sex-marital status group.

The remaining characteristics are assigned to each individual by sampling the appropriate conditional distribution for the specific age, sex and marital status of the individual. For example, the age of husband is assigned by generating a random observation from the conditional distribution of age of husband given age of wife.

2. Generation of Vital Event Histories (Phase II)

After creating an initial population of desired size and characteristics, a second program uses the Monte Carlo method to generate a vital event history or life pattern for each individual. This program advances the population forward through time in a series of time intervals or steps. At the end of each step, the program prints a series of tables and provides the user the option of updating the probabilities of the various events. The user must specify the total length of the simulation period and the time interval for each step. For example, one may simulate the vital events that occur to the initial population for a period of ten years in ten steps of one year each, or for a period of ten years in a single step, or 25 years in five steps of five years each.

The basic set of input data required by POPSIM for generating vital event histories includes:

1. Monthly birth probabilities for females by age group, marital status (married or not married) and parity (or number of living children).
2. Monthly divorce probabilities by interval since marriage (or by age).
3. Monthly death probabilities by age group, sex and marital status.
4. Annual marriage probabilities for females by age and marital status.
5. Parameters to determine the marital status of the groom given that of the bride.
6. Bivariate distributions of ages of brides and grooms for first marriages and remarriages.

At first glance, the list of required input parameters for both phases of POPSIM appears rather formidable. This is not a peculiarity of this or any other microsimulation model. The data required for any model increase rapidly with the degree of complexity of the model. However, POPSIM does not require all the input data for the model to be obtained entirely from one census or sample survey of the population of interest. Estimates of the various parameters used may be obtained from different prior studies on the population of interest and still be used in the model. Occasionally it may even be possible to use data from studies of different but similar populations and obtain valid results. Clearly, the data employed in a microsimulation model will have an important effect on the accuracy of predictions made with the model for particular populations. However, for most applications, the use of parameter values which have been estimated from data which do not pertain exactly to the population of interest will not hamper the usefulness of the model severely. In specific applications, simulation runs using upper and lower bounds for parameters whose values are in the questionable category may provide confidence bounds needed to accept the results from the model.

POPSIM generates marriages, births, divorces and deaths using stored matrices of monthly transition (event) probabilities. A conditional probability approach is used which permits these event probabilities to depend on the current characteristics and prior history of the individual. For example, the monthly birth probabilities vary by parity and marital status within five year age groups for females between 15 and 45 years of age. Births are not permitted unless the interval since the last live birth is at least nine months.

An event-sequenced simulation procedure is used in which an individual is processed only when an event occurs to him. The first step in this procedure is to generate the time interval (and hence the date) of the next vital event for each individual in the initial population. Since the type of event to occur next (i.e., a birth or a change in marital status or a death) is not known, POPSIM generates the time interval (or waiting time) separately for each of the competing events that can happen to the individual, under the assumption that nothing else does happen to him, and the event with the shortest generated time interval becomes the next event for that individual. Only this next event and its time of occurrence are carried in the record for each individual. This procedure for generating the next event is accurate under the assumption that the input parameters are independent probabilities for the competing events (or net rates) rather than crude rates.

The technique for generating the date of an event uses the inverse of the geometric distribution within time periods for which the monthly probabilities of the event remain constant. Consider, for example, the generation of the date of death of a married male who is exactly 42 years old. The monthly chance of dying for married males in the 40 to 44 age group is read from the stored table; this probability, denoted by $P_{mm}(11)$, refers to the 11-th age group for married males. A uniform random number r between zero and one is generated and the quantity

$$t = \frac{\ell n(1-r)}{\ell n(1-P_{mm}(11))}$$

is computed. This is the randomly generated number of months (from the current month) until the death of the 42 year old male, provided $t \leq 36$. If $t > 36$, then the generated age at death of this man is greater than 45, in which case a new r and t must be generated using $P_{mm}(12)$, the monthly probability of dying for married males in the 45-49 age group. The month of death will have been determined if the new $t \leq 60$. Otherwise, the process is repeated again, using the appropriate monthly death probability, $P_{mm}(13)$, for the next age group (50-54). The process is continued until a time interval to death which does not extend into the next age group is obtained for this married male. The procedure is equivalent to using a sequence of independent, uniform (0, 1) random numbers and making a decision by comparing with the appropriate probability for each month. During intervals having constant probabilities, this assumes in effect, a Poisson process for the event, or a negative exponential distribution for the intervals between events.

An important point to note is that it is assumed that no other events happen to the individual in the interval. For example, it is assumed that the marriage of the 42 year old male does not end, due to divorce or death of his spouse, when calculating the time interval to his death. However, he is subject to having these competing events occur. To determine whether he becomes widowed before he dies, POPSIM computes the date of death of his wife independently, using the monthly death probability for married females of her age class, and then compares the two death dates to determine which of these competing events is to occur first. Since divorce is also a possibility, the date of this event must be computed and compared with the two death dates to determine the next event for this 42 year old married male. Again, the time interval for each of the competing events is computed under the assumption that the other events do not occur.

For a female between the ages of 15 and 45, a competing risk is that of giving birth to a child. Monthly probabilities of a live birth by age group, parity and marital status are used to generate the time interval to the next live birth.

After the point in time for the next event is generated, it is checked to see if it falls within the interval chosen for that simulation step. If it is, time is advanced to that point and the event processed. If not, the individual is stored and not processed again until the beginning of the next simulation interval. Processing consists of recording the essential facts concerning the event and changing the status of individual characteristics affected by the event.

If the event happens to be marriage, some further processing is required. First, a decision must be made with respect to the marital status (single, widowed or divorced) of the partner. This choice is determined by means of 3 x 3 arrays of probabilities for six age groups (age of the primary individual involved). Once this has been done, the age of the marriage partner is obtained from the appropriate (first marriage or remarriage) bivariate distribution of ages of brides and grooms.

The remaining event which requires special treatment is the birth of a child. The date of birth and sex of the child is recorded in the list of secondary children associated with the mother. The sex of the child is determined by using a random number. The parity and count of the number of living children are both updated. Finally a special note is made of the birth in order to supply a sample of births (primary individuals) for the population.

When the event has been processed, a new next event is generated for the updated individual. This is continued until finally an event is obtained which is beyond the time allotted for the step in the simulation or the individual dies.

POPSIM utilizes the series of births which occur to primary females as a random sample of births to add to the computer population. Thus, whenever a birth occurs to a primary female, the infant is entered as a member of the set of her associated children (secondary individuals), and the program also keeps a separate record of all these births. Then, after the female in question has been advanced beyond the time period specified for the simulation step and has been placed in a storage file, the program returns to the list of births and establishes a RECORD for a new primary individual for each birth. Each of these new primary individuals is immediately subjected to the various risks from the date of birth to the end of the simulation step, before being placed in the storage file. In subsequent steps of the simulation, these individuals are not distinguishable from other individuals who were generated for the initial sample population. It is important to note that no contact is maintained between the mother and her child, except as a secondary individual. Future events for the child as a primary individual are generated completely independently of those generated for the mother through the presence of the child as a secondary individual in her RECORD.

Three successive arrays of information are retained in the population file for each individual, when applicable. The first array is the basic set of characteristics given in Figure 1. The second array contains the date of birth, sex and date of death for each of the secondary children associated with a primary female. The third array contains the history of events which have occurred to the individual during the simulation. The entries which are possible in the history of events array are shown in Figure 2. It is significant to note that the population or history file produced in any simulation run can be used as an initial population file in a subsequent run.

A series of tables are printed at the end of each step of the simulation. Those which have been retained for use with the hospital utilization model include:

1. The distribution of the population at the end of the interval by age, sex and marital status.
2. The distribution of the deaths in the population (primary individuals only) by age (at death), sex and marital status.
3. The distribution of births during the interval by age of mother (at the time of the birth) and the marital status of the mother.
4. The distribution of marriages by age, sex and marital status of the primary individual involved.

The data in the population file at the end of a simulation period essentially constitutes a set of vital event histories for each individual in the computer population. The data in the file (stored on a disk or on magnetic tape) are equivalent to data collected from a sample of individuals in a longitudinal survey. This history feature of the POPSIM output is very useful for special tabulation and analysis of the simulation data.

B. Some Modifications of POPSIM

1. <u>Race</u>

The original version of the POPSIM computer program was not designed to distinguish between segments of the population, such as ethnic or racial groups, in a single simulation run. The program has been modified so that two segments or subpopulations can be processed simultaneously and, at the user's option, either separate or combined tables are printed in the standard tabular output. For the hospital utilization simulation the white and non-white subpopulations are recognized. Separate parameters for births, death and marriages are required for POPSIM.

2. <u>Residence</u>

POPSIM has also been modified to classify each individual in the computer population according to whether he resides in a metropolitan area (i.e. a Standard Metropolitan Statistical Area or SMSA) or in a non-metropolitan area (non-SMSA). The residence assignment is made by age, race (white and non-white) and sex on a stochastic basis by the program which creates the initial program. Further, the program which advances the population through time determines those changes of residence which occur between an SMSA and a non-SMSA in either direction. Changes of residence are generated as events using monthly transitional probabilities on a competitive risk basis and in the same manner as the other eligible events are generated. The propensity to move is based on the age, race, sex and present residence of the individual. The event "change of residence" has been added to the history of events array, recording the current residence status, the age of the individual at the time of the move as well as the date of the event.

3. <u>Family Income</u>

Each individual in the initial population is assigned a family income according to his race, residence, type of household and age of the head of household. Four types of households to which individuals over the age of 18 and also ever married individuals 18 and under can belong have

Figure 2

Detailed Illustration of Possible Entries
in the History of Events Array

Type of event	Code for event	Date of event	Marital status of primary individual	Descriptive information
Birth of a child	1	date in months	mother's marital status	interval since previous birth
Divorce	2	date in months	married	length of marriage in months
Death	3	date in months	marital status at death	age at death in months
Marriage	4	date in months	marital status prior to marriage	interval since last change in marital status
Widowhood	5	date in months	married	interval since marriage
Death of a child	6	date in months	(not assigned)	(not assigned)

been defined for purposes of determining family income. They are:

a. If an individual is married, that individual belongs to a "husband and wife" household.
b. If a woman is divorced or widowed and has children under the age of 18, she belongs to a "female head" household.
c. If a woman is widowed or divorced with no children under 18 or is single (never-married) and over the age of 18, she belongs to an "unrelated female" household.
d. If a man is widowed or divorced or is single and over the age of 18 he belongs to an "unrelated male" household.

The model assumes that for a given race, residence, type of household and age of head (19-34, 35-44, 45-64, 65 and over) the log-normal distribution adequately represents the distribution of family income. It is further assumed that the mean family income can change over time, but the variance is not assumed to vary with time, an assumption for which there is empirical support and which is consistent with the assumption that the distribution is log-normal. A computer routine is used by the Phase II program to update family income (consistent with the initial family income) whenever it is called and also at the end of each simulation step. The updated family income and the time of the last update are retained in the individuals RECORD (items 19 and 20).

III. THE HOSPITAL EPISODES MODEL (HOSPEP)

A. Introduction

In two previous studies [3, 14], a probability model for generating hospital admissions and durations of stay for the U.S. civilian non-institutional population, based primarily on age and sex, was developed and refined. Using this model, a computer program was written to simulate hospital episodes for each individual in a given population. In the present study the earlier probability model has been extended to include a more detailed hospitalization history than just admissions and duration of stay and to take into account more demographic characteristics of each individual than just his age and sex. In addition, a computer program has been written to simulate hospital episodes using this extended model.

Specifically, the present study has extended the earlier hospital episodes model in the following ways:

1. For each individual in a given population the model will generate in addition to hospital admissions and duration of stay, the reason for hospitalization (diagnosis), whether or not surgery is performed and the bedsize of the hospital the individual is admitted to.
2. The model considers not only the age and sex of each individual in the population but also his race, family income, hospital insurance status and whether

his residence is located in a metropolitan area (SMSA) or in a non-metropolitan area (non-SMSA).

The computer program for the extended hospital episodes model has been set up to run either with input data from a POPSIM history file which gives the demographic characteristics of a simulated population or with data similar to that which POPSIM generates. The output of the program gives detailed tables on the hospital utilization of the simulated population and also a hospital utilization data tape that can be read as input into a hospital manpower model computer program (cf. [1]).

B. Summary of Earlier Model

A detailed description of the original probability model for hospital episodes is given in [3]. This model was developed to generate hospital admissions and durations of stay for a given population. The hospital admissions model assumes that the number of admissions annually, X, for an individual in a particular age-sex class follows the Poisson distribution. That is,

$$f(X|\lambda) = \frac{e^{-\lambda t}(\lambda t)^X}{X!} \quad X = 0,1,2,... \tag{1}$$

where the parameter λ is the expected number of hospital episodes during time t for the individual. It also assumes that λ varies over the population and has a Gamma distribution. That is, the distribution of λ for all individuals in the population in a particular age-sex class is assumed to be

$$g(\lambda) = \frac{\beta}{\Gamma(\alpha)} (\beta\lambda)^{\alpha-1}e^{-\beta\lambda}, \alpha>0,\ \beta>0\ . \tag{2}$$

It follows that

$$\begin{aligned} f(X) &= \sum_{\lambda} f(X|\lambda)g(\lambda)d\lambda \\ &= \binom{\alpha+X-1}{X}\left(\frac{\beta}{1+\beta}\right)^{\alpha}\left(\frac{1}{1+\beta}\right)^{X},\ X=0,1,2,... \end{aligned} \tag{3}$$

which is the Negative Binomial distribution. Thus, the model assumes that the Negative Binomial distribution provides a reasonable description of observed data on number of hospital episodes annually for a given age-sex group. The duration of stay model assumes that the distribution of the number of days spent in the hospital by an individual is distributed as a log-normal variate with parameters depending on the age-sex class of the individual.

C. Modification of Earlier Version of HOSPEP

There are six basic steps in the extended model for generating the hospital episodes for each individual during a simulation period. They are:

Step 1. Assign hospital insurance status to the individual depending on his race, age and family income.

Step 2. Generate the next date of admission to a hospital depending on his race, age, sex, family income and hospital insurance status. If this date is beyond the end of the simulation period, the next individual is processed beginning with Step 1.

Step 3. Assign a diagnosis for the generated admission depending on his age and sex.

Step 4. Determine whether surgery was performed or not depending on the age, hospital insurance status, diagnosis and residence of the individual.

Step 5. Assign the bedsize of the hospital depending on the diagnosis and surgery status of the individual.

Step 6. Generate the date of discharge of the individual from the hospital depending on his age, family income, hospital insurance status, diagnosis, surgery status and size of hospital. If this date is still within the simulation period, the program returns to Step 2 and continues to process the same individual. If the generated date of discharge is beyond the end of the simulation period, the next individual is processed beginning with Step 1.

In addition to these steps the HOSPEP program has special subroutines for delivery episodes and for generating the hospital episodes for persons in their last year of life.

Since the particular factors which explain the variation in the conditional probabilities appropriate to the occurrence of each event or to the distribution of each characteristic to be assigned in the six steps of the model were not known, considerable statistical analysis of available data was necessary before the appropriate probability distributions to use in the model could be specified. The factors listed above for each step were derived from a larger set of factors in each instance.

Although data exist from which the needed joint distributions of the dependent and independent factors could have been computed (e.g. the distribution of number of hospital episodes annually by age, race, sex, family income, residence and hospital insurance status), but these data could not be obtained. Since tabulations of these factors were available only in two, three, four and five-way tables, it was necessary to devise a special procedure whereby the critical factors for each step in the model could be determined.

The special procedure consisted of using iterative proportional fitting to construct a k-dimensional contingency table from a set of lower order marginal tables involving k factors [15]. This procedure essentially disaggregates the data in the lower dimensional tables into the single higher dimensional table in a manner which preserves all of the information on main effects and interactions among the factors that are present in the same lower order tables. Further, any higher order interaction which is missing from all of the lower dimensional tables is not introduced into the k-dimensional table by this procedure. It is of interest to note that the marginal tables from several different studies can be used to derive the requisite k-dimensional table. A general computer program, TABLES, has been written to carry out the iterative procedure for constructing a k-dimensional contingency table from a set of tables of lower dimension.

Using data from the Health Interview Survey and the Hospital Discharge Survey, both conducted by the National Center for Health Statistics, and from the Bureau of the Census Current Population Survey, the program TABLES was used to estimate a higher order contingency table for each of the six basic steps. Analyses of variance on the natural logarithms of the cell frequencies in each table were then carried out to determine the factors which best determined the conditional probabilities for each step.

D. Generation of Hospital Episodes

The computer program for the extended HOSPEP model generates hospital episode histories for each individual on the input population file for a specified simulation period. If the population input is from POPSIM, a HOSPEP edit routine selects the appropriate data for each individual pertinent to the time period of interest, carefully checking for changes in age class, family income and residence during the time period. If there are any changes, the simulation period is broken up into subperiods for which the relevant characteristics of the individual remain the same. This is essential for proper use of the conditional probabilities which generate the hospital admissions and characteristics for each admission.

The basic input data required by HOSPEP, other than the information in the population file for each individual to be processed, includes:

1. Conditional probabilities of having hospital insurance by age, race and family income.
2. Parameters to determine hospital admission dates by race, age, family income and hospital insurance status.
3. Conditional probability distribution of diagnosis by age and sex.
4. Conditional probabilities of surgery status by age, hospital insurance status, diagnosis and residence.
5. Conditional probability distribution of bedsize of hospital by diagnosis and surgery status.
6. Parameters to determine length of stay in the hospital based on age, surgery, diagnosis, family income, hospital insurance status, and size of hospital.
7. Daily probabilities for hospital admission for persons in their last year of life.
8. Parameters to update family income by age, race, residence and type of household.

The categories used for each of the variables in the tables of parameters and conditional probabilities are shown below:

Race

White
Non-White

Sex

Male
Female

Age Group

1 <15 years
2 15-44 years
3 45-64 years
4 65+ years

Family Income Group

1 < \$4,000
2 \$4,000-\$6,999
3 \$7,000-\$9,999
4 \$10,000+

Hospital Insurance Status

Insured
Not Insured

Residence

SMSA
Non-SMSA

Surgery

Surgery
Nonsurgery

Size of Hospital (# beds)

1 < 50
2 50-74
3 75-99
4 100-149
5 150-199
6 200-299
7 300-399
8 >400

Diagnosis

1 Infective and parasitic diseases
2 Malignant neoplasms
3 Benign neoplasms and neoplasms of unspecified nature
4 Allergic, endocrine system, metabolic and nutritional diseases
5 Diseases of blood and blood-forming organs
6 Mental, psychoneurotic and personality disorders
7 Diseases of the nervous system and sense organs
8 Diseases of the circulatory system
9 Diseases of the respiratory system
10 Diseases of the digestive system
11 Diseases of the genitourinary system
12 Deliveries and complications of pregnancy, childbirth and puerperium
13 Diseases of the skin and cellular tissue
14 Diseases of the bones and organs of movement
15 Congenital malformations
16 Certain diseases of early infancy
17 Symptoms, senility, ill-defined conditions and special admissions
18 Injuries and adverse effects of chemical and other external causes

Each individual is processed by HOSPEP for the entire simulation period, one at a time. Hospital insurance status, diagnosis, surgery and size of hospital are assigned by generating a uniform random number (0,1) and comparing it with the appropriate conditional probability distribution.

The earlier model for hospital admissions has been simplified slightly by setting the parameter $\alpha = 1$ (Eq. 2 above). Thus the original Gamma distribution of the parameter λ (which reflects the variation among individuals of the same age, race, family income and hospital insurance status in their likelihood of having to be hospitalized) reduces to a Negative Exponential Distribution with parameter β where $1/\beta$ is the expected number of hospital episodes annually per person for the sub-population. In order to determine the next date of admission, HOSPEP first derives the daily admission probability P appropriate to the individual by computing

$$P = \frac{-\ell n\ R}{365\ \beta}$$

where $\ell n\ R$ is the natural logarithm of a uniform random number (0,1). The next date of admission is then generated by sampling the geometric distribution with parameter P . In order to derive the length of stay d for each admission, it is necessary first to compute the mean and standard deviation of the appropriate normal distribution for log d . For this purpose, linear additive models of the significant effects and interactions of the various levels of age, family income, hospital insurance status, diagnosis, surgery status and size of hospital on the means and standard deviations of the distributions of log d were derived (by analysis of variance) from the k-dimensional contingency table of these factors generated by the program TABLES. A random normal deviate, say z , is then generated and the corresponding random variable log d computed.

As each individual is processed, the HOSPEP program tabulates a series of tables. These tables are printed after all individuals have been processed, including:

1. The distribution of the average population for the simulation period by age, race, sex, family income, and hospital insurance status.
2. The distribution of hospital discharges by age (<15, 15-44, 45-64, 65 or older), race, sex and days in hospital (1, 2-4, 5 or more).
3. The distribution of hospital discharges by age, family income, hospital insurance status and days in hospital.
4. The distribution of hospital discharges by age, sex, diagnosis, surgery status and days in hospital.
5. The distribution of hospital discharges by diagnosis, surgery status, and days in hospital for each hospital bedsize class.
6. The distribution of the sample population by number of hospital admissions (0, 1, 2 or more), age, race and sex.
7. The distribution of the sample population by number of hospital admissions, age, family income and insurance status.
8. Hospital admission rates per 1000 persons by age, race and sex.
9. Hospital days rates per 1000 persons by age, race and sex.
10. Hospital days by age, race and sex.
11. Hospital days by age, family income and insurance status.
12. Hospital days by age, sex, diagnosis, and surgery status.
13. Hospital days by size of hospital, diagnosis, and surgery status.
14. Average length of stay by age, race and sex.

At present no record is retained at the end of the simulation period of the hospital episodes experienced by each individual in the population. It is recognized that such a record could be quite useful for further tabulations of the simulation results and hence provision for including this feature is now underway. The HOSPEP program does prepare an output data file for use as input to a Hospital Service Model designed to derive hospital requirements from hospital utilization data [1]. The information on this file is restricted to hospital episodes only and includes the admission and discharge dates, diagnosis, surgery status and size of hospital.

Parameters for POPSIM and HOSPEP for simulation of hospital utilization for the U.S. civilian non-institution populations have been developed for purposes of testing the model.

ACKNOWLEDGEMENTS

The authors appreciate the many useful discussions with Mr. Garrie J. Losee of the Bureau of Health Professions Education and Manpower Training, NIH. A number of other individuals have contributed important ideas to the development of POPSIM and HOSPEP. We wish to acknowledge specifically Dr. F. G. Giesbrecht, Dr. B. V. Shah and Mr. P. V. Piserchia.

REFERENCES

[1] F. D. Kennedy, Simulation of Health Manpower Requirements in a Hospital Setting, presented at the Conference on a Health Manpower Simulation Model, National Institutes of Health, Washington, August 1970.

[2] D. G. Horvitz, F. G. Giesbrecht, B. V. Shah, and P. A. Lachenbruch, POPSIM, A Demographic Micro-Simulation Model. Paper presented at the meetings of the International Union for the Scientific Study of Population, London, September 1969.

[3] National Center for Health Statistics: Computer Simulation of Hospital Discharges, Vital and Health Statistics, PHS Pub. No. 1000-Series 2-No. 13, Public Health Service, Washington, U. S. Government Printing Office, February 1966.

[4] B. V. Shah, On the Mathematics of Population Simulation Models, presented at the annual meetings of the Population Association of America, Atlanta, April 1970.

[5] B. V. Shah and F. G. Giesbrecht (1969), Mathematical Models for Changing Marriage Patterns. Paper presented at the meetings of the International Union for the Scientific Study of Population, London, September 1969.

[6] J. C. Barrett, A Monte Carlo Study of Reproduction. Presented at the Society for Human Biology Symposium, London, England, 1967.

[7] H. Hyrenius and I. Adolffson, A Fertility Simulation Model. Report #2, Demographic Institute, University of Göteborg, Sweden, 1964.

[8] H. Hyrenius, I. Holmberg, and M. Carlsson, Demographic Models: DM 3. Report #5. Demographic Institute, University of Göteborg, Sweden, 1967.

[9] I. Holmberg, Demographic Models: DM 4. Report #8, Demographic Institute, University of Göteborg, Sweden, 1968.

[10] G. H. Orcutt, M. Greenberg, J. Korbel, and A. Rivlin, Micro-analysis of Socioeconomic Systems. Harper, New York, 1961.

[11] J. C. Ridley and M. C. Sheps, An Analytic Simulation Model for Human Reproduction with Demographic and Biological Components. Population Studies, Vol. XIX, pp. 297-310, 1966.

[12] F. G. Giesbrecht, G. Ranney and J. R. Chromy, Demographic Microsimulation Model POPSIM I: Manual for Programs to Generate the Initial Population, Open Core Model. Technical Report #4, Project SU-285, Research Triangle Institute, September 1969.

[13] F. G. Giesbrecht and L. Field, Demographic Microsimulation Model POPSIM II: Manual for Programs to Generate Vital Events, Open Core Model. Technical Report No. 5, Project SU-285, Research Triangle Institute, September 1969.

[14] F. Giesbrecht, Further Development of Computer Simulation of NCHS Hospital Data. Final Report, Project SU-204, Research Triangle Institute, September 1966.

[15] B. V. Shah, On Predicting k-Dimensional Contingency Tables from Lower Dimensional Marginal Tables, presented at the Conference on a Health Manpower Simulation Model, National Institutes of Health, Washington, August 1970.

A MICROSIMULATION MODEL OF THE HEALTH CARE SYSTEM

Donald E. Yett,
Leonard Drabek,
Michael D. Intriligator,
Larry J. Kimbell,

Human Resources Research Center
Graduate School of Business Administration
University of Southern California[1]

I. An Overview of the Model

This paper reports on the conceptualization of a simulation model of the health care system which deals explicitly with the microanalytic behavior of the individuals and institutions comprising the system.[2] The scope of the model includes all aspects of the production and utilization of personal health care but excludes environmental, mental, and dental health, as well as drugs and biomedical research. Even with these limitations the system under consideration is extremely complex and rich in structural and institutional detail, involving, as it does, many different participants, typical institutions, and complex multi-party transactions, etc.

To cope with the complexities of the health care system we have employed a modular approach, organizing the system into three modules--health services, health manpower, and health professions education. These particular modules were chosen because interactions within each are manifold, while interactions among them are fewer and simpler. The modules and interactions, which summarize the conceptualization of the complete model, are displayed in the block diagram below. In this particular conceptualization there are five simulated populations. Three of the populations are composed of people namely: individuals, health manpower, and students. The remaining two populations are composed of institutions, namely: health services institutions and health professions education institutions.

The individuals and institutions in each of the simulated populations are described by certain attributes or characteristics. The characteristics for the population of individuals include age, sex, residence, ethnic group, family income, health insurance status, etc. The population of health manpower is subdivided into physicians, nurses, and allied health manpower, and individuals in each of these subpopulations are characterized by age, sex, length of training, specialty (for physicians), marital status and number of children (for nurses), etc. The population of students is characterized by age, sex, ethnic group, previous education, and marital status. With respect to the institutions, the population of health service institutions is divided into hospitals, nursing homes, outpatient clinics, and physicians' offices--each of which are, in turn, described by a variety of characteristics such as size, ownership or control, nature of payment, length of stay, etc. Similarly, health professions education institutions are divided into medical, nursing, and allied health personnel schools--each of which are described by size, university affiliation, ownership, accreditation, etc.

Each module contains markets corresponding to that particular sector of the health care system; and these markets are influenced by the simulated populations as illustrated in the block diagram. Thus, in the health services module the demand for health services is generated by the population of individuals, and the supply of health services is generated by the population of health service institutions. In the health manpower module the demand for health manpower is obtained from the population of health service institutions. It is derived from the underlying demand for health services. The supply of health manpower is obtained from the population of health manpower. Finally, in the health professions education module, demand for health professions education is generated from the population of potential students, while the supply is generated from the population of training institutions.

The markets in each of the modules are characterized by widespread and persistent disequilibria. This is reflected in the model by variations in waiting time, and allocations of shortages and surpluses according to the nature of the disequilibrium.

Perhaps the best way to describe the set of interactions that take place among the five populations is to summarize the events occuring in a typical simulation run. For this purpose we will refer to the linkages indicated above by the number given to each on the Block Diagram. The basic solution sequence is presented in Figure 1. Major clusters of these linkages are enclosed in the boxes lettered A through G.

Before simulation begins each of these five populations is generated, and initial conditions are specified for values of the lagged variables.

A. We begin with an individual from the general population, assign a diagnostic condition to him (1), determine his needs for patient visits (2) and bed days (5), then adjust his needs for patient visits by the prices he faces, giving his demands for patient visits (3). The demands for patient visits are then aggregated over all individuals to give market demands for outpatient services (4). The determination of bed days demanded is slightly more complex than that of patient visits demanded due to the usual requirement that a physician be seen before admission to a hospital is granted. The quantities needed are transformed into demands by taking account of prices and financial constraints (8), as in the case of patient visits. Additionally, estimates of the patient visits to be received are employed (7) to

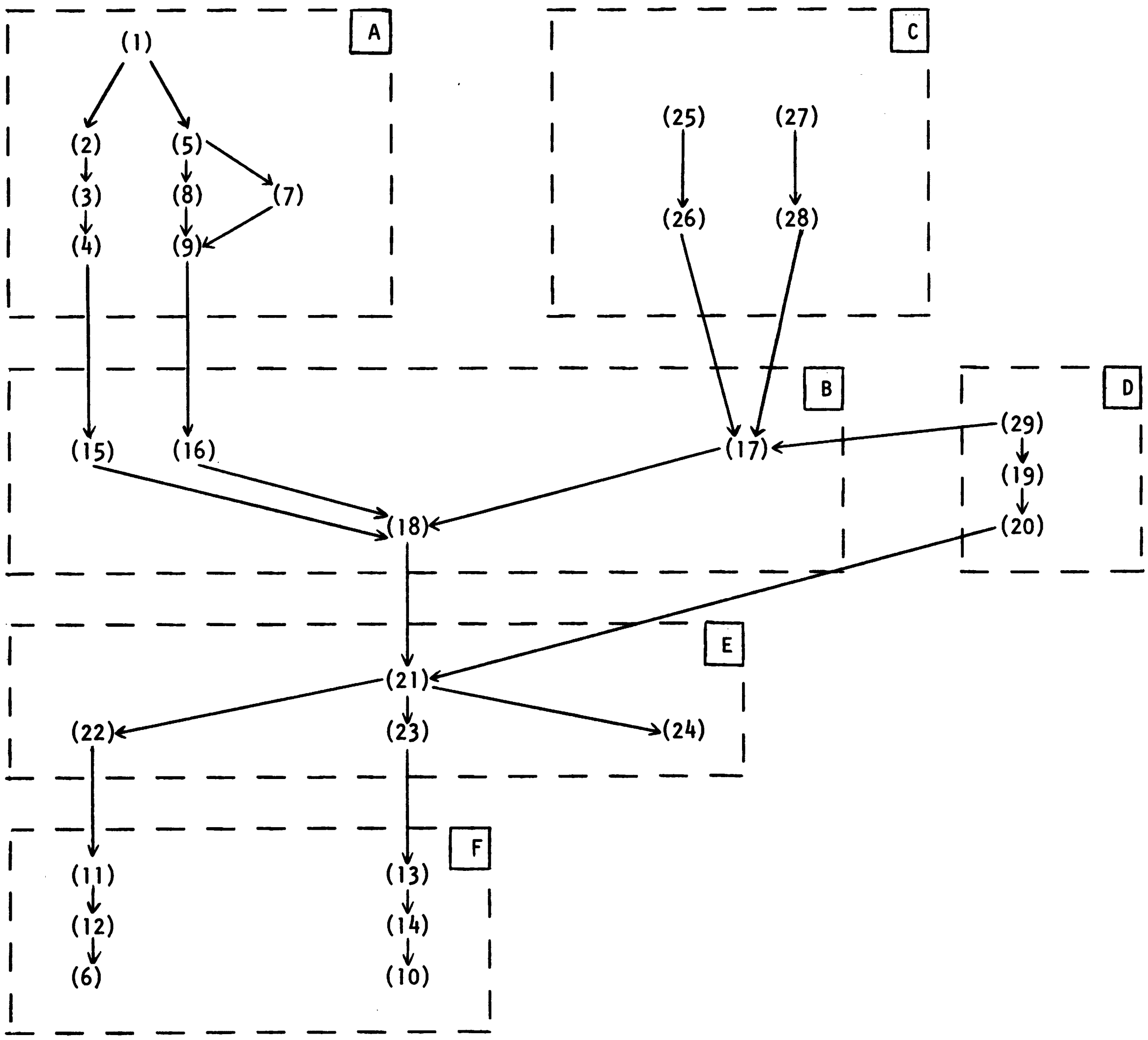


Figure 1: The Solution Sequence For One Period
(Numbers refer to linkages on the Block Diagram)

adjust further for the physician approval aspect of hospital admission. That is, if fewer patients manage to see physicians, then fewer patients will demand care from hospitals. Bed days demanded by individuals are then aggregated (9) to yield market demands for inpatient services.

B. On the basis of the demands faced by service institutions, demands for health manpower are derived separately for the various types of inpatient institutions (15) and outpatient institutions (16). These are aggregated to give part of the market demands for health manpower. Additional health manpower demands come from educational institutions for faculty (17). The demands for manpower from all three sources--outpatient, inpatient and health professions education institutions--are aggregated to give market demands for each category of health manpower (18).

C. The demands for faculty are derived by the following sequence of steps. First, the demand for health professions education by individual students is generated (25), then these demands are aggregated to yield market demands (26). Independently, the supplies of openings at individual institutions are developed (27), and aggregated to provide market supplies (28). If market demands for education exceed available openings, as is to be expected for physicians' education, then the scarce openings are allocated to the students. If the number of openings exceeds the number of qualified applicants, as is expected in nursing programs, then the vacancies are allocated among the programs. Demands for faculty by educational institutions (17) are derived from the numbers of students educated. Eventually students graduate (29) and enter the various pools of health manpower.

D. Having generated the manpower demand side in B, we turn next to the supply side. Hours of work supplied by each individual in the manpower pool are developed (19) and aggregated over all similar types of individuals to obtain market supplies, in terms of hours of work offered, by all types of health manpower (20).

E. The demands and supplies for health manpower interact in the health manpower market to determine wage adjustments, shortages, and the actual allocations of manpower to the various types of institutions (21), with corresponding feedbacks to the populations of outpatient service institutions (22), inpatient service institutions (23), and educational institutions (24).

Following this key solution step, the forces in motion branch outward again, with the manpower markets' influences on the outpatient, inpatient, and educational institutions proceeding to move through separate channels.

F. Outpatient institutions have now received the inputs which (through the production function) limit their capacity to provide patient visits to individuals. The next step determines the amounts of patient visits these institutions will seek to supply to individuals in the various diagnostic classes (11). These supplies (12) interact with demands in the outpatient services market to produce changes in prices and waiting times, and to result in allocations (6) of physician visits (if shortages occur) to the different categories of patients.

A similar pattern develops for inpatient service markets. The allocation of manpower to these institutions (23) sets limits on the amounts of bed days (13) they will seek to supply to individuals in the various diagnostic classes. These supplies (14) interact in the inpatient markets to generate changes in prices and waiting times, and, further, to produce allocations (10) of bed days (if shortages occur) to the various types of patients.

This completes the steps through a given simulation solution for one period. Additional periods would normally be run by updating all populations by the growth functions, and then repeating the above solution steps.

The foregoing conceptualization is, of course, highly simplified by contrast to the enormously complex real world system. Nevertheless, we feel that it represents a major step in the direction of adequately modeling the real system. Previous studies have focused on only one small part of this conceptualization looking at, for example, the demand for health services or "shortages" of physicians in isolation from the rest of the system. Such a partial equilibrium approach is obviously less desirable than a general equilibrium approach, which is the ultimate goal of our health care system modeling efforts.

Perhaps the most important reason for considering the entire health care system is that models of the complete system could be of tremendous value in designing and evaluating policy with respect to national health programs. Experimenting with the real world system is prohibitively expensive, both in terms of both costs and human suffering. In this regard, Medicaid, the major Federal program for providing health care services to low income families, has been criticized on the grounds that it amounts to an extremely expensive unsuccessful experiment which now should be replaced.

With a running microsimulation model of the health care system it would be possible to experiment on the model rather than on the real world system. Alternative policies could be studied by simulating the real world system modified to allow for these policies, and a particular policy or set of policies could be adopted on the basis of an examination of the simulated outcomes. An example is national health insurance. Several alternative plans have been proposed and a number are under active consideration by the Federal government.

These include: Federal payment of a certain stipulated percentage of hospital and physician costs; Federally sponsored major medical insurance which would pay for all health expenditures above a stipulated percent of family income lowering of the eligibility age for Medicare and, possibly, also covering preschool children under this program; and numerous other. These could all be analyzed using the model we have developed. The first impact typically would be in the health services market. The demand for such services would shift outward, increasing the amount of services supplied and prices of these services. In order to produce the increased services provided, the health service institutions would shift outward their demand for health manpower. The next impact would be on the health education module, shifting outward the demand for health professions education. The last impact would, in the long run, eventually result in an increase in the size of the manpower pool. All of these impacts, and their complex interactions, could be analyzed quantitatively using a running version of our microsimulation model.

II. Specific Features of the Model

In designing the model described above, we adopted a micro rather than a macro approach because of the requirement that the model be capable of simulating both the aggregate behavior observed in health care markets and the individual behavior of the many participants in these markets.[3]

A macro econometric model may be capable of predicting the gross magnitudes of certain indicators of health care activities--and it would clearly be less complex to operate and less expensive to build. However, such a model yields little or no information about the distribution of health care, yet the distribution as well as the total level of health care is of manifest policy concern. Medicare for the aged and Medicaid for the indigent are two important illustrations of programs designed to alter the distribution of care among types of consumers.

A microsimulation model is ideally suited for estimating the distributional, as well as the aggregate, impacts of major policy changes. However, there is a problem in moving from a macro approach to a microsimulation approach in building this type of model--namely, the temptation to focus narrowly on one component of the health care system in order to make microsimulation replicate as closely as possible, the intricacies of the real system being modeled. This effort to simulate intricate details almost necessitates a drastic restriction of the scope of the model, since to model many components, and to simulate each in great detail, becomes totally unmanageable. It is especially important to guard against this danger in the design of a model of a highly interdependent system in which changes in one sector have vital ramifications with respect to the variables in other sectors of the system. In the health care system, for example, changes in medical and hospital insurance for the aged will affect not only the utilization of services covered by insurance, but also services not covered. Additionally, the utilization of inpatient and outpatient care by the rest of the population is affected as a result of fee increases, longer waiting times, and, possibly, shortages in these markets. The challenge in developing a microsimulation model, therefore, is to balance the desire for comprehensive systems analysis against the temptation to reach for too much detail and intricacy in modeling individual behavior. Although recent technological advances, in the form of better computer software and hardware, greater data availability and more refined mathematical techniques, have enormously expanded capability to move in both directions, there still remains a tradeoff which, in the last analysis, cannot be avoided in constructing a microsimulation model of the complexity we have undertaken. In this regard, we feel that the terms "micro" and "macro" suggest a false dichotomy whereas, in reality, there exists a continuum of degrees of detail and aggregation. For example, we simulate the behavior of each patient and physician, yet the two interact only in aggregate markets. That is, the behavioral variables associated with individual patients and physicians are treated as "micro" in nature; but the fees for physician services are "macro" variables insofar as they are determined not for every pair of interacting patients and physician but only for the overall markets relevant to the various categories of physicians. This is illustrative of a large number of problems where we tried to design an optimal mix of detail and generality. As a consequence of the approach chosen our model is considerably more aggregative than is true of most previous microsimulation models, while, at the same time, it is much more finely disaggregated than any existing macroeconometric model of the health care system.[4]

The device of having populations not interact on a specific individual to individual basis has been used generally in our model.[5] That is, a patient is not matched with a specific physician, and sent to a specific hospital. Rather, aggregate demands for physicians services and hospital services interact with aggregate supplies. To have included multi-party matching would have yielded little in detailed information by comparison to the loss in not treating the types of variables that can be included using the adopted approach of interacting aggregate supply and demand equations. This approach has made it feasible to employ several populations simultaneously in our microsimulation model--which would have been prohibitively expensive if we had followed the lead of others and maintained multiparty matching.

In addition, the decision to use aggregate interactions between populations resulted in significant data storage and processing simplifications, since microsimulation data can then be stored according to what we have called the "cell", instead of the more typical "file", approach. In a file, each of say 20,000 individuals is represented by a set of variables characterizing that individual. The first file entry will, for example, contain an individual's age, marital status, sex, etc. In the cell approach one uses a complete joint distribution exhausting all possible types of individuals. While the cell and the file approaches are fully equivalent in terms of information contained, in the sense that except for an arbitrary permutation of file entries one can transform a data set organization from one into the other, their processing consequences are not the same. First, the cell approach is very sensitive to the subdivision of attributes and, thus, is ideal for coarse divisions. On the other hand, it is completely invariant to the sample size, which is, of course, a definite strategic advantage in the effort to control sampling error. By contrast, the file approach is invariant to the degree of subdivision of attributes, but very sensitive to the number of individuals in the population. Second, the cell approach has the advantage that a random number, drawn from the distribution of the sum of random events for a given set of individuals, can be used instead of the sum of many random drawings, one for every individual. It is clear that no essential information is lost in microsimulation by aggregating over individuals of identical characteristics. Furthermore, a certain amount of aggregation may be extremely efficient, giving rise to little loss of detail yet reducing the size of the simulation task significantly.

Other unique characteristics of our model result from the inclusion of salient institutional features of the health care system, rather than considerations of operational feasibility and efficiency. One such characteristic is the widespread use of non-price rationing when shortages occur in most, if not all, of the relevant markets. That is, in such situations we employ a two-price system, using variations in waiting time to ration services when (or for as long as) pecuniary price adjustments are insufficient to accomplish this task. Thus, for example, in the services market of our model admissions to hospitals are based partly on priorities which give preference to the critically ill when there are not sufficient beds to satisfy the total demand. By the same token, since it is apparent that prices (i.e., tuitions) are not set so as to clear the market for medical school openings, we developed an allocation mechanism which distributes the shortage of such slots among the demanders. A similar mechanism distributes the surplus of first-year openings in nursing schools among the suppliers.

Still another aspect of our model which represents a marked departure from its predecessors has to do with the specification of input-output relationships. Specifically, our model recognizes that many factors of production are employed in producing each of the major categories of health services. In this connection the popularly used production functions (e.g., the n-factor CES function) have a characteristic that is manifestly absurd if applied to the analysis of such finely subdivided input categories. This characteristic is the constancy of the partial elasticity of substitution (however defined) between all pairs of inputs. This means that if any two factors are made close substitutes then all factors must be close substitutes. Thus, for example, if registered nurses are estimated to be very close substitutes for licensed practical nurses in some institutional settings, the form of these production functions will force one to the further implication that RN's are very close substitutes for physicians. In order to overcome this problem, we have employed a multi-level CES function which permits tractable variation in the partial elasticity of substitution.[6] In this form of the production function, the direct partial elasticity of substitution between factors in the same subset is constant and easily variable. The Allen partial elasticity of substitution between factors not in the same subset is also constant and also easily variable. Thus, the multi-level CES production function permits certain factors to be close substitutes for each other while, at the same time, being poor substitutes for factors not in the same subset.

Finally, in order to handle the heterogeneous nature of the "outputs" produced by both outpatient and inpatient institutions, a similar modification was made with respect to the output side of the production functions employed in our model. That is, we have employed a multi-output production function with mathematical properties relating the various outputs that are similar to those relating inputs. Thus, output substitutability is also specific to pairs of outputs and can vary between diagnostic types of patients by institutional setting.

III. Implementation of the Model

The brief report presented above has, of necessity, been limited to a broad overview of the model, and a few remarks regarding its most unique features. A complete description of the model we are now seeking to implement is contained in the forthcoming proceedings of the National Conference on Health Manpower Simulation Models. It is our hope that within two to three years we will have an operational prototype of the model. Once the model is actually running it will be progressively modified by the addition of new relationships, and the respecification of those now included, until it has been judged acceptable as a forecasting tool by health planners and other public officials charged with responsibility for establishing and administering national health programs.

Footnotes

1. M. D. Intriligator and L. J. Kimbell are also on the faculty of the Department of Economics, UCLA.

2. This project was supported in part by Contract No. PH-108-69-69 from the Bureau of Health Professions Education and Manpower Training, National Institutes of Health. A more detailed description of the model is given in our final report ("The Development of a Microsimulation Model of Health Manpower Demand and Supply"), forthcoming in the Proceedings of the Conference on Health Manpower Simulation Models, August 31-September 1, 1970, Bethesda, Maryland.

3. The pioneering research on microsimulation models was performed by Orcutt and his associates. For a detailed description of their work see: Orcutt, G. H., Greenberger, M., Korbel, J., and Rivlin, A. Microanalysis of Socioeconomic Systems: A Simulation Study. New York: Harper & Brothers, 1961.

4. For examples of macroeconometric models of the health care system see:

   Feldstein, M. S. "An Aggregate Model of the Health Care Sector," Medical Care, V (November-December, 1967), pp. 369-81.
   Feldstein, M. S. "An Econometric Model of the Medicare System." Discussion Paper No. 103. Cambridge, Massachusetts: Harvard Institute of Economic Research, Harvard University, January, 1970. (Processed.)
   Feldstein, P. J., and Kelman, S. "A Framework for an Econometric Model of the Medical Care Sector," in Empirical Studies in Health Economics, Klarman, H. E. (ed.). Baltimore: The Johns Hopkins Press, 1970, pp. 171-90.

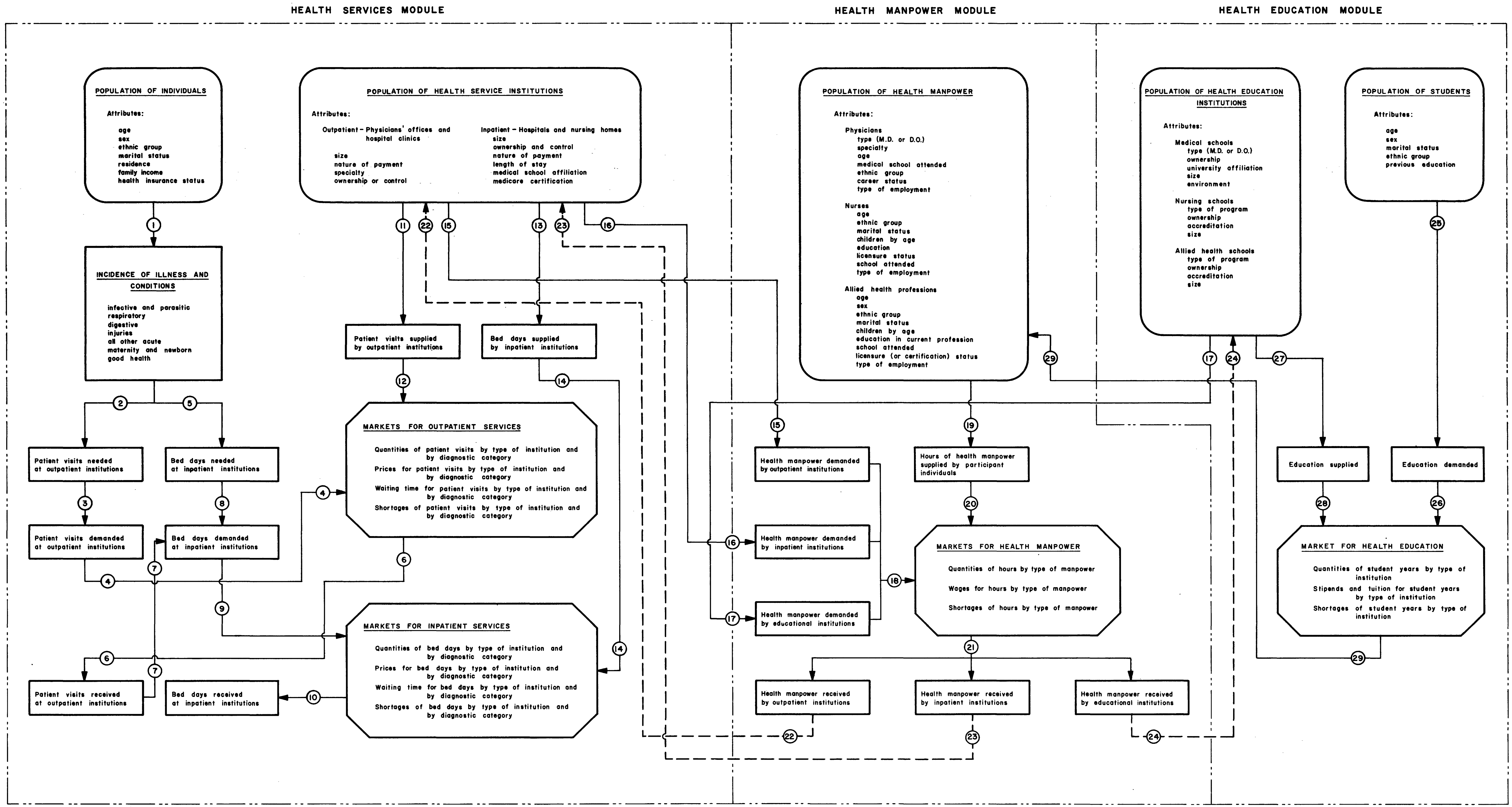
5. Investigators at the Research Triangle Institute compared the relative efficiency of "open" and "closed" microsimulation models, and concluded that "open" versions are much more efficient for purposes such as ours. A detailed description of their recent research is presented in the forthcoming Proceedings of the Conference on Health Manpower Simulation Models, August 31-September 1, 1970, Bethesda, Maryland.

6. The multi-output, multi-level CES production function takes the form:

$$\left((y_1^{\delta_1} + y_2^{\delta_1})^{\frac{\delta}{\delta_1}} + (y_3^{\delta_1} + y_4^{\delta_1})^{\frac{\delta}{\delta_2}}\right)^{\frac{1}{\delta}} = \left((x_1^{\rho_1} + x_2^{\rho_1})^{\frac{\rho}{\rho_1}} + (x_3^{\rho_2} + x_4^{\rho_2})^{\frac{\rho}{\rho_2}}\right)^{\frac{1}{\rho}}$$

   This function can be called a two level CES production function in both inputs and outputs.

BLOCK DIAGRAM OF A MICRO-SIMULATION MODEL OF HEALTH MANPOWER SUPPLY AND DEMAND



MICROANALYTIC SIMULATION MODELS

Robert Michielutte, Wake Forest University - J. Timothy Sprehe, Agency for International Development

Introduction

In order to place our discussion of microanalytic simulation models in proper perspective, it is helpful to make explicit a definition of simulation which best fits the models to be discussed.

In its broadest sense, simulation is "essentially a technique that involves setting up a model of a real situation and then performing experiments on the model" (Naylor, et al., 1966). The types of simulation we are concerned with, however, are more restricted in scope. They are all computer simulations, based on models that have been developed through some combination of logical and statistical information, which attempt to describe the functioning of specific systems over time. Naylor and others (1966) have suggested a working definition of simulation, which explicitly incorporates these characteristics. It is as follows:

> Simulation is a numerical technique for conducting experiments on a digital computer, which involves certain types of mathematical and logical models that describe the behavior of a business or economic system (or some component thereof) over extended periods of real time.

We would extend this definition to include any system or subsystem of interest to the social scientist that is amenable to the development of logical and mathematical models.

Within the framework of this definition, it seems essential, both in designing and evaluating computer simulation models, to ask the question: why do we simulate in the first place? A number of justifications for the use of simulation techniques have been suggested:

1) It may be either impossible or too costly to observe or experiment with certain processes in the real world. It is, for example, impossible to control the birth rate of even a small sample of women for the purposes of observing its impact on other processes and of formulating hypotheses.

2) The observed system may be too complex to describe mathematically in such a way that analytic solutions could be obtained.

3) In some situations, it is possible to develop a mathematical model, but it cannot be solved by straightforward analytic methods.

4) It may be possible to derive a set of mathematical equations which can be solved to describe a system, but validating experiments to test the model cannot be performed. This differs from the first problem only in the manner in which the simulated data would be utilized. "In the first case simulated data are necessary to formulate hypotheses whereas in the latter case simulated data are required to test hypotheses" (Naylor et al., 1966: 7).

It would appear that the basic rationale for the simulation models to be discussed here lies in points one and four. In attempting to answer questions concerned with the future distribution of income, hospital utilization, or occupational prestige, one approach is to develop simulation models which will allow the researcher to project current trends and to conduct hypothetical experiments on the simulated populations.

In this brief discussion, we propose to examine the methods and objectives of four simulation projects--those of the Urban Institute, The Research Triangle Institute, USC's Human Resources Center, and our own--in the light of some general facing simulations of this type.

The Simulation Models

I. The Urban Institute has undertaken the construction of a microanalytic model for the simulation of income activity at the family level (Orcutt et al., 1970). The model provides a dynamic description of the processes of birth, death, and changes in marital status. In addition, it has been designed to simulate education as a first step in the simulation of income variables.

The present stage of development operates on what the authors term a "persons" model, with a "family" model projected for the future. This distinction appears to be analogous to the open and closed populations discussed in the other models. Basically, the distinction involves the grouping and identification of individuals in the simulated population. In the "family" or "closed" model, marriage mates are selected from within the population and the households/families are treated as a unit. In the "person" or "open" model, marriage mates are in a sense hypothetical and exist only as information on an individual's record (one in the smulation population). In addition, no attempt is made to keep household the family members together. The relative merits of the two models will be discussed in a later section.

One particularly interesting feature of the Urban Institute simulation is the fact that an explicitly defined macro-model is incorporated as a means of obtaining closure in a simulation run. The authors recognize the fact that a limited model based on person and families cannot present a picture of an entire economic system. Thus the macro-model is utilized to supply parameters such as total employment, price level changes, and GNP information to the micro-model.

The use of a macro model such as this one is not unique. In fact, all the models under consideration supply limiting parameters to the simulations in varying degrees. The Urban Institute model, however, makes the supplementary model most explicit and, by doing so, focuses attention on an important principle. No microanalytic simulation can be expected to represent an entire system or subsystem. Certain types of external information must be provided to the model before either projections or experiments can be run.

II. The POPSIM model developed at the Research Triangle Institute is a microanalytic model which, in its basic form, will simulate the basic demographic processes of birth, death, and changes in marital status (Horvitz et al., 1970). The simulated population generated from this model can then be used as input for more specific simulations.

One valuable feature of the basic POPSIM model is its ability to generate an input population for the purposes of simulation. Therefore, the POPSIM user is afforded the capability of generating a population with characteristics suitable to the user's research interests. For example, one could generate a population which is representative, on a number of measures, of the entire United States population at a given time, or one can work with a more specialized population, say, males aged 20 to 45. The major limitation, of course, is the availability of suitable aggregated tables which contain the information desired.

The version of POPSIM that is currently being utilized appears to be an open (person) model where no attempt is made to keep families intact. The record for each individual in the simulation population contains information on the individual, secondary individuals (children), and family characteristics such as family income.

POPSIM has been applied to the problem of hospital utilization in the United States. The basic model was modified to include the variables of race, residence, and family income for the purpose of simulating hospital usage. In this application, inputs (a simulated population)from POPSIM were introduced into a hospital-episodes (HOSPEP) model which simulates admission, duration of stay, bedsize of hospital, insurance status and residence for each individual in the population. The purpose has been to estimate the effects of changes in population composition on utilization of hospitals, and correspondingly, the magnitude and kinds of hospital manpower which would be required under various projected conditions.

III. The simulation model in the process of development at the University of Southern California's Human Resources Research Center (Yett, et al., 1970) has been aimed at creating an analytical model which could cope with the national health-care crisis. The investigators intend that their ideal model, the Mark I, could be used for studying the structure of the health care system, for forecasting the effects of demographic changes on the system, for policy analysis and evaluation, and as a guide for research on the health care system.

The Mark I model would consist of three interacting modules: health services, health manpower, and health education. Technical problems and data deficiencies, however, have led the authors to conclude that the Mark I could not be developed in the near future. Thus, a simplified model called the Mark II has been proposed. This model is based on an open (persons) population rather than a closed model. The notion of an open population can be conceptualized in exactly the same manner as for the previous models. The closed population, however, means something slightly different in this case. The other models define a closed population as one in which marriage mates are selected from within the population and individuals in households are treated as a unit. A closed population in this context would be one which matches specific individuals to specific medical institutions.

Since the model is fundamentally concerned with simulating health care, less emphasis has been placed on the demographic processes than in the case of the other models. Births, deaths, and changes in marital status are treated as exogenous events in the model and are generally based on few variables. Death, for example, is treated as a function of age only. Since an accurate simulation of changes in health services, health manpower, and health education is dependent upon accurately projecting the population, it may be that this model does not devote enough attention to these basic processes.

One additional feature of the Mark II is of particular interest. All of the other models under consideration are based on what is termed a "files" approach by Yett and his associates. Data based on this approach are stored internally by individual records. The Mark II, on the other hand, is based on a "cell" approach. Data based on this approach are stored internally as contingency tables which tabulate the number of individuals with particular combinations of characteristics (e.g. age, sex, race). As long as the cross-tabulations are by all variables included in the simulation, the amount of information retained is exactly the same for both approaches.

In general, the superiority of one approach over the other appears to be based on the computer storage required and the number of decision making operations required in the course of the simulation. Briefly stated, it is suggested that a situation with many variables (age, sex, race, education, etc.) and few cases is best examined by a files approach, while a situation with many cases and few levels is most efficiently handled by the cell approach.

It would appear, however, that another factor must be entered into the choice of best method. If the researcher has developed the simulation model in such a way as to require entry simultaneously of all input information on all individuals in the simulation population, storage space will become a problem regardless of which approach is selected. Twenty variables for 10,000 individuals would require 200,000 words for the files approach and much more space for the cell approach. Either approach will far exceed the capacity of most computers.

However, if the input data are stored on tape or disks and one individual at a time entered into memory and operated upon, the type of approach becomes almost irrelevant. Memory storage becomes more a matter of the length of the program, the operations to be performed, and the output desired. In general, it would appear that the files approach is most appropriate for microanalytic simulations that operate in this manner.

In fact, most simulations that would use a cell approach are generally macroanalytic as opposed to microanalytic models. The exact manner in which the simulation under development by the Human Resources Institute will utilize the cell approach is not completely clear. It may be, however, that the Mark II could be more appropriately termed a macroanalytic simulation. The Mark I, on the other hand, proposes to use a files approach and is clearly microanalytic in nature.

IV. The fourth computer simulation model to be considered is one we have developed within the framework of social accounting (Michielutte and Sprehe, 1970). The model represents an initial step in the development of a set of "social indicators" which could be used to monitor changes in American society.

The computer simulation program that has been developed is based on a microanalytic model which will simulate the basic demographic processes of birth, death, marriage and divorce. In a limited way, the model also incorporates international migration into the United States by adding in a sample of migrants each simulation year. In addition, the program in its present form will simulate changes in the educational and occupational structure (males only) of the United States population.

The simulation model deals with individuals as they pass through a life-cycle. Events occur within households principally on the basis of where individuals are located in relation to their time in life and accompanying circumstances. A set of data records representing characteristics of individuals in households in passed through the computer simulation program. Each pass represents a time period of one year and the simulation is repeated over many passes to effect the lapse of a number of years. At the end of each simulated year, a census of the population is taken, and the basic demographic changes in the population monitored. The simulated population is also stored on magnetic tape and becomes the input for the next year of simulation.

The data base for the simulation is a 1/10,000 census sample of the United States population. The population is defined as being closed, which means that marriage mates are selected and matched from within the population, and households are kept intact as units of analysis.

The simulation model can be used for both projection and experimentation. Basic twenty-year projections have been run by estimating probabilities for birth, death, and marital status, and by projecting these probabilities according to part trends. Hypothetical experiments have also been conducted by altering probabilities (e.g. probability of birth for women with two or more children) and observing the consequences for population growth.

Basic output for the program in its current form includes population counts by age, race, and sex; current marital status by age and previous marital status; vital rates (birth and death) by age, race, and sex; educational distributions by age, race, and sex; and occupational prestige distributions for employed males by age and race. This output is presented for each simulation year. In addition, individual records can be an output if desired.

Problems in Microanalytic Simulation

Although this brief summary of each model does not present a full picture of the different approaches, it does provide the background for a discussion of some of the more important problems in microanalytic simulation.

Monte Carlo vs. Analytic Techniques. All of the simulation models under consideration use or propose to use Monte Carlo techniques in varying degrees. Briefly stated, this means that in a dynamic simulation model, the occurence of events (e.g. death, hospital admission) is controlled by (1) finding a probability value which represents the likelihood of the event's occurring, (2) comparing this probability with a random number that has been generated, and (3) depending on the comparison, causing the event either to occur or not to occur.

Monte Carlo methods, while extremely valuable, have at least two major limitations. First they represent a stochastic process and thus are subject to sampling error. This means that one must either conduct repeated simulations to determine the degree of sampling error, or have a sample large enough to be assured of very small errors. Even in the latter case, repeated simulations are desirable.

Secondly, Monte Carlo methods are extremely time-consuming and thus expensive. In our own simulation, for example, it requires approximately five minutes of central processor time to simulate one year for an initial sample of approximately 17,000 persons. The computer is a very fast CDC 6400.

Essentially, Monte Carlo simulation is a brute force technique and should be recognized as such. Hammersley and Handscomb (1964) suggest as a general principle what wherever analytic techniques can be employed in place of Monte Carlo techniques, the analytic techniques are preferable. They suggest, in effect, that the researcher interested in simulations of the type under consideration be constantly alert for ways of doing thing analytically.

We suspect that all of the models previously discussed, our own included, could profit from the advice. From the information available to us, it appears that the model of income distribution under development at the Urban Institute, the POPSIM model, and our own all make relatively heavy use of Monte Carlo methods, while the Mark II model of hospital care will not rely as heavily on these techniques.

Each of the models makes use of analytic techniques at various stages in the simulation. The decision as to whether an analytic procedure can profitably be substituted for a Monte Carlo technique will depend on the particular variable being simulated and the output desired. However, we feel that all simulation models should present clear justification for the use of Monte Carlo methods, which means an explicit statement of why analytic procedures are not applicable.

In the case of our own model, the inefficiencies of the Monte Carlo methods led us to employ more analytic methods in the simulation of occupation. For this routine, occupation was defined as a function of race, education, marital status, age and the interaction between these variables. Occupational prestige is assigned to every eligible male by means of a regression equation which predicts a mean score for males with various combinations of characteristics. This procedure at present still has many flaws, but its application resulted in a considerable savings in time, and no loss of information for our particular problem.

<u>Open vs. Closed Populations</u>. Another problem related to the long running times associated with Monte Carlo simulations is the question of whether to simulate on a basis of an open or closed population. Three of the simulations discussed, the Urban Institute model of income distribution, the Human Resources Research Center model of Health Care, and the Research Triangle Institute's POPSIM are currently based on open populations. Our simulation, on the other hand, at present operates on a closed population.

Based on our experience with the closed population, we would suggest that the open population approach is superior for most purposes. The closed population has a number of disadvantages which hamper the efficiency of the simulation. In our closed population, it is necessary to match marriage partners from within the population and treat households as units for a number of decision-making processes. In the event of death, for example, a series of decisions must be made with respect to reorganizing the family and choosing a new head of household if the previous head is the one who dies. In the case of divorce, decisions must be made with respect to the children and to the creation of new households. Essentially, the problem becomes one of increasing effort for diminishing returns. A great deal of additional decision-making and programming is required which contributes nothing to the desired output, but which must be done in order to allow more important parts of the simulation to operate.

The increased number of decision steps and simple clean-up operations necessary to maintain intact households adds to both the running time of a simulation and the amount of memory required for the program.

A further limitation is the increased data requirements for the closed model. In the model of health care, for example, the authors point out that the use of a closed model would require knowledge of the attributes of individuals employed in each health occupation, by each type of institution. In our model, use of the closed population requires knowledge of relationships between variables for which data could not be obtained (e.g. are children institutionalized if both parents die?)

<u>Estimation Procedures</u>. The problem of unavailable data is one which affects every simulation model. There is nothing inherent in simulations, however, that make the data problems a particular failing of the technique itself. It simply reflects the fact that simulations tend to be complex and include a relatively large number of interacting variables. The simple addition of even one variable into the simulation problem geometrically increases the data required.

Ideally, a simulation model would be able to specify the exact relationship between every variable included in the simulation. This is nearly impossible in even a moderately complex simulation and the limiting assumptions must be explicitly recognized when interpreting the output.

It is important to emphasize that the quality of the data, with respect to both accuracy and complexity, determines what output is feasible. If, for example, one simulates the occurrence of death on the basis of age and sex, death rates computed by age, race, and sex would be accurate only if there were no relationship between death rates and race. In general, any time a researcher excludes a variable such as race from one part of the simulation and then produces output which includes that variable, he is implicitly assuming that no relationship exists or that it is so small as to have little effect.

The four simulation models vary considerably with respect to the complexity of estimation procedures. The Urban Institute model has defined death as a function of year (trends over time), age, race, sex marital status, education, and parity. The POPSIM model bases mortality on year, age, race, sex, and marital status. Our simulation model defines death as a function of year, age, race, and sex. Finally, the model of health care appears to base death only on age.

The implications of omitting marital status and education from the simulation of mortality in our model extend beyond the results pertaining to death. We could (and do) have the output of death rates by age, race, and sex with a relatively high degree of accuracy; no attempt is made to examine death rates by marital status. However, if marital status and death are related, then the results obtained for the simulation of marital status are likely to be somewhat inaccurate. Since the relationship tends to be in the direction of lower death rates for married people, omission of marital status from the mortality routine will result in too many married people dying in the course of a simulation year. This will introduce some error into the results for percent married, etc.

Accuracy of the Models. In general, surprisingly little mention is made of the manner in which the simulation models under consideration will be tested. The process of model verification requires data collection for at least two points in time. The first set of data will be used either as the input population or as a base to generate the input population. The simulation results should then be compared to data obtained at later points in time in order to obtain some estimate of the accuracy with which the different processes have been simulated.

All of the models discussed propose to simulate events based on real data and which have some relevance for policy decisions in future years. Since this is the case, confidence in the results of projections and hypothetical experiments must be based on the results of earlier verification procedures.

A Strategy for Simulation

An important direction for further development of microanalytic simulation models would seem to be with respect to causal models. Implicit in the models discussed here are causal relationships between many of the variables in a given system. Although in many cases specification of the causal links would be different or uncertain, such a development would aid in the evaluation and improvement of existing models.

Once simulation models are viewed within the framework of a set of causally related variables, the possibilities for development of expanded models should become apparent. The models viewed in this paper tend to focus on the simulation of variables which are basically demographic in nature. There is no reason, however, why simulation models could not also include measures of attitudes and values. This has already been done to some extent in the Urban Institute model. Birth is defined partially as a function of contraceptive usage in the simulation.

The first step in the development of a simulation model would be to hypothesize a set of causal links between the variables to be included in the simulation. If the causal scheme could be expressed in terms of a set of regression equations, one could then start with an initial population and simulate the consequences of the particular causal structure. Processes such a mortality would continue to operate and each individual in the population would be simulated through the set of equations. Ideally, the basic demographic events such as mortality would be incorporated into the causal model. For example, one would be able to specify the effect of a father's death on the son's educational chances.

The difficulties in implementing this approach have not been underestimated. It is likely that the ideal situation outlined here simply will not be possible for some time. However, computer simulation models are frequently accused of being atheoretical and having little relationship to the real world. The approach suggested here would provide both a theoretical orientation to the simulation and aid in interpretation of the results.

References:

1964 - Hammersley, J.M. and D.C. Handscomb, Monte Carlo Methods, New York: John Wiley & Sons.

1970 - Horvitz, D.C., et al., "Simulation of Hospital Utilization." paper presented to the Conference on a Health Manpower Simulation Model.

1970 - Michielutte, Robert and J. Timothy Sprehe, Simulation of Large-Scale Social Mobility: Toward the Development of a System of Social Accounts. Final Report: National Science Foundation Grant #GS-2311.

1966 Naylor, Thomas H., et al., Computer Simulation Techniques. New York: John Wiley & Sons.

1970 - Orcutt, Guy H. et al., "Some Demographic Results of a Dynamic Simulation." mimeographed paper.

1970 - Yett, Donald, et al., The Development of a Microsimulation Model of Health Manpower Demand and Supply. Final Report: U.S. Public Health Service Grant #PH-108-69-69.

Note: Work on this paper and on the Michielutte - Sprehe simulation model was carried out under NSF Grant No. GS-2311.

STATISTICAL MODELS FOR ASSESSING COLLEGE IMPACTS

John A. Creager, American Council on Education

Studying the impact of college attendance on students should supply information for making useful and noncatastrophic decisions. We need to know the structure of the system of higher education, the flow of students through various kinds of colleges, the effects of different matches between students and colleges upon a wide variety of outcomes of the educational process, and the effects of changes in system parameters on such outcomes.

Complex research strategies are required to study the highly complex and heterogeneous system of higher education. Moreover, such research cannot possibly be the bailiwick of any special discipline; multidisciplinary attack is required both substantively and methodologically. It is therefore important that we have means, such as this symposium, for communication and mutual criticism of what we are doing and thinking.

Recently, Feldman (1970) reviewed the present state of the art of studying college impact. He has also suggested how to choose methods and models by noting differences in the investigative purpose and correlative features of the models that might make one more or less appropriate than another. The new directions in studying college impact will probably build upon selection, elaboration, and combination of the models noted by Feldman rather than upon the development of some radically new approach to the whole business.

In deciding which proposed model to use for studying college impact, one must consider the problem of multicollinearity. Our measurements of input, of the demography of students and colleges, environments or treatments, and outcomes simply are not orthogonal to each other. Moreover, multicollinearity is more severe than is indicated by observed correlations, which are attenuated by random error of measurement.

Some, but not all, of the multicollinearity of a system results from our inability to assign students at random to colleges, thus forcing us in practice to use the natural and quasi-experimental designs, especially the longitudinal study of well-defined cohort groups. Some investigators have tended to regard this matter of nonrandom input as a nuisance: i.e., something to be corrected for. This view is not necessarily wrong, but it needs to be supplemented by analyzing the multicollinearity itself. Otherwise, we will continue to be confounded in our judgments of the relative importance and interplay of variables purporting to measure some aspects of the system of higher education. Such attention to the nature of nonrandom input and to other sources of multicollinearity in the system should produce information useful in its own right and possibly alert us to matters requiring caution in our conclusions and recommendations.

Since the literature of psychology, sociology, and education includes extensive discussion of multicollinearity and of nonrandom input problems, we can dismiss from further discussion such models as analysis of variance and analysis of covariance, in either their univariate or multivariate forms, and most partial correlation techniques. As Feldman noted, these techniques are usually inappropriate and misleading for studying college impact. Either they simply fail to cope adequately with nonrandom input, often removing college effects along with the confounding input, or they presume some inappropriate distribution of multicollinear sources of variation among the input, treatment, and output variances. We are left then with three methodologies: (1) regression analysis, with subsequent partitioning of explained variance; (2) path analysis, with its representation and testing of causal theory; and (3) stochastic approaches to understanding the flow of students through the system.

Those of us involved with regression methods in studying college impact are <u>not</u> solely interested in prediction, useful as that may be. Regression analysis can give us much information about the status quo of the system and about those relationships in the system that can be used to formulate empirically based theory. This view assumes that certain realities of the educational system are reflected in regression analyses. Moreover, generalized regression procedures are applicable to ordinal and nominal, as well as interval data and to the examination of nonlinear relations. Thus, we have the powerful practical tool we need to explore the domain of higher education; i.e., to break through the folklore with hard data. The problem is one of interpreting the results in terms of sets of variables of interest. Bottenberg and Ward (1963) proposed omitting a given set of variables and observing the effect on the predictive efficiency of the system. What such a procedure indicates is just how much the variable-set in question adds to what is <u>already</u> present in the system to account for variance in the dependent variable. This is often useful to know, especially if certain variables have already been identified as optimal and irreplaceable for inclusion in analysis. Repeated application of this approach to various subsets of variables in the full regression model leads to the uniqueness-commonality model, which is essentially a multiple-part correlation method. I question this procedure for studying college impact because it can stumble on the multicollinearity problem in subtle and potentially misleading ways.

The closely related two-step regression model is also a multiple part-correlation procedure, one which reverses the Bottenberg-Ward procedure by first computing a reduced regression model and then building up a fuller model. Most investigators would agree with Feldman's summary of the criticisms of this procedure, especially with regard to its assignment of input-treatment relations to input. However, in one study, Astin (1968) reversed the order of the stepwise entry of input and treatment sets of variables to see whether his substantive conclusions were vitiated by the alleged error: that is, he assigned the input-treatment dependencies to the treatments. This does, indeed, give some protection against error in concluding that input accounted for more of the outcome variance in achievement than did the measured college environmental characteristics. In some cases, however, when input and treatment are more closely balanced or the input-treatment dependencies are sufficiently large, the reversal procedure would result in ambiguity rather than confirmation of the findings that emerge in doing the initial regression against input.

Some have defended the two-step procedure on grounds of temporal asymmetry; their argument is plausible and pragmatic, provided that one is reasonably sure that the input-treatment dependencies are caused by input variables or their antecedents and not by some feedback of information about college characteristics. This matter can be studied operationally by including among the input variables measures of why the student chose that college, and by including among the college characteristics variables measures of college admissions policies. The way in which these variables are distributed among the factors defined by the total system should throw more light on the question.

One also needs to be assured that the input-treatment dependencies are not a function of socioeconomic or of cultural factors which influence both input and treatment variables. A reanalysis by the orthogonal decomposition model (Creager and Boruch, 1969) of some of Werts' (1968) data about the effects of home and school variables on achievement yielded just such a result. It makes no sense to assign such sources of variance or hypothesized causes to either input or treatment sets of variables. This kind of experience has led me to prefer methods which isolate multicollinear dependencies and force examination of them, especially those associated with the confounding of input and treatment through nonrandom admission of students to colleges. Meanwhile, the two-step method continues to be used extensively (Astin, in press, b) with considerable insight and judgment.

Recently, Astin (in press, a) has been concerned with another problem that is critical regardless of the models used: the effect of errors of measurement, not only on our analyses, but also on our interpretation of results. Beyond the usual lip service, inadequate attention has been paid to this matter. In fact, the college effects literature and much of the sociological literature lack empirical estimates of reliability even for commonly used variables. We have, therefore, been obtaining some empirical estimates for many of the variables used in the Council's Cooperative Institutional Research Program.

The various procedures based on multiple part-correlation methods attempt to account for college impact through partitioning explained variance in the dependent variable; this accounting is done by comparing proportions of outcome variance explained by full and reduced regression models. The alternative approach involves the examination of variance in the full regression composite. This examination is sometimes done by looking at the algebraic terms of the standard formula for linear composite variance; it is annoying to see that this procedure continues to be used in spite of widespread criticism about its generally fallacious nature. Because it too stumbles over the multicollinearity problem, it is useful only for making rough qualitative judgments of the rank of importance in systems with low correlations and large differences in the weights.

There remains the procedure of complete orthogonal decomposition of the regression system in terms of <u>interpretable</u> common and unique factors (Creager and Boruch, 1969). Here again, the aim is to account for predictable variance in the dependent variables or outcomes, because such variance is directly and pragmatically a measure of differences in outcomes that one may ultimately be interested in modifying. This procedure is applicable to canonical and discriminant composites and to bipolar systems. Moreover, the results are invariant under reflection of either variables or factors. A simple computer program is now available (Creager, 1971) for inputting the weights on up to six linear composites, and inputting the results of the factor analysis of the system under study, to obtain the orthogonal decomposition of composite variance.

Two comments should be made about this method. The first has to do with attempting to analyze very small systems where clear factor definition may not be attainable. Such a situation may also occur in somewhat larger systems in which only those variables selected by stepwise regression are included in the analysis and thus adequate factor definition and interpretability may become a problem. Fortunately it is not necessary to confine the definition of factors to just those variables in some ad hoc system of immediate interest. One may include variables which were permitted to enter freely into regression but did not in fact do so, with no adverse effect on the analysis of one or more composites derived from the total system. Moreover, this kind of operation has advantages in a large-scale research program where a single factor analysis can serve as the basis for analyzing many derived composites. The second comment is that we do not yet know the effects of measurement errors on the outcome of such analysis. The problem can be virtually eliminated, however, by correcting correlations for attenuation prior to doing either the regression or the factor analyses.

With the orthogonal decomposition model, one can say: If I do something to change values on this variable, it will have certain effects on the values of the dependent variable and on any other variable in the system. This procedure does not preclude the use of previously specified hypotheses which may or may not be confirmed by the analysis. Thus we are interested in more than prediction and description of the status quo. Since the method is new, most of the effort so far has been one of understanding the limitations of the method, of perfecting it, and of providing computer software to handle it (Creager, 1971). Recently, the method has been put to its most severe test in a reanalysis of the data for Astin's study of undergraduate achievement and institutional excellence (1968). Three full model regression composites developed by free entry were analyzed by orthogonal decomposition and the results compared with those of Astin, who used two-step regression and the uniqueness-commonality account of variance. The conclusions were very similar, in that both strategies indicated that achievement is more strongly determined by input than by environment, and the agreement was partly a function of the fact that the multicollinearity within input and treatment sets was stronger than that between sets. However, the reanalysis also indicated that free-entry regression defines a more efficient and parsimonious predictor system and is less biased toward input. Orthogonal decomposition also provides more detailed information about variances, which may be pooled in various ways for heuristic interpretation. Further demonstration of the scientific value of this strategy remains on the agenda for the study of college impact.

About path analysis, I claim no expertise. Since we have others on the panel speaking on this topic, I shall limit myself to some comments and some questions regarding this procedure for studying college impact.

I do not think path analysis and the orthogonal analysis of prediction systems are interchangeable, although at several points, they are similar and possibly compatible. It is more likely that the two methodologies are mutually complementary in the sense that both can contribute substantive information for a synthesis of the larger picture of higher education. The two methods differ in how they define direct and indirect effects, how they express these effects in empirically derived numbers, and how they add up to some kind of total effect. In the orthogonal decomposition method, orthogonal portions of variance represent sources of variance that can be explicitly subtotaled across inputs, treatments, and other sets of variables in the system, and totaled to the predictable variance of the dependent variable. In path analysis, path coefficients -- which are usually regression weights -- and their products add up, under certain conditions, to the zero-order correlations with the dependent variable. Thus, it partitions and accounts for the slopes of the individual regressions in terms of the slopes of partial regressions. Much of my objection to interpreting regression weights as anything but a particular set of weights that maximizes prediction has resulted from attempts to interpret them, their squares, and their products as independent portions of variance, which they are not. Regression weights treated as slopes of particular partial regressions, and correlations treated as slopes of total regressions, indicate change in the dependent variable produced by change in the independent variables. This alternative to describing these changes in terms of variance is a quite legitimate and useful one. However, regression weights -- and therefore path coefficients -- are not orthogonal to each other so that the meaning of what adds up by the internal algebraic consistency of path models is not entirely clear. But the path analyst cannot be accused of completely ignoring multicollinearity. External to the algebra is the causal model embodied in the path diagrams which imply the observed relationships. In some path models, however, correlations among exogenous variables are left unexplained, although they are used in computing the coefficients for other paths. Doesn't this imply only a partial attention to multicollinearity, and does it not invite surprising side effects when results are applied in practice?

Path analysis is a class of methodologies consisting of models implied by external hypotheses about causal relations. Any realistic study of college effects would seem to require recursive systems from multistage, multivariate models. When one has many input, cultural, and treatment variables, the resulting path analysis would be extremely complex, to say the least. In orthogonal decomposition, this complexity is advantageous in the sense that one can obtain a finer factor resolution of the system. If the path analyst solves the problem of complexity by doing many analyses of small pieces of the system, can the results of the many analyses be synthesized in a logically and statistically consistent way? Perhaps so, although I have only seen very simple systems studied by path analysis in college effects studies. Another new direction in college impact studies with path analysis might well demonstrate the consistency of causal relations confirmed in two or more fractionated studies with those confirmed by a single and more complex analysis.

Concerning stochastic models, there is little to add to Feldman's remark that they are promising but have to be employed in actual analysis of data before they can be properly evaluated. I simply note here that classical Markov models are not generally realistic for college impact studies. The Cornell models developed by McGinnis and his associates (1968) try to achieve more realism by changing axioms: e.g., to permit more than one-step dependency. This approach is clearly a relevant line of future development. A different line of development assumes that rectangular, rather than square transition matrices will be required for realistic applications to college impact studies. The states meaningfully defined at college entry and at later stages are not necessarily the same.

Both the stochastic and path methods have the problem of keeping the number of transitions from getting out of hand, and if this problem is handled by fractionation, then synthesis of results is required. Prior regression analyses and application of the hierarchical grouping algorithm may help to solve this problem.

So far, stochastic approaches to college impact studies are little more than extensive cross-tabulations with chi-square tests to detect nonrandom input, which we already know exists, and to detect the existence of college effects. There seems to be a kind of built-in separation of input and treatment, and obviously the description of flow of students is useful for purposes of manpower and resource allocation. Beyond this, it is not yet clear whether this approach can be developed to produce any real understanding of the system of higher education.

The extension of college impact methodology to the study of multiple outcomes and their interrelations is available by canonical regression with orthogonal decomposition of the canonical variates and as multiple outcome states in the stochastic methods. I see no reason why path analysis could not be similarly extended to study multiple dependent variables. However, the additional complexity involved may present problems.

At the present state of the art, we need to press forward with all three methodologies, not only to detect, but also to estimate, the extent of college impacts. We must also press forward to polish our methods and ensure their integrity, which involves not only their internal logical and statistical consistency, but the soundness of any interpretations of substantive results. We hope, and expect, results to be taken seriously in practical decision making. We had, therefore, better be right! On the other hand, if we wait too long to be sure, the world will have passed us by and our results will no longer be relevant.

REFERENCES

[1] Astin, A. W., "Undergraduate achievement and institutional 'excellence,'" Science, 161 (1968), 661-668.

[2] Astin, A. W., "The methodology of research on college impact," Part I, Sociology of Education, 43 (1970), 223-254; Part II, in press. (a)

[3] Astin, A. W. Predicting Academic Achievement in College. New York: The Free Press, in press. (b)

[4] Bottenberg, R. A., and Ward, J. H., Jr. Applied Multiple Linear Regression. Technical Documentary Report PRL-TDR-63-6. Lackland Air Force Base, Texas: 6570th Personnel Research Laboratory, Aerospace Medical Division, (1963).

[5] Creager, J. A., "A fortran program for the analysis of linear composite variables," in press.

[6] Creager, J. A., and Boruch, R. F., "Orthogonal analysis of linear composite variance," Proceedings, 77th Annual Convention, American Psychological Association, (1969), 113-114.

[7] Cronbach, L. J., and Furby, L., "How we should measure 'change'--or should we," Psychological Bulletin, 74 (1970), 68-80.

[8] Feldman, K. A., "Research strategies in studying college impact," ACT Research Report, 34 (1970).

[9] McGinnis, R. A., "A stochastic model of social mobility," American Sociological Review, 33 (1968), 712-722.

[10] Werts, C. E., "The partitioning of variance in school effects studies," American Educational Research Journal, 5 (1968), 311-318.

THE 'CHARTER' AND SOCIALIZATION IN COLLEGE: SOME EFFECTS OF PRESTIGE AND COLLEGE SIZE ON OCCUPATIONAL CHOICE AND DROPOUT[1]

David H. Kamens, Northeastern University

Introduction

In this paper we offer a new conceptualization of college characteristics, such as size and complexity, and examine some empirical findings on college dropout and changes in occupational choice during college which are readily explained by this theory, but which are difficult to account for on the basis of traditional interpretations of effects of college.

Organizational attributes, such as quality and size, are usually seen as indicators of two kinds of resources available to colleges: 1) student inputs, such as measures of the social class composition, academic aptitude and intellectualism of students recruited to different colleges. College quality, for example, is usually explicitly conceptualized as an aggregate measure of the academic talent colleges have to work with. (Spaeth, 1968; Davis, 1966; Astin, 1964). 2) Secondly, organizational attributes such as size and complexity have been viewed as measures of the socializing resources that are differentially distributed among colleges. Size and faculty-student ratio are frequently viewed as measures of internal resources per student, which in turn have a number of subsequent consequences that affect the shape of the internal structure of relations; e.g., frequency of interaction, stratum isolation, social distance between the socializing agents and the socializees, etc. (Newcomb, 1943; Feldman and Newcomb, 1969).

Both sets of interpretations have a common denominator in that they focus on measures of inputs; student inputs and available socializing resources. Thus, as Meyer has pointed out (1970), discussions of college quality are typically conceptualized solely in terms of differential social resources with no reference to outcomes.

What is missing from such conceptualizations and associated measurements is the relation between colleges and the wider social order. Obsession with inputs has obscured the idea that colleges have constituencies whose demands they must be responsive to if they are to survive. These outside clients set constraints not only on the resources available but also on the kinds of products that are socially acceptable; i.e., what kinds of people colleges should be producing. It is very clear, for example, that the production of student radicals has become unacceptable by many important clients of American universities, most notably state legislators. Colleges that continue to produce them will be penalized. In short, the clients of higher education impose cultural constraints on socialization in colleges as well as resource limitations. It is our contention that such constraints have a powerful effect on the socialization process in college.

This idea leads us to urge a new interpretation of college characteristics, and to argue for new types of measures of colleges. We suggest that we need measures that focus on the connection between colleges and the larger social order.

Once we accept the idea of attempting to characterize colleges in terms of the expectations of their external constituencies (as well as in terms of inputs, such as internal resources and socializing structures), it may be possible to re-interpret some widely used measures of college structures, such as size, complexity and prestige, and to offer some predictions about their socializing capacity that are different from those ordinarily made from theories of 'inputs.'

Our major point is that college characteristics such as prestige and size may have direct effects on students by indicating the kinds of social changes that the organization is 'licensed' to produce by the wider society. Meyer has called this mandate the organizational charter, and we will adopt this term. (Meyer, 1970) Meyer argues that, "Any socializing organization has crucial features which lie largely outside its own structure and which constitute its relationship with its social setting. One such feature--perhaps the most important--is the social definitions of the products of the organization. If, for example, everyone knows that a particular school or class of schools (i.e. colleges) produces successful people, and if they know that others--employers, professional gatekeepers--know and accept this, then the school has acquired an invaluable resource in transforming its products." (Meyer, 1970, 9)

We hypothesize that schools differ in the kinds of charters they have, which in turn affects their socializing ability. This argument emphasizes that colleges and other socializing agencies vary not only in the social characteristics of their clientele and other inputs, but also in the kinds of future statuses to which they are able to allocate their recruits. The point that needs to be stressed is that the strength and credibility of the guarantee colleges can offer regarding status placement may vary among schools.

There are two well known ways of characterizing status locations that are relevant to his general argument. These dimensions are: first, the horizontal differentiation of statuses or division of labor; and secondly, the vertical stratification of statuses, where the relative rank of statuses is the central feature.

The point is that colleges may differ in their status allocating capacity on both dimensions. Thus, from this view the strength of large, public institutions rests on the diversity of middle class occupational roles that students perceive graduates of such colleges assume. On the other hand, the socializing capacity of elite universities is based on the diversity of high status professional roles that the typical graduate is perceived to move into. Thus the socializing capacity of institutions is linked in this argument, to the variety of career outcomes that its external constituencies will validate.

Now we must deal with the question of how these expectancies are transmitted to students. We will suggest how two college attributes, prestige and size, directly signify the character of the organization's social license to students. Secondly, we will consider how this affects two aspects of the socialization process: 1) the ability of organizations to maintain members' commitment, as indicated by dropout; and 2) the effects of colleges on changes in occupational choice. Occupational choice is used as a direct measure of status allocation. Dropout is used as a measure of the value students attach to membership in the college. Value here refers to the perceived economic and social benefits, i.e., motivation, economic resources, etc. To remove the influence of other variables and thus 'purify' dropout as a measure of the value of college, we will control for a variety of variables known to be related to dropout and thus attempt to remove from dropout the effects of other causes of attrition.

The Effects of College Size

The case for size as a 'chartering' mechanism is intriguing and also more difficult to make, primarily because of the weight of negative evidence and theory that emphasizes the dysfunctions of size on student socialization. (Cf. Feldman, and Newcomb, 1969; Panos and Astin, 1968; for counter evidence, cf. Kamens, 1970). We suggest that large schools have a special charter; but this function of size may be peculiar to American higher education. The chartering aspect of size seems to depend on two characteristics of the larger American stratification system; 1) the large size of the elites, to which college graduates are recruited; (Lipset, 1960, 330ff.), and 2) secondly, the fact that the American stratification system is organized around a system of functionally differentiated occupational groups rather than highly organized social classes. (Cf. R. Brown, 1965, 113ff for some evidence).

Given these two conditions socialization in American higher education must proceed on different cultural assumptions than that in more aristocratic cultures. In the United States higher educational institutions are expected to train students for high status, but functionally specific occupational roles rather than to act as upholders of traditional values and aristocratic culture. (Turner, 1961) Thus they are chartered to accomplish very limited purposes. In this kind of cultural and institutional setting size may have positive consequences for student socialization that it would not have in countries characterized by a small elite and well defined social classes.

Given the mission of occupational socialization as their primary responsibility vis à vis students in the United States, the socializing capacity of colleges is apt to depend on the diversity of career roles that they incorporate within their ecological and social organization. Consider the plight of small colleges in this social and cultural context. They integrate in their structural and curriculum few of the occupational roles, activities and identities that exist in the wider society. College teaching, especially in the humanities, is one of the few activities that students are exposed to and which college can train them for. Furthermore, small colleges typically represent only limited functions of even academic roles. They expose students to academics as teachers but typically not to the research functions of the role. On the basis of interviews with transferees to Berkeley Lipset argues that this emphasis on teaching functions causes many intellectually oriented students to leave small colleges like Reed and go to much larger universities like Berkeley. He asserts that having been taught the value and excitement of research, bright students are often willing to pay the price of anonymity and little contact to hear major scholars in person, and to be in a research setting where the action is. (Lipset, letter)

Big colleges and universities, on the other hand, integrate a large number and diversity of values and activities in their social structures. This results from the fact that in American higher education size and structural complexity have historically been closely associated. (Scott and El-Assal, 1969; Meyer, 1970). Rather than forming separate organizations for new educational and intellectual innovations, as in Europe, these have been fused with existing university structures to form bigger and more complex structures, i.e., the 'multiversity.' This development, we suggest, has two consequences for student socialization: 1) Students at large universities are exposed to more diverse models of what graduates of the college actually become; 2) secondly, the ecological proximity and structural connection of the college with graduate and professional schools may act as a guarantee that the college is actually expected to produce diverse kinds of professionals and specialists. If not, why would these training facilities exist on the same campus?

The easily visible flow of average undergraduates from college to graduate and professional schools may also function as an indicator to students of the value outside agencies typically attach to the undergraduate experience.

We stress the symbolic function of institutional arrangements characteristic of large universities because of our concern with chartering mechanisms. We recognize that there may also be structural arrangements at larger colleges which make it easier for students to assume future occupational identities. For example, undergraduates are likely to have more contact with graduate students, and faculty who teach in graduate schools. This exposure may give them more concrete knowledge of standards, the kinds of people who become professionals, and may lead to a clearer picture of the meaning and value of these careers. In short, large colleges may have a number of structural features which serve to articulate undergraduate and graduate status sequences. This argument suggests that large universities may produce more continuity in status socialization than small colleges. Thus, large schools may reduce anxiety over future role assumption, though they may breed other kinds of anxiety.

This view of size leads to a number of hypotheses, that run contrary to usual expectations. First, the idea that large schools produce more continuity in status sequences because of their charter leads us to hypothesize that large colleges will have important positive effects on students' occupational ambitions, despite the fact that they have fewer socializing resources per student. For instance, one measure of student faculty-interaction, the faculty-student ratio, correlates -.34 (Tau-B) with size in our data.

We can test this idea with our panel data on occupational choice by looking at the effects of size on the general status of students' aspirations and the kinds of occupational choices students make, when appropriate individual characteristics are held constant. Meyer has shown elsewhere with this data that size, along with many other college characteristics, has no effect on the general social status of students' occupational choices. Large schools do not increase students' general occupational ambitions (Meyer, 1970; 36). However, school size does affect the allocation of students to different sectors of the occupational structure. Tables 1 and 2, from Meyer's report (1970), show the effects of size on 1) students' ambition to enter high status academic occupations; and 2) students' aspirations to enter the high status professions. The tables together show that large colleges shift students toward the traditional professions at the expense of high status academic occupations. Conversely, small colleges shift students toward academic careers and away from the professions. This evidence partially supports the early findings of Knapp and Greenbaum on the effectiveness of small, often mediocre, colleges in producing scientists and scholars. (Knapp and Greenbaum, 1953) Unfortunately, because there were too few cases for analysis, we cannot tell from this data whether small colleges are more effective at maintaining the commitment of students who initially aspire to become scientists and scholars than large schools.

TABLE 1.--Students with Non-Academic Freshman Choices Only: Senior Academic Occupational Choice by School Size, Sex and Ability Index

	School Size	
Ability	Large (1000+)	Small (Under 1000)
High	22% (59)	39% (33)
Medium	14% (59)	17% (35)
Low	7% (73)	19% (48)
High	14% (50)	18% (66)
Medium	3% (65)	11% (72)
Low	10% (51)	5% (95)

Freshman Choice = Non-Academic Professional, low status, undecided and other.

TABLE 2.--Senior Professional Occupational Choices by Freshman Choice, Ability, Index Score and School Size -- Males

		School Size	
Ability	Freshman Occup'l Choice	Large (1000+)	Small (Under 1000)
High	Prof'l	69% (32)	50% (18)
	Other	19% (36)	15% (34)
Medium	Prof'l	68% (34)	47% (19)
	Other	20% (41)	10% (20)
Low	Prof'l	70% (30)	47% (19)
	Other	6% (50)	8% (40)

All NAs excluded. All entries are per cent Choosing Professional Occupations as Seniors.
Freshman Choice:
Professional = high status professional occupations
Other = Academic choices, low status and undecided

We now consider some consequences that this variation in status allocating capacity between large and small schools may have on dropout.

This leads us to the second hypothesis about size. Larger colleges are likely to increase the average value attached to membership in the college and lower dropout, even when individual characteristics, known to affect dropout, are held constant. The problem is basically demographic. Small colleges are better than large colleges at attaching students to academic careers and less able to connect them with careers in the non-academic professional occupational sector. Yet, very large proportions of the student body aspire to non-academic professional occupations at both large and small colleges. The basic idea here is that the special charter of small colleges may reduce the original commitment of many students who do not aspire to academic careers. Conversely, large colleges may lead more students to believe that the college will connect them with professional occupations.

The charter of larger colleges may, however, have negative effects on students with academic aspirations. Large colleges may reduce the commitment of such students and increase their dropout.

Table 3 presents the data on the effect of size on dropout when students' social class and the ability index are held constant.

TABLE 3.--Per Cent Dropout by College Size, Standardized on Students' Social Status, Verbal Ability and High School Grades[2]

Size	Per Cent Dropout
0-999	39% (492)
1000-4999	29% (322)
5000+	24% (111)

Size reduces dropout and the effect is linear over the three categories.

Table 4 presents the data when sex is introduced as an added control.

TABLE 4.--Per Cent Dropout by Size and Sex, Standardized on Ability Index and Social Class Index

	Sex	
Size	Male	Female
0-999	37% (225)	39% (267)
1000-4999	24% (197)	39% (125)
5000+	27% (84)	10% (26)*

*Standardized per cent unreliable because all cell sizes too small.

The evidence confirms the main argument on the effect of size on dropout. Size, however, has a differential impact on men and women. Male dropout is higher in small colleges than in either of the other two size categories. In contrast, women have lower dropout only in the largest schools. Because of the small number of women sampled from the largest colleges, we do not place much confidence in this finding.

A separate analysis, not presented here, examined the possible negative effects of size on students with academic career aspirations in their freshman year. The data show that small colleges reduce dropout among students with academic and scientific career aspirations but increase dropout among all other students. Conversely, large schools lower dropout among students with professional, semi-professional and business aspirations but increase it among those with scientific and college teaching career choices. This evidence provides tentative support for the idea that the special charters of large and small colleges may reduce organizational commitment among students whose aspirations do not coincide with the college's charter.

Our last hypothesis on size concerns the effects of academic achievement on dropout. We expect that the status allocating capacity of large colleges will increase commitment to continue, independently of students' academic success. Whatever their grades, students at large colleges should be more dependent on membership than those in smaller schools. Furthermore, low academic achievement may be less deprivational in large colleges, given the impersonality and anonymity of these contexts. Students see teachers less outside class and are probably less visible, and therefore less easily 'typed' by faculty and other students on the basis of their academic performance for large schools. (Kamens, 1970) Grades are thus less likely to become the basis for diffuse status in the community.

Table 5 presents the data, showing the effect of size on male and female dropout when grades have been controlled

TABLE 5.--Respondents and Non-Respondents: Per Cent Dropout by Size and Sex, Standardized on College Grades--Institutional Record

	Sex	
Size	Male	Female
0-999	39% (438)	47% (577)
1000-4999	30% (454)	46% (302)
5000+	36% (144)	45% (63)

through standardization. Both respondents and non-respondents to the original questionnaire are included in this cross tabulation. Size lowers dropout among males, independently of grades; though the relation is curvilinear. Size, however, has no effect on women's dropout; when grades are controlled. The non-standardized table shows the same effect.

with one exception. In the largest colleges, women drop out less at all but the highest grade level. Again only the largest colleges reduce female dropout. Further research with larger samples of big colleges and students is needed to determine the validity of this finding.

Conclusion

This paper has argued that colleges achieve much of their effects on students because of their connection with the wider social order. This idea led us to hypothesize that colleges differ in their status allocating capacity because of differences in the structural networks colleges have with occupational and economic groups in the larger society. Two characteristics of colleges were chosen as indicators of different kinds of structural connections with the larger social order: college prestige and size-complexity. Empirical analysis showed that each has effects on occupational allocation and dropout, when individual characteristics of students are held constant. Two qualifications of these results need to be mentioned. First, the results in the case of women indicate that a separate theory is needed to account for female dropout. Secondly, while the effects of prestige-selectivity on dropout are relatively clear, the effects of size must remain tentative. This is due to the fact that we under-sampled the largest colleges so that conclusions about the effects of the very largest schools are unreliable. Further, given the complexity of the effects of size, we feel that more powerful analytical tools are necessary that can consider: a) the effects of a large number of variables simultaneously; and b) possible interaction effects between size and individual characteristics.

In conclusion, we urge that future research focus on two interrelated problems: 1) developing models of the ways that colleges are related to the wider social order and how this may affect different aspects of student socialization and 2) developing measures of colleges that reflect student outcomes as well as inputs. In regard to the measurement of colleges we have as yet almost no measures that characterize colleges both in terms of the types of students they recruit, e.g. average ability, and the kinds of finished products they produce, e.g. occupational choices. One interesting measure of college academic quality, for example, might be the ratio of student academic career choices to the proportion of high ability students.

Bibliography

Astin, Alexander W.
1964 "Personal and Environmental Factors Associated with College Dropouts among High Aptitude Students," Journal of Educational Psychology. 55: 219-227.
1969 "Institutional Prestige and Dropout," College and University," Summer, 341-356.

Brown, Roger
1965 Social Psychology. Free Press.

Davis, James
1965 Undergraduate Career Decisions. Aldine Publishing Company, Chicago.

Feldman, Kenneth A. and Newcomb, Theodore
1969 The Impact of College on Students. San francisco: Jossey-Bass.

Kamens, David H.
1968 "Institutional Stratification and Institutional Commitment: Contextual Effects on College Dropout," Unpublished Ph.D. dissertation. Columbia University.
1970 "The 'Charter' and Socialization in College: Some Effects of Prestige and College Size on Occupational Choice and Dropout," Presented at the American Statistical Assoc. meetings, Detroit, Michigan.

Knapp, R. H. and Greenbaum, J. H.
1953 The Younger American Scholar: His Collegiate Origins. University of Chicago Press.

Lipset, Seymour M.
1960 Political Man. Doubleday.
1969 Letter to the Hon. Henry Friendly, privately circulated.

Meyer, John W.
1970 "The Charter: Conditions of Diffuse Socialization in Schools" in W. Richard Scott. ed., Social Processes and Social Structures: An Introduction to Sociology. New York: Holt, 564-578.

The Effects of College Quality and Size on Student Occupational Choice. Final Report to U.S. Office of Education.

Newcomb, Theodore
1943 Personality and Social Change. Holt, Rinehart and Winston.

Panos, Robert J., and Astin, Alexander W.
1968 "Attrition Among College Students," American Educational Research Journal, 5, 57-72.

Scott, J. W. and El-Assal, M.
1969 "Multiversity, University Size, University Quality and Student Protest: An Empirical Study," American Sociological Review, 34: 702-709.

Spaeth, Joe L.
1968 "The Allocation of College Graduates to Graduate and Professional Schools," Sociology of Education, 41: 342-350.

Turner, Ralph
1961 "Modes of Social Ascent through Education: Sponsored and Contest Mobility" in Halsey, Floud, and Anderson, eds. Education, Economy and Society. Free Press: Glencoe, Ill.

Footnotes

1. The section on prestige has been deleted due to limitations of space.

2. Our ability index is composed of two items: verbal ability scores on a number of college entrance tests and high school grades. Test scores were standardized on the basis of published distributions and were converted to equivalent scores on the CEEB verbal ability test. The resulting distribution was then trichotomized: low = 200-499; medium = 500-599; and high = 600-800. High school grades were also trichotomized: high = A, A-; medium = B+, B; and low = B- and lower. These two items were then combined into a three point index of ability.

   The social class measure is an index that combines the status of fathers' occupation and mothers' and fathers' education level into a three point scale.

THE MULTI-DIRECTIONALITY OF STUDENT CHANGE[1]

Arthur W. Chickering, American Council on Education

From 1965 through 1970, thirteen small colleges with enrollments of 1500 or less cooperated in a study of institutional characteristics, student characteristics, and student development. Diverse instruments and multi-level studies documented dramatic differences among the colleges--in goals; climate; rules and regulations; curriculum, teaching, and evaluation; sense of community and student-faculty relationships, and in the concrete experiences and behaviors of students and faculty. Students also differed widely from campus to campus: in their orientations to college, intellectual interests, attitudes and values, religious beliefs, autonomy and personal integration, complexity of ambiguity, and desire for material success.

In addition, the characteristics of the students and of the colleges fit comfortably together. Apparently self-selection and admissions practices operated with both precision and power so that certain kinds of students entered certain colleges and not others, and little overlap occurred among colleges that were substantially different from one another.

What developmental outcomes follow from these associations between particular kinds of institutions and particular kinds of students?

Does Change Occur? In What Areas? At What Colleges? [2]

The students changed on all but one of the fourteen scales of the Omnibus Personality Inventory. More significantly, despite the major differences among institutions and among the entering students, the direction of change was basically the same in all colleges.

Table 1

Mean Scores for Students Pooled from Twelve Colleges
Omnibus Personality Inventory
1965 - 1969

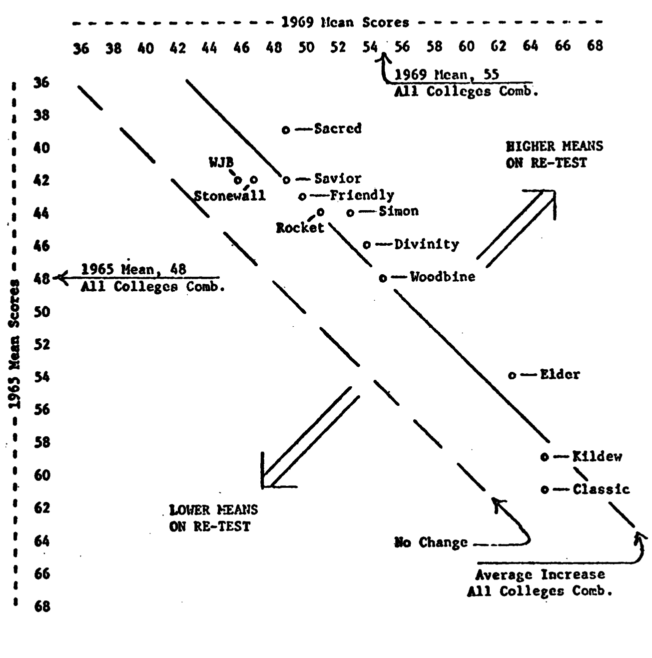
Scale	All Students N = 585		Men N = 254		Women N = 269	
	Fall 1965	Spring 1969	Fall 1965	Spring 1969	Fall 1965	Spring 1969
Autonomy	48	55	48	55	47	55
Practical Outlook	52	47	52	47	51	47
Impulse Expression	48	52	49	54	45	50
Complexity	48	50	48	51	47	49
Estheticism	49	52	46	50	52	55
Thinking Introversion	48	50	47	50	49	50
Masculinity-Femininity	49	48	55	53	43	42
Personal Integration	50	54	51	54	50	54
Anxiety Level	49	50	50	52	48	49
Religious Orientation	46	50	47	51	45	49
Theoretical Orientation	46	47	48	49	43	44
Altruism	51	52	49	50	54	54
Social Extroversion	47	47	46	46	48	48
Response Bias	47	48	48	49	46	47

NOTE: Underlining indicates that the differences in scores were significant at the .05 level or beyond and therefore likely to occur by chance less than one time in twenty.

Pooling the individual colleges yields 168 pairs of mean scores--12 colleges times 14 scales. The differences between sixty-eight of these pairs was statistically significant beyond the .01 level, and seventeen more fell between the .05 and .01 levels. Of these eighty-five cases, only two were contrary to the typical direction, and different scales were involved in each case. Ignoring statistical significance, and setting aside the Social Extroversion scale on which no general change occurred, among 154 comparisons, only 16 were contrary to the usual direction. Seven of these atypical changes occurred at Kildew, where the mean scores of entering students were often near the extremes, leaving little room for change in the directions typical of most other students.(see Table 2, Mean Scores for Individual Colleges).

All the colleges changed in the same direction on Autonomy, Practical Outlook, and Impulse Expression, and those changes were consistent when men and women were analyzed separately. Eleven of the twelve colleges changed in the same direction on Personal Integration and Estheticism, and ten reflected similar change on Complexity, Thinking Introversion, and Religious Orientation (Liberalism).

Figures 1 and 2 present the findings for Autonomy and Estheticism and show the relative positions of the twelve colleges, mean scores and average change when all colleges are combined. The results for Estheticism vividly illustrate the consistency among the institutions. Even though institutional means spanned a wide range, and even though the extent of change was small, change at each college--except at Kildew where no change occurred--was close to the average.



Note.--Numbers indicate standard score intervals. Mean scores for norms group of 7,283 freshmen = 50.

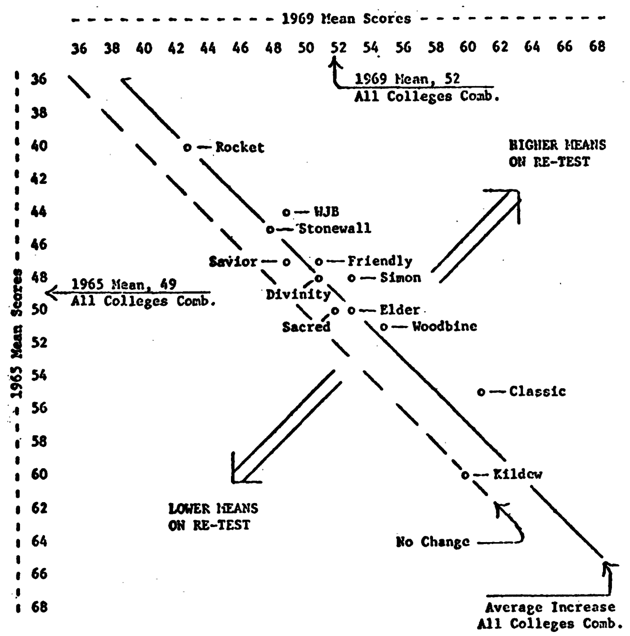
Figure 1. AUTONOMY (OPI)
1965-1969 Mean Scores for Individual Colleges

Table 2

Mean Scores for Individual Colleges
(Same Students Tested in 1965 and 1969)
Omnibus Personality Inventory

College	N	Autonomy	Practical Outlook	Impulse Expression	Complexity	Estheticism	Thinking Introversion	Masculinity Femininity	Personal Integration	Anxiety Level	Religious Orientation	Theoretical Orientation	Altruism	Social Extroversion	Response Bias
		65 69	65 69	65 69	65 69	65 69	65 69	65 69	65 69	65 69	65 69	65 69	65 69	65 69	65 69
WJB	33	42 46	55 51	42 45	43 46	44 49	44 49	50 50	54 56	52 53	37 37	43 45	51 53	47 49	47 50
Savior	38	42 49	57 52	42 49	44 45	47 49	45 44	51 49	49 52	49 49	39 41	45 41	50 49	43 46	48 47
Sacred	29	39 49	57 50	46 53	42 48	50 52	45 51	47 46	49 51	48 47	41 47	43 47	51 53	51 50	47 49
Stonewall	49	42 47	56 54	53 55	47 48	45 48	43 47	53 52	45 53	46 50	48 49	45 47	43 48	46 49	43 48
Simon	76	44 53	54 48	42 47	43 49	48 53	48 51	48 47	54 57	51 52	38 44	42 45	55 54	49 48	48 50
Divinity	51	46 54	52 46	43 48	46 49	48 51	47 48	49 49	53 58	52 54	39 43	45 44	53 53	49 48	49 51
Friendly	61	43 50	54 51	44 46	45 45	47 51	45 46	47 46	50 55	49 51	42 45	42 43	52 52	48 48	46 47
Kildew	36	59 65	41 39	57 61	62 62	60 60	60 55	43 44	48 53	48 50	62 60	52 52	54 53	47 47	46 47
Classic	13	61 65	42 39	54 61	58 62	55 61	58 61	54 48	51 52	50 49	55 58	54 56	51 52	44 47	50 51
Elder	123	54 63	47 42	49 55	50 54	50 53	51 52	49 47	50 52	50 50	52 58	49 48	53 54	46 45	47 46
Woodbine	56	48 55	53 45	55 57	50 51	51 55	45 49	47 44	46 51	44 48	53 58	45 49	48 50	47 48	43 46
Rocket	20	44 51	54 52	50 55	47 50	40 43	41 45	60 57	51 50	49 48	49 53	51 51	43 46	44 47	51 51

Note: Underlining indicates that the differences in scores were significant at the .05 level or beyond and therefore likely to occur by chance less than one time in twenty.



Note.--Numbers indicate standard score intervals. Mean scores for norms group of 7,283 freshmen = 50.

Figure 2. ESTHETICISM (OPI)
1965–1969 Mean Scores for Individual Colleges

The consistencies across the scales and the colleges are somewhat exaggerated because these measures are not entirely discrete and independent; some of the scales have items in common and many of them are inter-correlated. Nevertheless, the data appear sufficiently clear to justify the general conclusions. Several major areas of change were shared by virtually all the colleges. Students became more autonomous, more aware of their emotions and impulses and willing to express them, more integrated personally, more esthetically sensitive and interested in the arts and humanities, more tolerant of ambiguity and of complexity, more liberal in their religious views, and less concerned about material possessions and practical achievements.

These changes occurred among authoritarian students attending highly structured institutions where there were many rules and regulations and where adults kept a close eye on the students. They occurred among anti-authoritarian students attending loosely structured institutions where the rules and regulations were few and where students were left fairly much on their own. Two traditional colleges--one of them relatively unknown, financially poor, lacking facilities, and the other prestigious, affluent, having ample facilities and resources--changed in the same ways; so did two nontraditional colleges, one which gave the student considerable freedom in selecting his courses and carrying out independent study, the other which had a formal curriculum, many required courses, and a complex system of comprehensive examinations.

Do Specific Changes in Attitudes, Beliefs, and Behaviors Underlie the Main Scores?

Analysis of individual items indicated the proportions of students at each college whose responses as seniors had, or had not, changed. For each college, items on which 27 percent or more of the students had changed their responses were culled from the 390 items of the Inventory. For most colleges, this cut-off point produced between fifteen and twenty-five "high-change" items--particular attitudes, beliefs, and behaviors which had changed for substantial numbers of students at each college.

A number of these high-change items were common to several colleges. For example, at eight colleges 27 percent or more of the freshmen who agreed that "no man of character would ask his fiancee to have sexual intercourse with him before marriage" no longer agreed with that statement as seniors. At six colleges a similar shift occurred for the item "I believe it is the responsibility of intelligent leadership to maintain the established order of things." At five colleges there was an increase of 27 percent or more in the number of students who agreed that "there is nothing wrong with the idea of intermarriage between different races" and that "women ought to have as much sexual freedom as men."

Table 3
High-Change Items Common to Several Colleges

Common Items	Direction of Change[1]
Common to Eight Colleges	
No man of character would ask his fiancee to have sexual intercourse with him before marriage.	-
Common to Six Colleges	
I have never done any heavy drinking.	-
I believe it is a responsibility of intelligent leadership to maintain the established order of things.	-
Common to Five Colleges	
There is nothing wrong with the idea of intermarriage between different races.	+
I believe women ought to have as much sexual freedom as men.	+
Common to Four Colleges	
I prefer people who are never profane.	-
The surest way to a peaceful world is to improve people's morals.	-
Perfect balance is the essence of all good composition.	-
When science contradicts religion it is because of scientific hypotheses that have not been and cannot be tested.	-
In the final analysis, parents generally turn out to be right about things.	-
I like modern art.	+
I go to church or temple almost every week.	-
Common to Three Colleges	
Our way of doing things in this nation would be best for the world.	-
I like short, factual questions in an examination better than questions which require the organization and interpretation of a large body of material.	-
At times I have been so entertained by the cleverness of a crook that I have hoped he would get by with it.	+
I often feel that the people I meet are not interested in me.	-
Young people sometimes get rebellious ideas, but as they grow up they ought to get over them and settle down.	-
Every person should have complete faith in a supernatural power whose decisions are obeyed without question.	-
It is a pretty callous person who does not feel love and gratitude for his parents.	-
Every person ought to be a booster for his own home town.	-
Nothing about communism is any good.	-
I dislike test questions in which the information being tested is in a form different from that in which it was learned.	-
I dislike women who disregard the usual social or moral conventions.	-
Trends toward abstractionism and the distortion of reality have corrupted much art in recent years.	-
I like worldliness in people.	+
I never attend a sexy show if I can avoid it.	-
We should respect the work of our forefathers and not think that we know better than they did.	-
I like to talk about sex.	+
When I go to a strange city I visit art galleries.	+
I like to work late at night.	+

[1] + indicates more frequent agreement over time; - indicates less frequent agreement

These shifts on particular items are thought-provoking and worth further scrutiny and reflection. The significant point for our present purposes, however, is this: Not only are the general changes among the colleges consistent at the level of scale score means, but also changes in particular attitudes, behaviors, and beliefs are common to many diverse types of students attending diverse types of institutions.

Do Students Become More Similar?

The standard deviations associated with initial and final testing show whether the distributions of individual scores on the various scales becomes narrower or wider after four years. If students become more similar, re-testing will show that individual scores are less widely dispersed--standard deviations will have become smaller. If students become increasingly different from one another, individual scores are more widely dispersed and standard deviations will increase. Table 4 indicates that on every scale, standard deviations were larger after four years. Therefore, even though students at Project colleges spanned a wide range as freshmen they had become even more diverse as seniors.

Changes in the standard deviations within each college are generally consistent with the changes when students were pooled. Table 5 reports 168 pairs of standard deviations--12 colleges times fourteen scales. In one hundred and four pairs standard deviations are higher after four years, 39 show no change, and 24 are lower. The twenty-four cases where lower standard deviations occur are scattered across all the scales, so increased similarity is not concentrated in a single area. Ten of the 24 cases occur at Kildew and Classic, the two colleges where students scored closest to the extremes as freshmen. Individuals scoring close to the extremes at entrance could not move further out, but because change typically is in their direction, more moderate classmates could move toward them, decreasing the differences reflected by the Inventory. Under these conditions of measurement it cannot be determined whether "true" change toward similarity occurred or whether the smaller standard deviations result from the limitations of the instrument. Standard deviations did increase on seven scales at Kildew and on six at Classic, so in any event there was no highly generalized trend toward increasing similarity at these two colleges.

In general, students did not become more similar during their four years of college. Diversity increased for the total group and more often than not, diversity increased within each college.

No Mean Change ≠ No Change

But it was hard to believe that such dramatic institutional differences did not affect at least some of the students. Because mean scores might mask underlying changes for certain kinds of persons it seemed desirable to examine individual changes within some of the groups.

Several OPI scales were selected for more detailed study; on some, statistically significant change had occurred; on others, mean scores were identical at both testings. A person's scores at first testing were subtracted from his retest scores, and the resulting distributions of individual change scores were examined.

Table 4

Standard Deviations, 1965 - 1969
Students Pooled from Twelve Colleges

Scale	F/65	S/69	Difference
Autonomy	9.71	10.24	+.53
Practical Outlook	8.60	9.34	+.74
Impulse Expression	10.82	11.43	+.61
Complexity	10.01	11.00	+.99
Estheticism	9.55	10.13	+.58
Thinking Introversion	10.05	10.38	+.33
Masculinity-Femininity	9.70	9.90	+.20
Personal Integration	9.85	10.72	+.87
Anxiety Level	9.76	10.16	+.40
Religious Orientation	10.03	10.29	+.26
Theoretical Orientation	9.37	9.87	+.50
Altruism	9.51	9.87	+.36
Social Extroversion	10.14	10.53	+.39
Response Bias	9.13	9.64	+.51

Table 5

Standard Deviations for Individual Colleges
Same Students Tested in 1965 - 1969

College	N	Autonomy 65 69	Practical Outlook 65 69	Impulse Expression 65 69	Complexity 65 69	Estheticism 65 69	Thinking Introversion 65 69	Masculinity Femininity 65 69	Personal Integration 65 69	Anxiety Level 65 69	Religious Orientation 65 69	Theoretical Orientation 65 69	Altruism 65 69	Social Extroversion 65 69	Response Bias 65 69
WJB	33	8 8	7 9	8 8	8 9	9 10	9 10	9 11	9 8	9 9	3 5	7 8	9 7	11 10	8 8
Savior	38	6 8	7 9	9 11	9 9	9 9	10 11	10 10	11 12	10 12	6 7	9 10	9 10	11 11	9 10
Sacred	29	6 9	7 9	10 14	7 11	8 10	8 8	8 10	10 11	10 9	4 7	7 9	8 8	9 13	9 11
Stonewall	49	8 9	9 9	7 10	9 9	9 10	9 9	8 8	9 11	10 11	8 8	9 10	9 10	9 11	7 11
Simon	76	7 9	7 9	9 12	8 11	9 11	9 10	9 10	9 9	9 10	4 7	9 9	7 9	10 10	9 9
Divinity	51	7 7	7 7	10 11	8 11	9 10	9 10	10 9	8 9	9 8	5 7	9 10	10 12	9 10	8 9
Friendly	61	8 9	8 9	10 10	8 10	8 9	9 11	10 10	10 11	9 10	6 7	9 9	10 10	11 10	9 10
Kildew	36	11 9	6 7	11 9	11 8	8 8	8 9	7 8	8 12	9 10	6 6	10 9	9 11	10 9	6 7
Classic	13	7 5	8 7	8 9	10 12	8 9	10 7	11 11	9 11	10 13	10 7	11 9	10 10	9 11	8 8
Elder	123	8 7	8 7	10 10	9 10	9 9	9 10	10 10	10 10	11 11	9 7	9 9	9 9	10 10	10 10
Woodbine	56	8 10	7 8	11 12	9 12	8 10	10 11	8 10	8 11	9 10	8 8	8 11	9 11	9 10	9 10
Rocket	20	5 6	5 7	11 9	7 10	7 8	9 11	7 6	12 13	11 10	9 8	9 10	9 9	12 14	9 10

The distributions of individual changes underlying the mean differences revealed four major patterns:

1. On some scales where significant increases had occurred, scores for practically all the students rose, but usually only a small amount.
2. On other scales where significant increases occurred, scores of a substantial number of students dropped, but the relatively large increases among the majority

outweighed the decreases among the contrary minority.

3. On some scales where the mean scores were identical at both testings, very few students made identical scores at both testings; many scores increased substantially, and many others dropped. Though the net effect was zero, many students had changed as much as half a standard deviation or more. (See Figure 3)
4. Finally, of course, on several scales where means were identical, individual scores changed little.

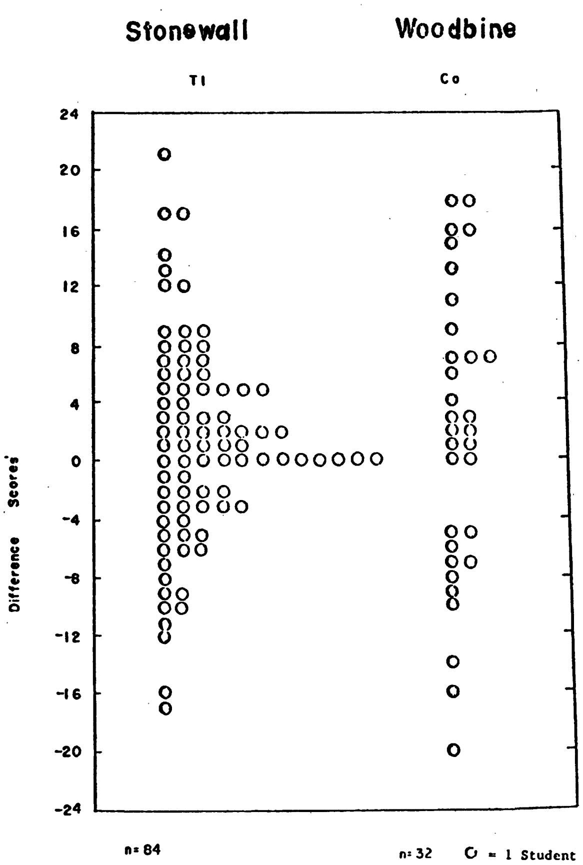


Figure 3. Distributions of Difference Scores

In effect, three different patterns of change lay behind the mean scores. First, on some scales all students changed in the same direction. Second, most students changed in the same direction but a substantial minority changed in the opposite direction. Third, most students changed in one direction or the other, but increases and decreases cancelled each other out, yielding similar mean scores.[3]

These exploratory studies suggested that despite the appearance of similarity suggested by mean differences, changes at one college might be quite different from those at another.

Do Similar Students at Different Colleges Change Differently?

Two studies examined change in subgroups of students who had similar OPI scores at entrance but who attended different colleges. Attention focused on the five scales which reflected the most substantial change: Autonomy, Impulse Expression, Practical Outloqk, Complexity, and Estheticism. For each scale, groups of students with similar scores at entrance were selected. Because entering students differed so widely from one college to another and because scores were relatively homogeneous within each college, the analyses were necessarily restricted to six of the twelve colleges, and even then the number of students from each college was small. Examining change at both two-year and four-year intervals further complicated matters. At both intervals students having similar scores at entrance were chosen, but different students and a slightly different group of colleges were involved.

Once the subgroups had been selected, retest means were computed for each group to see whether the extent and direction of change would once again be similar in diverse colleges. Table 4 gives the score intervals, the colleges, the mean scores, and the differences, for two- and four-year intervals.

These analyses revealed substantial differences among the colleges, both in the directions and the extent of change. Furthermore, the colleges maintained roughly similar relationships to one another on each of the scales and over the two-year and the four-year periods. Students at Kildew showed the greatest increases in autonomy, awareness of impulses and ability to express them, and tolerance of ambiguity and complexity; they also manifested the greatest decrease in drive for material success. Elder students, and--for the two-year interval--Classic students, changed in similar ways but to a lesser degree. Simon students, in contrast, consistently changed only

Table 6
Change for Similar Students at Different Colleges
(Selected OBI Scores)

Two-Year Change for Selected Subgroups: Scales and Score Intervals

College	Estheticism 52-57				Complexity 57-62				Autonomy 53-58				Impulse Expression 52-57				Practical Outlook 42-47			
	N	'65	'67	X_1-X_2	N	'65	'67	X_1-X_2	N	'65	'67	X_1-X_2	N	'65	'67	X_1-X_2	N	'65	'67	X_1-X_2
Kildew	5	55	59	4	5	60	68	8	5	56	66	10	8	55	58	3	5	44	38	-6
Classic	5	54	59	5	5	59	57	-2	6	55	63	8	6	54	57	3	5	46	45	-1
Elder	14	55	58	3	10	59	57	-2	14	56	60	4	15	55	56	1	17	45	44	-1
Stonewall	11	55	54	-1	11	59	52	-7	10	54	50	-4	26	55	57	2	6	46	48	2
Friendly	12	55	57	2	7	60	62	2	5	55	55	0	12	54	54	0	11	45	46	1
Simon	25	55	56	1	7	59	55	-4	7	54	55	1	9	54	50	-4	17	46	47	1
Four-Year Change for Selected Subgroups: (Same scales & scores intervals, different students)																				
	N	'65	'69	X_1-X_2	N	'65	'69	X_1-X_2	N	'65	'69	X_1-X_2	N	'65	'69	X_1-X_2	N	'65	'69	X_1-X_2
Kildew	5	56	56	0	8	60	64	4	11	56	65	9	9	55	62	7	11	44	38	-6
Elder	35	54	60	6	25	58	58	0	32	55	62	7	21	54	58	4	29	44	42	-2
Woodbine	25	55	57	2	8	59	59	0	13	54	58	4	7	55	58	3	8	46	40	-6
Friendly	17	54	57	3	5	60	50	-10	6	55	56	1	8	54	56	2	10	45	40	-1
Divinity	12	54	56	2	7	58	58	0	7	54	56	2	4	56	54	-2	9	45	43	-2
Simon	20	55	60	5	4	60	56	-4	9	55	57	2	6	55	56	1	10	46	43	-3

Note.--The study of two-year change was carried out by Kenneth Carter for his doctoral dissertation. Though membership overlaps, the student groups differed from scale to scale.

slightly and sometimes in the opposite direction. At Stonewall, Friendly, and Divinity, the pattern of change was closer to the Simon pattern than to the Kildew, Elder, or Classic pattern. No clear patterns of change emerged for the Estheticism scale.

Perhaps these changes simply reflect regression effects for the sub-groups of students within the different score intervals. Several considerations suggest, however, that such effects could not operate with much force. Regression effects represent changes from initial responses attributable to "chance." The assumption is that extreme initial scores are most likely to have been influenced by chance errors and that later scores for the same persons are less likely to include these chance errors. But the fact is that the score intervals were not initially extreme; in only eight of the thirty scale-college comparisons do the sub-groups differ from their college means by as much as one standard deviation. Furthermore, the three colleges in which differences between subgroup means and college means are greatest show least change in the regressive direction. These considerations do not mean that regression effects were entirely absent, but they do suggest that such effects did not seriously distort the results.

These results, then, suggest (a) that the extent and directions of change for selected subgroups of students vary from college to college, and (b) that these intercollegiate differences are consistent for both two-year and four-year change and for several different dimensions of development. Apparently a college's characteristics do make a difference to student development. Even though mean changes are similar for diverse students and diverse institutions, similar students who enter different kinds of colleges change differently.

Discussion

The varied findings indicate that most students develop along the same general lines during the college years. But they also demonstrate that for particular groups of persons at particular colleges, such development may be accelerated or retarded. The results make it clear that research concerning student development and institutional impacts must go beyond simple measures of central tendency and simple examination of net change. Averages obliterate individuals and fail to reveal the complex interactions which influence events and their developmental consequences. Both standard deviations and underlying frequency distributions are required for accurate understanding and sound judgements.

It is clear that the college student is no _tabula rasa_; he is no clay for the potter, no vessel to be filled, no lamp to be lighted. He's already lit. When he moves into college as a freshman he brings with him--along with his Webster's Collegiate Dictionary and his Tensor lamp--his mother, father, two older friends of the family, a girl friend, and a set of high school buddies. He also brings strengths and weaknesses, prides and prejudices, clarities and confusions, and a lot of unfinished business. The unfinished business typically includes not only improving intellectual and interpersonal competence, but also achieving autonomy, learning better ways to manage sexual and aggressive impulses, becoming freer with diverse kinds of persons, clarifying identity, sharpening purposes and developing integrity.[4]

Most students move toward institutions whose purposes and programs fit their own interests and inclinations and whose students and faculty hold values and attitudes similar to their own. In many cases, the processes of self-selection and institutional admissions practices creates a comfortable fit between the college and the person. Though the college may no longer act _in loco parentis_, it does act _in loco uterus_, providing a warm and supportive setting which insulates students from unduly disruptive outside influences.

Under these comfortable conditions, personal development proceeds along the vectors of change set by the general cultural and genetic forces operating in our society; most students in most colleges work on the seven major areas of unfinished business mentioned above. Across the country there are a variety of student-college types; in each type these common developmental tasks are pursued in somewhat different fashion. But because the institutional differences correspond with differences among the students, the patterns of general development are about the same for the different types.

One major model for college influence, therefore, is the womb. The diverse colleges provide safe havens and proper nourishment for the diverse students in our pluralistic society. Persons who attend college become more autonomous, more flexible, more complex, less materialistic, more aware of their own emotions, and better able to express them in thought and action; more tolerant of ambiguity, less dogmatic, more intellectually curious. Persons who do not attend college change less, and sometimes even move in contrary directions. So wombs are good things. Without them, most of us would not be here. And without the protection and nourishment many colleges offer, most seniors would not have become what they are at graduation.

But there is another kind of college influence, and for it the term _impact_ is appropriate. Every institution has two different kinds of deviants, two kinds of uncommon or atypical students. The first kind is the student whose development has not yet reached the general level of his peers and of the college; the second kind is the student whose development has gone beyond the level of the college and of the other students. Deviation can occur in many areas. A student's intellectual competence and breadth of information may be so limited that he finds it difficult to cope with the academic program. Or his competence and store of information may be so high that he gets little stimulation from classes, study requirements, and his fellow students. Some students may be more liberal than most students at his college, others may be more conservative; some may be more culturally sophisticated or less, more autonomous or more dependent.

By examining subgroups of similar students at different colleges, we were, in effect, studying persons who were deviants at some colleges but not at others. Students who belong to these subgroups, who deviate from the norms within the different colleges, change according to the relationships between their characteristics and those of the college. These students are apparently influenced by institutional differences in general atmosphere and student characteristics, in educational practices, in student-faculty relationships, and in the nature of relationships among friends and acquaintances.[5] For these persons, the choice of a college and the subsequent experiences may have significant consequences.

Footnotes

[1]This research was undertaken in the context of the Project on Student Development In Small Colleges, supported by PHS Research Grant #MH14780-05, National Institute of Mental Health. Credit is also due the American Council on Education, Office of Research, for critical comments and secretarial assistance. Education and Identity (San Francisco: Jossey-Bass, 1969), describes the theoretical framework behind these studies and summarizes prior pertinent research.

[2]Although the findings reported here come only from the Omnibus Personality Inventory, the general areas of change and the general principles suggested by these results are supported by test-retest data from other Project instruments. Some of these other results are reported in College Impacts on Political Liberalism and College Impacts on Cultural Sophistication, listed among the Project publications.

[3]For more detailed information, see A.W. Chickering, "FD's and SD's: Neglected Data in Institutional Research" (paper presented at the Eighth Annual Forum of the Association for Institutional Research, Detroit, 1968).

[4]For further information about these major areas of development in college, see A.W. Chickering, Education and Identity (San Francisco: Jossey-Bass, 1969).

[5]For evidence concerning these relationships see Chickering, A.W., College Experience and Student Development, paper presented at the 137th Meeting of the American Association for the Advancement of Science, December, 1970.

DISCUSSION

Kenneth A. Feldman, State University of New York at Stony Brook

The papers that have just been presented ("New Directions in the Study of College Impact on Students") vary with respect to such matters as methodological sophistication, the extent to which data are presented, the kinds of theory underlying the data presentation, and so forth. Yet they are alike in one respect: they all make distinctive contributions to the study of college impacts, and in this sense each of them represents a "new" direction. Thus my remarks intentionally are appreciative in tone. I shall discuss the papers in an order different from that of their presentation.

David Kamens ("Size of College and Its 'Charter' as Determinants of College Effects") makes no methodological breakthroughs in his study--his technique of analysis is to compare percentages--but he does offer important theoretical and conceptual contributions. In effect, the typical college impact study conceives of the college as a more or less self-contained unit, as an environment unto itself. Kamens points out the need to probe past this point of view, to look for the relationship of a college to the wider social order of which it is a part. Indeed, the way in which the college is related to the larger social structure can and does have impacts on students quite apart from the effects of its internal environments.

Kamens' framing of his problem and his analysis of colleges clearly manifests a sociological approach to the study of college impacts--as, to some extent, does his choice of dependent variables (shifts in occupational commitment and commitment to the college). On the other hand, Arthur Chickering ("The Multi-Directionality of Student Change") is a psychologist (or, more correctly, a social psychologist). He is interested in explaining student changes in personality traits. Like Kamens, he takes a multi-institutional approach. He offers a variety of interesting analyses: (1) he finds out which particular scale items show high change across colleges; (2) he discusses the importance of comparing standard deviations of college classes (freshmen and seniors) as well as average scale scores; (3) he demonstrates how average change scores can camouflage the direction and amount of individual change; (4) he controls for initial position on the variable for which change is being measured. None of these procedures are world-shaking in themselves in any absolute sense. But they are so seldom done in the research on college students that they add up to a definite methodological contribution.

Chickering notes that student self-selection and college admissions practices operate in such a way that certain kinds of students tend to enter certain kinds of colleges. One implication of this is that differential changes by students in different colleges may be due to the personality and background differences of the students rather than to the environmental differences of the colleges. Chickering does not control or adjust for this non-random distribution of students among the colleges he studied. Alternately put, he does not deal with the multicollinearity of student input and environmental variables. It is in fact not easy to do so. Assessing the differential nature and amount of the impacts on students of different college environments--while at the same time taking into consideration the background, attitudinal, and personality differences among students within and between colleges--has turned out to be a particularly vexing assignment for researchers interested in the effects of higher education.

John Creager ("Statistical Models for Assessment of College Impacts") has presented a number of methods and models that have been suggested and are currently being used to handle this distinctive challenge in assessing college impacts. I think he has been particularly adept in comparing and contrasting these methods and in pointing out the advantages, disadvantages, and problems with each of them. He has been somewhat too modest about the model he himself has helped to develop--namely, the orthogonal decomposition of composite variance--which promises to be very useful.

Charles Bidwell ("The Use of Path Analysis in the Study of College Impacts"), in his study of Harvard students, is actually applying one of the methods reviewed by Creager: path analysis. Indeed, to the best of my knowledge, the efforts of Bidwell and his associate, Rebecca Vreeland, represent the first large-scale and systematic use of path analysis in the study of college impacts. As such, they are helping to answer some of Creager's questions and uncertainties about the method and its applicability.

INCOME FINDINGS FROM THE SOCIAL SECURITY ADMINISTRATION'S SURVEY OF NEW BENEFICIARIES

William Rabin and Alan Fox, Division of Retirement and Survivor Studies,
Office of Research and Statistics, Social Security Administration,
Department of Health, Education, and Welfare

Amendments to the Social Security Act effective in 1956 permitted women to become entitled to reduced benefits as retired workers the month they become 62. In 1961, this choice was extended to men. For both men and women the monthly benefit rate is reduced five-ninths of 1 percent for every month of benefits they receive before they are 65. The response to the early entitlement option has been very substantial. Currently, about half of the men awarded retired worker benefits and two-thirds of the women are choosing to become entitled before they reach 65.

To answer questions about the circumstances surrounding the decision to become entitled, and especially the decision to become entitled before age 65, the Social Security Administration has undertaken the Survey of Newly Entitled Beneficiaries (SNEB). This Survey provides a unique opportunity to study a large sample of new social security beneficiaries close to the time they file for benefits and, presumably, make the decision to begin or not to begin the process of retirement. Obviously, the income resources of these older people are a major factor in such a decision.1/

Each month, starting with July 1968, a sample of persons was selected by a random process from awards made during the preceding month. A questionnaire was mailed by the Bureau of the Census, as collection agent for the Social Security Administration, to each person selected. Second and third mailings were made if necessary, and a personal interview follow-up was made to reduce residual nonresponse.

In all, about 75 percent of the questionnaires mailed out were returned and met the minimum acceptance criteria. An additional 5 percent refused to answer, were reported as deceased, were returned as undeliverable, or failed to meet the acceptance criteria. The remaining 20 percent were not returned.

After the personal interview follow-up was completed the response rate rose to 90 percent. The follow-up also resulted in the conversion of a large number of nonresponses into firm refusals when contact was made.

1/ Other studies that have dealt with this question include the Social Security Administration's 1963 and 1968 Surveys of the Aged, and Richard Barfield and James Morgan: Early Retirement: The Decision and the Experience (Survey Research Center, The University of Michigan, 1969).

The survey population consists of two distinct groups of persons recently awarded retired worker benefits:

1. Workers with benefits payable at award. About 85 percent of them became entitled before age 65 to reduced benefits.

2. Workers with benefit payments postponed because they have substantial earnings and, therefore, are not eligible to receive social security benefits. These are mostly workers whose main reason for filing was to obtain Medicare coverage. About 85 percent of them became entitled at age 65, and we can assume that in the absence of Medicare, most of them would have delayed filing beyond age 65.

Since the population includes only those persons who were awarded benefits as retired workers during the survey period, it excludes persons who were awarded widow's, children's, wife's or other types of survivor's and dependent's benefits as well as those awarded disability benefits.

The Survey of Newly Entitled Beneficiaries is one of a series of interlocking studies of retirement currently being conducted by the Social Security Administration. Other studies include the Longitudinal Retirement History Survey which follows a sample of persons from pre-retirement to post-retirement, the 1968 Survey of the Demographic and Economic Characteristics of the Aged, a periodic cross-sectional study of persons 65 and over, and additionally some in-depth analyses of social security earnings and benefit records.

The SNEB questionnaire requested a nonmarried beneficiary to fully report on the receipt and amount of each of seventeen separate sources of income. In addition, a married beneficiary was asked to report for his spouse on eleven of these sources. Total income was accumulated from the information reported about each separate source. If a report on even one single source of income was incomplete, after editing, it caused a nonresponse on total income.

Nonresponse on total income was very high for the first year of SNEB, ranging from about 25 to 40 percent depending on marital status, sex and benefit payment status of the new beneficiary (see table 1). Earnings and asset income were the most poorly reported, with nonresponse ranging from 11 to 25 percent. Nonresponse on earnings was higher among married couples where two reports (one for the husband and one for the wife) provided two opportunities for an incomplete response. The nonresponse rates for the other two major sources of income, social security payments and private pension payments, were also higher among married couples, but not as high as those for earnings and asset income.

In order to reduce the nonresponse rates for total income to the point where information on the size of income could be analyzed, special procedures were developed starting with the July 1969 awards. All questionnaires returned by mail were screened clerically for completeness of income response. The defective cases were followed up by telephone or personal interview.

The reduction in nonresponse on total income for the last six months of 1969 varied between 11 and 20 percentage points, bringing the income response rate for most groups to 80 percent or higher. The only exception was for married women with benefits payable where the response rate was increased from 62 percent to only 76 percent.

There are two types of nonresponse: nonresponse on receipt (where the beneficiary does not report whether or not he is receiving a specified source of income), and nonresponse on amount (where the beneficiary reports receipt of income from a specified source but does not report the amount of such income). By far, the major reason for the decrease in total nonresponse has been the reduction of nonresponse on receipt of all sources except earnings. Since editing procedures imputed employment status, no nonresponse on receipt of earnings is possible.

The survey was designed to probe some of the economic characteristics and other reasons which probably influence the beneficiary's decision to claim social security benefits. It was reasoned, therefore, that the sources and amount of income that the worker was actually receiving at the time of his award, or would receive in the immediate future, was more likely to affect his desicion to claim benefits than the amount of income during the previous calendar year or even in the twelve month period prior to the survey.

For this reason new beneficiaries were asked to report the monthly income they were currently receiving for sources such as private pensions and workmen's compensation payments (which are often paid monthly or weekly), and to report an annual rate for asset income and other sources often paid quarterly or semi-annually. Earnings for the beneficiary's spouse were requested annually, while the earnings of the new beneficiary himself were reported at an hourly, weekly, monthly, or annual rate at his option.

All income information not reported as an annual rate was converted to an annual rate. This process probably yields levels close to the actual amounts for types of income which are received regularly, such as pensions, but may overstate income that is not received every month. On the other hand, income which a beneficiary starts to receive a short time after he responds would be understated.

Preliminary income findings from SNEB confirm conclusions drawn from other studies that only a few types of income contribute prominently to the total income picture of older people (see table 2). As might be expected, earnings dominate the total income of persons with benefits postponed. Among persons with benefits payable at award earnings are also important, but share prominence with social security payments which, although generally providing a lower income than earnings, are received by many more respondents. Income from other pension plans, both public and private, was reported by 15 to 41 percent of the respondents, with married men most likely to receive such income, and unmarried men and women least likely. For those who received income from these sources, the amount was quite substantial. Income from assets, reported by roughly half the respondents, was considerably smaller than the amounts of the other major types of income.

Sources of income often associated with the elderly, such as public assistance and contributions from relatives, were reported by only a very small proportion of respondents.

The social security earnings test allows beneficiaries to earn as much as $1,680 a year without any loss of benefits. Those who earn over $1,680 have one dollar withheld from benefits for every two dollars earned between $1,681 and $2,880, and one dollar withheld for every dollar earned above $2,880. However, no benefits are withheld for any month in which the person does not earn $140 or more or does not render substantial services as a self-employed worker. For this reason a person may receive benefits for part of the year even though his annual earnings are high.

In addition, there may be a considerable interval of time between the date of filing for benefits and the completion of the SNEB questionnaire--as much as six months. During this interval the employment status and earnings of a beneficiary may have changed. Since benefit payment status at award is generally determined from the statements of the beneficiary at filing, when changes do occur there may be some apparent inconsistencies between earnings and benefit payment status.

As stated before, income from earnings is an extremely important component of total income even for workers receiving benefits and presumably "retired." Among persons with benefits payable at award about the same proportion of men and women, approximately one-third, was employed. About half the employed men and slightly more than half the employed women reported earnings of $1,680 a year or less, thereby suffering no loss of benefits due to the earnings test. Their median annual earnings were about $1,700 and $1,600 respectively (see table 3).

On the other hand, most with benefits postponed reported current employment, with median earnings of $7,400 for men workers and $5,100 for women. In fact, nearly 30 percent of these employed men workers reported earnings of $10,000 or more.

Generally the income position of nonmarried persons with benefits payable at award is quite poor. The median annual income reported by such persons, men and women, is about $2,200. About 40 percent had incomes below the poverty level (see tables 4 and 5). The median total income for those nonmarried beneficiaries who were not currently employed was about $1,800 for men and $1,500 for women. About one-third of these non-workers reported no income other than social security payments. A little less than half of all nonmarried men and women reported some income from assets or from other pension plans.

The income picture for married couples with benefits payable is brighter than for nonmarried persons due mostly to the presence in the family of two potential recipients of income rather than one. Relatively few of these couples reported income below the poverty level--in fact 60 percent of them had income greater than the Labor Department's "moderate" standard of $4,230 a year.

Since the employment of married persons with benefits payable is subject to the earnings test, the earnings of these persons would tend to be low. Their spouses, on the other hand, do not necessarily have such an inhibition on earnings unless they too happen to be receiving social security benefits. This would favor higher earnings for the couple when the newly entitled beneficiary is a woman rather than a man because a husband, usually having a greater earning capacity, tends to work more frequently even though he is generally older. This is supported by findings that the median earnings for couples is $3,300 when it is the husband who receives benefits and $5,400 when it is the wife (see table 2).

About one-third of all these couples reported that they had no asset income and no pension income other than social security benefits. The median annual income from both assets and other pension plans, for those who did receive such income, was about $2,200 for married men respondents and their wives and about $1,400 for married women respondents and their husbands.

Persons with benefits postponed at award, two-thirds of whom are married men, reported relatively high incomes. Almost 9 out of 10 had income above the "moderate" standard. This income was composed mainly of earnings. If they stopped working or reduced their earnings, the additional income they might receive from social security and other types of pensions would only partially offset the loss.

The survey did not collect information about the financial assets of beneficiaries, only about their income from such assets. Beneficiaries with substantial assets might deplete them over their remaining lifetime and provide themselves with additional funds for subsistence. However, since roughly half the respondents did not report any income from assets at all it is reasonable to assume that such assets do not exist in great abundance among the beneficiary population.

The income information available at this time from the survey provides further confirmation that a large proportion of our elderly population continues to work even if it is entitled to retired worker benefits. Earnings, not pension or asset income, appear to be the major supplement to social security payments for those receiving benefits at the time of award. As the ability to work declines with age, the total income of beneficiaries will also probably decline and for many of them, without substantial income from assets or other pensions, it will drop even below the poverty level.

* * *

So far the discussion has concerned total income, its distribution, and some of its components. Now factors associated with total income will be briefly examined, using the Automatic Interaction Detector (A.I.D.) program developed at the University of Michigan's Survey Research Center. This program takes total income as the dependent variable, and divides the sample into a series of subgroups that maximize the ability to explain the variation in income.

The analysis was restricted to married men and their wives classified by whether or not they were employed at the time of the survey. Figure 1 describes A.I.D. findings for couples where the husband was employed, and Figure 2 for those where he was not employed. Other groups--married women and their husbands, and nonmarried men and women--will be analyzed in the near future.

Married Men, Currently Employed (Figure 1)

For currently employed married men, benefit payment status, occupation, and employment status of their wives were the most important factors associated with income. The first step splits those whose benefits were postponed at award from those who were receiving benefits but continuing to work: those whose benefits had been postponed had higher average incomes. For both Groups 2 and 3, current occupation was a significant determinant of income: those in the higher-status occupations had higher average incomes than the rest (Groups 5 and 9).2/ For respondents in the lower-status occupations (Groups 4 and 8), whether or not their wives were working was a significant factor (Groups 10, 11, 12, and 13). It is interesting that the employment status of wives was not a significant determinant of income where the respondents were in higher-status occupations; perhaps for such couples wives tend to have earnings so much lower than their husbands that their

2/ The term "high-status occupations" used throughout this section varies from case to case: Figures 1 and 2 give more precise definitions.

contribution to family income is relatively small. Place of residence was significant only for respondents whose benefits had been postponed and who were in high-status occupations (Group 5): those who resided inside an SMSA had higher incomes than those who resided outside of an SMSA (Groups 6 and 7).

Predictors that did not prove to be significant for currently employed men included: race (because of the small number of Negro men in the sample), region, health-disability status, home ownership, and whether or not the respondent was receiving another pension. Retirement status (whether the respondent considered himself to be retired, partly retired, or not retired at all) also was not significant; this might be due to the close relationship between retirement status and employment status, and since the A.I.D. runs were split according to employment status, the impact of retirement status would tend to disappear. Class of worker among wage and salary workers probably was overshadowed by current occupation, and proved not be significant. Employment-tenure (whether the respondent had been employed more than five years in his current job) also was not significant. This is not to say that these variables did not behave as expected (for instance, workers reporting work limitation due to poor health or disability tended to have lower incomes than those who reported no limitation); what is meant is that these variables did not reduce the total sum of squares by a significant amount.

The predictors isolated in this A.I.D. run explain about one-quarter of the total variation in income. That is, the coefficient of multiple determination, R^2, is about .25.

Married Men, Not Currently Employed (Figure 2)

The second run was restricted to married men who were not currently employed. This group has a far lower average income than the previous one (nonemployed married men have an average income of $5,600 as opposed to the employed married men's average of $10,300. Again, occupational status (this time the occupation of the respondent in the last job he held) is a significant variable, with those who had been in high-status occupations having an average income of $9,000 (Group 3), as opposed to the remainder, whose average income was only $5,100 (Group 2). Following out the low-status occupation branch, last occupation appears again at the end: for respondents in low-status occupations whose wives were not employed, and who were not receiving any pension other than social security, those who were self-employed had a higher average income than the rest (Groups 12 and 13). Also, receipt of a pension other than social security was important only for respondents whose last occupation had been a comparatively low-status one (Groups 6, 7, 10, and 11). Residence in the South was significant for one group (Groups 14 and 15). For respondents in high-status occupations (Group 3), the only significant variable was health limitation (Groups 8 and 9). The probable reason for the shortness of this branch is that the high occupational status group is rather small, and few statistically significant splits can occur.

Predictors that were nowhere significant in this run were: race (again, due to small sample size), home ownership, retirement status, benefit-payment status, class of worker in previous job, and employment-tenure in previous job.

For these nonemployed married men, the predictors isolated by the A.I.D. technique explain approximately one-third of the total variation in income.

The characteristics used as predictors in the A.I.D. process just described by no means exhaust the abundant supply that the survey has produced. Further use of the A.I.D. technique is planned, not only to identify significant income factors for the population as a whole but also to isolate discrete groups of workers for separate analysis.

Table 1.--Nonresponse on total income and selected sources of income: Percent of persons initially entitled to retired-worker benefits by sex, marital status and benefit-payment status, July-December 1969 and July-December 1968 awards

	Benefits payable at award				Benefits Postponed at award			
	Non-married men	Non-married women	Married men and their wives	Married women and their husbands	Non-married men	Non-married women	Married men and their wives	Married women and their husbands
	July through December 1969							
Total number (thousands)......	41	68	151	112	16	30	115	14
Nonresponse rate:								
Total income................	15	19	20	24	15	17	18	19
Earnings....................	6	8	9	12	5	6	10	6
Social security payments....	1	1	3	4	1	*	2	3
Private pensions............	1	2	4	2	2	1	3	3
Income from assets..........	7	11	7	11	10	12	8	12
	July through December 1968							
Total number (thousands)......	44	67	152	108	17	28	112	13
Nonresponse rate:								
Total income................	28	32	39	38	26	33	33	39
Earnings....................	12	13	18	20	12	11	16	24
Social security payments....	3	4	12	12	6	6	11	10
Private pensions............	3	4	12	12	7	9	11	10
Income from assets..........	15	20	16	20	15	25	16	23

* Less than 0.5 percent.

Table 2.--Selected sources of income: Percent receiving and median annual rate of persons initially entitled to retired-worker benefits by sex, marital status and benefit-payment status, July-December 1969 awards

	Benefits payable at award				Benefits postponed at award			
	Non-married men	Non-married women	Married men and their wives	Married women and their husbands	Non-married men	Non-married women	Married men and their wives	Married women and their husbands
Total number (thousands).......	41	68	151	112	16	30	115	14
	Percent reporting income from specified sources							
Earnings......................	38	44	52	60	77	84	88	83
Social security payments.......	90	85	89	91	32	24	32	66
Private pensions...............	15	11	28	21	14	11	15	22
Assets.........................	36	46	51	51	51	57	62	64
	Median annual rate (in dollars) for recipients of specified sources							
Earnings......................	1,520	1,640	3,330	5,450	6,400	5,130	8,660	7,470
Social security payments.......	990	1,000	1,520	1,470	1,790	1,620	1,780	1,640
Private pensions...............	1,560	1,110	2,200	1,480	1/	1/	1,800	1,220
Assets.........................	490	340	660	500	430	340	650	360

1/ Less than 50 sample cases.

Table 3.--Annual rate of income from earnings: Percentage distribution of persons initially entitled to retired worker benefits by sex and benefit-payment status, July-December 1969 awards

	Benefits payable at award		Benefits postponed at award	
	Men	Women	Men	Women
Total number (thousands)....................	192	181	131	44
Number reporting employment (thousands)......	72	60	109	35
Number reporting amount (thousands)..........	61	51	101	33
Total percent..............................	100	100	100	100
Under $1,000..............................	22	27	1	1
$1,000 - $1,999...........................	34	35	3	4
$2,000 - $2,999...........................	10	12	2	3
$3,000 - $3,999...........................	9	9	6	20
$4,000 - $4,999...........................	5	5	8	21
$5,000 - $5,999...........................	4	4	11	16
$6,000 - $6,999...........................	3	1	14	7
$7,000 - $7,999...........................	3	1	11	11
$8,000 - $8,999...........................	3	1	9	5
$9,000 - $9,999...........................	1	1	6	3
$10,000 - $14,999.........................	4	2	15	8
$15,000 and over..........................	2	2	14	2
Under $1,680..............................	49	55	3	4
Median annual rate of earnings (dollars).....	1,710	1,570	7,430	5,060

Table 4.--Annual rate of total income: Percentage distribution of persons initially entitled to retired-worker benefits by sex, marital status and benefit-payment status, July-December 1969 awards

	Benefits payable at award				Benefits postponed at award			
	Non-married men	Non-married women	Married men and their wives	Married women and their husbands	Non-married men	Non-married women	Married men and their wives	Married women and their husbands
Number (thousands)............	41	68	151	112	16	30	115	14
Reporting total income (thousands)......................	35	55	121	85	14	24	94	11
Total percent.................	100	100	100	100	100	100	100	100
Under $1,000.................	19	19	2	2	1	1	*	---
$1,000 - $1,999.............	28	28	9	5	8	5	3	1
$2,000 - $2,999.............	17	19	12	12	7	5	3	6
$3,000 - $3,999.............	13	12	13	12	8	17	4	8
$4,000 - $4,999.............	9	9	14	12	6	18	5	11
$5,000 - $5,999.............	5	5	11	11	13	16	7	6
$6,000 - $6,999.............	2	2	8	9	15	9	9	7
$7,000 - $7,999.............	3	1	6	8	9	9	8	7
$8,000 - $8,999.............	2	2	6	6	7	7	8	11
$9,000 - $9,999.............	1	1	4	5	6	3	6	5
$10,000 - $14,999...........	1	2	9	14	9	8	23	19
$15,000 and over............	1	2	6	6	10	3	22	20
Median annual rate (dollars)...	2,180	2,180	4,930	5,660	6,570	5,320	9,270	8,390

* Less than 0.5 percent.

Table 5.--Annual rate of total income by income level standards: Percentage distributions of persons initially entitled to retired-worker benefits by sex, marital status and benefit-payment status, July-December 1969 awards

	Benefits payable at award				Benefits postponed at award			
	Non-married men	Non-married women	Married men and their wives	Married women and their husbands	Non-married men	Non-married women	Married men and their wives	Married women their husbands
	Income level standards (in dollars)							
Poor (Social Security Administration......................	1,773	1,751	2,217	2,217	1,773	1,751	2,217	2,217
Moderate (Department of Labor).	2,342	2,314	4,230	4,230	2,342	2,314	4,230	4,230
	Percentage distributions							
Total number (thousands).......	41	68	151	112	16	30	115	14
Number reporting income (thousands)......................	35	55	121	85	14	24	94	11
Total percent.................	100	100	100	100	100	100	100	100
Poor.........................	40	40	14	9	6	4	3	1
Poor to moderate............	14	13	26	25	5	4	8	16
Above moderate..............	46	47	60	66	89	92	89	82

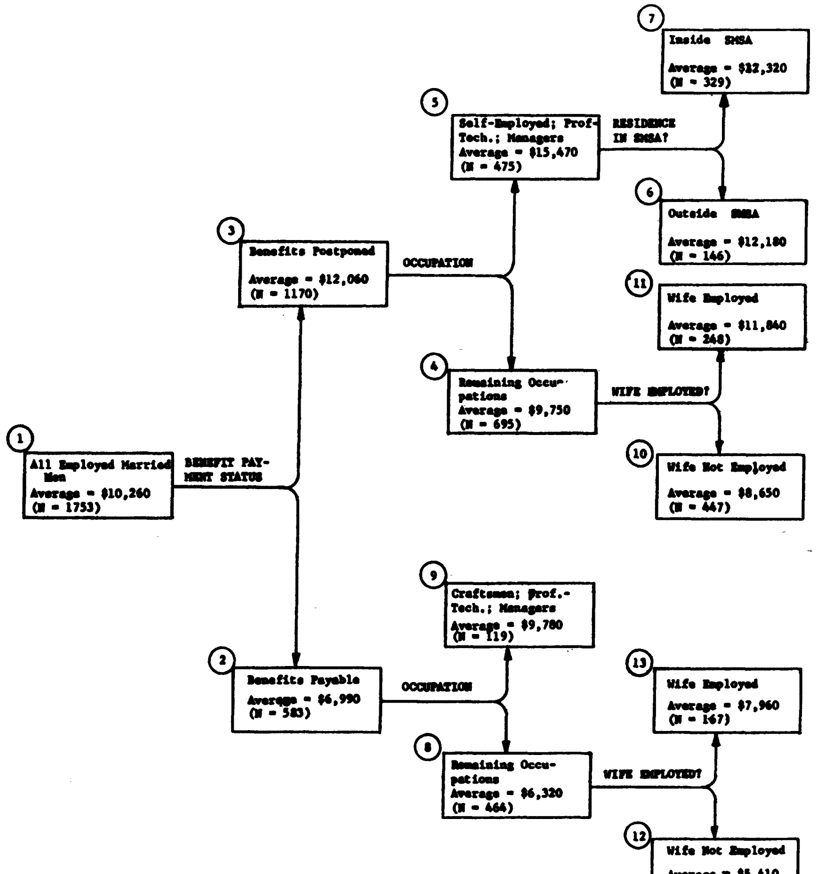


FIGURE 1:

Married Men and their Wives -- Currently Employed

R^2 = .254

Source: A.I.D. 10/21/70 Run 8

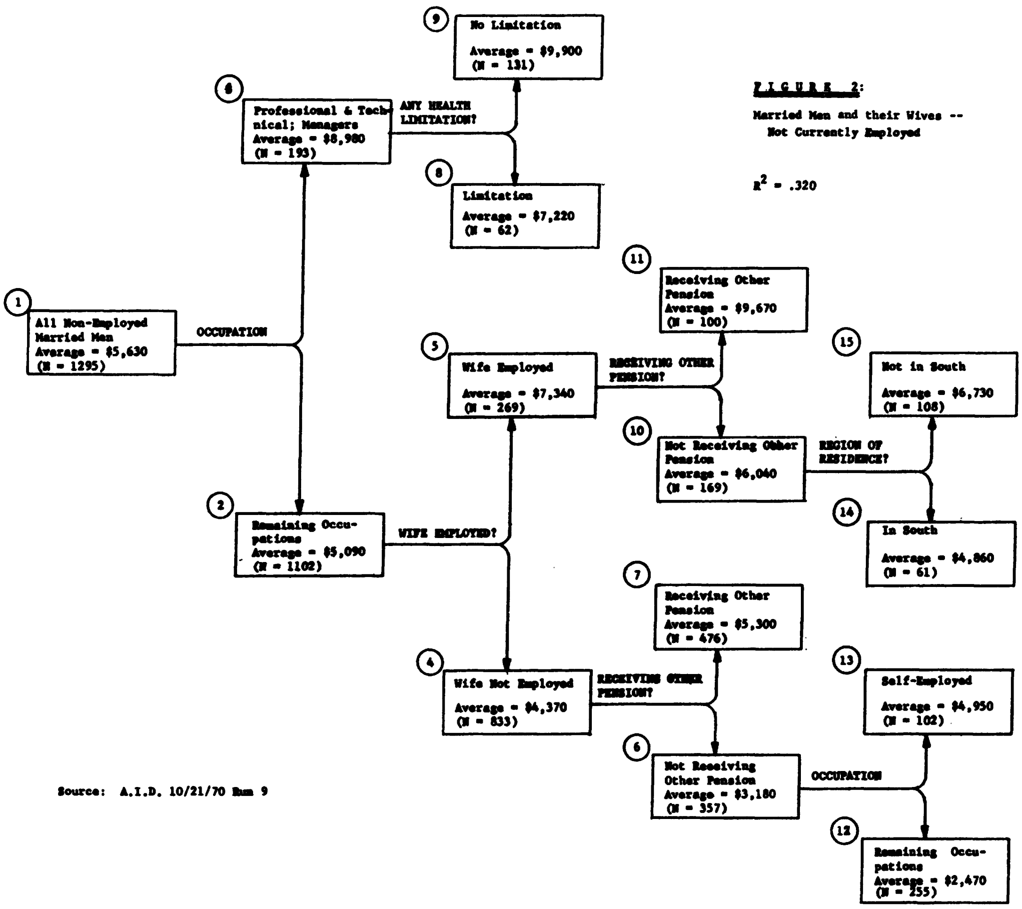


FIGURE 2:

Married Men and their Wives -- Not Currently Employed

R^2 = .320

Source: A.I.D. 10/21/70 Run 9

USE OF MEDICAL CARE SERVICES AND THEIR RELATION TO ECONOMIC AND SOCIAL CHARACTERISTICS

Jerome Green and Jack Scharff, Social Security Administration

Medicare consists of two related health insurance programs for persons aged 65 and over--a basic compulsory program of hospital insurance and a voluntary program of supplementary medical insurance. The hospital insurance program pays for a large part of the cost of hospital and skilled nursing home services in facilities qualifying for Medicare participation, and under certain conditions, the cost of home health services. The medical insurance program provides payment for 80 percent of "allowable" charges for physicians' and other covered services after the enrollee pays the first $50 of such charges during the calendar year. In addition to the payment of physicians' services regardless of place of service, this program also pays for diagnostic tests and therapeutic treatment either in the physician's office or in an outpatient department of a hospital. It also pays for rental or purchase of durable medical equipment, ambulance use, and home health services.

This paper deals only with this latter program, supplementary medical insurance, and will present data from a survey designed to collect information on use of and charges for medical services. After reviewing the survey, we will summarize findings for calendar year 1967. The discussion will be divided into three general areas: (1) characteristics of persons who use medical services; (2) sources of payment which enrollees use in paying the deductible and coinsurance amounts not paid by Medicare; and (3) use of prescription drugs, not covered under the medical insurance program, but about which data have been collected in the Current Medicare Survey.

The basic Medicare data system has certain limitation because medical bills are its sole source of information. Perhaps the major limitation is the considerable delay in the reporting of current information. Final data for the supplementary medical insurance program do not become available until SSA receives the medical bills sent to and paid by carriers throughout the country. Another disadvantage is the inability to collect information on any thing not specified on the bills which are needed to effect payment. In addition, physicians or enrollees may decide to accumulate a number of bills before filing for reimbursement. In fact, initially they were encouraged not to send in their bills until they had met the $50 deductible. As a result, it was originally estimated that there would be a 6-month delay from the date of service to receipt of the bill by SSA. In fact, the lag has turned out to be much longer than the anticipated 6-months. Recognizing the need for current data, the Social Security Administration initiated a continuing monthly Current Medicare Survey. A series of household interviews with persons enrolled in the supplementary medical insurance program was initiated in July 1966 when Medicare began.

Sample Design

The medical insurance sample includes about 4,500 persons selected from the 5-percent statistical sample of persons enrolled in the supplementary medical insurance program.

Each year, a new panel is selected and these individuals remain in the sample for 15 months. This 15-month cycle was determined by the fact that any covered medical expenses incurred by an individual in the last 3 months of a calendar year and applied to the deductible for that year may be carried over and applied to the deductible for the next calendar year.

The Current Medicare Survey includes persons in the final annual estimates even though they may have actually been in the sample only part of the year. For example, persons aging into the universe in December of the survey year are considered a part of the sample even though only one month's data may be collected. Likewise, a person who either terminates medical insurance or dies at the beginning of the survey year is included in the cumulative estimates for the complete year even though interviewing stops.

Limitations of the Data

The Current Medicare Survey excludes charges made and services provided for hospital patients by hospital-based pathologists and radiologists which are reimbursable under the medical insurance program because the hospital patient is frequently unaware of them and may not be billed separately for them. This exclusion results in some understatement of total charges. General under-reporting as a result of faulty recall by the aged may also depress charges but the use of a diary form helps to minimize this problem because the enrollee is encouraged to record his use of medical care services in advance of the interviewer's visit.

The reimbursable charges derived from the Current Medicare Survey are, of course, estimates. Reimbursement for covered services is predicated upon the individual's meeting the $50 deductible. Each person's deductible status in the survey is calculated as charges occur. We also assume that all charges are "allowable" as determined by the Medicare carrier, who has this responsibility and makes payment. Finally, the Survey uses the term "potentially reimbursable" because there is no certainty that all persons entitled to reimbursement will, in fact, file for benefits. This estimating procedure leads perhaps to some overestimation of reimbursable charges.

While a 15-month time period for the monthly interviews of a panel of eligible persons offers many advantages, it is also realized that the person possibly becomes better acquainted with Medicare so that the data from the sample could become biased. However, Census interviewers are instructed not to answer any questions about the Medicare program; they refer persons with genuine questions to a Social Security district office.

Before discussing some of the new findings from the survey on the use by the aged of medical care services and their relation to economic and social characteristics, it is essential to note that the data include only those services covered under the medical insurance program. Charges and services covered by the hospital insurance program are excluded but the services of physicians during a hospital stay are included. Also excluded are charges for the services of noncovered practitioners such as dentists and chiropractors. Some data on prescription drugs will be presented separately from the data for covered services under the program.

It should also be mentioned that unknowns on some demographic characteristics constituted about 10 percent of the sample in 1967. In that year, the demographic information was collected near the end of the year thereby eliminating some persons who had been enrolled earlier in the year but had terminated coverage due to death, or disenrollment, or had moved out of the sampling area. In subsequent years these questions were asked near the beginning of the survey period which will result in an improved response rate. In addition to nonresponses, many who answered most questions could not or would not supply complete income information. Data on family income were reported in full for about three-fifths of all persons who were ever enrolled in 1967. Proportions cited by income, therefore, only apply to those who reported on income, and do not represent the entire enrolled population.

Previously published data from the Current Medicare Survey has examined the use of medical services by age, sex, and region. This report represents the first time that data have been presented that relate the use of medical services to a larger number of social and economic characteristics. In the data that follow, all differences discussed are significant at the .05 level. Within the allotted time, I can only highlight a few findings. The tables we have distributed will enable you to follow our discussion.

In 1967--the first full year of Medicare--there were about 19 million persons enrolled at some time during the year in the supplementary medical insurance program. This represented about 93 percent of all persons entitled to the hospital insurance portion of Medicare at any time during the year.

This ever-enrolled population (those enrolled at any time during the year) is used as a base for determining the proportion of persons who used services during the year.

Use of Covered Medical Services

In calendar year 1967 about 15 million persons, or almost four-fifths of all enrollees used some covered medical services. Total charges of almost $2.2 billion were incurred during the year, representing an annual average of $152 per person using medical services. Various types of services are covered under the medical insurance program including physician visits in and out of the hospital, services of other medical personnel such as nurses, physical therapists, and those providing ambulance services and other medical services and supplies. All types of medical services numbered 238 million in 1967, of which 221 million were physicians' visits. Physicians' visits are defined as the visits or services of physicians to patients, performed in the hospital and in out-of-hospital settings (i.e. extended care facilities, clinics, offices, private residences, etc.).

The number of physician visits averaged 16 per person using this service. On a per person enrolled basis, the average is 14 visits per person. For those familiar with the data on physician visits from the health interview survey conducted by the National Center for Health Statistics (NCHS), this figure would appear to be much too high. A total of 6 physician visits per person aged 65 and over is reported by NCHS for the year ending June 30, 1967. 1/ Most of the difference between the two numbers--14 visits per person from the Current Medicare Survey and 6 visits per person from the National Health Survey--is that the former counts the surgical and medical visits of physicians in hospitals and nursing homes whereas the latter omits such visits from their total count. When the inhospital and nursing home visits are omitted from the total reported by the Current Medicare Survey, the number is reduced substantially and compares closely to the National Health Survey figure. 2/

Let's look now at the use of and charges for covered medical care services in terms of some of the demographic, economic, and social characteristics of the aged persons as presented on table 1.

Age, Race, and Sex. By age, the proportion of the population using covered medical services increased with advancing age from 75 percent for persons aged 65-69 to 82 percent for persons aged 75 years and over. There was also a rising trend in the average number of physicians' visits per person served--14 visits for persons aged 65-74 and 18 visits for persons aged 75 and older. These differences are not unexpected since the aging process is generally coupled with a rise in chronic health conditions which require medical attention.

By sex, a larger percentage of women (82 percent) than of men (75 percent) used medical care services during the year, but differences in average charges and the average number of physicians' visits were not significant.

For race, it is interesting that there was no significant difference in the proportion of persons using services. Average charges and visits per person using services, however, were significantly higher for white persons than for persons of all other races. Average charges for the former were almost twice that for the latter group--$157 compared with $86. The average number of physician visits per person using services was also significantly higher for white persons--16 visits compared with 12 visits for persons of all other races. The larger number of visits for white persons partly accounts for the higher charges.

Health Limitations. The impact of poor health on use of medical services is clearly demonstrated by the data on health limitations of the sample population. Twice a year the sample person is asked to rate his health in terms of his mobility inside and outside of his living quarters. The proportion of persons confined to bed (not shown separately on table 1), and who used services is about one-third higher than for persons who had no health limitations. Seventy-four percent of persons with no limitations used some medical service while 99 percent of persons confined to bed used services. It is clear and not surprising that persons with no health limitations are least likely to use medical services, while those at the other end of the severity of limitation spectrum are the most likely to seek medical care.

Health limitations also have a significant impact on the charges and frequency of use of services. For persons confined to bed or house, the average number of physician visits per person served was about $3\frac{1}{2}$ times that of persons who reported no health limitation--38 visits compared with 11 visits. The large number of physician visits per person using services who report severest health limitations account for the very high average for this group. Average charges per person confined to bed or house was $338 during 1967. This amount decreased proportionately with the severity of limitation to $119 for those reporting no health limitations.

You will note that the response rate is high for this question. The question was asked only once in 1967, at the end of the year. It is clear that those not responding in 1967 had died prior to the interview and these persons might have had health limitations resulting in a high rate of use of medical services--91 percent of the persons not reporting health limitation used medical services in 1967, a rate of use comparable to that for persons reporting the severest limitations.

Work Status and Family Income. An individual's employment status influences the amount of income and other resources which accrue to the family unit. Particularly for persons 65 years and older, this factor also may be a partial indication of the ability of the individual to work. While many factors must be considered, including the availability of, and the desire to work, it is probably true that people with major health problems are less likely to be employed. The data on use of services by employment status confirm this hypothesis. Persons who worked full or part time had a lower rate of utilization and had fewer physicians' visits per person served than did those who did not work. Almost four-fifths who did not work during the year used medical services, compared with 71 percent for part-time workers and 67 percent for full-time workers. The average number of physicians' visits for those who did not work was about $1\frac{1}{2}$ that for workers.

Welfare Status. For purposes of the survey, persons were classified as receiving some welfare if--(1) the central records at the Social Security Administration showed that a State where the person resided had agreed with the Administration to pay the premium for the individual or (2) in response to a question in the survey about sources of income, a person listed some public assistance or other welfare payments or (3) he indicated when reporting a medical service that welfare would pay either the entire amount or, at least, the part not covered by Medicare.

About 17 percent of all medical insurance enrollees received some welfare services during 1967. Welfare recipients had higher utilization rates and a higher average number of physicians' visits per person served, although differences in average charges per person served were not significant.

Region and Size of Community. There were slight variations reported in the use of and charges for medical services among the four Census regions. There were differences between the North Central region and the West in the percent utilizing services and the average charges per person using services--77 percent for the North Central region compared with 82 percent for the West. Average charges per person served were highest in the West, while the average number of physicians' visits per person served was highest in the Northeast. Differences among the regions in the use of and charges for medical care services reflect differences in age composition, illness rates, type of illness, and level of charges for medical care.

Size of community apparently is an important factor in the use of medical services. Persons residing in urban areas had higher utilization rates, higher average charges, and more physicians' visits per person served than persons in rural areas. The availability of a larger number and variety of medical services perhaps contributes to the higher use of services and the higher physician visits for persons in urban areas.

Source of Payment

Medicare is not responsible for the first $50 of covered charges incurred by the enrollee nor for more than 80 percent of charges after this deductible is paid. These amounts may be paid by the enrollee or the spouse, by private insurance plans, by welfare, or by combinations of these and other sources.

While the relative importance of payment sources can be measured in various ways, we have chosen to classify enrollees in mutually exclusive categories which may represent one or a combination of payment sources for all services in the period. Only selected categories have been tabulated and shown on table 2. In 1967, about half the enrollees using services indicated that they or their spouses paid all deductible and coinsurance charges. One-fifth of the enrollees indicated the participation of a private insurance plan either alone or with some payment by the person. One-tenth indicated that welfare paid all of the coinsurance and deductible in combination with some self-payment. The remaining fifth of the enrollees used other combinations of sources of payment.

As expected, welfare recipients as a group are likely to have welfare pay a major share of their expenses not covered by Medicare. It is not surprising, therefore, to find that 50 percent of the enrollees indicated that welfare had paid all or part of the deductible and coinsurance amounts. It is also clear that this group has very little private health insurance--only 3 percent indicated private insurance as a payment source.

Another group indicating high use of welfare as a source of payment is the institutional population--30 percent had their deductible and coinsurance paid for by welfare.

Younger persons are more likely than older persons to pay their own bills for the deductible and coinsurance amounts not covered by the medical insurance program. Younger persons are also somewhat more likely to have private insurance as a payment source.

There are differentials by race in payment source. While about the same proportion of white persons and persons of all other races indicated self or spouse as the sole payment source for deductibles and coinsurance, there is an inverse relationship between insurance and welfare payments for these two groups.

By level of education, again there are no significant differences in the proportion of persons paying their own deductibles and coinsurance bills. But there are differences between insurance and welfare payments. There is more reliance on private insurance and less on welfare as a source of payment with higher levels of education. Undoubtedly, this pattern of decreasing reliance on welfare as education level rises is interrelated with income.

As expected, the proportion of persons who rely upon welfare was inversely proportional to reported family income. Persons with lower incomes are less likely than those with higher incomes to have used private insurance as a means to cover deductible and coinsurance amounts.

With respect to health limitations, when insurance or welfare is a payment source there are differences between persons without limitations and all others. The proportion of enrollees who pay the deductible and coinsurance amounts themselves, however, increases from a low of 28 percent for persons confined to the bed or house, to 42 percent for those with other limitations, and, finally, it reaches 55 percent for persons with no limitations.

It is interesting to compare for persons with and without health limitations the proportion relying on themselves as the sole payment source with the average charges per person served shown in table 1. As average charges rise from $119 for persons with no limitations to $338 for persons confined to bed or the house, the percent who use themselves solely as the payment source falls from 55 percent to 28 percent. In addition to higher charges for these persons, it is possible that many of the persons who were limited in physical activity in 1967 also may have suffered a reduction in income and/or assets and, therefore, became less able to rely upon themselves as a payment source.

Use of Prescription Drugs

While the medical insurance program does not cover the cost of prescription drugs, charges for drugs constitute a sizeable portion of a person's total medical expenses.

About 14.8 million persons, representing almost four-fifths of the medical insurance enrollees, had prescriptions filled during 1967. This represented an average of 14.1 prescriptions per person acquiring drugs. Charges per prescription averaged $3.96, or $56 per person acquiring drugs during 1967.

About the same proportion of persons used prescription drugs and medical services (78 percent), but average charges for drugs per person served was about one-third of the average charges for medical services. Highlighting some of the data, it is interesting to note that while differences existed between urban and rural areas in regard to the use of medical care services, no such differences are indicated in the use of and charges for prescription drugs among rural and urban residents. This probably reflects the general availability of drugs regardless of size of community. On the other hand, rural areas generally do not have the more sophisticated and costly medical services which are available in urban areas, possibly accounting for the lower utilization rates and charges in rural areas.

As with the use of medical services, there were significant differences between persons confined to bed or house and persons with no limitations, in the use of drugs, average charges, and average number of drugs. Workers were less likely to use prescription drugs than nonworkers.

Average charges and the average number of prescriptions per person served also were higher for nonworkers than for workers. Finally, relatively more persons with some welfare services acquired prescription drugs than persons not receiving welfare. More women acquired drugs than men.

Summary

In summary, the Current Medicare Survey was initiated to provide current information on the utilization of and charges for medical care since lags in processing bills prevented the regular system for providing these data on a current basis. It has also proven useful in collecting types of information not obtainable from other sources, such as the use of noncovered services, and utilization and charges at levels below $50 as well as above.

In examining the data for 1967, a rough profile of medical care users emerges. Persons more likely to utilize services at higher rates tend to be older, confined to the bed or house, and to reside in urban areas. They are further inclined to be nonworkers, be alone in the household, and have received some welfare services in the year. Persons using services at lower rates tend to be relatively younger, have no health limitation, be rural residents, and are more likely to be employed.

To pay deductible and coinsurance amounts, individuals frequently rely on themselves, a private health insurance plan, and welfare as primary sources of payment. These data tend to support a fact commonly accepted--that often the persons who must use medical care the most, have the least resources to pay for it.

Although prescription drugs are not covered under the medical insurance program, data on use of and charges for them are collected in the survey. About four-fifths of all enrollees acquired prescription drugs, the same proportion as used covered medical care services. Older persons, those with some health limitations, and persons not working were more likely to acquire drugs.

1/ "Volume of Physician Visits, United States, July 1966-June 1967," National Center for Health Statistics, Series 10, Number 49.

2/ See "Current Medicare Survey Report, Medical Insurance Sample, January-December 1968," R&S Health Insurance Statistics, Social Security Administration, CMS-12, January 27, 1970.

TABLE 1.--Estimated use of and charges for covered medical services under the supplementary medical insurance program (SMI), by selected characteristics, 1967

Characteristic	Persons ever enrolled: Total (in thousands)	Persons ever enrolled: Using SMI services: Number (in thousands)	Persons ever enrolled: Using SMI services: Percent of total	SMI charges: Total (in millions)	SMI charges: Average per person using services	Physicians' visits (ambulatory and hospital): Number (in thousands)	Physicians' visits (ambulatory and hospital): Average per person using services
Total...........	18,960	14,946	79	$2,159	$152	220,972	16
Age:							
65-69.............	5,933	4,448	75	552	136	55,961	14
70-74.............	5,528	4,350	79	637	151	62,333	15
75 and over........	7,499	6,148	82	971	165	102,678	18
Sex:							
Men................	8,090	6,039	75	893	157	92,263	16
Women..............	10,870	8,907	82	1,267	149	128,710	15
Race:							
White..............	17,248	13,605	79	2,032	157	204,081	16
All other..........	1,462	1,134	78	91	86	13,009	12
Unknown............	251	207	82	37	219	3,883	23
Education:							
Less than 7 years..	4,622	3,503	76	429	125	48,154	14
7-8 years..........	4,920	3,795	77	545	147	51,987	14
9 years and over...	7,007	5,456	78	829	156	72,724	14
Not reported.......	2,411	2,191	91	357	206	48,107	28
Health limitations:							
Confined to bed or house...........	880	825	94	277	338	30,980	38
Other limitations..	2,190	1,931	88	403	211	41,078	22
No limitation......	13,782	10,266	74	1,191	119	114,044	11
Not reported.......	2,109	1,923	91	289	197	34,870	24
Marital status:							
Nonmarried.........	8,582	6,757	79	923	140	99,561	15
Married............	8,896	6,814	77	1,038	158	95,244	14
Not reported.......	1,482	1,375	93	198	196	26,168	26
Living arrangement:							
Institution........	869	812	93	272	343	39,080	49
Living alone.......	4,076	3,162	78	386	124	38,837	12
Living with spouse only......	6,914	5,274	76	775	151	69,132	14
Living with other..	5,291	3,997	76	498	128	46,485	12
Not reported.......	1,809	1,700	94	229	180	27,438	22
Household size:							
1 person...........	4,879	3,907	80	646	168	76,598	20
2 persons..........	8,744	6,670	76	918	142	83,823	13
3 or more persons..	3,524	2,663	76	367	142	33,013	13
Not reported.......	1,813	1,704	94	229	180	27,539	22

TABLE 1.--Estimated use of and charges for covered medical services under the supplementary medical insurance program (SMI), by selected characteristics, 1967--Continued

Characteristic	Persons ever enrolled			SMI charges		Physicians' visits (ambulatory and hospital)	
	Total (in thousands)	Using SMI services: Number (in thousands)	Using SMI services: Percent of total	Total (in millions)	Average per person using services	Number (in thousands)	Average per person using services
Work status:							
None................	12,537	9,914	79	1,449	148	149,164	15
Part time...........	1,631	1,164	71	115	103	10,186	9
Full time...........	1,857	1,236	67	140	122	11,044	10
Not reported........	2,935	2,632	90	455	211	50,578	23
Family income:							
Less than $3,000....	6,130	4,674	76	632	137	70,150	15
3,000-4,999.........	2,539	1,886	74	258	141	24,605	13
5,000 and over......	2,746	2,072	75	305	152	26,491	13
Not reported........	7,545	6,314	84	965	168	99,727	17
Private health insurance coverage:							
No plan at all......	8,284	6,210	75	876	144	95,974	16
Hospital care only..	1,333	1,046	78	147	145	13,025	13
Hospital and surgical care only...	1,993	1,611	81	237	150	20,748	13
Hospital, surgical and physicians' care..............	4,849	3,809	79	570	154	51,194	14
Other combinations..	307	268	87	32	123	2,984	11
Not reported........	2,196	2,001	91	297	192	37,047	24
Welfare status:							
No welfare..........	15,696	12,077	77	1,682	147	160,008	14
Some welfare........	3,265	2,869	88	477	174	60,964	22
Region:							
Northeast...........	5,170	4,130	80	587	150	67,654	17
North Central.......	5,587	4,313	77	569	139	62,203	15
South...............	5,386	4,190	78	590	147	57,280	14
West................	2,817	2,313	82	414	189	33,836	15
Size of community:							
Urban...............	16,641	13,242	80	1,973	158	200,460	16
Rural...............	2,315	1,699	73	186	112	20,491	12

1/ Less than $500,000.

Note: Small numbers are subject to relatively large sampling variability. They are shown here only to assist the users of data should they wish to form aggregates, and not because they possess reliability in and of themselves.

TABLE 2.--Estimated number and percent of supplementary medical insurance enrollees using covered medical services, by source of payment of deductibles and coinsurance amounts and selected characteristics

Characteristic	Enrollees using services (in thousands)	Percent by source of payment				
		Total	Self or spouse	Insurance 1/	Welfare 1/	Other 2/
Total................	14,946	100	49	20	10	21
Age:						
65-69....................	4,448	100	53	23	7	17
70-74....................	4,350	100	51	20	9	20
75 and over..............	6,148	100	44	17	12	27
Sex:						
Men......................	6,039	100	50	20	9	21
Women....................	8,907	100	48	20	10	22
Race:						
White....................	13,605	100	49	21	9	21
All other................	1,134	100	48	7	19	26
Unknown..................	207	100	36	14	18	32
Education:						
Less than 7 years.........	3,503	100	52	10	15	23
7-8 years................	3,795	100	51	23	9	17
9 years and over..........	5,456	100	53	25	4	18
Not reported.............	2,191	100	29	18	17	36
Health limitations:						
Confined to bed or house..	825	100	28	14	17	41
Other limitations.........	1,931	100	42	14	15	29
No limitation............	10,266	100	55	22	7	16
Not reported.............	1,923	100	30	19	15	36
Marital status:						
Nonmarried...............	6,757	100	47	18	12	23
Married..................	6,814	100	54	23	6	17
Not reported.............	1,375	100	29	16	19	36
Living arrangement:						
Institution..............	812	100	16	10	30	43
Living alone.............	3,162	100	50	20	12	18
Living with spouse only...	5,274	100	55	25	4	16
Living with other.........	3,997	100	55	16	8	21
Not reported.............	1,700	100	29	19	16	36
Household size:						
1 person.................	3,907	100	43	18	16	23
2 persons................	6,670	100	55	23	5	17
3 or more persons.........	2,663	100	53	16	8	23
Not reported.............	1,704	100	29	19	16	36

TABLE 2.--Estimated number and percent of supplementary medical insurance enrollees using covered medical services, by source of payment of deductibles and coinsurance amounts and selected characteristics, 1967--Continued

Characteristic	Enrollees using services (in thousands)	Percent of source of payment				
		Total	Self or spouse	Insurance 1/	Welfare 1/	Other 2/
Work status:						
None	9,914	100	50	19	10	21
Part time	1,164	100	60	22	4	14
Full time	1,236	100	59	24	1	16
Not reported	2,632	100	34	20	14	32
Family income:						
Less than $3,000	4,674	100	53	14	14	19
3,000-4,999	1,886	100	51	26	7	16
5,000 and over	2,072	100	50	23	4	23
Not reported	6,314	100	44	22	9	25
Private health insurance coverage:						
No plan at all	6,210	100	57	4	20	19
Hospital care only	1,046	100	57	22	2	19
Hospital and surgical care only	1,611	100	52	30	1	17
Hospital, surgical and physicians' care	3,809	100	42	42	1	15
Other combinations	268	100	41	36	6	17
Not reported	2,001	100	30	18	15	37
Welfare status:						
No welfare	12,077	100	55	24	---	21
Some welfare	2,869	100	20	3	50	27
Region:						
Northeast	4,130	100	43	23	10	24
North Central	4,313	100	50	26	7	17
South	4,190	100	56	14	8	22
West	2,313	100	45	14	17	24
Size of community:						
Urban	13,242	100	47	21	10	22
Rural	1,699	100	64	13	10	13

1/ Either alone or in combination with payment by self or spouse.
2/ Includes other combinations of sources of payment and unknowns.

Note: Small numbers are subject to relatively large sampling variability. They are shown here only to assist the users of data should they wish to form aggregates and not because they possess reliability in and of themselves.

TABLE 3.--Estimated use of and charges for prescription drugs, supplementary medical insurance enrollees, by selected characteristics, 1967

Characteristic	Persons acquiring prescription drugs		Number of prescriptions		Charges		
	Total (in thousands)	Percent of total enrolled	Total (in thousands)	Per person acquiring drugs	Total (in millions)	Per person acquiring drugs	Per prescription
Total	14,780	78	199,007	14.1	$788	$56	$3.96
Age:							
65-69	4,430	75	51,997	12.8	211	52	4.05
70-74	4,348	79	58,261	13.7	230	54	3.94
75 and over	6,002	80	88,748	15.3	347	60	3.91
Sex:							
Men	5,914	73	72,723	13.0	297	53	4.08
Women	8,865	82	126,284	14.8	491	58	3.89
Race:							
White	13,481	78	185,351	14.4	737	57	3.98
All other	1,110	76	11,227	10.7	40	39	3.59
Unknown	189	75	2,428	15.8	10	67	4.22
Education:							
Less than 7 years	3,613	78	51,438	14.6	191	54	3.70
7-8 years	3,857	78	52,487	13.9	204	54	3.89
9 years and over	5,486	78	74,191	13.9	311	58	4.19
Not reported	1,823	76	20,891	14.3	83	56	3.95
Health limitations:							
Confined to bed or house	766	87	19,804	26.1	81	107	4.09
Other limitations	1,981	90	37,945	19.5	146	75	3.84
No limitation	10,454	76	124,660	12.2	496	49	3.98
Not reported	1,580	75	16,598	13.6	65	53	3.90
Marital status:							
Nonmarried	6,783	79	94,498	14.3	359	54	3.80
Married	6,892	77	92,796	13.9	382	57	4.12
Not reported	1,105	75	11,714	14.2	47	57	4.03
Living arrangement:							
Institution	686	79	13,797	20.6	57	86	4.15
Living alone	3,193	78	40,163	12.8	156	50	3.88
Living with spouse only	5,387	78	72,729	13.9	297	57	4.09
Living with other	4,139	78	57,933	14.4	221	55	3.81
Not reported	1,375	76	14,385	13.8	57	54	3.94
Household size:							
1 person	3,829	78	53,373	14.2	211	56	3.95
2 persons	6,809	78	92,285	13.9	372	56	4.03
3 or more persons	2,762	78	38,904	14.5	148	55	3.81
Not reported	1,379	76	14,445	13.8	57	54	3.94

TABLE 3.--Estimated use of and charges for prescription drugs, supplementary medical insurance enrollees, by selected characteristics, 1967--Continued

Characteristic	Persons acquiring prescription drugs		Number of prescriptions		Charges		
	Total (in thousands)	Percent of total enrolled	Total (in thousands	Per person acquiring drugs	Total (in millions)	Per person acquiring drugs	Per prescription
Work status:							
None	10,119	81	149,456	15.0	589	59	3.94
Part time	1,194	73	12,024	10.5	48	42	4.00
Full time	1,189	64	10,594	9.7	44	40	4.17
Not reported	2,277	78	26,931	14.1	107	56	3.97
Family income:							
Less than $3,000	4,796	78	66,965	14.2	257	55	3.84
3,000-4,999	1,962	77	25,946	13.6	100	52	3.86
5,000 and over	2,133	78	30,087	14.5	125	60	4.16
Not reported	5,889	78	76,009	14.0	305	56	4.02
Private health insurance coverage:							
No plan at all	6,415	77	86,860	13.8	336	54	3.87
Hospital care only	1,090	82	14,916	14.0	58	54	3.87
Hospital and surgical care only	1,585	80	23,054	14.9	93	60	4.01
Hospital, surgical and physicians' care	3,767	78	52,150	14.3	215	59	4.12
Other combinations	260	85	4,150	16.2	17	67	4.15
Not reported	1,662	76	17,878	13.7	70	53	3.89
Welfare status:							
No welfare	12,102	77	157,672	13.7	634	55	4.02
Some welfare	2,677	82	41,334	16.0	154	60	3.73
Region:							
Northeast	4,024	78	48,746	12.7	186	48	3.81
North Central	4,250	76	56,521	13.9	230	56	4.07
South	4,281	79	63,717	15.6	235	58	3.69
West	2,224	79	30,023	14.2	137	65	4.57
Size of community:							
Urban	12,984	78	173,592	14.0	691	56	3.98
Rural	1,791	77	25,376	14.5	97	56	3.83
Not reported	4	100	38	9.1	1/	30	3.34

1/ Less than 500,000.

Note: Small numbers are subject to relatively large sampling variability. They are shown here only to assist the users of data should they wish to form aggregates and not because they possess reliability in and of themselves.

DISCUSSION

Regina Loewenstein, Columbia University

The paper just presented is an excellent talk about an important ongoing survey.

The Current Medicare Survey gives trend data about services and charges before this information could be available from records of claims. There would be delays in obtaining these facts because of time-lags in filing of claims by respondents and in payments of claims by carriers. In addition, the Current Medicare Survey obtains information about charges incurred for which claims were not filed because the deductible was not met or because claims were not filed. Charges incurred for selected services not covered by the Medicare law are also included. The various types of data from the Current Medicare Survey are valuable for program planning.

The methodology of the Current Medicare Survey is excellent. One outstanding feature is the use of the diary method to facilitate recall about utilization and charges during the past month. The talk did not discuss whether or not records of respondents are used. Interviews of the same respondents for fifteen months is a clever way to obtain data about meeting the deductible, considering the carry-over feature of the Medicare law.

The reliability of reporting in-institution visits was not discussed. When persons are very ill they often do not know the frequency of visits by doctors. The use of records of Medicare payments or of doctors' bills probably improved these data.

This talk is the first presentation of the relationship of utilization and charges to social and economic characteristics, other than age, sex and region, of beneficiaries of Supplementary Medical Insurance.

Since utilization increases with age and since many other characteristics are related to age, it is important to consider the relationship of each characteristic with age when the findings are interpreted. For example, age-sex specific indices would indicate whether higher percentages and means for women are related to the older age of women and/or higher utilization by women than men of the same age.

Similarly, the relationship of income to other characteristics would be valuable.

There is a question about the definition of the term "Using SMI services" in Table I. Does this refer only to services with charges incurred that are covered by SMI, or does it also include medical services without charges incurred, such as care by clinics of local health departments and mobile units?

In conclusion, this paper was a clear and precise discussion about an important survey with valuable information about the relationship of utilization and charges for medical services to many social and economic characteristics of Medicare beneficiaries.

A MATRIX MODEL OF POPULATION REDISTRIBUTION

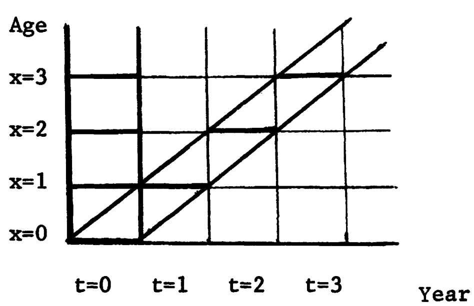
W. L. Li
Ohio State University*

This paper employs matrix theories in constructing a probabilistic model of interregional population redistribution. The analysis can be viewed as an extension of the previous contribution made by Leslie (1945), Keyfitz (1964), Tarver (1965), Roger (1968), and many others. In this paper, regional population is treated as an open system and the interregional flows of population are considered. By drawing on the merits and eliminating the defects of various previous models, this paper can present, it is hoped, a more realistic projection of regional distribution.

Population Projection as a Cohort Approach

Demographic projections are typically made by calculating the future consequences of past conditions. In this sense, a population projection is merely one way of regarding the fertility, mortality, and migration of a given period in the past. Its usefulness does not necessarily depend on whether it constitutes a good prediction, for it can be interesting to analyze a hypothetical trend based on the assumption that given rates would continue into the future.

To illustrate a general demographic analysis, a Lexis diagram is perhaps the most easily comprehensible device (see Pressat, 1961). In the diagram time is along the horizontal axis, and age along the vertical. A group of individuals that starts at x = 0 at the moment of birth in time t, and is followed up to a terminating point corresponding to both their age and time of death, is generally referred as a cohort.



As the analysis of actual cohort requires many years of observation, a very useful construct in demographic analysis is the synthetic cohort. This is formed by taking the cross-sectional experience of one age category of individuals and applying it as a constant to an actual cohort. The experience could be mortality, fertility, migration, school enrollments, or any other social events that can be treasured numerically. If mortality is analyzed the construct is a life table; if fertility, mortality, and age distribution, it could be a stable population model. In other words, a Lexis diagram illustrates a fundamental mode of demographic analysis that is the consequence of a particular social process if an actual cohort followed exactly the same series of probabilities that various cross-sectional age groups at a given date experienced. It is based on the postulates that population projections should be regarded as a legitimate exercise, and that demographers ought to seek "more accurate" population projection by basing them on more realistic assumption.

The Leslie-Keyfitz Model

In the Leslie-Keyfitz model, the procedure of projecting a population is separable from its original age distribution, following the principle sometimes referred as the Markov process. The projection operator is a matrix whose elements reflect the range in mortality and fertility of the given population in a given year. The matrix can be repeatedly multiplied by an age distribution, and it can carry the numbers at the several ages into successive time periods. In other words, to denote the age distribution after n periods from the starting point, one raises the matrix to the nth power and then multiplies it by the initial age distribution.

As an illustration, let us denote:

K_x^t= the number of females alive in age group x at time t.

S_x= the survival ratio; that is, the probability that a female aged x at time t will be live to age (x + 1) at time (t + 1).

F_x= the fertility rate of a surviving female from aged x to (x + 1).

For simplicity, let us illustrate the principal with a population of only three age groups, which can be denoted as x = 1, 2, and 3. Then, the Leslie-Keyfitz model can be expressed as follows:

$$\begin{bmatrix} S_o \cdot F_1 & S_o \cdot F_2 & S_o \cdot F_3 \\ S_1 & 0 & 0 \\ 0 & S_2 & 0 \end{bmatrix} x \begin{bmatrix} K_1^t \\ K_2^t \\ K_3^t \end{bmatrix} = \begin{bmatrix} K_1^{t+1} \\ K_2^{t+1} \\ K_3^{t+1} \end{bmatrix} \qquad (1)$$

Or it can simply expressed as:

$$S \cdot K_t = K_{t+1} \qquad (2)$$

* This research has been supported in part by the National Sciences Foundation under contract NSF - GS2630.

The projection matrix S is obtained from the cross-sectional data of a population at a given time. It is treated as constant in order to project the population to its future size.

The empirical implementation of this matrix, as suggested by Keyfitz (1968: 30-31), is rather simple for those diagonal elements, S_1 and S_2. For instance,

$$S_x = L_{x+1}/L_x \qquad (3)$$

where L_x is adopted from the life-table notation of the survivorship function. However, it is relatively more complicated to implement the elements in the first row of the matrix. S_o denotes the survival function of those newly born during the first years. Thus,

$$S_o = L_o/l_o \qquad (4)$$

and F_x can be implemented by:

$$F_x = 1/2 \left[f_x + (L_{x+1}/L_x)\cdot f_{x+1}\right] \qquad (5)$$

where f_x is the age specific birth rate at age x.

Two Interregional Models

The Leslie-Keyfitz model is dynamic but essentially aspatial, based on the assumption that population is a closed system with no migration. The model has to be modified to be useful in studies of interregional distribution of a population.

A Markov-chain model similar to Equation (2) was used by Tarver and Gurley to analyze the 1955-60 interregional mobility in order to make long-range projections of the populations of the nine census divisions of the United States. In their model the projection matrix is implemented by the transition probabilities calculated from data on the interdivisional migration streams, Let us call it a M matrix, where the element m_{ij} denotes the probability of moving from region i to region j. Thus, in a n-region model, the matrix is:

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix} \qquad (6)$$

Then, the population in various regions is projected by raising the power of this matrix and multiplied by the regional populations in 1960.

Despite their success in analyzing a process of theoretical interest, Tarver and Gurley (1965: 139) expressed their dissatisfaction with the stringent assumptions they had to make. They concluded:

> The Markov process is a rather elegant model to specify the future population distribution of various geographic areas which will result, providing the migration patterns observed in a recent interval continue indefinitely. It is concise, for it includes only one of the three components (births, deaths, and migrants) affecting future population numbers. One limitation is that the model excludes children born and deaths after the observed period, since it applies to a constant total population. A modification in the model to allow for varying natality and mortality processes in each area would make the technique much more realistic, for natural population increase tends to be highest in areas having the highest out-migration tendencies.

A model proposed by Andrei Rogers (1968) can be viewed as an attempt to remedy these limitations of Tarver and Gurley's model. It is relatively a more realistic approach in that birth and death are incorporated with migration. The three components are integrated into what he calls a "multi-regional matrix growth operator." The procedure of constructing the model is basically as follows:

1). The Leslie-Keyfitz approach is extended by adding an in-migration component to their model. Thus,

$$K_{t+1} = S\cdot K_{t+1} + N_t \qquad (7)$$

where Nt is a vertor denoting the number of migrants into each region.

2). The in-migration component is obtained by:

$$N_t = M\cdot K_t \qquad (8)$$

where M is the migration transition matrix, as defined by Tarver and Gurley.

3). Substituting Equation (8) in Equation (7), an interregional matrix growth operator G is obtained:

$$K_{t+1} = S\cdot K_t + M\cdot K_t = G\cdot K_t \qquad (9)$$

The operator G is a supermatrix with the number of regions as its dimension, and within each region there is a submatrix with the number of age groups as its dimension. For instance, in a model of two regions (e.g., California and rest of the United States) and of three age groups within each region, the supermatrix can be expressed as:

$$\begin{array}{cc} & \begin{array}{cccccc} & \text{Region 1} & & & \text{Region 2} & \end{array} \\ \begin{array}{c} \\ \text{Region 1} \\ \\ \\ \text{Region 2} \\ \\ \end{array} & \left[\begin{array}{ccc:ccc} S_o\cdot F_1 & S_o\cdot F_2 & S_o\cdot F_3 & 0 & 0 & 0 \\ S_1 & 0 & 0 & m_1 & 0 & 0 \\ 0 & S_2 & 0 & 0 & m_2 & 0 \\ \hdashline 0 & 0 & 0 & S_oF_1 & S_oF_2 & S_oF_3 \\ m_1 & 0 & 0 & S_1 & 0 & 0 \\ 0 & m_2 & 0 & 0 & S_2 & 0 \end{array}\right] \end{array} \qquad (10)$$

$$X \begin{bmatrix} K_1^t \\ K_2^t \\ K_3^t \\ \hline K_1^t \\ K_2^t \\ K_3^t \end{bmatrix} = \begin{bmatrix} K_1^{t+1} \\ K_2^{t+1} \\ K_3^{t+1} \\ \hline K_1^{t+1} \\ K_2^{t+1} \\ K_3^{t+1} \end{bmatrix} \qquad \text{(10 cont.)}$$

In structure Roger has made a mechanical synthesis of the Leslie-Keyfitz model, on the one hand, and Tarver and Gurley's model, on the other. The regional populations in each age group are estimated by using the original regional populations as distribution vector, post-multiplying it to the fertility-mortality matrix S and migration-transition matrix M, respectively, and summing the results of the two multiplications. An obvious weakness of this calculation is that it does not allow for the interaction between the numbers of survivors and of migrants. The population of any region, it is true, comprises two components, survivors and in-migrants, but survivors also migrate during the period, and migrants follow a certain mortality schedule. Moreover, it would seem that Rogers's model takes into consideration only the input of migration, which is relatively acceptable in a two-region model. But in a multi-regional model Rogers's formulation is rather cumbersome. It is theoretically more appropriate to consider simultaneously the in-migration to and the out-migration from each region.

An Alternative Model

It appears to be preferrable to adopt a multiplicative interregional model, rather than the additive model that Rogers suggests. The idea has been discussed by Feeney (1970) and independently by Li (1970). Following the Leslie-Keyfitz formulation, an interregional model can be constructed that is structurally based on the cohort-survival analysis, as depicted by a Lexis diagram--in other words, on the postulate that whenever a cohort reaches a certain age, it will assume that age's rates of fertility, mortality, and interregional migration. Consequently, the population at any given time can be estimated by summing the various cohorts at that time.

In distinction from Rogers's formulation, our supermatrix is arranged according to the number of age groups, and any element in the supermatrix is a submatrix whose dimensionality is determined by the number of regions. The elements in the first row of the supermatrix designate the probabilities that in each region any surviving non-migrant, in-migrant, and out-migrant will bear a live child. According to the multiplicative model, a migrant woman tends to follow the fertility-mortality schedule of the region of her origin, an assumption that can be substantiated by demographic evidence from a number of sources.

Thus, within a given age cohort x, the submatrix of two-region model can be mathematically expressed as follows:

$$R_x = \begin{bmatrix} S_oF_x^1 & 0 \\ 0 & S_oF_x^2 \end{bmatrix} \begin{bmatrix} m_{11}^x & m_{12}^x \\ m_{21}^x & m_{22}^x \end{bmatrix} = \begin{bmatrix} m_{11}^x S_oF_x^1 & m_{12}^x S_oF_x^1 \\ m_{21}^x S_oF_x^2 & m_{22}^x S_oF_x^2 \end{bmatrix} \qquad (11)$$

where m_{ij} is an element of the migration transition matrix; $S_oF_x$is the number of surviving births per female, as before. Also note that the sub- or super-script x denotes age, and the subscripts 1 and 2 denote the regions.

The diagonal elements of the supermatrix are constituted by the joint probabilities of migration transition and survivorship. Thus, for a given age cohort x, the two-region submatrix will be:

$$Q_x = \begin{bmatrix} S_x^1 & 0 \\ 0 & S_x^2 \end{bmatrix} \times \begin{bmatrix} m_{11}^x & m_{12}^x \\ m_{21}^x & m_{22}^x \end{bmatrix} = \begin{bmatrix} m_{11}^x S_x^1 & m_{12}^x S_x^1 \\ m_{21}^x S_x^2 & m_{22}^x S_x^2 \end{bmatrix} \qquad (12)$$

Incorporating these two components into a cohort surviving supermatrix, we obtain a model with which we can project over a given future period the regional populations by age, sex, and color. The model can be expressed as:

$$K_{t+1} = G \cdot K_t \qquad (13)$$

where G is a supermatrix that includes such submatrices as:

$$G = \left[\begin{array}{c|c|c} R_1 & R_2 & R_3 \\ \hline Q_1 & 0 & 0 \\ \hline 0 & Q_2 & 0 \end{array}\right] \qquad (14)$$

and K is a supervector with three elements (e.g., age groups) within each subvector of two elements (e.g., regions).

An Application

The model presented in Equation (13) can be tested using the U. S. data for 1960. Census and vital statistics publications give the following information, necessary for the implementation:

1). interdivisional migration transition matrix by age, sex, and color, computed from 1960 census data.

2). age-specific birth rates by color for each division, available from U. S. National Center of Health Statistics (1968).

3). survival ratios by age, sex, and color, obtained from the life table of each division.

4). population by age, sex, and color for each division, from the 1960 census.

These data represent the necessary inputs for the model. Table 1 illustrates for white females aged 20-24 the results of implementing the matrix of R_x in Equation (11). A 9 x 9 matrix is derived because there are 9 census divisions in the United States.

An application of the matrix R_x (such as shown in Table 1) to the female population yeilds estimates of total births, rather than only female births, since age-specific total birth rates are used, (rather than age-specific female birth rates). Consequently, to obtain separately the estimated female births and male births, the figures have to be adjusted with a constant sex ratio at birth. The estimated number of male births can then be substituted in the first row of the supermatrix for males.

Obviously, the model presented at Equation (13) can be carried out as an iterative process, for each division's population by age, sex, and color can be projected for every 5 years. Table 2 shows how the model can contribute to a presentation of the population redistribution among the various regions. For instance, with respect to this particular age group, 20-24, the North East Central, Mountain, and Pacific divisions are growing relatively and the other divisions are declining or, with fluctuations, remaining more or less constant.

A Suggestion

Our model does not take international migration into account. Since the United States has a relatively stable inflow of international migrants, we should add this variable to the analysis of interregional population redistribution.

Assume that immigrants follow the same pattern of fertility, mortality, and interregional flow as natives. In other words, the interregional growth operator G can be applied to both the immigrants and natives. Second, assume that the regional distribution of the immigrants is known for every age-sex-color group, and constant for every time period (which may not be too far from the truth in any country with a restricted immigration policy.)

Thus, the immigrant distribution can be represented in a supervector structurally similar to K in Equation (13). Le us denote it as I. Then the distribution of total regional population, combining the natives and immigrants, for the nth period from the starting point t will be:

$$K_{t+n} = G^n \cdot K_t + \sum_{j=0}^{n} G^j \cdot I_t \qquad (15)$$

The empirical implementation of this model is currently in progress. It is obvious that the results can yield a more realistic projection of regional population, especially in those countries where a great amount of immigration occurs.

References

Feeney, Griffith M.
1970 "Stable Age by Region Distributions," Demography 7: 341-348.

Keyfitz, Nathan
1964 "The Population Projection as a Matrix Operator," Demography 1: 57-73.

Table 1:
Interregional Growth Submatrix
White Females, 20-24

	1	2	3	4	5	6	7	8	9
1	0.195	0.009	0.004	0.001	0.007	0.001	0.002	0.001	0.006
2	0.004	0.192	0.006	0.001	0.008	0.001	0.001	0.001	0.004
3	0.001	0.003	0.224	0.005	0.005	0.002	0.002	0.002	0.007
4	0.001	0.001	0.011	0.204	0.003	0.001	0.006	0.009	0.014
5	0.002	0.005	0.009	0.002	0.184	0.004	0.003	0.002	0.005
6	0.001	0.002	0.021	0.002	0.016	0.175	0.008	0.002	0.005
7	0.001	0.001	0.005	0.006	0.005	0.003	0.214	0.006	0.010
8	0.002	0.003	0.010	0.013	0.007	0.002	0.016	0.129	0.059
9	0.002	0.003	0.008	0.007	0.006	0.002	0.008	0.015	0.185

Sources: Vital Statistics Rates in the United States, 1940-1960, Table 20, pp. 142-145, Life Tables for the Geographic Divisions of the United States: 1959-61, Tables 1 to 36. Lifetime and Recent Migration, Table 6.

Table 2:
Population Distribution by Census Division,
White Females, Age 20-24

Census Division	1965	1970	1975	1980
1	6.1	6.1	6.3	6.3
2	17.9	18.6	19.4	19.3
3	20.8	21.8	23.5	22.5
4	9.3	9.4	10.1	9.6
5	12.8	12.6	12.1	11.2
6	15.1	7.0	6.8	6.5
7	9.2	14.0	10.0	9.4
8	2.4	2.7	3.1	4.0
9	6.3	7.8	8.6	11.0
Total	100.0	100.0	100.0	100.0
Population	5,936,038	6,996,386	7,238,838	7,014,382

Sources: 1960 Census of Population: U. S. Summary, Part I, Table 59.

1968 Introduction to the Mathematics of Population, Reading, Mass.: Addison-Wesley Publishing Co.

Leslie, P. H.
1945 "On the Use of Matrices in Certain Population Mathematics," Biometrika, 33: 183-212.

Li, Wen L.
1970 "Matrix Analysis of Migration Streams," Paper presented at the Population Association of America Meetings, April, 1970.

Pressat, Roland
1961 L'analyse démographique, Paris: Presses Universitaires de France.

Rogers, Andrei
1968 Matrix Analysis of Interregional Population Growth and Distribution. Berkeley, Calif.: University of California Press.

Tarver, James D., and William R. Gurley
1965 "A Stochastic Analysis of Geographic Mobility and Population Projections of the Census Divisions in the United States," Demography 32: 134-139

U. S. Bureau of the Census
1963 1960 Census of Population: United States Summary. Washington, D.C.: U. S. Government Printing Office.

1963 1960 Census of Population: Lifetime and Recent Migration. Washington, D.C.: U. S. Government Printing Office.

U. S. National Center for Health Statistics
1965 Life Tables for the Geographic Divisions of the United States: 1959-61, Public Health Service, Publication No. 1252. Washington, D.C.

1968 Vital Statistics Rates in the United States, 1940-1960. Washington, D.C.: U. S. Government Printing Office.

DISCUSSION

Richard B. McHugh, University of Minnesota

My comments will be confined to two of the three papers: That of Professor Fredrickson and of Professor Masnick, et al. The paper by Professor Li is as new to you as to me.

Fredrickson's paper, in my view, may well prove to be of fundamental importance to mathematical demography. First, a few notational and semantic issues:

1) The standard notation for distribution and density function has been reversed.

2) The term "rate" is used synonymously with total frequency (for example, B(t) in (F) and $\rho(a, a', t)$ in (G)).

3) The phrase "random mating" may be ill advised in the present context. As employed by geneticists, panmixia has a specific probabilistic interpretation, viz, that with respect to a specific locus, the relative frequencies of mating types are determined by the product of the probabilities of independent events. It appears in his presentation that Fredrickson uses "random mating" as a euphemism for "complete sexual promiscuity" (to use Karmel's phrase). Actually, in a more detailed version of his work, to be published in the Winter 1971 issue of the journal Mathematical Biosciences, Professor Fredrickson does introduce a random mating mechanism. How critical this may be to his model remains problematic.

More fundamental matters:

1) How would appreciable illegitimacy affect the Fredrickson approach?

2) As Fredrickson points out, the basic issue in the "Random Mating Model" is the choice of the Birth Rate, and in the "Monogamous Model" the choice of the Marriage Rate. However, these may not be unique to the present approach, which employs the Kendall-Von Foerster equations to incorporate age into a two sex model. It may be that a "Lotka-like" extension is also possible.

Finally, it should be remarked that

1) The present model is concerned only with the first moments of the frequencies. A stochastic approach might be tried in the Goodman-Kendall manner;

2) The present model deals with a continuous scale for time and age. A corresponding study might be formulated in which these scales are discrete--the Bernardelli, Lewis, Leslie approach. This in turn suggests a "reconciliation" of discrete and continuous approaches, ala Goodman and Keyfitz.

In Professor Masnick's interesting paper, a number of questions arise:

1) Why are the quantities α and β taken as constant? This question intrudes because the exponential distribution implies that, for example, α is the probability at a point in time that a woman will pass from FN to PN, so that constancy of α assumes the probability is the same for the first month as for the twelfth which may be doubtful.

2) Similarly, β may be length dependent, i.e., how long a woman is in PN may depend on how long she was in FN.

3) Infant mortality in the model may be a problem. What if the child survives until ovulation?

4) The technical problem of actually completing the estimation of the parameters λ_1, λ_2, λ_3, α, β by the Method of Moments appears formidable. A less efficient method may prove necessary.

(The papers by Frederickson and Masnick, discussed above by Mr. McHugh, were not sent for inclusion in this Proceedings volume.)

MEMORIAL TO JOHN B. LANSING AND HIS CONTRIBUTIONS

Angus Campbell, University of Michigan

John Lansing was born in Geneva, New York, the son of a professor of chemistry at Hobart College. His family was of Massachusetts origin. John took his undergraduate work at Hobart, graduating in economics in 1940. He spent the next year at Columbia University and the following year at the University of North Carolina studying sociology. In the Summer of 1942 he joined Rensis Likert and his staff in the Bureau of Agricultural Economics and worked for two years on the national surveys of saving and spending which later developed into the Survey of Consumer Finances. After the war he went to Harvard University to complete his graduate work and took his Ph.D. in economics at Harvard in 1949. He and Marjorie Tillis were married in 1945.

In the fall of 1949 John and Marjorie Lansing came to The University of Michigan to join John's former Department of Agriculture colleagues who were establishing the Survey Research Center in Ann Arbor. He took an equal responsibility with George Katona and James Morgan in carrying forward the series of studies which have become the Center's Economic Behavior Program. During the 1950's he developed a special interest in the travel market and his numerous studies in this area culminated in his book Transportation and Economic Policy published in 1966. John was offered a joint appointment in the department of economics at Michigan shortly after joining the Center staff and he taught in that Department for nearly 20 years. In the Spring of 1970 the Department of Economics asked him to accept the position of Chairman for the next three years and he formally assumed that title on July 1 while maintaining a part-time association with the Survey Research Center.

John Lansing was a man who enjoyed the respect of all who came to know him. His simplicity and sincerity of manner were the outward signs of the faultless integrity which lay beneath. He was a reserved man, reflecting perhaps something of his New England ancestry, but he was also a man of quiet good humor who enjoyed his many personal associations in the Center and the sociability of his immediate work group. He was a conscientious teacher and a patient and understanding mentor to his graduate students. He was a totally admirable colleague. None of us ever heard him speak in anger, he was unfailingly thoughtful and cooperative, his obligations were invariably fulfilled. He treated everyone, of high station or low, with the same gentle manner and the same equalitarian spirit. His good judgment, his steadiness in moments of stress, his unselfish devotion to the Center--these were traits we all recognized and valued.

While the entire University of Michigan is diminished by the loss of John Lansing, it is the Survey Research Center which is most grievously injured. As one of the earliest members of the Center staff he had become a central and familiar part of everyday life at the Center. Those of us who were his close associates over these many years realize that we have lost a friend whom we cannot replace. We also realize how much we depended on John to carry forward the program of research with which he was concerned. Over two decades he had built an international reputation as an imaginative and productive scholar which made it possible for him to command support for his extensive research program.

It is especially painful to all of us at the Institute for Social Research that John Lansing's life should have been terminated at the height of his career. He had contributed much to his profession and to society and he had much more to give. In recent years he had moved vigorously into a new series of studies of the quality of life in the cities. In 1969 he served as chairman of a Study Conference on Environmental Research for the National Academy of Sciences. His last book, published in October, is entitled Planned Residential Environments. The total number of persons in this country who are capable of carrying out the kind of research with which John was increasingly concerned is so small that we can ill afford to give up the 20 years of productive scholarship which we confidently expected from him.

John Lansing made a great contribution to the building of the Survey Research Center and he will be sorely missed. He enriched the lives of all those around him. We were fortunate to have him with us for the past 21 years and we prize his memory.

DISCUSSION

Michael J. Flax, The Urban Institute

Mr. Chairman: I want to begin by stating that I am doing research at one of the institutions that was formed in an attempt to cope with the rapid changes occurring in information technology which Dean Kozmetsky was talking about. The Urban Institute is a problem oriented, multidisciplinary research organization investigating urban problems. Many of the staff, including myself, have had experience in the "systems approach" which was used so extensively by defense research agencies.

In general, I don't go along with the pessimistic tone of Dean Kozmetsky's paper. Even though it is true that we need a lot more long term research, I think we must also seek short-term, interim improvements and keep constantly making incremental adjustments to our institutions. This is necessary for two reasons: First, with our imperfect knowledge, we are only capable of "muddling through" from crisis to crisis in order to keep our society from going under. Secondly, we have to keep making attempts at improvement in order to keep our citizens from giving up hope and turning in despair to the radical methods of the left or the right.

I think some progress, imperfect though it is, is taking place in many areas. If there were more time it would be possible to give numerous examples.

With regard to the Johnson-Ward paper, I believe they are suggesting an initial approach to a very important area. We are only just beginning to attempt to collect and interpret data on the physical aspects of the socio-economic conditions of cities.1/ The work in measuring attitudinal data with regard to different cities is in an even earlier stage of development.2/ So there is no doubt, in my mind, that this is an area that needs further innovative exploration. The paper rightly points out the need for more information by the Lindsay Administration, for example, as well as the fact that there is reason to believe that the attitudes of citizens do not correspond with either those of the activists that claim to represent them or the planners who feel they know what is good for them.

The approach suggested is an innovative one. A panel which will permit longitudinal data to be gathered from a representative cross section of the population. Also, a feeding back to the panel of their responses in order to change the perceptions of the panel members themselves. I'd like to see this concept tested experimentally. Can peoples attitudes be changed in this manner? Can the results be replicated more than once? Will this procedure result in constructive improvements or merely raise the expectations and the frustration level of the participants?

While I think the concept is an interesting one, I want to offer several suggestions regarding future work in this area. My suggestions are in three areas.

> First, concerning the details of the approach suggested.
>
> Second, regarding the use of computers and other elements of the new information technology, and
>
> Thirdly, regarding the explaining of the proposal to the public.

First, I would like to see much more detailed information concerning this proposal. How large would the panel be? How would you involve a representative sample of the community? How would this compare with sample survey work in this area? Who would conduct this panel and who would pay for it? Would it be a research tool or an aid to city administrators?

Such information is needed in order for me to understand exactly what is being proposed and how, when, and by whom it might be implemented.

Secondly, I would opt for a more gradual approach to the use of computers. I have worked with complex computer systems in the defense department and I know how time consuming and expensive the implementation of even simple systems can be. Implementing mixed, conditional and adaptive models into a system that can be used by poor folks, the young, and middle America, seems to me a herculean task.

I don't believe adequate models exist, so their generation would comprise a substantial project. Even computerizing such models once they were developed and tested would be a difficult endeavor. Design of the man-machine interface necessary for their use by the proposed inner city population would be a third very difficult task. I'd prefer a much more modest and incremental approach. Specifically, I'd like to advance step-by-step, gradually increasing the sophistication of the computer techniques used. I'd also go slow with the multi-media approach. If we carefully define what we wish to communicate we can often select the most efficient medium for our particular message. Multi-media may not be the most practical method of reaching the proposed audience.

Regarding the Delphi method, I think you should distinguish between the "Delphi Exploration" and the "Delphi Technique" developed at RAND. The second, consults experts and feeds their opinions back to them. It is not computer based. The "Delphi Exploration" being developed at the University of Illinois, is probably what was referred to in the Johnson-Ward paper. A detailed study of this preliminary experimental application of computerized sampling and subsequent feedback to the participants may indicate that it could be modified for use by inner-city citizens instead of University students. It is worth further future exploration.

Finally, I'd like to say a word about the relationship of this proposal to the public. Insisting that all results be available to the public does not assure that this activity would be carried out in "goldfish bowl". The way it is presently written, I'm sure the public would not understand what is going on. I think much more clarity in presentation is needed so that city officials, let alone the average citizen, can understand just what you are proposing. In order to be useful, I believe that research proposals and results must be written in a manner that is clearly understandable to the people who must use them as well as the people who must pay for them.

Your report is titled "An Exploration." I hope you will continue your explorations in some of the directions I have suggested.

REFERENCES

1. Jones, Martin V. and Flax, Michael J, "The Quality of Life in Metropolitan Washington (D.C.): Some Statistical Benchmarks" Working Paper 136-1, The Urban Institute, Washington, D.C. 20037, March 1970.

2. Lester Milbrath, of the New York State University at Buffalo and Winston Franklin at the Sloan-Kettering Foundation in Dayton, Ohio, are doing work in this area.

(The papers by Kosmetsky and Johnson-Ward, discussed above by Mr. Flax, were not sent for inclusion in this Proceedings volume.)

EVALUATION OF AN ASPECT OF ENVIRONMENTAL QUALITY

John V. Krutilla, Resources for the Future, Inc.*

I

Environmental factors which have commanded attention of social scientists used to imply a social or cultural context. In more recent times, however, environmental factors to which we give consideration are increasingly involving aspects of the physical environment. Prominently in public discussion, political debate and legislative actions are the effects of industrial pollutants on the ambient environment. From the first law of thermal dynamics we know that all that goes into an industrial process by way of mass and energy also comes out, while changed in proportions and form, remaining unchanged in quantity (making allowances for transformations among solid, liquid and gaseous states).[1] Accordingly, the output of industrial activities is the combination of industrial goods, the object of production, and the residual industrial "bads" or "wastes." Waste disposal involves the attempt to distribute the residuals somewhere within the ambient environment and this gives rise to the degradation of the environment whether in the form of pollution of the earth mantle, air mantle or the hydrosphere.

A related aspect of environmental pollution is the degradation of the visual or aesthetic attributes of the environment. This may be the landscape, or through habitat modification, the biological diversity and ecological integrity of a site or an area. The latter represents a quasi-aesthetic dimension for those who comprehend and appreciate ecological homeostasis. This paper will address an example arising in the latter context. It deals with the comparative valuation of a geomorphologic-hydrologic phenomenon of unique characteristics, use of which is being contested by advocates of incompatible alternatives.[2] The area in question involves the Hells Canyon of the Snake River, forming the boundary between northeastern Oregon and southwestern Washington on the one hand, and a portion of the boundary of Idaho on the other. The Hells Canyon represents the deepest gorge on the North American Continent. Because of the variation in elevation from Canyon floor to the Seven Devils mountain peaks, through which the Canyon is cut, the varied ecology represents virtually all of the life zones found on the North American Continent. For this as well as other reasons it qualifies for preservation under the Wild and Scenic Rivers Act.[3] At the same time, because of narrowness of the gorge, steepness of its walls and the volume of flow in this reach of river, the Canyon provides exceptionally fine sites for hydro-electric development. The former use we shall refer to as the preservation alternative; the latter, as the developmental alternative.

This problem has several rather special features. The developmental alternative has an irreversible consequence for the ecology of the region. Once development occurs, reconsidering the decision is meaningless; i.e., no effective option remains. If the site is preserved in its present conditions all options are retained so long as no decision with an ecologically irreversible consequence is taken. Secondly, and perhaps related to the first, is that since electrical energy is so standardized a commodity, many competing prime movers can be substituted for kinetic energy of falling water in its production. From a scenic and/or recreational point of view, the services yielded by the Canyon in its preservation alternative are not a standard commodity; we find here, if not an absolute uniqueness, then certainly, a highly differentiated product for which demand is likely to be inelastic relative to the demand for services of the Canyon in its developmental alternative. Moreover, advances in technology are likely to have a differential effect on the valuation of the site for the two alternatives considered. We know that there have been rapid advances in the production of electrical energy from thermal sources, whether fossil fuel combustion or nuclear reaction. This will have a bearing on the value of the site if used for energy production. Since the Canyon, and its related ecology, is a gift of nature not produced or reproducible by man, advances in technology are not likely to affect adversely its value if dedicated to the preservation alternative. In fact, if advances in technology continue to achieve gains in productivity in the "produced goods" sector, thereby increasing real per capita income, the value of income elastic irreproducible assets is likely to be favorably affected by technological advance.[4] Accordingly we need to recognize an asymmetry in the implications of technological advance for the valuation of the site when considering the two alternatives.

In Section II, I investigate the significance of these observations in more detail. In Section III, I present quantitative results of an evaluation of the incompatible alternatives when asymmetry in the implications of technological progress is introduced in the evaluation procedure.

II

The Developmental Alternative

Consider now the proposed hydro-electric development. The technology of a given time is incorporated in the dam and powerhouse in such a facility at the time it is built, and will fix the costs of generation over the economic life of the facility. The benefit, on the other hand, being governed by the cost of the most economical

---

*I gratefully acknowledge the assistance of Charles Cicchetti and Clifford Russell with the research on which this paper is based and the helpful comments by Blair Bower, Charles Cicchetti, Pierre Crosson, Allen Kneese and Clifford Russell on an earlier draft of this paper.

[1] See [1] for elaboration of this point.

[2] See [4].

[3] *Ibid.*, Exhibit No. R-624.

[4] See for example [6, 7].

alternative source,[5] does not remain constant over the life of the hydro-electric facility. The cost of thermal power generation has declined progressively over the past half century, and by about 4.5% per year over the past two decades. Part of this was due to the decrease in capital investment per kilowatt of capacity (capacity costs); part was due to the increased efficiency in the utilization of fuel (energy costs). If the life of the alternative source is shorter than that of the hydro-electric facility (and the real cost of the more technologically advanced replacement capacity is lower than at the time of hydro-project construction), then the capacity benefits of the hydro-electric facility will be lower upon the hypothesized retirement and replacement of the thermal alternative with which the hydro is being compared.

The effects of advances in technology of thermal generation, however, have not been restricted to the capacity component of costs. Gains in efficiency with which fuel is used have occurred and also have implications for the valuation of the hydro facility. As the plant factor on technologically advanced new plants will be higher than the system load factor, the difference in factors represents the percentage of a new plant's capacity which can generate "economy energy" to displace energy produced by the most uneconomic plant in the system during off-peak periods. A given plant, when new, will enter the system at, say, 90 percent plant factor. As it ages, it will be used a progressively smaller proportion of the time so that by the twentieth year it may operate only 30 percent of the time.[6] Accordingly, the relevant energy cost will be that given by the weighted average of today's and tomorrow's technology, with the costs related to future technology figuring progressively more significantly as the relevant annual energy costs until the original thermal alternative is replaced (say, in the thirtieth year). At that time both the energy and capacity values would be governed by the state of technology of the thirty-first, not the original year. Thereafter, the capacity value will remain constant from the thirty-first to the fiftieth year,[7] but the energy values will again begin to decline because the relevant costs used for evaluation purposes will again become a blend of the level of technology of the thirty-first year, and the advances in technology from the thirty-first to the fiftieth year.[8]

The Preservation Alternative

Consider next the preservation alternative. The value of any quantity of service consumed per unit time is measured by the area under the demand schedule. When the facility providing the service is a reusable, non-depreciating asset, such as a natural environment protected against destruction or degradation, the value of benefits is the area under the demand curve for each time period the natural area is used. If time is given the customary value of one year, the gross benefit of the natural area would be approximated by the sum of discounted annual benefits. The present value can then be compared with the capital investment (if any) the present value of annual operating costs (if any) and also the opportunity cost, or net present value of the most economical alternative use precluded by retention of the area for uses compatible with existing environmental conditions in the Canyon.[9]

---

[5] The benefit from hydro-electric development can be represented as below:

$$b_d = B_d - C_d - B_a + C_a$$

Where:

- b_d = net benefit from hydro-electric development
- B_d = gross benefit from hydro-electric development
- C_d = cost of hydro-electric production
- B_a = gross benefit from alternative source of power
- C_a = cost of alternative source of power production.

Since the alternative to the hydro-electric development, for comparative purposes, is designed to produce identical services, $B_d = B_a$. Accordingly, the net benefit, b_d is equal to $C_a - C_d$, or the resource savings, if any, from development of the hydro-electric resource. See [10].

[6] Federal Power Commission studies indicate that historically, for fossil fuel plants, the plant factor has fallen to 20 percent by the twentieth year. For computational convenience I use an initial plant factor of 90 percent, a 3 percentage point per year plant factor decay to give us a plant factor of 30 percent in the twentieth year and retirement in the thirtieth year.

[7] The relevant planning horizon for the hydro-electric facility.

[8] Allotted space does not permit presentation of the technological change computational model. The interested reader may refer to [4] Exhibit No. R-670 for details of the model.

[9] Note, to establish the consistency in the treatment of the developmental and preservation benefits, we represent the benefit derivation model for the preservation alternative as below:

$$b_p = B_p - C_p - B'_a + C'_a$$

Where:

- b_p = net benefit from preservation alternative
- B_p = gross benefit from preservation alternative
- C_p = cost of providing services from the preservation alternative
- B'_a = gross recreation benefit from alternative to preservation
- C'_a = cost of providing recreational services alternative to the services provided by the Canyon.

Now, since the Canyon in an undeveloped state is a gift of nature, the costs (other than opportunity costs accounted for in b_d, footnote 5) are

(continued on next page)

If the demand for the services of the area grows, a point may be reached beyond which the use of the area by one more individual per unit time either results in a lessening of the utility obtained by others due to the well known congestion phenomenon, or to the destruction of the environmental characteristics of the area. In the case of Hells Canyon, it must be recognized that a recreational carrying capacity, for example, will be reached in time and if a given quality of recreational experience is to be maintained, resort to rationing is imperative.

Growth in the demand for services of the area and a capacity constraint introduce some complexity in analysis. First, income and population change through time, reflecting increases in the demand for services of the Canyon, other things remaining equal. But as the supply is not augmentable, the Canyon being an irreproducible asset, we would expect the annual value of the services to grow as the demand curve in conventional analysis shifts outward, reflecting income and population growth. Such growth in annual value of services must be incorporated in the benefit estimation procedure. Secondly, the capacity constraint adds to the complexity in quantitative evaluation, since it sets a limit on the range over which the quantity demanded can be summed without adjustment.

The analytic and computational models developed to deal with this problem are too involved to permit their treatment in the space allotted here. Accordingly, only a rough schematic of the argument is presented below to indicate the rationale underlying the analysis.[10] In Figure 1 we have the conventional price-quantity axis, with D_oD_o' the initial period's demand for the recreational services of the Canyon. The vertical SS' represents the non-augmentable supply of services of a constant quality. In the initial period there is an excess supply, relative to quantity demanded at zero price, and all who seek the services can be accommodated without utility-diminishing congestion externalities. The annual benefit, therefore, would be equal to the total area (b_o) under the initial period's demand curve D_oD_o'. At some time (t+n), the quantity demanded at zero price will exceed the supply and to retain quality of the service, rationing must be introduced. P_{t+n}, P_{t+n}' represents the schedule which

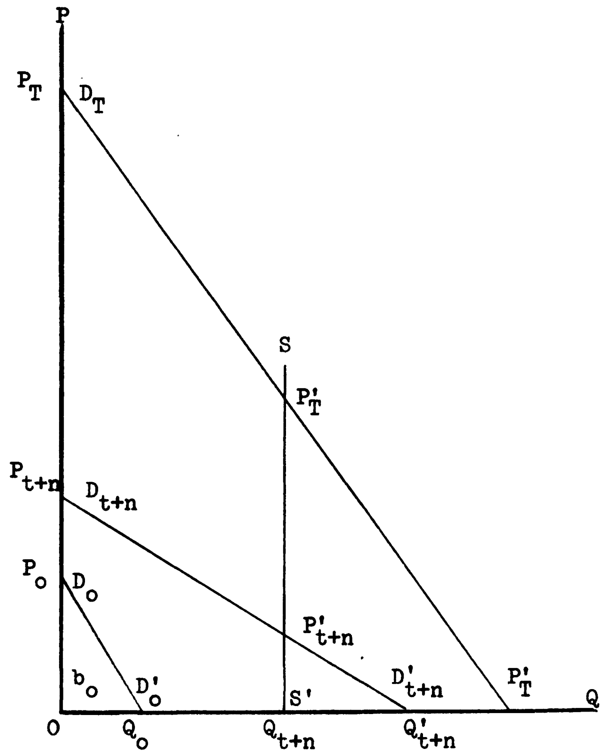


Figure 1.

a discriminating monopolist could exact as prices, and the total value under the demand curve D_{t+n}, D_{t+n}', less that represented by the area under the excess demand portion of the schedule (Q_{t+n}, P_{t+n}', Q_{t+n}') represents the annual value for the transaction period given the schedules in question.

A simplifying assumption would be that the demand curve shifts out uniformly from the origin, but investigation suggests this assumption should be modified in the interest of greater realism. As a result, taking what evidence we have on the growth in demand for primitive area recreation generally, and the income elasticity and related phenomena for this type of service,[11] we relate the shift in the demand function intercept of the price axis (r_y) to the projected growth in real per capita income. We relate the shift along the horizontal, or quantity, axis intercept to the recorded rate of growth in quantity demanded at zero price (γ) dampened to eventually equal only the rate of growth of population. The resulting shifts will produce both demand schedules with changing slopes and also, given the capacity constraint, changing geometric shapes in the relevant areas under the demand curve. These observations are illustrated in the three time-dated demand schedules in Figure 1.

So much for the outline of the argument and computational models. One additional point merits mention before the quantitative results of analysis are presented. Ideally one would wish to develop a demand schedule for each of the several recreational activities which one could anticipate

---

9 (continued from previous page) taken to be negligible. Accordingly, we have:

$$b_p = B_p - B_a' + C_a'$$

However, since we look to produced assets' services as alternatives, and assuming free entry into the recreational services industry, we would expect that the leisure formerly consumed in Hells Canyon facilities would be distributed across the alternatives impinging at the margins. Now, since the benefits at the margin under the circumstance would equal the costs at the margin, B_a' and C_a' would be equal. Accordingly, $b_p = B_p$, which corresponds to the results presented in Table IV, section III below.

10 The interested reader may consult [6,7] for the details of the model.

11 See [4]. Witness Krutilla Transcript, pp. R-5859-69.

being enjoyed in the area, e.g., fishing, white-water boating, hunting, backpacking, etc. These demand functions could be estimated whether jointly where merited, or independently, by procedures developed in the evolving literature in recreational demand estimation.[12] Were information available, the behavior of such schedules for each separable activity could be projected and the specific present worths computed, taking into account congestion costs, if any, of two or more distinctly different recreational activities indulged in simultaneously by different individuals. Unfortunately, time was not available to undertake this kind of analysis. Instead, a "composite demand function" was contrived so that, as implied in Figure 1, only one shifting demand schedule was employed as a proxy for the combination of independent and related demand functions. Moreover, since no less time would have been required to estimate such a hybrid function than to estimate the individual demand functions, an alternative strategem was adopted. The question was asked, in effect, "What would the benefit from preservation need to be to be equal to, or exceed, the developmental benefit?"

This question can be answered by determining what the initial year's benefit from preservation would need to be, growing at the rate (α) implied by the annual shifts in the composite demand function, to be equal to the present value of the developmental alternative. This would be desirable, for example, if we were not able to obtain any adequate estimates of the initial year's preservation benefits and would need to have some threshold value on which to base a judgment. We could obtain such a threshold value by computing the present value of a dollar's worth of initial year's benefits growing at the rate of α and discounted appropriately for time. The result of such a present value computation, divided into the present value of the hydro-electric development, would yield the estimate of the initial year's preservation benefit which would be required to justify economically, the preservation alternative; i.e., would be equal to the opportunity cost of foregoing the development. The results of performing such an exercise are given in section III.

III

In this section I display the results obtained when the asymmetry in the implications of technological progress is considered explicitly in the evaluation of the two incompatible alternative uses of Hells Canyon.

In the case of introducing technological advance in thermal alternatives to hydro-electric development, the quantitative results will depend on investment per <u>kilowatt capacity</u>[13] of the alternative thermal source, itself partly depending on the interest rate. In addition, the results will depend on the cost per <u>kilowatt hour</u> of thermal <u>energy</u>.[14] Finally, the rate of advance in technical efficiency itself enters into the calculation of the difference between the results obtained when technological advance is, and when it is not, introduced explicitly into the analysis. For our purposes, we have relied on construction cost data provided by a Federal Power Commission expert;[15] have used opportunity cost of capital of 9 percent, but with estimates provided alternatively using 8 percent and 10 percent for purposes of sensitivity analyses; rates of technological progress of between 3 percent and 5 percent per year, to bracket what is believed to be the relevant range;[16] and energy costs, again supplied by FPC staff witnesses, of 0.98 mills per kilowatt hour in the early stage and ranging up to 1.28 mills per kilowatt hour in the later period of analysis.[17] The adjustment factors for introducing the influence of technological change into the analysis are given in Table I following.

Accordingly, for any given interest rate, rate of technological change, and energy costs in mills per kilowatt hour; the generation costs estimated by traditional methods (sum of capacity and energy costs) would be divided by the values given in Table I to obtain the adjusted alternative costs -- hence, the benefits of the proposed hydro-electric development. While the gross benefits of hydro appear to be only marginally affected by introducing technological change into the analysis, i.e., are reduced by only five to ten percent, the net benefits and hence present value of the site for hydro development is reduced to a half. This result follows from the fact that the thermal alternative to hydro was a close cost competitor; thus a five to ten percent change in <u>gross benefits</u> had a large effect on the <u>net value</u> of the developmental alternative.[18]

In connection with the benefit computations for the preservation alternative, the present value of a dollar's worth of initial year's benefit is a function of both the rate of growth in annual benefits, α, and the discount rate, i. But, annual benefits grow at a non-uniform rate over time depending on the values which are taken by γ, r_y, k and m. (See Table II for definition of terms.) Since k represents the "recreational carrying capacity" which is given by the capacity of the area to accommodate recreation seekers without eroding the quality of the recreational experience, the k's and γ's are related.

[12] For a survey of the literature as of 1969, see [2, 3, 9].

[13] A fixed cost for capacity to meet peak requirements.

[14] A variable cost for fuel, supplies, etc., related to the production of energy invariably at a rate below system peak capacity.

[15] See [4]. Witness Jessell's Exhibit No. R-54-B.

[16] See [13] for basis of computing technological progress, 1950-1968.

[17] See [4]. Witness Chavez's Exhibit No. R-107-B.

[18] See [4]. Witness Krutilla, Transcript pp. R-5842-43 and Exhibits No. R-671 and R-671-A, for detailed explanation of the derivation of benefits using technological change model.

Table I

OVERSTATEMENT OF HYDRO-ELECTRIC CAPACITY AND ENERGY VALUES BY NEGLECTING INFLUENCE OF TECHNOLOGICAL ADVANCES

Discount Rate/Year	Technological Advance Rate/Year	Estimated Capacity Values	Conventionally Estimated Benefits as a Percentage of Actual Benefits when Adjusted for Influence of Technological Advance, for Various Capacity and Energy Costs		
$i =$	$r_t =$	$/KW	Percent at 0.98 mills per kwh	Percent at 1.22 mills per kwh	Percent at 1.28 mills per kwh
	0.03		107.4	107.9	108.0
0.08	0.04	$27.43	109.0	109.6	109.7
	0.05		110.2	110.9	111.1
	0.03		105.9	106.4	106.5
0.09	0.04	$30.08	107.2	107.7	107.8
	0.05		108.2	108.8	108.9
	0.03		104.8	105.1	105.2
0.10	0.04	$32.89	105.8	106.2	106.3
	0.05		106.5	107.1	107.2

Source: [4], Exhibit R-670, Table I, page 3.

Table II

PRESENT VALUE OF ONE DOLLAR'S WORTH OF INITIAL YEAR'S BENEFITS GROWING AT α

i = 8%, m = 50 years			
r_y	γ=7.5% k=25 yrs.	γ=10% k=20 yrs.	γ=12.5% k=15 yrs.
0.04	$ 134.08	$ 169.86	$ 173.90
0.05	211.72	263.49	262.12
0.06	385.10	467.30	449.00

i = 9%, m = 50 years			
r_y	γ=7.5% k=25 yrs.	γ=10% k=20 yrs.	γ=12.5% k=15 yrs.
0.04	$ 93.67	$ 120.07	$ 125.89
0.05	136.12	172.35	176.25
0.06	214.76	267.10	264.49

i = 10%, m = 50 yrs.			
r_y	γ=7.5% k=25 yrs.	γ=10% k=20 yrs.	γ=12.5% k=15 yrs.
0.04	$ 69.28	$ 89.45	$ 95.71
0.05	95.15	121.91	127.68
0.06	138.17	174.85	178.66

Where:

i = discount rate

r_y = annual rate of growth of price per recreation day

γ = annual rate of growth of quantity demanded at given price

k = number of years after initial year in which carrying capacity constraint becomes effective

m = number of years after initial year in which gamma falls to rate of growth of population.

The particular values taken, i.e., γ of 10 percent and k of 20 years, with alternative assumptions for purposes of sensitivity analyses, were chosen for reasons given elsewhere.[19] A discount rate of 9 percent, with alternatives of 8 and 10 percent was the result of independent study.[20] The selection of the value for m of 50 years, with alternative assumptions of 40 and 60, was governed by both the rate of growth of general demand for wilderness or primitive area recreation, and the estimated "saturation level" for such recreational participation for the population as a whole. Finally, the range of values for r_y was taken from what we know about the income elasticity of demand for this kind of recreation activity[21] and growth in per capita income over the past two or three decades.

The results of our "preferred" values, with alternatives given for changes in assumptions are displayed in Table II. Each of these present value computations can be divided into the net present value of the water resource development project -- i.e., the hydro-electric power value, along with incidental flood control and related multi-purpose development benefits -- to yield the initial year's preservation benefit which (growing at α and discounted at i) would have a present value equal to the present value of development. The corresponding initial year's preservation benefits are displayed in Table III.

Now what does this tell us which the traditional analysis of comparable situations requiring the allocation of "gifts of nature" between two incompatible alternatives does not?

Let us take for illustration, subject later to sensitivity analysis, the computed initial year's preservation benefit corresponding to i of

---

19 *Ibid.* Transcript pp. R-5864-66 and R-5872.

20 See [5, 8].

21 See [3,7].

9 percent, r_t of 0.04, γ of 10 percent and k of 20 years, m of 50 years and r_y of 0.05; namely, \$80,122. Is this a preservation benefit we might expect to be equaled or exceeded by the first year the hydro-electric project would otherwise go into operation? In many cases we would have only the sketchiest information and would have to make such a comparison on a judgmental basis. In the case of Hells Canyon, we obtained rather better information and shall return to the matter subsequently. But for now, we have \$80,000 as threshold value which we feel is necessary to justify, on economic grounds, allocation of the resource to uses compatible with retention of the area in its present condition. This sum of \$80,000 compares with the sum of \$2.9 million, which represents the "levelized" annual benefit from the hydro-electric development, when neither adjustments for technological progress have been made in hydro-electric power value computations, nor any site value (i.e., present value of opportunity returns foreclosed by altering the present use of the Canyon) is imputed to costs.[22] Typically then, the question would be raised whether or not the preservation value is equal to or greater than the \$2.9 million annual benefits from development.

Let us consider the readily quantifiable benefits from the existing uses of the Canyon. These are based on studies conducted by the Oregon and Idaho Fish and Game Departments, in collaboration with the U.S. Forest Service, and are displayed along with my imputation of values per user day in Table IV below. From Table IV one could argue, for example, that the preservation benefits shown are roughly only a third as large as would be required based on traditional analysis of similar cases. By introducing differential incidence of technological progress on the mutually exclusive alternatives for Hells Canyon, we have a different conclusion. The initial year's preservation benefit (\$900,000), subject to re-evaluation on the basis of sensitivity tests, appears to be an order of magnitude larger than it needs to be to have a present value equal to or exceeding that of the development alternative. Thus introducing differential incidence of technological progress affects the conclusions in a significant way.

What about the sensitivity of these conclusions to the particular values of the variables used in our two simulation models? Sensitivity tests can be performed with the data contained in Tables II and III, along with additional information available from computer runs performed. Some of these checks are displayed in Table V.

Given the estimated visitor days and imputed value per visitor day, it follows that the conclusions regarding the relative economic values of the two alternatives are not sensitive, within a reasonable range, to the particular values chosen for the variables and parameters used in the two computational models.

There is need, however, for another set of tests when geometric growth rates are being used. We might regard these as "plausibility analyses." They would test, for example, the plausibility of the ratio of the implicit price to the projected per capita income in the terminal year, to ensure credibility of the results. Similarly for the plausibility of the ratio of the terminal year's preservation benefit, say, to the GNP in the terminal year. The year at which the growth rate in quantity of wilderness type outdoor recreation services demanded falls to the rate of growth of the population must also be checked to ensure that the implicit population participation rate is something one would regard as reasonable. Such tests were performed in connection with the Hells Canyon case in order to avoid problems which otherwise would stem from use of unbounded estimates.

Finally, since the readily observed initial year's benefits appeared to be in excess of the minimum which would be required to have such preservation benefits equal or exceed in present worth the developmental benefits, the analysis

[22] Derived from Exhibit No. R-671 of [4] (exclusive of adjustments for technological change).

TABLE III

INITIAL YEAR'S PRESERVATION BENEFITS (GROWING AT THE RATE α) REQUIRED IN ORDER TO HAVE PRESENT VALUE EQUAL TO DEVELOPMENT

r_y	γ=7.5% k=25 yrs.	γ=10% k=20 yrs.	γ=12.5% k=15 yrs.
i=8%, m=50 years, r_t=0.04, PV_d=\$18,540,000*			
0.04	\$138,276	\$109,149	\$106,613
0.05	87,568	70,363	70,731
0.06	48,143	39,674	41,292
i=9%, m=50 years, r_t=0.04, PV_d=\$13,809,000*			
0.04	\$147,422	\$115,008	\$109,691
0.05	101,447	80,122	78,336
0.06	64,300	51,700	52,210
i=10%, m=50 years, r_t=0.04, PV_d=\$9,861,000*			
0.04	\$142,335	\$110,240	\$103,030
0.05	103,626	80,888	77,232
0.06	71,369	56,397	55,194

* Source: [4], Exhibit No. R-671.

Where:

i = discount rate

r_y = annual rate of growth in price per user day

γ = annual rate of growth of quantity demanded at given price

k = number of years following initial year upon which carrying capacity constraint becomes effective

m = number of years after initial year upon which gamma falls to rate of growth of population

PV_d = present value of development

r_t = annual rate of technological progress.

TABLE IV

ILLUSTRATIVE OPPORTUNITY COSTS OF ALTERING FREE FLOWING RIVER AND RELATED CANYON ENVIRONMENT BY DEVELOPMENT OF HIGH MOUNTAIN SHEEP

Quantified losses	Visitor Days 1969[2]	Visitor Days 1976
Stream Based Recreation:[1]		
Total of boat counter survey	28,132	51,000
Upstream of Salmon-Snake confluence	14,439	26,000
Non-boat access:		
Imnaha-Dug Bar	14,517	26,000
Pittsburgh Landing	14,464	26,000
Hells Canyon Downstream:		
Boat anglers	1,000	1,800
Bank anglers	2,333	4,000
Total stream use above Salmon River	46,753 plus[3]	84,000 at $5.00/day=$420,000
Hunting, Canyon Area[4]		
Big Game	7,050	7,000 at 25.00/day= 175,000
Upland Birds	1,110	1,000 at 10.00/day= 10,000
Diminished value of hunting experience[5]	18,000	29,000 at 10.00/day= 290,000
Total Quantified losses . . . . $895,000 ± 25%		

Unevaluated Losses:

A. Unmitigated anadromous fish losses outside impact area.
B. Unmitigated resident fish losses:
   1) Stream fishing downstream from High Mountain Sheep.
C. Option Value of rare geomorphological-biological-ecological phenomena.
D. Others.

1 Source: An Evaluation of Recreational Use on the Snake River in the High Mountain Sheep Impact Area, Survey by Oregon State Game Commission and Idaho State Fish and Game Department in cooperation with U.S. Forest Service, Report dated January 1970 and Memorandum, W.B. Hall, Liaison Officer, Wallowa-Whitman National Forest, dated January 20, 1970.

2 "Visitor Day" corresponds to the President's Recreational Advisory Council (now, Environmental Quality Council) Coordination Bulletin No. 6 definition of a visitor day as a twelve-hour day. Operationally, the total number of hours, divided by twelve, will give the appropriate "visitor day" estimate.

3 Not included in the survey were scenic flights, nor trail use via Saddle Creek and Battle Creek Trails. Thus, estimates given represent an under-reporting of an unevaluated amount.

4 "Middle Snake River Study, Idaho, Oregon and Washington," Joint Report of the Bureau of Commercial Fisheries and Bureau of Sports Fisheries and Wildlife in Department of the Interior Resource Study of the Middle Snake, Tables 10, and 11.

5 The figure 18,000 hunter days is based on Witness Pitney's estimate of 15,000 big game hunter days on the Oregon side, and estimated 10,000 hunter days on the Idaho side (provided in letter from Monte Richards, Coordinator, Idaho Basin Investigations, Idaho Fish and Game Department, dated February 13, 1970), for a total of 25,000 hunter days (excluding small game, i.e., principally upland birds) in the Canyon area, less estimated losses of 7,000 hunter days. This provides the estimated 18,000 hunter day, 1969 total, which growing at estimated 5 percent per year for deer hunting and 9 percent per year for elk hunting would total 29,000 hunter days by 1976.

was terminated. Following Weisbrod,[23] however, while an excess of benefits as estimated above from the preservation of an irreplaceable asset is sufficient to justify its retention on economic grounds, it need not be necessary. Two reasons can be given; one relates to the matter of option value, i.e., the value of retaining an effective option when faced with a decision having irreversible consequences. This value was not included in the above estimation procedure.

The second relates to the particular measure of consumer surplus used in estimating the benefit, i.e., whether the aggregate willingness of users to pay for the services of the Canyon preserved in its present condition -- the measure implied in the analysis above -- or the aggregate sum which would need to be provided users of the Canyon retained in an unaltered condition to have them voluntarily relinquish their claims to its use. These measures are not identical except in a special case, and the one used in the analysis results in only a lower bound estimate. Since

23 See [12].

TABLE V

SENSITIVITY OF ESTIMATED INITIAL YEAR'S REQUIRED PRESERVATION BENEFITS TO CHANGES IN VALUE OF VARIABLES AND PARAMETERS (AT i = 9%)

Variable	Variation in Variable From	To	Percent Change	Percent Change in Preservation Benefit
r_y	0.04	0.05	25	39 - 49
r_t	0.04	0.05	25	25
k^*	20 yrs.	25 yrs.	25	30 - 40
γ	10%	12.5%	25	-4 to +7
m	40 yrs.	50 yrs.	25	3

* The 25 percent change in years before capacity is reached translates into a 40 percent change in carrying capacity at the growth rate of 10% used here.

these considerations were not essential to the analysis, i.e., the lower bound estimate exceeded the required total, I mention then only in passing.[24]

IV

In this paper I have reported on a study directed toward aiding a resource allocation decision involving amenity aspects of the environment. The problem contains a number of considerations which are either novel, or at least considered only for the first time in any quantitative sense. Perhaps the reason the heretofore elusive elements were considered at all in this case relates to the Federal Power Commission's interest in responding to the Supreme Court's directive to give the visual and related aesthetic aspects of the environment explicit consideration in reaching a decision as to whether the remaining portion of the Hells Canyon should or should not be licensed for development.[25]

As a first venture in this area there is no reason to pretend that it represents the ultimate development of analytic means for dealing with problems of this sort. The sensitivity tests have revealed in fact, that while the conclusions would not be reversed were the assumed values of the parameters to be changed within any reasonable range in the Hells Canyon case, there is evidence that in cases where the results of analysis would fall within a narrower range, the particular values which the parameters were assumed to take could be critical to the outcome. Accordingly, there is need to investigate, both theoretically and empirically, a number of problems to further sharpen the analysis for cases in the future where the problem of choice would be less clear cut.

Among problems rating high priority would be the further investigation of the asymmetric effects of technological progress particularly as they influence the value of the r_y parameter (note Table V). Another problem demanding additional attention is the problem of developing an operational measure for optimal recreation capacity for such low density recreational resources. Now while an estimate of option value was not necessary in the Hells Canyon case, the results in its absence being sufficient to justify retaining the Canyon in its present state, in future cases the value of retaining an option when faced with a decision having an irreversible result might be the critical element on which the decision would turn. Accordingly, additional work in the area of developing operational measures for the value of such options ranks among the priority research tasks to aid making similar decisions in this general area in the future. Finally, since the Supreme Court in recent decisions appears to have granted the aggrieved public "standing" in court when common property resources are being used to the detriment of the general public, the measure of the damages stemming from a change in the natural environment deserve careful consideration. Typical of traditional benefit-cost analysis, as well as in the measure employed in the study reported on above, has been an estimate of the willingness of beneficiaries of the unaltered environment to pay the prospective developer to dissuade him from modifying the status quo. With the standing accorded the public in such cases the nature of the measure changes. It now becomes the amount which the party proposing to alter the environment must pay the aggrieved public to just compensate it for losses it suffers in altering the environment. As this measure (price equivalent measure of consumer surplus) is normally greater than the conventional measure used (price compensating measure of consumer surplus) the difference in measures employed may become critical in future cases where the outcome from traditional analysis is insufficient to support preservation of the existing environment in unaltered form. This problem merits joint economic and legal investigation in order that consistency in legal and economic doctrine be achieved and methods of measurement consistent with this be developed for application in future cases of the nature reported on in this paper.

[24] See [7].

[25] See [11].

REFERENCES

[1] Ayres, R.U. and Kneese, A.V. "Production, Consumption and Externalities," American Economic Review 49 (1969) 282-97.

[2] Burton, T.L. and Fulcher, M.N. "Measurement of Recreational Benefits -- A Survey," The Journal of Economic Studies (1967) 35-48.

[3] Cicchetti, C.J., Seneca, J.J. and Davidson, P. The Demand and Supply of Outdoor Recreation, Rutgers University, Bureau of Business Research 1969.

[4] Federal Power Commission, In the Matter of Pacific Northwest Power Company and Washington Public Power Supply System Project Nos. 2243/2273.

[5] Joint Economic Committee, 90th Congress, 2nd Session, Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis, Washington, D.C. GPO, 1968.

[6] Krutilla, J.V. "Conservation Reconsidered," American Economic Review 47 (1967), 777-86.

[7] Krutilla, J.V., Cicchetti, C.J., Freeman, A.M., and Russell, C.S. "Observations on the Economics of Irreplaceable Assets," in Environmental Quality Analysis: Research Studies in the Social Sciences, edited by Kneese, A.V. and Bower, B.T., 1971.

[8] Seagraves, J.A. "More on the Social Rate of Discount," Quarterly Journal of Economics, 84 (1970), 430-50.

[9] Smith, R.J. "The Evaluation of Recreational Benefits: Some Problems of the Clawson Method," Studies of Recreational Demand No. 5, Faculty of Commerce and Social Science Discussion Paper Series B., University of Birmingham, 1970.

[10] Steiner, P.O. "The Role of Alternative Costs in Project Design and Selection," Quarterly Journal of Economics, 79 (1965) 417-30.

[11] Udall v. Federal Power Commission, Nos. 463 and 462, June 5, 1967 (October Term 1966 United States Supreme Court).

[12] Weisbrod, B. "Collective Consumption Services of Individual Consumption Goods," Quarterly Journal of Economics, 78 (1964), 471-77.

[13] "Steam Station Cost Survey," Electrical World, November 3, 1969 and October 18, 1965.

JOINT-OUTPUTS OF LAND AND WATER WASTES IN A LINEAR PROGRAMMING MODEL FOR AIR POLLUTION CONTROL

Robert E. Kohn, Southern Illinois University at Edwardsville*

A linear programming model has been developed for the St. Louis Airshed. This model is based on the simplifying assumption that ambient air quality goals can be achieved by reducing total emissions of each pollutant in an airshed to a predetermined allowable level for that pollutant. The reduction in emissions is obtained by instituting a least cost set of air pollution control method activity levels. A mathematical formulation of the model and the type of data used to characterize emission sources and the abatement technology is contained in an appendix to this paper. A more complete description of the model and some of the results have been reported elsewhere [5, 6, 7, 8, 9]. Although it was apparent that certain control methods for air pollution alter the flow of land, water, thermal, and other wastes, these interrelated pollutants were ignored in the original model.

In a 1969 article in the *American Economic Review,* Ayres and Kneese argue that the "primary interdependence between the various waste streams . . . casts into doubt the traditional classification of air, water, and land pollution as individual categories for purposes of planning and control policy." [1, p. 286] They warn that a partial equilibrium approach, "while more tractable, may lead to serious errors." [1, pp. 295, 6] This paper examines the nature of the errors which may have been introduced into the linear programming model by ignoring three specific external waste streams created by air pollution abatement measures. These are (a) liquid wastes, (b) landfill waste, and (c) thermal discharge to rivers. The nature and concentration of the contaminant in waste water or the type of solid waste that is landfilled is ignored; it is assumed that each external waste flow is homogeneous.

The solutions for four different models will be examined:

(I) a model for air pollution control in which the three external waste flows associated with the optimal solution are totaled but not constrained,

(II) as Model I, but with the provision that the resultant external waste flows be less-than-or-equal-to zero,

(III) the air pollution control model augmented to include net costs for reprocessing the joint-output waste flows,

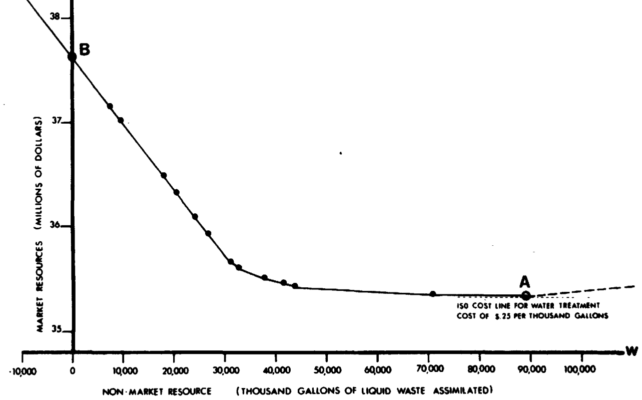
(IV) as Model III, but including feed-backs on the air pollution model associated with the treatment of the external wastes.

This sequence of models, which is presented in mathematical form in the appendix, may be related to Figure 1. The graph, derived from the original model by parametric programming, is an isoquant for the given set of air quality standards for the St. Louis Airshed in 1975.[1] The goals can be achieved with varying combinations of market resources, z, and a non-market resource, w, the disposal capacity of the river system. The latter is measured in thousands of gallons of disposed liquid waste. Joint-outputs of landfill and thermal wastes are ignored in this diagram. If waste water output is unconstrained (Model I), the optimal solution is at point A, where total cost of air pollution abatement is minimized. If waste water output must be less-than-or-equal-to zero (Model II), the optimal solution is at B. If a shadow price for w is introduced, say $.25 per thousand gallons (Model III), an isocost line is defined; in this particular example, the least cost solution is still at point A. If there are feed-backs between water treatment and air pollution abatement (Model IV), the isoquant shifts with the quantity of water treated and a solution cannot be graphed in two dimensions.

*This research is supported by Grant No. GS-2892 from the National Science Foundation. The writer is grateful to Professors H. J. Barnett, M. Friedlander, R. Hoeke, and G. Schramm for their constructive criticism of an earlier version of the paper.

[1]The non-economic, upward sloping segment of the isoquant (distinguished by dashes) may be questioned. If the technology in the model were to include possibilities for increasing liquid waste at zero cost (say, by reduced recirculation of scrubbing water), the isoquant would very likely contain a horizontal facet to the right of point A.

FIGURE 1. COMBINATIONS OF MARKET RESOURCES AND A NON-MARKET RESOURCE FOR ATTAINING A SPECIFIC SET OF AIR QUALITY GOALS IN THE ST. LOUIS AIRSHED IN 1975



The air quality goals used in the model are related, in part, to those adopted by the Missouri Air Conservation Commission. They are annual average ambient air concentrations of 5 ppm for carbon monoxide, 3.1 ppm for total hydrocarbons, .069 ppm for nitrogen oxides, .02 ppm for sulfur dixoide, and 75 $\mu g/m^3$ for particulates. The solution of each model includes non-zero activity levels for more than 50 out of 215 control methods. To facilitate comparison of the four solutions, the following will be examined:

a) total cost of air pollution abatement (This includes the market value of labor and materials, the depreciation of equipment, the opportunity cost of invested capital, the economic value of substitute fuels, less credits for recovered by-products.)
b) total cost of reprocessing the external wastes generated by air pollution abatement
c) the quantities of untreated external wastes
d) shadow prices for the external waste constraints
e) quantities of external wastes reprocessed
f) quantity of natural gas replacing coal in the least cost solution (The fluctuations in this quantity provide insight on what is happening in the model.)
g) marginal cost of abatement (the dual values) for sulfur dioxide and particulates.

The values for these indicators are listed in Table 1.

Model I: The Original Model

The solution of Model I indicates that the cost of achieving the air quality goals for the St. Louis Airshed in 1975 is $35,000,000 plus 89,000 thousand gallons of disposed liquid waste, 450,000 tons of landfill waste, and 1,400,000 million Btu's of heat discharged to rivers. It should be noted these are the incremental costs of air pollution abatement and do not include the cost and waste outputs associated with the pre-regulation, or 1963 base-year level of air pollution control. The external waste flows which would be created are in addition to an estimated 435,000,000 thousand gallons of waste water, 865,000 tons of landfill waste, and 120,000,000 million Btu's of thermal discharge from all human activites unrelated to air pollution control in the St. Louis Airshed in 1975.

TABLE 1. AIR POLLUTION ABATEMENT MODELS

Indicators	I No Constraints on External Waste Streams (solid, liquid, thermal wastes)	II Less-than-or-equal-to zero Constraints on External Waste Streams	III Reprocessing Prices on the External Waste Outputs	IV Prices on the External Wastes and Feedbacks on the Model from Reprocessing
Cost of Air Pollution Abatement	$35,337,283.	$56,106,262.	$35,398,680.	$35,405,008.
Cost of Reprocessing the External Wastes	0	0	$514,832.	$517,013.
Thousand Gallons of Liquid Waste Generated	88,865	0	0	0
Tons of Solid Waste Generated	450,082	0	0	0
Million Btu's of Heat Discharged to Rivers	1,363,366	0	0	0
Shadow Price per Thousand Gallons of the Liquid Waste Constraint	0	$92.33	$.25	$.27
Shadow Price per Ton of the Solid Waste Constraint	0	$17.94	$1.40	$1.40
Shadow Price per Million Btu's of the Thermal Waste Constraint	0	$20.07	$.04348	$.04729
Thousand Gallons of Liquid Waste Treated	0	0	94,975	94,975
Tons of Solid Waste Reprocessed	0	0	539,474	538,892
Million Btu's of Thermal Waste Diverted	0	0	1,368,699	1,429,583
Optimal Quantity (in Millions of Cubic Feet) of Natural Gas Replacing Coal	14,193	72,948	14,288	14,370
Marginal Cost of Abatement for One Pound of Sulfur Dioxide	$.02193	$.06819	$.02193	$.02193
Marginal Cost of Abatement for One Pound of Particulates	$.07748	$.17357	$.09206	$.09208

TABLE 2. REPROCESSING METHODS FOR THE EXTERNAL WASTE FLOWS

	Primary, secondary, and tertiary treatment of municipal waste water	Combustion of processed refuse as supplementary power plant fuel	Hyperbolic natural draft cooling tower
Activity Unit	Thousand gallons	Ton of refuse	Thousand Kilowatt Hours
Cost per Activity Unit	\$.25	\$.80	\$.20
Output of Liquid Waste	−1. thousand gallons	0	0
Output of Land Waste	.00045 tons	−.7 tons	0
Direct or Indirect Output of Thermal Waste	.046 million Btu	.1035 million Btu	−4.6 million Btu
Coal Burned at the Sioux Power Plant to Produce Required Electricity	.004 tons	.009 tons	.004 tons
Reduced Emissions of Carbon monoxide	−.0012 pounds	−.4027 pounds	−.0012 pounds
Reduced Emissions of Hydrocarbons	−.0005 pounds	−.1011 pounds	−.0005 pounds
Reduced Emissions of Nitrogen oxides	−.0472 pounds	−.1062 pounds	−.0472 pounds
Reduced Emissions of Sulfur dioxide	−.2511 pounds	1.5350 pounds	−.2511 pounds
Reduced Emissions of Particulates	−.0009 pounds	−.0022 pounds	−.0009 pounds

The incremental amount of liquid waste from air pollution abatement is an insignificant percentage of the estimated volume of waste water. This may be attributed to the fact that only five of the optimal air pollution control methods affect water pollution. The major contributor is the dolomite wet scrubbing process for power plant desulfurization, which involves a discharge of 70 gallons of waste water per ton of coal burned. However, it is possible that the present model does not include an adequate representation of scrubber-type control methods, so that the volume of waste water associated with an optimal solution for air pollution control may be understated.[2]

The solid waste that must be landfilled as a result of air pollution abatement represents more than a 50% increase in the projected landfill tonnage. This includes recovered, unsalvageable particulate matter and the solid waste from control methods which replace open burning or incineration with landfill disposal.[3]

The incremental thermal discharge is associated with the generation of 300 million kilowatt hours, the annual electrical requirements for the set of optimal control methods.[4] This represents approximately one percent of the projected electrical output of the utility power plants in the St. Louis Airshed in 1975. [6, pp. 239-303] It is assumed that there are .0046 million Btu's of heat discharged to the cooling water for each kilowatt-hour generated.[5]

[2]The insufficient representation of water using, air pollution control methods in this model may be attributed to the unavailability of data on local chemical processing industries, which would be major users of scrubbing water. As a consequence of this lack of data, a single source is proxy for all hydrocarbon chemical processing. Furthermore, it was assumed when the model was originally set up, that this source was initially equipped with scrubbers and that the optional control method would be an afterburner in combination with the scrubber. As a result of this simplification, a number of incremental-water-using control methods may be missing from the model.

[3]It should be emphasized that the external waste flows are joint-outputs of an optimal solution of a specific model. The air pollution regulations being enforced in St. Louis are more stringent with regard to open burning and incineration than the solution of the present model. As a result, the increase in land waste for the St. Louis Airshed as a result of air pollution control will be larger than the figure suggested here.

[4]The original model did not account for the additional coal combustion that would be required to generate this electricity. In the present study, only the incremental thermal pollution is considered. In a subsequent study, Model I was rerun to investigate the feedback effect of any additional, required coal combustion. Except for a 1% increase in total abatement cost, the optimal set of control methods and dual values was essentially unchanged.

[5]This is an average figure based on data in reference 15, p. 291.

Model II: The Original Model with Constraints on External Wastes

Model II incorporates the constraint that there be no incremental external waste flows as a consequence of controlling air pollution. Such a requirement, although stringent, is less binding than the stipulation sometimes expressed, that any control methods which reduce one waste flow while increasing another should never be used.[6] Model II permits the use of such air pollution control methods but requires that any incremental joint-wastes be offset within the technological framework of the model. The resulting solution indicates that air pollution abatement with zero external waste outputs would cost over \$56 million in 1975. This solution corresponds to point B in the two dimensional Figure 1. The shadow prices of the external waste flows are the marginal costs associated with the constraints. If, for example, an incremental joint-output of one thousand gallons of liquid waste were allowed, the total cost of air pollution abatement would decline by \$92.33.

For the most part, the waste flows are offset within the linear programming model by conversion of certain types of furnaces from coal to natural gas. An example is a control method which represents the conversion of travelling grate stokers with mechanical dust collectors to natural gas.[7] This eliminates 170 pounds of bottom ash and 40 pounds of collected fly ash per ton of coal burned. In addition, the retirement of the stoker and mechanical collector eliminates power consumption of about 12 kilowatt hours per ton of coal. In the case of travelling grate stokers that are equipped with wet scrubbers, conversion to natural gas reduces the output of waste water as well as bottom ash. It will be noted in Table 1 that because of the imposition of constraints on external wastes, the marginal costs of controlling sulfur dioxide and particulates are considerably higher for Model II than Model I.

Model III: The Simple, Augmented Model

The high shadow costs for the three external waste flows in Model II suggest that it would be inefficient to require that an air pollution control program involve no incremental external wastes. There are methods to reprocess equivalent quantities of these wastes that cost less than the

[6]For example, the dolomite wet scrubbing process for power plants received some criticism because in the process of eliminating sulfur dioxide and particulate air pollution, it adds some salt content to the discharged water.

[7]The air pollution coefficients for this control method, identified as Control Method 21F, are discussed in reference 8. A complete description of the priority of conversion, by type of stoker, is contained in reference 7.

shadow prices indicated. In Model III, three reprocessing methods are added to the model. Only the cost and output (waste flow reduction) of these methods are included in Model III, so that in effect, a single opportunity cost of reprocessing each waste stream is introduced. The estimated coefficients for the three reprocessing methods are presented in Table 2.

The cost of water purification depends on the nature of the waste content, and an accurate model would have a range of water treatment prices. For simplicity, the present model uses a single reprocessing method; primary, secondary, and tertiary treatment of municipal waste water.[8]

It is assumed here that a substantial quantity of solid waste can be used to generate electricity. The City of St. Louis, with partial support of a federal grant, is constructing pilot facilities to prepare municipal refuse for use as fuel by Union Electric Co. The project is based on a feasibility study prepared by Horner and Shifrin, Inc. [16] The cost for this reprocessing method is the estimated cost of preparing, transporting, and firing the refuse in the Labadie power plant, less the value of recovered heat and less the avoided costs of landfill disposal.

It is assumed that any land waste generated by air pollution abatement could be offset by the diversion of municipal waste from landfill to utilization as fuel. The land waste output for this reprocessing method is −.7 tons.[9] This is based on the assumption that the metallic and ash content, approximately 30% by weight, must still be buried. Actually, there is a good possibility that this residue will be recycled, in which case the land waste output coefficient should be −1.

The social anxiety over solid waste is based in part on (1) the mounting relative costs of pollution-free disposal, (2) the possibility that resources are being too rapidly depleted and that wastes should therefore be recycled,[10] and (3) concern that landfill limits the future use of land and frequently destroys so-called wastelands, which have an important ecological role. The use of solid waste as fuel is a partial answer to each of the above. A ton of prepared refuse replaces .4 tons of coal, thereby conserving a resource[11] and slowing the rate of environmental conversion by strip mining.[12]

The reprocessing method for thermal pollution is a hyperbolic natural draft cooling tower for the Labadie power plant. It is assumed here that the problem of thermal pollution is eliminated if the waste heat from power production is discharged to the atmosphere.[13]

If the unit cost of reprocessing is divided by the output coefficient, the shadow price for the external waste constraint, which is a part of the computer program output, can be independently determined.[14] For thermal pollution, the shadow price, using data from Table 2, is \$.20/4.6 = \$.04 per million Btu's. The shadow cost for solid waste is probably an upper bound. It is likely that there are situations where producers could effect substitutions resulting in reduced solid waste for less than \$1.40 per ton.

The cost of reprocessing the external waste joint-outputs in the solution of Model III exceeds \$500,000. By imposing costs on what were free environmental services in Model I, some substitution away from the external waste outputs would be expected in Model III. This substitution is not readily apparent in Table 1; in fact, the quantities of liquid and thermal wastes reprocessed are somewhat greater than their outputs in Model I. However, the land waste joint-output of air pollution abatement drops to 378,000 tons. (Reprocessing 539,474 tons of solid waste eliminates 378,000 tons of land waste, assuming a metallic and ash content of 30%). Valuing the joint-outputs at their shadow prices in Table 1, there is a decline in the value of the three joint-outputs of air pollution abatement in Model III.

The reduction in solid waste output in Model III can be traced to a change in the optimal set of control methods. A control method in which on-site open burning is replaced by sanitary landfill disposal is no longer optimal in Model III, as it was in Model I. The cost imposed on incremental landfill waste now offsets the air pollution benefits for this particular control method.

The marginal cost of controlling particulate matter is higher in Model III than Model I. The optimal solution, as a consequence, includes three additional control methods for abatement of particulates. It is of interest that two of these control methods do not increase solid wastes; they apply to grain processing and cement manufacturing, where the recovered particulate matter is saleable output.

The optimal quantity of natural gas increases very little in the model with costs on external wastes over that in Model I. In view of the fact that conversion to gas reduces the external wastes (note the great increase in natural gas usage in Model II), one might expect an increased emphasis on this fuel in Model III. However, it appears to be more economical to reduce external wastes by reprocessing them, rather than by converting additional boilers to natural gas.

The fact that there are some changes in the optimal set of air pollution control methods, as a consequence of imposing costs on what were formerly free services, suggests that the cost of air pollution abatement will be higher in Model III than Model I. Table 1 indicates that the cost increase is \$60,000. This relatively small increase suggests that the solution of Model III corresponds, figuratively, to a tangency solution close to point A in Figure 1.

Model IV: The Augmented Model with Feedbacks

In Model IV, the feedbacks associated with reprocessing the external wastes are introduced. This adds more complexity to the model than the simple inclusion of opportunity costs. In the case of water treatment, some solids remain after digestion of sludge, and there is additional air pollution and thermal waste associated with the electric power requirements. It is estimated that the power to reprocess one thousand gallons of water requires the additional combustion of .004 tons of coal at the Sioux power plant.[15]

There is a net increase in electric power requirements associated with the grinding and shredding to reprocess solid wastes. However, the increased air pollutants reflect not only the incremental power generation at the Sioux plant but also the difference in emissions when four-tenths of a ton of coal are replaced in the Labadie power plant by a ton of prepared refuse. In the case of sulfur dioxide, the low sulfur content of refuse as compared to coal results in this reprocessing method being, on balance, a

[8] The coefficients for this reprocessing method are based principally on data in references 11 and 14. It is by no means clear that this is an optimal reprocessing method; Kardos, for example, suggests that municipal waste water should receive only primary and secondary treatment and then be recycled for use as enriched irrigation water for farmlands. See reference 4.

[9] Note that coefficients which indicate a reduction in air pollution are positive whereas those indicating a reduction in the external wastes are negative.

[10] The contention of increasing resource scarcity is examined and challenged by Barnett and Morse. See reference 2, chapters 1, 8, and 9.

[11] It is conceivable that the present value of prepared municipal refuse is *more* valuable than coal. Hart advocates that municipal refuse be composted and utilized to maintain and improve agricultural land. See reference 3, pp. 29-32.

[12] Krutilla suggests, in effect, that there is a socially optimal rate of environmental conversion for the production of consumption goods. See reference 10, p. 785.

[13] There may be serious ecological problems associated with the extensive evaporation of water from wet cooling towers. More research on this subject is needed, if cooling towers are to be used on a large scale basis. The coefficients for this reprocessing method are based on data in reference 15, p. 302.

[14] This is not quite true for the solid waste reprocessing method. The base level of pollution control in the St. Louis Airshed included the burning of approximately 5% of landfill waste. [6, p. 334] The \$.80 cost for the reprocessing method is based on a credit for the cost of sanitary landfilling, which is more expensive than landfilling in which part of the bulk is reduced by burning. The way the model is set up, when landfill is diverted to fuel, there is a corresponding elimination of the 5% open burning. In effect, the cost of the reprocessing method includes, at the margin, an additional \$.72, the cost to eliminate the open burning less \$.54 which is the value of the foregone pollutants from landfill burning times the duals of the pollutants. The shadow price of the land waste constraint is therefore,

$$(\$.80 + \$.72 - \$.54)/(.7) = \$1.40.$$

[15] The incremental air pollution associated with the combustion of .004 tons of coal at the Sioux power plant is actually more than the quantities indicated in Table 2. The original and subsequent models incorporate a 41% reduction factor for emissions from stacks 600 feet high. This correction to "effective" emissions is based on data in reference 12.

method of reducing sulfur dioxide. The decrease is relatively small, however, because the refuse is burned in combination with coal, and a stack gas desulfurization process is simultaneously optimal.

The added pumping power and reduced turbine efficiency associated with the hyperbolic natural draft cooling tower involve increased power generation of one percent of plant capacity. [15, p. 302] Assuming that one ton of coal generates 2500 kilowatt hours, this is equivalent to .004 tons of coal combustion for each thousand kilowatt hours of electrical output in a power plant equipped with the cooling tower.

The major feedback effect is the additional power required for the three reprocessing methods. This results in additional combustion of 6600 tons of coal at the Sioux power plant. The consequent addition to the burden of air pollution control is reflected in the slightly higher cost of air pollution abatement for Model IV as compared to Model III.[16]

The cost of reprocessing external wastes increases in Model IV, mainly because the power requirements for the additional air pollution abatement activity increases the net quantity of thermal discharge. This increase in the cost of reprocessing external wastes is moderated by a decline in the joint output of solid waste; this decline is a by-product of the increase in natural gas replacing coal for air pollution abatement in Model IV.[17]

It is somewhat surprising that the relatively substantial feedbacks in Model IV produced very little change in the optimal solution. This may be due to the above observations that some of the feedbacks tended to offset each other. The marginal costs of sulfur dioxide and particulate abatement are virtually unchanged in Model IV, and a comparison of optimal control methods for Models III and IV indicate that with one minor exception, they are identical.

Conclusions

The reader is cautioned that the results in this paper are based on a least cost solution of a specific model for a single airshed, given a particular set of air quality goals. Any general conclusions based on these results must be considered in the light of this limitation.

The solution of Model I indicates that air pollution abatement is likely to generate alternative waste flows. The most significant of these was found to be incremental land wastes, representing more than a 50% increase in the projected landfill tonage in the St. Louis Airshed in 1975.

To arbitrarily prohibit incremental waste joint-outputs would increase substantially the cost of air pollution abatement; this is borne out by the results in Model II. A conclusion for policy making is that air pollution control strategy should take into account the possibilities for reprocessing external waste outputs.

The Ayres and Kneese warning of serious errors in a partial equilibrium approach may be tentatively challenged in view of the results for Model III. The increase in total cost of the optimal air pollution control methods, when external wastes are priced, is barely .2%, while the additional costs for reprocessing are less than 2% of total costs.

However, the models which incorporate the reprocessing methods are an improvement over the original model. The solid waste problem that is intensified when open burning or incineration is prohibited is given quantitative significance in the augmented models. The solutions for Models III and IV give some preference for air pollution control methods in which recovered particulate matter is recycled. In addition, the problem of relatively minor external wastes associated with certain sophisticated control methods, such as the dolomite wet scrubbing process for power plants, is put into proper perspective.

An unexpected finding was that, while conversion from coal to natural gas is of considerable economic value for air pollution abatement, its marginal value in reducing the external wastes was small. It costs less to reprocess incremental wastes than to reduce them by further conversions to natural gas. While this finding can hardly be generalized, it does serve to illustrate how more fully quantified models can avoid errors. On the basis of Model I and an intuitive weighting of the secondary benefits of waste reduction for natural gas, one might have been inclined to arbitrarily augment the optimal quantity of natural gas.

The similarity in results for Models III and IV suggests that the complicated feedbacks for the reprocessing methods might be ignored in future models, and simple opportunity costs for the external wastes suffice. However, considerably more environmental planning is necessary to determine the proper opportunity costs to be used.

Although the present model is a long way from the general equilibrium, total environment model envisaged by Ayers and Kneese, it is an advance beyond the isolated air pollution model, which these authors have challenged. [1]

MATHEMATICAL APPENDIX

There are M pollution sources, $s_1, s_2, \ldots, s_M$, in an airshed. Table 3 pertains to one such source; the combustion of refuse in flue-fed incinerators. It is projected that 34,000 tons of refuse will be burned in this type of incinerator in the St. Louis Airshed in 1975 [6, p. 318]. Based on the emission factors in the first column of Table 3, this source would generate 918,000 pounds of carbon dioxide, 68,000 pounds of hydrocarbons, etc. in that year.

The model contains N variables, $x_1, x_2, \ldots, x_N$, which represent the activity levels of air pollution control methods. These variables are characterized by cost, source (or input), and pollution abatement coefficients such as those illustrated in Table 3. Because of considerable variation in operating rates and types of refuse burned in flue-fed incinerators, the cost and pollutant coefficients for this particular source are crude averages at best. Although this is a relatively minor source of pollution in the St. Louis Airshed and a particularly difficult one to quantify as well, it is useful for illustrating the model.

The dollar cost for any set of control method activity levels is

$$1) \qquad \sum_{j=1}^{N} c_j x_j \ ,$$

where c_j is the unit cost of control method j. For example, the \$2.80 unit cost of converting from incineration to landfill disposal represents the capital and labor expended to collect and dispose of a ton of refuse in a sanitary landfill, less the avoided costs of incineration. A limitation of the model is the assumption of constant costs; *i.e.* that unit cost, c_j, is independent of the corresponding activity level, x_j.

The source, or input, constraint precludes the sum of activity levels of control methods from exceeding the magnitude of the source for which they are defined. Thus the combined activity levels of the four control methods in Table 3 cannot exceed 34,000 tons in the 1975 solution. This constraint is generalized by the equations,

$$2) \qquad \sum_{j=1}^{N} a_{ij} x_j \leqslant s_i \qquad\qquad (i = 1, 2, \ldots, M) \ ,$$

where a_{ij} is unity when control method j is defined for source i and zero otherwise.

[16]This cost reflects not only the increased air pollution control activity for the Sioux power plant but for other sources in the airshed as well. The model is based on an assumption of maximum allowable total emission flows for each pollutant in the airshed, and since the collection efficiencies of the optimal controls for the Sioux plant are less than 100%, the residual incremental emissions must be offset by increased abatement activity by other sources. In Model IV, this is accomplished by the conversion of additional boilers from coal to natural gas, as evidenced by the increase in the natural gas indicator in Table 1. The net increase in abatement cost would be at least \$20,000 more than shown, were it not for the offsetting credit from the reduction in sulfur dioxide associated with the burning of prepared municipal refuse in place of coal.

[17]This reduction of solid waste joint-output would account for the fact that the shadow price for the solid waste constraint does not increase in Model IV, as it does for the liquid and thermal constraints. These rises reflect the additional costs for the feedbacks associated with the reprocessing methods.

TABLE 3. COEFFICIENTS CHARACTERIZING A SET OF AIR POLLUTION CONTROL METHODS FOR FLUE-FED INCINERATORS

(the control method activity unit is one ton of refuse throughput controlled)*

Emissions and Waste per ton of refuse burned, without air pollution control	Description and symbol of the coefficient	Wet scrubber	Afterburner	Combination of wet scrubber and afterburner	Substitution of landfill disposal for incineration
	Abatement cost per ton of refuse controlled (c_j)	$4.00	$3.50	$7.50	$2.80
	Refuse throughput per unit of control method activity (a_{ij})	1. ton	1. ton	1. ton	1. ton
27. pounds	Abatement of carbon monoxide (b_{pj})	0 pounds	27. pounds	27. pounds	27. pounds
2. pounds	Abatement of hydrocarbons (b_{pj})	0 pounds	2. pounds	2. pounds	2. pounds
.3 pounds	Abatement of nitrogen oxides (b_{pj})	0 pounds	−9.7 pounds	−9.7 pounds	.3 pounds
.2 pounds	Abatement of sulfur dioxide (b_{pj})	0 pounds	0 pounds	0 pounds	.2 pounds
28. pounds	Abatement of particulates (b_{pj})	23.8 pounds	21.0 pounds	26.95 pounds	28. pounds
	Waste water generated per activity unit (d_{kj})	.43 thousand gallons	0 thousand gallons	.43 thousand gallons	0 thousand gallons
.235 tons of landfill ash	Land waste generated per activity unit (d_{kj})	0 tons	0 tons	0 tons	.765 tons
	Thermal discharge associated with electrical requirements of the control method (d_{kj})	.0184 million Btu	.0166 million Btu	.035 million Btu	0 million Btu

*Sources: Duprey, R. L., *Compilation of Air Pollutant Emission Factors,* N.A.P.C.A., Durham, N. C., 1968, pp. 9, 10.

Kaiser, E. R., et. al. "Modifications to Reduce Emissions from a Flue-Fed Incinerator," *A.P.C.A.J., 10,* June 1960, Table V, p. 190.

Kaiser, E. R., "Refuse Reduction Processes" in *Proceedings, The Surgeon General's Conference on Solid Waste Management for Metropolitan Washington, July 19-20, 1967,* U.S.P.H.S., Cincinnati, Ohio, p. 98.

Zinn, R. E. and Niessen, W. R., "Commercial Incinerator Design Criteria," *Proceedings of the 1968 National Incinerator Conference,* May 5-8, A.S.M.E., New York, 1968, p. 343.

Control Techniques for Particulate Air Pollutants, N.A.P.C.A., Washington, D. C., 1969, p. 165.

Reference 6, pp. 317-323; reference 15, p. 291.

Associated with each control method is a set of pollutant abatement coefficients. For example, the particulate coefficients contained in Table 3 are based on assumed reduction efficiencies of 85% for the wet scrubber, 75% for the afterburner, 85% plus 75% X 15% for the two control methods combined, and 100% for landfill disposal.[18] The negative coefficients shown in the table indicate that *additional* nitrogen oxides are formed in the more intense heat of the afterburner. The inequality,

$$3)\quad \sum_{j=1}^{N} b_{pj}x_j \geq r_p \qquad (p = 1, 2, \ldots, P),$$

where b_{pj} is the reduction in pounds of pollutant p obtained with a unit of control method j activity, requires total pollution abatement for each of P pollutants to be no less than some specified quantity.

The required reduction, r_p, for any pollutant is the excess of anticipated annual emissions of that pollutant in the airshed over an allowable flow. The anticipated emission flows for 1975 were calculated by projecting each of 94 pollution source magnitudes (these include 10 categories of transportation, 7 classifications of power plants, 32 types of stationary fuel burning installations, 9 refuse burning activities, and 36 industrial processes) to the year 1975 and multiplying each quantity times a corresponding set of emission factors, such as those illustrated in the first column of Table 3.

The concept of an allowable flow is based on a simplifying assumption by Zimmer and Larsen that the annual average ambient air concentration of a pollutant at some central monitoring station, less the background concentration, is directly proportional to total emissions of that pollutant in the airshed [13]. The allowable flows in the present model are calculated from air quality goals in Table 4 and from data for 1963 (or 1964 where 1963 data are not available) according to the formula in a footnote to the table.

It is a limitation of this model that all sources are considered equivalent regardless of their location in the airshed. Models are now being developed which relate emissions at specific locations in an airshed to ambient air concentrations at various receptor stations. Such models, incorporating meteorological characteristics and atmospheric chemical interactions, should eventually provide more definitive economic solutions than does the present model.

The K external waste outputs associated with a set of air pollution control method activity levels are given by the equations,

$$4)\quad \sum_{j=1}^{N} d_{kj}x_j = w_k \qquad (k = 1, 2, \ldots, K),$$

where d_{kj} represents the quantity of waste k associated with one unit of control method j activity. Thus the wet scrubber in Table 3 generates .43 thousand gallons of waste water per ton of refuse throughput, and an additional power requirement of 4 kilowatt hours for the scrubber system involves an incremental thermal discharge of 18,400 Btu to the cooling waters at the power plant. Because incineration generates approximately .235 tons of ash for landfill disposal, the incremental solid waste when the incinerator is discontinued is less than one, or an estimated .765 tons.

[18]The latter control method involves emissions associated with the collection and transport of refuse. Unfortunately, these were neglected when the model was set up.

The complete linear programming model, with the non-negative constraint, has the form

$$5)\quad \begin{array}{lll} \text{Minimize} & \sum_{j=1}^{N} c_j x_j & \\ \text{Subject to} & \sum_{j=1}^{N} a_{ij}x_j \leq s_i & (i = 1, \ldots, M) \\ & \sum_{j=1}^{N} b_{pj}x_j \geq r_p & (p = 1, \ldots, P) \\ & \sum_{j=1}^{N} d_{kj}x_j = w_k & (k = 1, \ldots, K) \\ & x_j \geq 0 & (j = 1, \ldots, N). \end{array}$$

In Model I of this paper, the w_k are unconstrained variables and the least cost combination of control methods is found which satisfies the remaining right hand side constraints, including the values of r_p from the final column of Table 4. In Model II, the w_k are set equal to zero.

The augmented models, in which the external wastes are reprocessed, contains K additional activity variables, one for each of the external waste outputs. The augmented formulation is

$$6)\quad \begin{array}{lll} \text{Minimize} & \sum_{j=1}^{N} c_j x_j + \sum_{j=N+1}^{N+K} c_j x_j & \\ \text{Subject to} & \sum_{j=1}^{N} a_{ij}x_j - \sum_{j=N+1}^{N+K} a_{ij}x_j \leq s_i & (i = 1, \ldots, M) \\ & \sum_{j=1}^{N} b_{pj}x_j - \sum_{j=N+1}^{N+K} b_{pj}x_j \geq r_p & (p = 1, \ldots, P) \\ & \sum_{j=1}^{N} d_{kj}x_j - \sum_{j=N+1}^{N+K} d_{kj}x_j = 0 & (k = 1, \ldots, K) \\ & x_j \geq 0 & (j = 1, \ldots, N) \end{array}$$

In Model III, the a_{ij} and b_{pj} for $j = N+1, \ldots, N+K$, and the non-negative d_{kj} for $j = N+1, \ldots, N+K$, are set equal to zero. In Model IV, the feedback relationships for the K reprocessing methods are introduced. For the liquid waste reprocessing method, the increased coal consumption at the Sioux power plant is expressed by the coefficient,

$$7)\quad a_{ij} = .004,$$

the increased hydrocarbon emissions, by the coefficient

$$8)\quad b_{pj} = -.0005,$$

and the increased solid waste, by the coefficient

$$9)\quad d_{kj} = .00045.$$

TABLE 4. ALLOWABLE ANNUAL EMISSION FLOWS, ANTICIPATED FLOWS IN THE ST. LOUIS AIRSHED IN 1975, AND REQUIRED REDUCTIONS*

Pollutant	Emissions in 1963 in million pounds	Annual average ambient air concentration in 1963 (or 1964)	Annual average ambient air quality goal	Background concentration	Allowable annual flow in million pounds	Anticipated emission flow in million pounds in 1975	Required abatement in million pounds in 1975
	f_p^0	q_p^0	q_p^*	b_p	f_p^*	f_p	r_p
carbon monoxide	2920	6.3 ppm	5. ppm	0	2335[a]	4200	1865
hydrocarbons	995	3.1 ppm	3.1 ppm	1.5 ppm	995	1520	525
nitrogen oxides	305	.069 ppm	.069 ppm	0	305	415	110
sulfur dioxide	1180	.059 ppm	.02 ppm	0	400	1390	990
particulates	300	128 $\mu g/m^3$	75 $\mu g/m^3$	31 $\mu g/m^3$	135	300	165

Sources: Reference 6, pp. 445-457; reference 18, p. 24; also, *Air Quality Data from the National Air Sampling Networks and Contributing State and Local Networks, 1964-1965,* U.S.P.H.S., Cincinnati, Ohio, 1966, pp. 4, 25.

Formulas: $f_p^* = f_p^0 (q_p^* - b_p) / (q_p^0 - b)$

$r_p = f_p - f_p^*$

[a]The allowable flow for carbon monoxide is based on the concentrations, $q_p^0 = 75$ ppm and $q_p^* = 60$ ppm, which represent maximum one hour concentrations in traffic. Although the measures for carbon monoxide were subsequently changed to the above annual averages, resulting in a slightly different value for f_p^*, the original allowable level was used.

REFERENCES

[1] Ayres, R. U., and Kneese, A. V., "Production, Consumption, and Externalities," *American Economic Review,* 59 (1969) 282-297.

[2] Barnett, H. J., and Morse, C., *Scarcity and Growth: The Economics of Natural Resource Availability,* Baltimore: Johns Hopkins, 1963.

[3] Hart, S. A., *Solid Waste Management/Composting; European Activity and American Potential,* Cincinnati, Ohio: U. S. Department of Health, Education, and Welfare, 1968.

[4] Kardos, L. T., "A New Prospect," *Environment,* 12 (1970) 10-27.

[5] Kohn, R. E., "Abatement Strategy and Optimal Air Quality Standards," in A. Atkisson, R. S. Gaines, ed., *Development of Air Quality Standards,* Columbus, Ohio: Merrill Publishing Co., 1970, pp. 103-122.

[6] ——, *A Linear Programming Model for Air Pollution Control in the St. Louis Airshed,* unpublished doctoral dissertation, Washington University, St. Louis, Missouri, 1969.

[7] ——, "Application of Linear Programming to a Controversy on Air Pollution Control," *Management Science: Applications,* June, 1971.

[8] ——, "Linear Programming Model for Air Pollution Control: A Pilot Study of the St. Louis Airshed," *Air Pollution Control Association Journal,* 20 (1970) 78-82.

[9] ——, "Optimal Air Quality Standards," *Econometrica,* November, 1971 (tentative date).

[10] Krutilla, J. V., "Conservation Reconsidered," *American Economic Review,* 57 (1967) 777-786.

[11] Slechta, A. F., et. al., "Water Reclamation Studies of the South Tahoe Public Utility District," *Journal of Water Pollution Control Federation,* 39 (1967) 811-812.

[12] Smith, M. E., "Reduction of Ambient Air Concentrations of Pollutants by Dispersion from High Stocks," *Combustion,* (April, 1967) 23-27.

[13] Zimmer, C. E. and Larsen, R. I., "Calculating Air Quality and its Control," *Air Pollution Control Association Journal,* 15 (1965) 565-572.

[14] *Cleaning Our Environment: The Chemical Basis for Action,* Washington, D. C.: American Chemical Society, 1969.

[15] "Problems in Disposal of Waste Heat from Steam-Electric Plants," (Federal Power Commission Staff Study - 1969),*Selected Materials on Environmental Effects of Producing Electric Power,* Joint Committee on Atomic Energy, Congress of the United States, Washington, D. C., (August, 1969) 285-335.

[16] *Study of Refuse as Supplementary Fuel for Power Plants,* St. Louis, Missouri: Horner and Shifrin, Inc., Consulting Engineers, 1970.

DISCUSSION

Peter H. Pearse, Department of Economics, University of British Columbia

The Hell's Canyon issue, analysed by Krutilla, exemplifies the growing problem of conflict between industrial exploitation of natural resources and preservation of their aesthetic and recreational values. This particular example embodies two special features important to the analysis; (i) the alternatives between "development" and "preservation" are mutually exclusive, and (ii) the decision to "develop" (though not to "preserve") is irreversible. These features give the problem its special interest, because the first does not lend itself easily to traditional marginal analysis, and the second is inconsistent with the usual assumptions in optimal growth theory wherein both investment and disinvestment are continuously possible.

Krutilla's contribution is substantial: it is an important extension of the literature on benefit-cost analysis (which already owes a good deal to him); it clarifies some of the perplexing conceptual problems in evaluating non-priced services of resources; and it is an ingenious demonstration of how careful identification of threshold values can short-cut some of the worrisome difficulties in evaluating environmental benefits. In view of the increasing frequency with which development proposals of this kind are being encountered, it is an eminently timely contribution.

Much of Krutilla's paper deals with how the assymetric time trends in the values for for the two alternative allocation regimes can be identified and worked into the benefit-cost framework in the context of the Hell's Canyon proposal. I should like to comment here on a couple of theoretical issues involved in the analysis of preservation values.

Krutilla postulates a conventional demand curve for recreation days at the site; "... the schedule which a discriminating monopolist could exact as prices ..." Under free access, the total value accrues to the recreationists as consumer surplus.

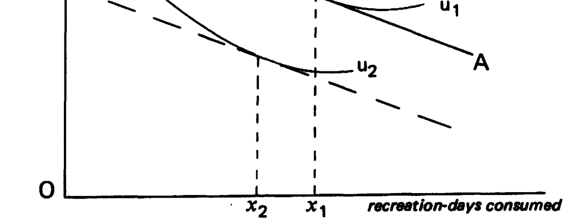


Figure 1.

In Figure 1, OY is the recreationists' total income, YA is a budget line the slope of which reflects the marginal cost of a recreation day (assumed constant), and u_1 and u_2 are indifference curves. Standard analysis suggests that the recreationist would consume Ox_1 recreation days.

Following Hicks, we can identify at least two measures of consumer surplus. The amount by which the consumer's income would have to be increased to fully compensate him for exclusion from the recreational opportunity, YZ, is the "price equivalent" measure of consumer surplus. The maximum amount the consumer would be willing to pay to retain the privilege of using the recreational site is YV, the "price compensating" measure.[1] Krutilla chooses the latter (which for normal goods is smaller) although it is probably the less appropriate for evaluating an existing asset.[2]

The difficulty arises in integrating this concept of consumer surplus into an aggregate demand curve, and inferring from it the effect of pricing. Krutilla has, on the horizontal axis of his demand curve, the quantity of recreation consumed — presumably measured in recreation-days. But note that if all or part of the recreationists' consumer surplus — YV in Figure 1 — was actually appropriated through some levy, the recreationist would alter his consumption from x_1 to x_2. The extent to which he would adjust consumption depends upon the marginal cost of a recreation-day and the shape of the indifference curves, but (unless recreation is a Giffen good) he will consume less.

To be consistent, then, the units of demand should be expressed not in numbers of recreation-days but in numbers of recreationists. Such a demand curve would represent the aggregate of recreationists' consumer surpluses YV in Figure 1 and would thus measure the maximum amount they would collectively be prepared to pay to retain the recreational opportunity. Conceptually, a perfectly discriminating monopolist could exact this amount in sales of visitor licenses (as opposed to a per-day charge) without eliminating any visitors, although each would consume less. This is important, because Krutilla goes on to discuss the implications of crowding and the control of crowding through price-rationing. Crowding, however, is a function of the number of recreation-days consumed.

Krutilla's concept of the carrying capacity of a recreational site implies that the quality of the recreational experience is adversely affected by crowding externalities. This is an awkward phenomenon to deal with in terms of conventional demand analysis. A given demand curve must relate to a product of constant quality, and hence any change in the degree of congestion calls for a new demand curve.

In Figure 2, D_1 represents the demand curve for recreation at an uncrowded (high-quality) site. D_2, D_3 and D_4

represent the demand curves for the same site under successively more crowded conditions. Here, demand is expressed in units of consumption — say visitor days per year — and quality is measured by some index of crowding — such as visitors per mile of trail.[3]

Krutilla implies only one level of quality, constant up to the fixed limit of the site's capacity. But while the site might be assumed fixed (in the Ricardian sense) it is more realistic to regard the capacity of the site to accommodate visitors as amenable to expansion at additional cost (e.g., by building more miles of trail).

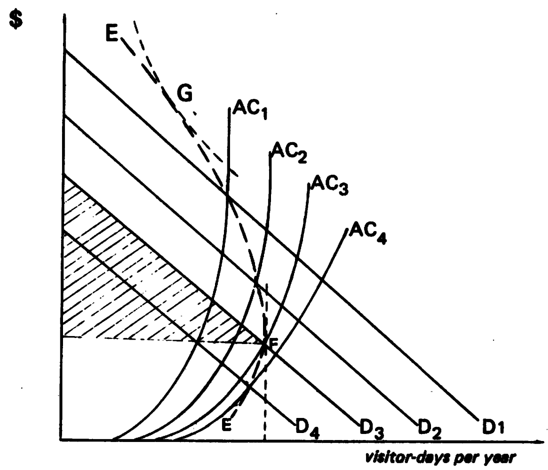


Figure 2.

Thus we might impose on Figure 2 a sheaf of upward-sloping average cost curves ($AC_1 \ldots AC_4$); one corresponding to the quality underlying each of the demand curves. There is now a pair of average cost and revenue curves for each quality standard, and their intersection indicates the price that would be dictated by average cost pricing.[4] If both sets of curves are symetrically shaped, a line joining these equilibrium prices will curve outward from the vertical axis like EE in Figure 2.[5]

Now consider alternative objectives, and the implications for choice among the various prices and quality standards. To maximize gross revenues, we would choose the regime indicated by the tangency of a rectangular hyperbola with EE (i.e., point G). To maximize total use, we would choose that indicated by the point of tangency of a vertical line with EE (i.e., point F). If the demand curves were parallel, we would choose the same point (F) to maximize net benefit: since the price just covers costs for all points along EE, net gain is in the form of consumer surplus, and the area under the demand curve above price will be maximized by the triangle that extends furthest to the right (shaded in Figure 2). It seems likely, however, that the demand for higher quality recreation would be less price-elastic, in which case consumer surplus would be maximized at a price higher than that consistent with the densest use.

Space prevents further exploration of this approach here, but the question of congestion (and other aspects of quality) deserves explicit analysis in assessing the benefits of particular recreational resources.

FOOTNOTES

1 See Peter H. Pearse, "A New Approach to the Evaluation of Non-Priced Recreational Resources," *Land Economics* 44(1):87-99.

2 See Krutilla *et al*, "Observations on the Economics of Irreplaceable Assets" (unpublished manuscript: Resources for the Future Inc., September 1970)

3 It may appear paradoxical that the quantity demanded at any price is lower under crowded conditions when it is greater consumption which causes the crowding itself. But the demand schedules along tell us nothing, of course, about what is possible to attain in the way of quantity and quality.

4 Alternatively, of course, we could demonstrate marginal cost prices.

5 I am indebted to Gideon Rosenbluth for this demonstration.

DISCUSSION

Paul B. Downing, University of California, Riverside

I find Mr. Kohn's paper of considerable interest. It is a substantial step away from the partial equilibrium analysis that has been all too prevalent in studies of environmental quality issues. Lack of recognition of the effects of control of the flow of pollutants to one resource on the production of pollutants released to another source can (and has) lead to unsatisfactory results. For example, a few years ago stringent controls were placed on water pollution on the Wisconsin River. These controls stimulated some paper manufacturers to change processes from one which polluted the water to one which polluted the air. The result is obvious to anyone who has passed within a few miles of one of those plants.

As with any initial break with traditional methods, the author leaves several questions unanswered or incompletely answered. There are three areas I wish to emphasize: factors affecting Mr. Kohn's estimates of the cost of control, the trade-off between air pollution and other forms of pollution, and the desired level of air quality.

I. Cost of Control

The first point I would like to stress is the apparent lack of inclusion of possible major process shifts by the manufacturing sources. Mr. Kohn has included such minor process shifts as changes in energy source within the existing plant but has not as far as I can tell included alternatives which would require substantial change-over costs or completely new construction. This is especially relevant in the context of an analysis in 1970 of how to meet the air quality standards in 1975. The planning period allows sufficient time for such process changes. If, for example, the power industry were contemplating the construction of new or expanded facilities in 1971, these facilities could be designed to meet the standards. Inclusion of air pollution control in the design process is likely to be cheaper than the exclusive use of add-on devices envisioned here. To the extent that these possibilities are not included the cost estimates presented represent an upper bound.

The assumptions of the model allow for the use of one or another control method on a source or possibly no control at all but they do not allow for the complete elimination of all or part of a source. That is, there is not a set of control levels (b_{pj}) and corresponding costs (c_j) which allow for 100 percent control via closing down one or more plants. The problem of estimating the c_j's would be complex and this set of alternatives may not be economically or politically viable but the matter deserves some attention. These alternatives may be more efficient than controls on existing sources. This fact provides another argument for the cost estimates being over estimated.

An additional cost not included in Mr. Kohn's estimates is the cost of inspecting and enforcing control efforts. Enforcement is often costly and measurement of compliance may be difficult or impossible. Furthermore, these measurement problems and enforcement costs vary with the type of source being controlled. Recognition of this problem would cause an upward adjustment in the c_j's and to the extent enforcement is not perfect some decrease in the b_{pj}'s.[1/]

It is likely that the controls instituted will increase costs of production and hence, through prices, the output and level of pollution production of the industries. This effect is not included. Estimating the extent of this type of feedback effect would be difficult and time consuming and may not be worth the effort but it ought to be given some consideration.

A version of the model might be developed which introduces institutional and political constraints. Such constraints might be in terms of maximum expenditures for one or another source or unequal air quality standards in different political jurisdictions. The effect of such constraints on the cost of control and the sources controlled might provide useful information to policy makers and to the public. An example of such a constraint in California is the legal requirement that retrofit control devices for used cars cost less than $60.00 to purchase and install. This requirement restricts mobile source control alternatives sharply and is probably increasing the cost of control substantially.[2/]

Finally, there is no account taken of the timing and location of pollutant release. The damages of air pollutants are more a function of the intensity and duration of exposure to a pollutant than its average annual level. Since the presumed goal of control is to reduce these damages, some model building efforts along these lines would be very helpful. To the extent control efforts are to reduce peaks rather than averages, the cost and level of control is likely to be greater than that determined by the model presented here. Taking all the locational, chemical, timing, and meteorological factors into account requires resources and knowledge far beyond that available to an individual researcher. A well funded multidisciplinary effort is required to make substantial progress along these lines. Mr. Kohn can certainly be excused from not taking this task on by himself.

II. The Desired Level of Air Quality

The air quality standards used in this paper and set by the Missouri Air Conservation Commission are not likely to be set at the most desirable level because they were set in ignorance of the results of this paper. It is my belief that if standards are to be used some recognition of their economic effects and the cost of reaching the standards should be included in the decision-making process. An expanded version of the model developed by Mr. Kohn would serve this purpose well. Different levels of control (air quality standards) could be imposed on the model and their effects on the cost of control, the level of control by source, and their expected effect on economic activity and employment could be estimated. These effects could then be weighed by the decision-makers against the damages averted (both health and physical) by that level of control.

Also it should be explicitly realized that there is a possible trade-off among air pollutants. The damages caused by a pound of carbon monoxide (CO) is most likely not equal to that nitrogen oxides (NO_x) or hydrocarbons (HC). Furthermore, NO_x and HC combine in the presence of sunlight to produce ozone which itself causes damages. The amount of ozone produced depends in a non-linear way upon the proportion of NO_x to HC present as well as their levels. This fact makes the solution of the linear programming model much more complex. It also makes the setting of standards based on ambient air levels of chemically unstable compounds tenuous. Some recognition of these chemical reactions and the damages resulting from their secondary products will likely yield better standards than we now have and a different solution to Mr. Kohn's model.

III. The Trade-off Between Air Quality and Other Environmental Quality Parameters.

I would extend this standards setting concept further by adding an analysis of the trade-offs between air pollution and other forms of pollution. Mr. Kohn's work allows us to make such an analysis but he does not do so. Rather, in Models III and IV he uses the criterion that if it costs more to treat and render harmless (100 percent effective treatment?) a secondary pollutant than to stop its production in the air pollution control process, that control process should not be used. It is possible and may indeed be economically rational to say, for example, that the burning of solid wastes as fuel in an electric power generating station should not be done. This would be true if the dollar value (or subjective value felt by the decision-makers) of the land-pollution-caused damages and the cost of the sanitary land fill operation are less than the cost of using the wastes as fuel plus the cost of the additional air pollution control and the residual air-pollution-caused damages. Perhaps it is this trade-off analysis ability, which is the heart of the Ayres and Kneese argument, that is the major contribution of Mr. Kohn's model.

IV. Conclusion

In conclusion, I feel that Mr. Kohn's paper is a substantial step in the right direction. That it does not go far enough is perhaps the fault of the complexity of the problem. We certainly cannot expect one economist, or the whole profession for that matter, to provide all the answers. I feel that the solution to the air pollution problem will be found in the combined coordinated efforts of scientists from many disciplines. Past efforts which excluded social scientists have not been successful. Expanded efforts along the lines of Mr. Kohn's paper will hopefully make future efforts at a solution more successful.

Footnotes:

1/ For an extended discussion of the problems and costs of enforcement under various control schemes see Thomas D. Crocker and A. L. Rogers III, Environmental Economics, (Hinsdale, Ill.: Dryden Press, Inc., Feb., 1971).

2/ Paul B. Downing and Lytton Stoddard, "Benefit-Cost Analysis of Air Pollution Control Devices for Used Cars", University of California, Project Clean Air, Research Reports, Volume 3, September 1, 1970.

QUERIES, QUESTS, QU'EST-CE QUE C'EST?

C. Michael Lanphier, York University

After the first flush of success, the Survey Research Centre is confronted with an overriding query: How does a centre provide to academic and non-profit clients a complete survey research service on a continuing basis, economically, with precision, efficiency, and payoff to the Centre as well as to the client?

Such a query rarely reaches the lips or the pen of the Centre director when discussing business with a potential client, whose project occupies their undivided mutual attention. Such a system-maintenance-and-update function hardly quickens the pulse of a researcher, whose project must occupy the forefront of academic consciousness. On the other hand, members of the Centre express the concern incessantly.

Admittedly the query is complex and intricately woven into the fabric of social science research, funding policies of both sources both outside and within the university; and finally, it is beholden to the line of development to be traced over time by the Centre itself.

For the first, the fabric of social science research, the surveys must contain the dual function of adhering to a social science problem on the one hand and contemporary relevance for Canadian Society on the other. The notion of regionalism takes on great potency in survey research, for example. Two studies in the problems of absorption of immigrants have taken different form when replicated in Alberta metropolitan areas than when the study was conducted in Toronto. Again, the bilingual character of studies which are conducted in Québec and any other province represents cultural differences. A panoply of differences in legal and political structure not only reveals differences in electoral jurisdictional boundaries but also different conceptions of legal offences. The term, "assault & battery" does not exist in French-Canadian parlance, for example.

In the second instance, social science developments have been closely allied to the increasing use of social science data and funding of social science research by the Canadian federal government. Principally, the Canada Council has been charged with subsidizing both social science and the arts. The hospitality of this organization to large-scale survey projects has dramatically widened in recent years, so that most social scientists will seek funding under that aegis alone. Other government departments have allocated significant sums to survey research indirectly. Funding from private foundations has played only a minor rôle, although one which promises increasing significance. Universities have[1] received almost no funds particularly earmarked for research.

Thus, the original query finds its principal constraint in the character of social research in the Canadian scene -- a scene which involves the structures of experience, cultural complexity, and sources of funding. Although the ingredients are familiar to any North American context, the particular configuration assumes uniqueness.

QUESTS

The Centre has opted for the creation of probability sampling frames of households as the most robust and general-purpose type of sampling frame for initial operation, stratified by the federal divisions of electoral districts.[2] Although such an option appears traditional in the development of survey centres, additional factors contributed heavily to this decision. The Census tabulates these areas by household, classified by small Enumeration Areas, which are used by enumerators for convenience. It is thereby possible to "blow up" all sample estimates to population values. Because these electoral districts may be sub-divided into units of just over 300 households, they serve as excellent proxy units to indicate other variables, such as degree of urbanization, ethnic composition or political orientation. Such frames may aggregate when these units are employed. That is, the first sampling frame comprehended 95% of the population in the Province of Ontario. We anticipated that other frames would be drawn in the future for other provinces and for metropolitan areas within the provinces as well. The Electoral District is common to all such frames.

How does such strategy of development affect linkage with another survey unit? In practical terms, this question related to the interest of York University to join forces with an academic survey centre in the province of Québec to provide co-ordinated bi-lingual capability and to allow simultaneously an independence of the two centres.

---

[1] Fred Schindeler and C. Michael Lanphier, "Social Science Research and Participatory Democracy in Canada," Canadian Public Administration, 12 (1969), 481-498.

[2] Eighty-five such districts are found in the Province of Ontario (population more than seven millions). The choice was dictated as well by consideration of temporal continuity of survey data. The task of re-adjustment of boundaries after the decennial census is fairly simple, if time-consuming.

In the fall of 1969, York University and Université de Montréal formulated an accord which arranged for such co-ordination between the Survey Research Centre and the newly-created Centre de Sondage. The pertinent features of the accord include the following:

1. Sampling frames and all pertinent materials for the Province of Québec are developed by Centre de Sondage. Correspondingly, the Survey Research Centre would develop sampling frames for all other provinces, providing access to all such materials on a free-exchange basis.

2. Interviewing and supervision would be geographically divided, with Centre de Sondage in charge of all interviewing in Québec. Standards of fieldwork and supervision are conventually agreed upon. Field supervision is local, regardless of language of interview.

3. Translation and co-ordination of bilingual aspects of surveys fall primarily to Université de Montréal.

4. Coding is centralized at whichever centre originates the study in question, in order to facilitate update of bilingual codebooks.

5. Data are deposited in the data bank of the originating centre, with a copy of the data provided the other centre at cost.

6. Either centre may initiate a contract with a client with work involving the other centre, over which the latter Centre retains the right of first refusal.

7. All time not reserved for co-operative studies will be used at the discretion of the particular centre.

While the first fruits have been reaped from such a union, the initial yield has been small, if promising. That is, the organizations have been chastened in terms of questionnaire design, code co-ordination, and the deposition of data in machine readable form. It is yet to be seen whether such lessons in research conduct generalize beyond the first year of the accord. Yet is is possible to trace significant advances, such as the development of a program of estimation for survey data with appropriate weights attached to each record. This tabulation program has drawn upon the talents of statisticians at both Centres.

QU'EST-CE QUE C'EST?

What organizational implications exist from the specification of a particular case of one survey centre? Among the general conclusions, the following stand prominent:

1. A standard, household sampling frame model, based upon geographical subdivisions has proved its robustness for nine studies in the past two years. Each of these nine required either some modification, or it initiated a new geographical area to be so included. Another ten studies, however, could not take advantage of any such frame, even though five of them have employed probability sampling from some other frame (student lists, lists of universities, assessment rolls, lists of recipients). The remainder of the studies have followed essentially a quota sampling procedure, primarily because of the rarefied nature of the target population (e.g., families who move into selected geographical areas within a city, classified by type of dwelling).

2. The standards developed for presumably fixed items in any survey repertoire shift from study to study. The accompanying list of occupational codes describes what is probably the most uniform question appearing throughout all instruments. The decisions to employ one as opposed to another code depend largely upon the intent of the principal investigator, who intends to compare the data with some other corpus. A standard form of coding would involve a double effort and likely a lower reliability as a result of interference of two similar codes.

3. The accord with Université de Montréal Centre de Sondage stands without precedent, although it sets one which implies a variety of alternative modes of liaison between two universities and geographically-based organizations. There is little question that co-operation in the present instance could have taken no other form. Whether all universities could formulate such bi-lateral (multiple?) accords calls into question the establishment of infrastructure which continues the liaison after the glow of good feeling subsides.

It can be argued in the present case that this accord is unique because of its bilingual-bicultural dimension which squares with geographical districts. It should be emphasized as well that the agreement is bi-lateral, so that co-ordination of field and coding schedules may be arranged with minimal difficulty. Such arrangements might well prove difficult if the co-ordinating task involves the combined forces of several organizations.

4. A recent governmental task force, charged with inquiring into the problems of dissemination of information about the federal government, recommended that a governmental social survey unit be constructed, along the lines of that established in the United Kingdom. While the proposal has met with varying degrees of warmth in reception, a director of the unit is currently being sought. It is likely that any such development of a governmental social survey unit of approximately 100 persons with an infrastructure budget of $\$2\frac{1}{2}$ millions per year would seriously hamper academic developments, since the former would compete for the scarce human resources, whatever long-run advantages might accrue to either or to survey research in general.

5. The challenges for academically-based survey research units lie principally in the development of survey research as itself a legitimate area of academic inquiry. At the same time, the obvious practical advantages which accrue to the conduct of surveys in providing descriptive data, from which all theoretical advances would doubtless originate, means that survey research organizations will probably continue to "bootleg" research under the guise of the quest on behalf of a given principal investigator for purity of his research findings! Although the goals of principal investigators are rarely at odds with those of a survey organization, the hierarchy of benefits clearly differs.

After the initial problem of survival, the more goal-oriented question of research contribution assumes paramount stature for a survey research organization. Because of the intimate connection between the practical and theoretical contributions which survey research makes, the direction of the latter course awaits the interchange between the two sets of researchers: principal investigator and survey researchers. Without the contribution of either, the advances will be halting.

APPENDIX

OCCUPATIONAL CODES USED FROM 1968-1970

I. Project 101, "An Analysis of Attitudes toward Unemployment Insurance Compensation"
Project 106, "Attitudes toward Government Information"
Project 109, "Study of Caloric Intake"
Project 117, "Evaluation Research for Unemployment Insurance Commission"

---

1. Professional
2. Managerial, owners
3. Sales
4. Clerical
5. Skilled Labour
6. Unskilled Labour
7. Farming
8. Retired, Unemployed

II. Project 105, "Political Socialization of Non-party Elites in Canada"

---

0. Professional, technical, and kindred workers
1. Farmers and farm managers
2. Managers, officials, and proprietors, exc. farm
3. Clerical and kindred workers
4. Sales workers
5. Craftsmen, foremen, and kindred workers
6. Operatives and kindred semi-skilled workers
7. Service workers
8. Farm labourers and foremen
9. Labourers, exc. farm & mine

(From 1960 Census of Population, United States)

III. Project 112, "Community Participation in Bowmanville, Ontario"

---

1. Managers
2. Intellectuals
3. White collar workers
4. Skilled blue-collar workers
5. Semi-skilled blue-collar workers
6. Unskilled blue-collar workers
7. Farmers

IV. Project 107, "Attitudes toward Crime and the Police in Toronto"

---

1. Professional
2. Semi-professional
3. Proprietors, managers, officials: large organizations
4. Proprietors, managers, officials: small organizations
5. Clerical and sales
6. Skilled labour
7. Semi-skilled labour
8. Unskilled labour
9. Farming

V. Project 102, "Political Attitudes in Ontario"
Project 110, "Residential Allocations and Preferences"
Project 111, "Social Effects of Housing"
Project 114, "Survey on Neighbourhoods and Population Movement"
Project 116, "Non-medical Use of Drugs"

001-009	1.	Managerial	Administrateurs
101-199	2.	Professional and technical	Professions libérales et techniciens
201-249	3.	Clerical	Employés de bureau
301-339	4.	Sales	Vendeurs
401-459	5.	Service and recreation	Travailleurs des services et des activités récréatives
510-588	6.	Transportation and communication	Travailleurs des transports et communications
601-609	7.	Farmers and farm workers	Agriculteurs et travailleurs agricoles
611-615	8.	Loggers and related workers	Bûcherons et travailleurs forestiers
631-633	9.	Fishermen, trappers, and hunters	Pêcheurs, trappeurs et chasseurs
651-659	10.	Miners, quarrymen, and related workers	Mineurs, carriers et travailleurs assimilés
701-919	11.	Craftsmen, production process and related workers	Ouvriers à la production et travailleurs assimilés
921-970	12.	Labourers, n.e.c.	Manoeuvres, n.c.a.
980	13.	Occupation not stated	Professions non déclarées

(From Dominion Bureau of Statistics Occupation Code, 1961 Census of Canada)

COORDINATION AND COOPERATION AMONG SURVEY RESEARCH GROUPS

Morris Axelrod, Joint Center for Urban Studies of M.I.T. and Harvard University

A. Introduction

We have a rare assemblage here of survey researchers who are gathered to talk not about the substantive findings of our surveys, but instead about operational matters - how to conduct our affairs more effectively - administratively, technically and methodologically.

Most of us are representatives of survey research groups which are typically non-profit and university-connected, and whose research is in the public interest or with scholarly pretensions. We seem to speak the same language and share a sub-culture to a large degree which distinguishes us from other survey groups which are part of the business world.

There are differences which characterize our organizations, and they are many (size, programmatic, emphases, geographic scope, etc.). Despite these, we have a unique opportunity to consider how our organizations may more effectively relate to each other.

I suffer from an arrested Boy Scout mentality coupled with quixotic ideals. I like to think of us as following the time-honored prescriptions of the scientific and scholarly fraternity - that we are dedicated to the furtherance of knowledge, that we are not beset by destructive competition (a little rivalry, perhaps), and that we are prepared to share our experiences and our techniques with our colleagues for the common good (especially when we have more to learn from than to offer our colleagues). We are non-profit, non-competitive. I think we share a serious commitment to advance the state of the art. We (presumably) have no trade or proprietary secrets.

I'm here to make a plea for some degree of cooperation or collaboration beyond what exists now among us. This shouldn't be too difficult. Judging from conversations and discussions going back many years - apparently, no one disputes the importance of this. It's as sacred, honored, respected, unquestioned as motherhood - or, as motherhood used to be. (I'm afraid we'll have to find another symbol). Unlike motherhood, our problem may be sterility, not fertility.

Cooperation may be altruistic, but fortunately it doesn't have to be, because I think that ultimately all of our interests are better served. And the gains from sharing information, practices and services are mutually advantageous.

How many times has each survey organization conducted an independent review of the literature and how often have they scanned scores of questionnaires to see what others have done to measure important variables - conceptually elusive ones like happiness or more objective but still elusive ones like occupation or ethnic origin. There's a tremendous amount of waste and duplicative effort; often what is current and in progress, but not yet reported, is of major interest in survey work. Yet this information is hard to gain.

B. Why Cooperate?

Well, what are some of the plusses from either pooling our resources on occasion, or from sharing information and experiences at all stages of the research process - not just in the final substantive reports?

There are three orders of "sharing". One is the pooling of <u>information</u>; a second is pooling of <u>materials</u>; and the sharing or pooling of <u>personnel</u>, using each other's facilities (information exchange, data exchange, service exchange).

I find it very useful to be kept informed of what others are doing in the survey field - what studies are in progress and what are their principal parameters. More and more often we coincide with other studies in the central city.

In our own community we launched a study of corporate giving. We found some resistance in our sample of large corporations. The business community was being surveyed simultaneously by several different organizations. When we encountered resistance and confusion among our respondents, we learned that they thought we were one and the same with a study being done by the Chamber of Commerce. Then we found we were treading the same ground as ABCD (the Boston Economic Opportunity agency) and thirdly and fourthly somewhat similar efforts by a local large insurance company and another financial institution located elsewhere. Several of our respondents suggested that we get the information from the other research groups since they felt they had answered the same questions, and in many cases they had.

A clearing house or information exchange would lessen duplication of effort. I also learned from Illinois' useful "Occasional Newsletter" of studies which had similar objectives to some we were planning. We can compare notes on procedures, particularly since they represent a departure from our usual household survey in using a telephone interview.

In addition to a report of studies in progress, I would like to see a freer exchange of housekeeping and administrative data and procedures - costs, response rates and experience (not just overall rates, but for varying size communities and samples - density of population), recruitment and turnover, pay strategies, attrition experience, etc. would be most useful to every university-based research organization.

And, in addition to wanting to know what you're up to, what studies you're doing, where, when and the content, and how you handle personnel and pay matters and other procedures - I would like access to the related forms and documents. I am an inveterate scavenger and eclectic.

We can also share our materials and documents and forms - questionnaires, instructional materials, codes, computer programs. I am less concerned with giving another organization access to my formidable superior products and my own discoveries than I am from the threat to me and all of us by debased currency in this field. At least, I think that good methods drive out bad. It's not only a matter of ethics; it's a question of gaining public support and acceptance and I am as much hurt by poor practices of my own staff as I am by those of any other organization. We have to answer for all surveys and polls and their conduct in the field.

This exchange of documents could be accomplished by a literal exchange of documents among us, or by the creation of a library or archives where they are catalogued in such a way as to provide rapid access and retrieval. For example, it should be possible to get all studies of crime or housing or voting behavior or outdoor recreation, or all questions on a particular - education, group membership, etc. quite quickly from such an archive.

None of the above diminish our own stock in trade - they enhance for each of us, I believe, our flexibility, versatility, skills and knowledge. It doesn't entail any departure from what we are doing now.

The kind of cooperation I have described above is sort of minimal in its implications for interaction among us. It does not involve any changes in our current modes of operation unless we choose to be influenced. We do our own thing and we tell each other about it, and if we like what someone else is doing, or someone likes what we do, we borrow or steal from each other.

Now I want to talk about a different order of cooperation - sharing of personnel, services or facilities. This does indeed introduce a new element (not new in the sense that it hasn't been done, but new in that it requires a different kind of interaction and accomodation between groups.)

There may be times when one's own resources are taxed temporarily or when it is more efficient to call upon another organization for help. The usual model is to sub-contract. For example, we are geared to a Metropolitan operation. There are times when we might want to interview in places where we do not have a resident staff or where our own staff is already fully occupied. In such situations, other survey organization field forces can be called upon for help.

I frequently receive requests from others who want to use our interviewers - to borrow them or hire them. In theory I subscribe to a free market concept. I do not feel that I should stand between my interviewers and other offers - the individual should be able to make his or her own decision without my acting as a gatekeeper - but in practice this works only if we all subscribe to this principle. Reciprocity is not as widespread as I would like.

Or someone else's sampling frame may be taken advantage of rather than making altogether new selections for the same area.

A division of labor in coding may have some efficiencies where each group can capitalize on its own specializations or existing manpower. And I suppose there may be times when centralization of data processing has its advantages.

All of our facilities could be regarded as a large pool; all would gain some degree of flexibility so that any study we contemplate carrying out does not become limited by our immediate resources and, at the same time, our sister group may welcome the opportunity to keep otherwise dormant staff busy. Of course there are risks. We have less control over the product. If we use similar procedures, conventions and practices, the risks are reduced.

Few organizations have a consistent and even flow of work. It's either feast or famine, or peak or valley. And our peaks don't necessarily coincide with those of other groups. Our average Boston field staff consists of 15 interviewers, but our studies always require 5 interviewers or 50 interviewers.Why maintain different field forces for each different national survey research organization in the very same metropolitan area?

And there are other times when access to another group would be helpful - for replication, for pretests, for comparability, for sample supplementation, for validation, to test a methodological procedure. Such services could be on a _quid pro quo_ basis.

In brief, then, our gains may be in 1) cost sharing, 2) greater flexibility, 3) quality, and 4) efficiency.

C. <u>Some Movement Toward These Goals</u>

It's not as though nothing has been done. There are many bilateral examples of collaboration or cooperation or sharing. Some examples are:

1. The University of Illinois, Survey Research Laboratory publishes an "Occasional Newsletter". Could be transformed into a more systematic information exchange?

2. For two successive years now, directors of university-based field organizations have met, first in Ann Arbor at the Survey Research Center, then in Chicago hosted by NORC and another is planned in this coming year at York - in Toronto. It is now an annual event. Although these have been tremendously useful, there is a nagging sense of déja-vu. Also, the focus is almost entirely on field issues, to the exclusion of coding, sampling, data processing or analysis.

3. Pooling of field resources, with the sample selected centrally, is not uncommon. There have been several instances in which one survey organization has subcontracted to several others to carry when the load was beyond the capacity of the existing staff.

4. Professional associations may present opportunities for communication and interaction both formally and informally. But professional staff is usually split among various disciplines and have difficulty finding each other at the same conventions.

5. The University of Michigan's political behavior consortium is another model for effective inter-institutional collaboration.

6. The several compendia of measures compiled by John Robinson, et.al.

D. Barriers

Barriers exist, but they are not insurmountable. These include apathy, inertia, failure to assume the initiative (minor, but chronic) and the competition for time and dollar.

Nor should institutional and bureaucratic inflexibilities be discounted. Vested interests, conceits, and arrogance (attributes which I share) stand in the way of yielding institutional sovereignty.

Perhaps too little credence is given to rivalry. After all, some of us are competing for the same dollars and prize our distinctive reputations, warranted or not. We may feel more comfortable in controlling our own quality and, in turn, being able to hide our weaknesses. and errors.

(Although standardization may be a good which we subscribe to, there is still the risk of stifling creativity or if not creativity, freezing a measure while still in the process of refinement or improvement and forever afterward being confronted with the perpetual need to maintain continuity and comparability - an argument we have all encountered or used).

But really I think our main hurdle is to devise an appropriate mechanism and to proceed cautiously and realistically.

It may be no small problem to reconcile the differences in style, format and methods of diverse survey organizations which are geographically dispersed, each with its own traditions, its own style, its own prejudices, and each imbedded in its own institutional bureaucracy - each with different needs, different emphases, and different levels of tolerance for departures from standards. I remain hopeful, despite these differences, there are large areas of agreement and we need not push beyond the exchange and clearing house function unless it progresses naturally to levels of greater commitment or involvement. Overly ambitious moves may scuttle any effort.

E. Proposed Model

There are a variety of possible models, ranging from a very informal understanding such as presently exists among us, to an elaborate formal organization. Our need is a rather specialized one and I think it would be useful if we could create a quasi-formal mechanism whose objectives are to reduce effort, reduce duplication, increase flexibility, increase knowledge at a cost, psychological, administratively or fiscally, which doesn't outweigh the gains. It could take the form of a loosely knit federation of survey organizations.

At this stage, why not let one of the major national survey groups take on these functions? Then, guided by a committee, a working secretariat could be created which would serve as a clearing house and information exchange leading to other levels of cooperation, such as a data bank.

And, the secretariat need not be a new creation, but conceivably could be an extension of the field directors conference, or the Illinois newsletter.

F. Conclusion

A council or a committee or small working group could establish some low level goals which are capable of achievement and should not be tremendously demanding; their value should be obvious, or at least testable in the short run, subject to discard or modification and not require a tremendous investment. It shouldn't flounder because of lack of commitment or resources.

As a minimum, let's try to agree to exchange some kinds of data and materials - but it would be more effective if we could manage the next step - that of organizing the materials in such a way that they are usable - so that we can retrieve from them what has been filed. The test will be in the use - and cumulatively, the value should increase at a greater rate than incremental additions to the file.

Clearly, much can be gained by academic units by integrating and organizing university-based survey research facilities.

MEASURING CHILDREN AS A MEANS OF EVALUATING PUBLIC NUTRITION PROGRAMS: A STUDY OF INDIAN SCHOOL FEEDING PROGRAM IN THE STATE OF ORISSA

David G. Mathiasen
United States Agency for International Development

I. Introduction

This paper reports on an unusual research effort which was designed to develop reliable data on the effects of a mass feeding program on children in one of India's poorest and most backward states--Orissa. The research was undertaken by an Indian organization, the Council for Social Development (CSD), under contract with the United States Agency for International Development (USAID) Mission in New Delhi. The research effort was unusual because:

1) The sample size and quality was unusually large for a rural survey, particularly one conducted in a backward area of India.

2) It appears to be the first time a sample survey has been used to evaluate the effects of a large operating nutrition program.

3) A new technique for developing data on nutrition intake of children was developed.

In the process of gathering data relevant to an understanding of the child feeding program, extensive data on nutrition, school attendance and enrollment, and other aspects of rural society were gathered which should be of interest to social scientists concerned with underdeveloped countries.

This paper focuses on the development of survey techniques as a means for evaluating a major program to combat malnutrition. The results of the survey are not emphasized because (1) the data are still being processed and analyzed, and (2) the Council for Social Development will shortly publish the results and it would be unfitting at this point to anticipate in detail what the Council's report will say.

II. The School Lunch Program, Nutrition, and Foreign Aid

In the United States, a school lunch program existed over 50 years ago. However, it was not until the interwar depression that school feeding developed into a mass effort to combat child malnutrition. Since World War II, federal funds approaching four billion dollars have been spent to subsidize school lunches. These funds have supported the distribution of surplus commodities under the National School Lunch Act of 1946.

Abroad, the concept of improving child nutrition through a school lunch has largely been spread by American-based voluntary agencies. In Japan, the program, initiated during the post-war American occupation as a relief program, is now fully supported by the Japanese and covers 90 percent of the school children.

In India, the school lunch program began in 1952 and has grown rapidly, largely under the auspices of CARE and Catholic Relief Services. About 18 million school and preschool children are now fed with U. S. foodstuffs provided on a grant basis. The food is brought into India by U. S. -based voluntary organizations such as CARE, and financed by the Department of Agriculture under authorization provided by Public Law 480. The Agency for International Development is authorized by the Foreign Assistance Act to finance the ocean freight cost of transporting the commodities to recipient countries. CARE's expenses are largely limited to those necessary to cover its administration of the program. In India, state governments pay for the freight within India and are responsible for getting the food to the schools. In some instances additional foods--primarily flavoring foods--are added by local contribution or through funds provided by the states.

III. Need for Evaluation

The shared financing of the program leads to unclear responsibility. From CARE's point of view, the program is largely free, since the food and its distribution is financed elsewhere. The Department of Agriculture has found programs under Title II of PL 480 (which includes not only school lunch programs but food for work and disaster relief) to represent a relatively small part of its total outlays for overseas food shipments. The program is popular with Congress and has been (like the domestic program) integrally connected with price support operations. From AID's point of view, the program appeared useful and since it was administered by CARE, largely financed by the Department of Agriculture, and popular with Congress, the payment of freight seemed a reasonable enough contribution. The Indian states' departments of education, although facing serious budget constraints, have generally found that the costs of local distribution are worth the benefits of having a school lunch program.

Under the circumstances, the opportunity cost of the program seemed very low to all those responsible: no single organization was faced with the full cost of the program. However, the program does represent substantial budgetary outflows for the U. S. as a whole. Although it is understood that to a substantial degree these expenditures may be fulfilling agricultural price support goals, increased pressure on U. S. budgetary expenditures resulting from the need to expand domestic antipoverty programs, the cost of the Vietnam war, and the desire to reduce inflationary pressures, resulted in serious questions being asked about the value of the program to developing countries. In August 1968, the USAID Mission in India decided to seek ways of stepping up earlier efforts to evaluate the usefulness of the school lunch program.

IV. Approach

There is ample evidence that if children who are malnourished are given a dietary supplement of three ounces of CSM[1/] and one-fourth ounce of vegetable oil each day they will be healthier, taller and heavier. This is the ration provided for each child (ideally) in the Orissa school feeding program. The usefulness of this supplement can be demonstrated by taking small samples of malnourished children and carefully controlling their entire diets. But the 18 million children who benefit from the school lunch program in India do not live in such a controlled environment. Indeed, how many are malnourished and to what extent is largely unknown. Furthermore, an important justification for the program is that it improves school attendance and reduces dropouts--obviously a hypothesis which cannot be tested in a nutrition laboratory.

It appeared, therefore, that a research approach which would evaluate what the program does for children in the field under actual uncontrolled (in the scientific sense) conditions would be required if a relevant evaluation of the program were to be obtained. Moreover, the urgency attached to developing some basic data on the benefits of the program, if any, precluded the use of a longitudinal study--typically the approach used in the nutrition field. The USAID Mission in India therefore opened discussions with a number of local research institutions on the feasibility of conducting a sample survey to gather basic data on the school lunch program.

The American aid program has in the field a small staff, particularly in relation to the large volume of resources involved in the U. S. program in India. It was out of the question for AID to do such a study itself. But even if it had been feasible, it would have been inappropriate, given the available research capabilities in India. AID's role was to finance the study, to provide consultants to the survey where needed, and to cooperate with the Council for Social Development on the structure of the study. The responsibility and credit for the study belong exclusively with the Council.

The goals of the study were:

1) to determine if children participating in the school lunch program are significantly different physically from those not in the program;

2) to determine if student enrollment and attendance are different in schools with a lunch program from schools without one;

3) to determine if health of students in the lunch program is better than those not in the program;

4) to determine if students in the program performed better academically than those not in the program;

5) to determine if the food supplement provided was in addition to or a substitute for food provided in the home.

The technique decided upon was to sample a known population of students attending schools participating in a school lunch program and compare it with a similar sample of students attending schools not in the program.

The approach was tested through a pilot study in the Indian State of Andhra Pradesh, using a sample of 2, 028 students. At that time various approaches to measuring students were attempted under field conditions in Indian villages. In addition, the pilot survey provided essential experience in questionnaire design, sampling problems, training of interviewers, and means of providing checks on the quality of data. Although the data resulting from the pilot survey were not sufficiently good to provide definitive answers to the hypotheses under consideration, it did prove essential to the success of the full-scale study, and many of the successes of the full-scale study can be directly attributed to the experience gained in the pilot study.

The sample survey in the State of Orissa was formally begun in November 1969. The Council for Social Development established a collaboration with the Department of Psychology of Utkal University, which is located in the capital of Orissa. The University provided a source of staffing to the project--both senior professionals and interviewers--and a base of operations.

V. The Orissa Survey

The State of Orissa is largely rural and one of the most backward areas of India with little industry and low per capita incomes. Of all Indian states, Orissa ranks first in infant mortality, last in the percentage of students passing matriculation examinations. It is India's least urban state, with 94 percent of the people classified as rural. It also has the highest percentage of "backward" people of any Indian state--that is, people recognized officially as belonging to scheduled (untouchable) castes, or to tribal groups. Orissa is estimated to have the second lowest per capita income (with Bihar having the lowest).[2/]

It is located on the East Coast of India, south of West Bengal and Calcutta. The coastal areas consist of extensively irrigated, rice-growing level land. Further inland, there are hills where most of the tribal people live. Many villages of the state cannot be reached by any vehicles at all; others can be reached by jeep roads during the dry seasons of the year. During the monsoons, communications are limited to those relatively few areas accessible by all-weather roads.[3/]

The state is a good target state for the CARE school lunch program because of its low incomes. Yet at the same time, it is a state in which such a program is very difficult to carry out; logistics alone are a formidable challenge.

From the point of view of surveying the effects of a mass-feeding program, the state had appeared to have real advantages. Among them:

-- The feeding program reaches enough schools to provide a large universe from which to sample, but not so many as to cause difficulties in developing the control sample.[4/]

-- The feeding program reached children from a variety of religious, ethnic and economic backgrounds.

-- The logistic problems of the state made it possible to judge the effects of the program under tough conditions; some schools were able to provide the frequency of feeding desired but others inevitably could not.

On the other hand, Orissa is probably the most difficult state in India in which to do social survey work. The number of trained and experienced field interviewers who speak Oriya is virtually nil and the problems of carrying out a carefully designed sample involved great effort on the part of field teams.

VI. Structure of the Survey

The major variables to be tested were as follows:

1) Weight of children (nearest one-tenth of a kilo)

2) Height of children (nearest one-half centimeter)

3) Chest of children (nearest millimeter)

4) Biceps of children (nearest millimeter)

5) Skinfold biceps of children (nearest millimeter)

6) Skinfold scapula point of children (nearest millimeter)

7) School attendance

8) School enrollment

9) Health (days absent because of illness)

10) Dropout rates

11) School grades

12) Total calories

13) Total protein.

Three questionnaires were used for the entire sample--one for the student, one for the teacher, and one for the student's parents. These questionnaires recorded data for variables 1-11, with various cross checks. In addition, they recorded data on attitudes toward the program, social and economic background of the child, age of the child, size of family, etc.

It is worth noting here that age represents a serious research problem because it is often not known. There is often a substantial difference between the age recorded on school records, the parents' estimate, and the child's estimate. The questionnaires were designed to cross check age, and parents were asked to identify all their children and rank them according to age. These procedures may have improved accuracy, but certainly did not insure it.

The sampling design was based on the administrative units of the State of Orissa. The state is divided into 13 districts and 314 blocks, 15 to 41 per district. Four districts are classified as predominantly tribal. Except in tribal areas, blocks have roughly the same population.

Before drawing the sample, a sampling frame was developed. To do this, each school sub-inspector was sent a set of cards--one card for each of the schools under his supervision. He was instructed to fill in a card for each school indicating location, enrollment, whether the school was in the feeding program, if so, how many days it actually served lunch during the past two years, and other basic data. The strength of the survey was helped by the prompt and accurate response that came from this mailed questionnaire. Within three weeks over 40 percent of the cards were returned and eventually 98.6 percent were returned. The importance of having these data for sampling is illustrated by the fact that at the outset there was no accurate count on even the number of schools in the state. Estimates varied between 22,000 and 26,000. These data provided a useful quality check by comparing them with overlapping data which interview teams gathered. Moreover, they provided substantial information not subject to sampling error.

The basic approach was to select a sample of schools in the feeding program and match it with schools not in the feeding program. This design had to be altered in four districts because the proportion of schools in the program was so great. These districts are predominantly tribal and represent the most backward areas of Orissa. In these districts virtually all the schools which are accessible are in the feeding program. No matching with nonfeeding schools was therefore possible.

At the first stage of sampling, the number of blocks to be selected from each district was determined in proportion to children in the district participating in the school lunch program. Once the number was determined, the sample plus two alternates were selected at random. Within the blocks, schools were selected proportional to enrollment. This was done by assigning serial numbers to the schools, based on enrollment. The first school with 200 students was assigned 1-200, the second with 160 students 201-360, the third with 120 students 361-480, etc. A table of random numbers was then used to select schools. Schools which were not within one or two miles of a jeep road were deselected for logistic reasons.

Within the schools, ten boys were selected at random from grades three and four. Boys were selected because few girls attend school and including them would have complicated data interpretation. The grades were selected because of the desire to test children who had been in the program at least two years.

The matching schools were selected according to the following criteria:

1) Similar size of school (plus or minus 20 percent;

2) Similar size of village (plus or minus 20 percent;

3) Similar representation from cultivator caste (plus or minus 10 percent);

4) Similar representation of scheduled castes or scheduled tribes (plus or minus 10 percent).

Children in this matched sample of nonfeeding schools were selected in the same way as for the feeding schools.

In the event, the sampling was somewhat more complicated than outlined here, but the basic approach was not altered.

The diet survey to determine the food consumed in the home (described below) was too elaborate to be conducted on the entire sample. Therefore a 40 percent subsample was selected for home diet surveys. Table 1 summarizes the sample.

Table 1

Sample Distribution Obtained in the Orissa School Lunch Survey

	Nontribal Areas	Tribal Areas
Basic Sample		
Districts	9	4
Blocks	35	15
Schools Selected		
Feeding	180	150
Nonfeeding	165	a/
Children Sampled		
Fed	1,388	808
Not Fed	1,222	a/
Nutrition Subsample		
Districts	9	4
Blocks	35	15
Schools Selected		
Feeding	72	59
Nonfeeding	68	a/
Children Sampled		
Fed	507	302
Not Fed	485	a/

a/ In tribal areas virtually all accessible schools had feeding programs; hence no nonfeeding sample could be selected.

VII. Substitution Feeding

The Diet Survey was designed to answer the question of whether the school lunch is a substitute for or an addition to food provided at home. The earlier pilot survey suggested that the usual recall method of determining diets was highly inaccurate for children in rural India. One problem was that since the subjects were young children, they were less responsive to questions on what they had eaten the day before than adults might be expected to be. A second was that in rural areas where food self-sufficiency is on a family basis, there is limited sophistication to measures of food. An urban consumer is more conscious of amounts because he purchases food frequently and pays for it in terms of some standard measures. But rural families are largely outside of the market economy and draw food from their own stores.

The alternative of having an observer in the family was also considered unsatisfactory since his presence would be likely to change eating habits. Moreover, for the large sample contemplated it was impractical.

The technique used in the Orissa survey was record keeping.[5/] The selected children were given a set of three aluminum bowls similar in shape to those used in Orissa. Each was marked with a line indicating the half-full point. Each was marked with a number. The child was then instructed to take the food after it was served to him and measure it in the appropriate bowl. He was also given a large illustrated questionnaire on which to record each meal. Adult family members were asked also to measure the food before they cooked it and after, so as to provide data for converting cooked food into basic nutrients.

The family got to keep the bowls, partly as an incentive and partly because it was unclear what the Council for Social Development would do with hundreds of used aluminum bowls. The taboos which are part of the Indian caste system precluded using bowls over again by other families. As it turned out, the results were well worth the cost.

This approach was used only after full pretesting. During the two months before the full survey began, the technique was worked out with a field sample of 100 families. Results suggested that the technique was considerably more accurate than one-day recall.

VIII. Training and Field Operations

Training of interviewers was of critical importance, since the pilot study in Andhra Pradesh had revealed a large degree of interviewer error. Eleven interview team leaders and 30 interviewers were selected and formed into teams of two senior and six junior investigators. Six reserve interviewers were also selected, all of whom eventually ended up in the field. The interviewers were in their midtwenties and most had degrees in the social sciences from Utkal University. The "senior investigators" were hired two months before the interviewers. They received training and participated in the organization of the survey, the pretesting of the questionnaire, and the development of the sample frame.

The interviewers began training on January 7. They were taught accurate techniques for taking physical measurements and given lots of practice. Similarly, they were trained for the diet survey by practicing it on themselves. A mechanic taught them how to take apart, adjust, and repair the balance scales used for measuring weight. After three weeks they were sent out to an area (not in the sample) for field practice. Overlapping interviews were arranged so that teams could compare results. Careful supervision by the project directors was an essential aspect of the training.

The field teams consisted of a leader and a coleader, plus six investigators. Each team had at least two female members for handling the nutrition work and for conducting parent interviews, since there was concern that mothers would be reluctant to talk to male investigators. In fact, this later proved not to be the case. The teams were divided into subteams, each with the assignment of handling a separate school. Each team was provided with a jeep, four of which had trailers. The logistics were complicated by the limited number of jeeps available, which in turn reflected the high cost of jeeps in India relative to the cost of other aspects of survey work. The jeeps delivered the subteams to their respective villages each day and carried the measuring instruments from one subteam to another. Removed from the survey by time and distance, the logistics seem to be an unnecessary preoccupation of those who worked on the survey. However, much of the success of this kind of work depends on handling purely mechanical problems. Ease of access to sampling units and the availability of cheap, adequate transportation for whatever personnel and equipment is assigned are luxuries which the social scientists of India do not enjoy. They need to worry about finding gasoline stations, keeping vehicles repaired, squeezing equipment and personnel into very restrictive vehicle space, and even finding responsible, reliable and motivated drivers.

Field work actually started on February 6, 1970, in six blocks fairly close to headquarters so that the teams could be within easy reach of supervision during the early stages. The result of the first month's work was brought in, coded and analyzed so that the teams could examine what kind of quality was being achieved. The Survey quality check team had been reinterviewing 10 percent of the sample. This 10 percent reinterview was then checked with the first interview. The results of this check during the first month of interviewing were extremely helpful from two points of view. First of all, the check indicated which teams were strong, which teams were weak, and what measurements were not being accurately taken. Secondly, the quality check turned out to provide a major element of motivation for the teams since they began competing to see which could achieve the highest degree of correspondence with the quality check team.

Field work was completed in May of 1970.

IX. Quality Control

The two elements of the quality check have already been mentioned but should be restressed because of their importance in reassuring the project directors of the validity of the data that were gathered. The first was in the design of the sample frame; many of the basic data on the schools were collected and sent in by mail during the frame development stage of the survey. Subsequent overlapping interview data provided by the field teams showed a very high degree of correspondence with the frame development data. The 10 percent reinterview provided an even more important quality check. The 10 percent quality check tested two separate things: (1) interviewer quality and (2) response error. Measurements of a child which were significantly different during the reinterview indicated clearly a problem of the quality of the field workers. However, age or attitudinal questions which turned out differently could have been an error on the part of the interviewers, or could have been because the respondents simply provided a different set of answers.

In the event, most physical measurements and the objective data which interviewers gathered from the records turned out to correlate highly with the quality check data. An exception was the skinfold thickness measurements. The problem in that case was probably due to the difficulty of using the only instruments available at the time of the survey. In addition, these measurements do require greater care and are inherently less reliable than height or weight.

There was considerably less correspondence between interviewers and quality checks in areas where respondents could be expected to give different answers at different times. As a result, a number of the questions, particularly those dealing with attitudes towards the school lunch program or with judgments (e.g., teacher evaluation of child's health) appeared to include too high a degree of response error for meaningful analysis. But for objective measurements the quality of the interviewing appears to have been very high indeed, which is a tribute both to the teams themselves and the training and leadership provided by the management of the survey.

X. Results of the Survey

As stated at the outset, the full analysis of the data has not yet been completed and the Council for Social Development, which is responsible for the research, has not completed its final report. It is not the purpose of this paper to prejudge what the survey in fact has to tell us, but rather to report on the status of what appears to be a rather interesting and unusual piece of research. However, a number of papers summarizing preliminary data analysis were prepared for a conference in New Delhi, which was held in October 1970. At that time, some interesting preliminary results were evident and are perhaps worth listing with the appropriate caveat that they are subject to later revision.

Overall, there were no statistically significant differences in physical measurements between those in the feeding program and those outside the feeding program. However, an analysis of students in schools which fed 200 or more days a year indicates significant differences in height, weight and chest circumference, with the children in the schools offering the school lunch being larger in these respects. This would appear to indicate that the feeding program does have a measurable effect if it is carried out according to specifications. More analysis is needed to see how the number of days fed correlates with differences in physical measurements. Moreover, the data show differences which are in some cases more striking than one would expect and need to be examined with some care. However, the fact that (1) all three measurements show statistically significant results supporting the hypothesis that supplemental feeding produces better physical development, and (2) the reliability of the data as indicated by the quality checks suggests there is reason to believe that a well-run feeding program produces measurable results.

Another tentative conclusion is that school feeding is additional to what the children receive at home. The diet survey indicated no significant difference in the protein and calorie intake at home between children who attended schools where lunches are provided and those who attended schools with no lunch program.

School attendance and enrollment analysis was complicated by the likelihood that the program is self-selecting. The frame analysis data show, for example, that enrollment in tribal areas is much higher in schools with a feeding program than in schools without it. But this is most probably explained by the fact that the larger schools are more accessible and therefore are selected for the school feeding program. Data on enrollment in the nontribal areas is inconclusive but does not suggest any strong tendency of the program to draw more students. Similarly, the first analysis of data on attendance does not reveal any reason to believe that the existence of a school lunch program decreases absenteeism due to better health or that the lunch program provides a greater incentive to attend school. However, considerably more subanalysis needs to be done before the hypothesis that the school lunch program increases attendance can be considered disproved. In particular, further data analysis may suggest that schools which achieve a highly successful feeding program, i.e., 200 days a year or more, do induce increased attendance.

School dropout rates appear to be higher in fed schools than nonfed schools, which was contrary to the hypothesis being tested. One explanation of this phenomenon would be that more students are drawn to the schools in the first instance by the feeding program and thus there is a greater tendency for dropouts after the children reach the age where they can assume family work responsibilities.

XI. Conclusion

The initial purpose and the basic purpose of the survey was to develop hard data on the basis of which those responsible for the financing and administration of the school lunch program could make judgments as to its usefulness and as to ways it could be improved. The use of this technique for program evaluation is uncommon, particularly in the area of nutrition and child measurement, which tends to be carried out using much smaller samples and a longitudinal research design.

The survey has produced a mass of data, many of which appear of good quality, about the effects of the school lunch program and how it operates. In addition, there was developed a great deal of data necessary for supporting analysis, but which also should be of major interest to those concerned with other aspects of social development. For example, the diet survey work resulted in an enormous store of information on who consumes how much of what varieties of food in the State or Orissa. These data can be analyzed in terms of geographical area and economic and social background. Similarly, a mass of data are available on school attendance, school dropout rates, and school enrollment, with the socio-economic background of the students again identified. This raises the hope that analysts interested in these questions will take advantage of the tremendous field effort which went into the survey and use the data for other kinds of analysis. Indeed, perhaps the most important function of this paper is to bring to the attention of the academic community the availability of this body of data. It is expected that a duplicate set of data will be recorded on magnetic tape and deposited in an American university so as to be available easily to researchers in the United States.

There is also a lot to be learned from this survey about conducting social science research in underdeveloped countries. The process of putting together and training a group of researchers was an exciting one to watch. The enthusiasm and interest of the young postgraduates in the field was also gratifying. As is often the case in underdeveloped countries, the university graduates were from the wealthier groups in society and were largely urban based. During the field work, many of them discovered their own state for the first time. Moreover, they were exposed to standards of sampling techniques and quality checks which were quite new to them.

It is worth noting as a final word that no survey such as the one described here, evaluating the effects of the school lunch program, has ever been undertaken in the United States.

1/ CSM is a blend of corn flour, soya flour and nonfat dried milk.

2/ Data compiled from various official sources by M. P. Srivastava, "U. P. in the Socio-Economic Map," Eastern Economist, April 24, 1970, pp. 794-801.

3/ As visitors to the subcontinent know, these are referred to as "pukka" roads, as opposed to "kucha." "Pukka" in Urdu or Hindi means durable or lasting. Pukka buildings are made of cement or brick; pukka roads are surfaced. It is a term those exposed to the subcontinent find hard to do without and one which could usefully be added to American English.

4/ This proved less true than hoped. As explained below, CARE and the Orissa State Government reached too high a proportion of the tribal areas to permit good sampling of nonfeeding schools.

5/ Record keeping is a new technique only in the sense of being adopted for use in rural Indian societies where literacy is limited and the concept of standard measures not part of the culture so far as home consumed food is concerned.

(The author is Assistant Director for Development Policy of the United States Agency for International Development Mission to Pakistan. The views expressed in this paper are his own and not necessarily those of A. I. D. or the United States Government.)

LABORATORY ASSESSMENT OF NUTRITIONAL STATUS

with Emphasis on Immunological Measurements

William C. Cooper, University of Minnesota

The ability to determine nutritional status is becoming more and more important as world population expands and food supply diminishes. Although the field of nutritional metabolism and physiology has achieved a high degree of sophistication, experts continue to be hampered by the lack of a single battery of tests which can be applied to populations throughout the world and in all environmental conditions. The ideal set of determinations would be one where all the tests are easily corrected for confounding variables. Since, however, a standardized human environment does not exist, it will probably never be possible to develop an ideal set of tests for the evaluation of nutritional status. Multiple factors are involved in the production of states of malnutrition; hence, a multidisciplinary approach is essential for the proper evaluation of each malnourished individual. Moreover, he should be compared with standards derived by testing normal members of his own population. This is rarely done. If a person is removed from his environment or if any form of treatment is given prior to testing him, his resulting nutritional "profile" will be biased beyond tolerable limits. Nutrient deficiencies in human populations are almost invariably multiple and they may affect any organ system of the body including the immune system, depending upon the specific deficiencies. Deficiency of one nutrient frequently alters the metabolism of another. The process of adaptation may alter or even obscure the manifestations of deficiencies which are real.

It is possible, despite all the above difficulties, rather accurately to evaluate human nutritional deficiency if one maintains a broad overview of the problems and correlates the results with each other. Tests are available which measure many deficiencies with great accuracy. Appraisal of data by a team of experts: clinicians, biochemists, physiologists, immunologists, epidemiologists, food technologists, sociologists, anthropologists, psychologists, economists, and statisticians - continues to improve our basic understanding of nutritional problems and has led to a greater degree of operational success in alleviating malnutrition. Although nutrient deficiencies are usually multiple, they commonly occur in well - recognized patterns, permitting a certain degree of selectivity in evaluational testing. Advances in the knowledge of nutritional metabolism also facilitates testing for the interactions involved when depletion of a specific nutrient affects metabolic pathways other than its own.

I shall attempt to present a brief survey of some of the more important biochemical and immunological tests for the assessment of human nutritional status. Unfortunately, space does not permit a commentary upon clinical evaluation, which should always constitute the first approach to the problem. By examining the subject clinically, one can usually get a general impression of the combination of nutrient deficiencies present, and can then pinpoint specific ones through biochemical and other tests. For example, prolonged niacin deficiency may result in the classical clinical stigmata of pellagra; dermatitis, diarrhea, and dementia. Since most low meat diets are also low in all the "B" vitamins, clinical pellagra would alert the examining physician to look for signs and symptoms of deficiencies of the other "B" vitamins, such as the cheilosis, etc. of riboflavin deficiency and the peripheral neuropathy of thiamine deficiency. Deficiencies of all these nutrients could then be assessed biochemically. Populations which consume corn as a staple food frequently display a high incidence of pellagra, as was detected in the Southeastern United States during the early part of this century.

Retardation of growth rate is the most common physical manifestation of malnutrition, and usually reflects combined nutrient deficiencies, protein being most important. Signs of deficiencies of nutrients other than protein often confuse the clinical picture, but it is possible to solve this problem by doing biochemical tests (Tables I and II) for the suspected specificdeficiencies. Seven important tests of protein deficiency are described in Table I. Facilities and available personnel are the usual factors in determining which tests are done. Most of these tests have been adapted for field use, so that samples can be rather easily collected and sent to a central laboratory for processing. The comments on each test given in the tables should indicate the feasibility of its performance in a given situation. To obtain an accurate biochemical profile of protein malnutrition, one should perform tests for: (1) plasma proteins (Table I, test 1); (2) basal urea nitrogen excretion, the best of which is urea nitrogen/creatinine (Table I, test 5); (3) body protein mass(Table I, test 6); and (4) the pattern of plasma amino acids (Table I, test 7). It should be stressed that individual results must always be compared with standards generated from the same population, and that the recent past history of the subject must be known, especially with respect to diet. For example, a low blood ascorbic acid level indicates that the tested individual absorbed very

TABLE I.

BIOCHEMICAL EVALUATION OF PROTEIN MALNUTRITION (PM)

some appropriate tests

	TEST	NORMAL RANGES for U.S. population	COMMENTS
1.	Plasma proteins (total; albumin, alpha, beta, and gamma globulins)	Total 6.5-6.9g/100ml Alb. 3.5-4.2g/100ml Alpha –1 4-7 Alpha –2 9-11 %of pl. Beta 11-15 protein Gamma 12-16	Levels of serum proteins only informative in severe P. M. Thus these tests only detect gross deficiency. Various methods are available for these determinations, some of which require simple apparatus.
2.	Hemoglobin and/ or Hematocrit	Hb Men 14-15g/100ml Hb Women 11-14g/100ml Hcrt Men 42-44% Hcrt Women 38-42%	Although anemia is usually present in moderate or severe PM, its diagnosis does not indicate causation. The anemia associated with protein-calorie malnutrition is of a mixed and complicated etiology, and may involve deficiencies of iron, protein, folic acid, vit. B_{12}, vit. E, and even selenium.

3. Urea excretion (urine) — Urea is the principal end product of protein metabolism in mammals. Its excretion is directly related to protein intake, but the amount of urea excreted must be expressed relatively, since its absolute amount in the urine is meaningless. For example, if a single high-protein meal is given to a protein-malnourished subject the night before his timed urine specimen is collected, the absolute amount of urea excreted might well be normal.

4. $\frac{\text{Urea N}}{\text{Total N}}$(urine) — Normally, urea comprises 80-90% of total urinary nitrogen. (Total urinary N about 13 g / day; urea N about 11 g / day). However, because even in PM urea N. comprises most of the total urinary N. changes in the numerator are not easily detectable in PM.

5. $\frac{\text{Urea N}}{\text{creatinine}}$ (urine) — This test is better than 4.because changes in the numerator are more perceptible. Creatinine N comprises a small percentage of total urinary N and tends to remain fixed except in severe stages of muscle catabolism. To use this test, must restrict water intake, because urea clearance is directly related to protein intake only at minimal urine volume. Approximate ratios if water is restricted are: Normal: about 20. PM: 7-9.

6. Creatinine-height index (C H I) — This is a modification of the creatinine coefficient used in well-nourished populations. The usual creat. coeff. is creat. excreted/kg. body wt. /24 hr., and in "rich" areas is an index of obesity, among other things. In protein malnourished populations, obesity is almost non-existent, and body height is much more meaningful than body weight. The CHI of Dr. Viteri is expressed as follows:

$$\frac{\text{mg. creat. exc'n. /cm. ht. /24hr. (in person tested)}}{\text{mg. creat. exc'n. /cm. ht. /24hr. (in person of same height, regardless of age)}}$$

This test is a useful measure of decrease in body protein mass. 8-11 A. M. is best time to collect a 3 hour urine sample for CHI.

7. Plasma aminogram — The ratio of non-essential (NE) to essential (E) plasma amino acids reflects protein nutrition in general, but the subject tested must not have been treated (diet-wise) prior to this determination. In PM this NE/E ratio increases, reflecting greatly reduced blood levels of essential amino acids, especially the ones with branched-chains. Leu, ileu, val, met, thr, and trp are all greatly reduced. The ratio may increase to 6 in kwashiorkor. Normal is 2 or below. Test is done on 200 mcl. of "fingerstick" blood, using paper chromatography, elution of the pink ninhydrin positive spots, and expression of results as optical density ratios.

little vitamin C during the day before his blood was taken. This finding does not necessarily mean that the subject is actually deficient in vitamin C, because blood levels of ascorbic acid do not reflect tissue concentration. In general, low blood levels of a nutrient do not necessarily indicate deficiency of that nutrient. This is especially true of the vitamins. A summary of some important concepts regarding the assessment of vitamin deficiencies is presented in Table II. Not included in this survey is a summary of methods by which depletion of essential mineral nutrients may be measured, because space does not permit it. While forming only a small portion of the total body weight, minerals are nevertheless vital to metabolism. The animal body requires seven principal minerals: Calcium, magnesium, sodium, potassium, phosphorus, sulfur, and chlorine. These minerals constitute 60 to 80 % of all the inorganic material in the body. Trace quantities of at least five other minerals are essential to human metabolism: iron, copper, zinc, manganese, and iodine. The essentiality of molybdenum and selenuim has not yet been established, but they are utilized in trace amounts. Cobalt is an essential nutrient as a constituent of Vitamin B_{12}, but this vitamin must be supplied preformed to monogastric mammals. Tests for mineral deficiencies are mostly indirect measurements; for example hypochromic, microcytic anemia usually indicates iron deficiency (without disclosing the cause of the deficiency), and enlargement of the thyroid (simple goiter) usually signifies iodine deficiency in underdeveloped areas. The seven major minerals function mainly in acid-base regulation, body structure, nerve transmission, and blood clotting; the trace minerals are usually enzyme activators or integral parts of important compounds, such as iron in hemoglobin, zinc in insulin, and iodine in thyroxin.

The importance of the immune response, in general, in protecting against infectious diseases was recognized historically at about the same time as was the importance of essential nutrients and of nutritional deficiency disease. Thus, medical literature during the first half of the twentieth century is replete with reports, largely conflicting, on the effects of nutrient deficiencies on the antibody response, both in experimental animals and man. The classical association of famine with pestilence leads to the assumption that malnutrition reduces resistance to infectious disease. It is not my purpose here to attempt an appraisal of this belief. Suffice it to say that there exists a well-documented interaction between nutrition and infection but that the details of this interaction are unclear. Almost no work has been published on the extent to which human populations actually suffer through failure of humoral antibody response resulting from protein depletion, the most important global nutrient deficiency. More importantly, the relationship between deficiency of protein and protective antibody production in man has not been studied. A systematic investigation of this problem of world health would be difficult to design because of the multitude of variables involved. Protein deficiency: (1) affects all organ systems and their interrelationships; (2) can rarely be quantitated as to degree; (3) produces sociocultural changes which cannot be measured, such as overcrowding, with its consequent effects on infectious loads, infant weaning time, and altered dietary practices; (4) may have a differential effect on humoral and cellular immunity; and, (5) has a different effect on the immune response depending upon the type of infectious agents. The nature of protective antibodies must be defined for each separate infection; all antigens evoke an antibody response which is heterogeneous. It is therefore difficult to determine which class of induced antibodies are actually protective and which are mere by-products of the heterogeneous antibody response. The antibodies most easily or most frequently measured are not necessarily the ones which are protective. Consequently, in a well-designed experiment concerning a specific infectious agent at least three things would require measurement: (1) the degree of protein depletion, (2) the specific protective antibody titer, and (3) some assay of presence or absence of the infection. To undertake such a study would omit, however, multiple factors concerned with aspects of the mammalian defense mechanism which are unrelated to antibody production, such as tissue integrity, cellular phagocytic activity, endocrine imbalance, and levels of protective enzymes such as lysozymes, the complement system, properdin and interferon, to name a few. In other words, one must be specific about: (1) the nutrient involved, (2) the host, (3) the infectious agent, and (4) the end points to be measured, for (4) varies with (1 - 3).

The foregoing omits from consideration the other major limb of the immune response: cell-mediated immunity (CMI). CMI is responsible for delayed hypersensitivity, allograft rejection, the graft-versus-host reaction, and defense against viruses, protozoa, fungi and some intracellular bacteria. The inclusion of tests of CMI would also be necessary in our hypothetical assessment of the effects of nutrient deficiency on the immune responses.

Thus, attention to proper experimental design is of paramount importance in assessing the effects of malnutrition on bodily metabolism and immunocompetence. A computerized approach in an idealized setting is clearly impossible when dealing with human populations. Therefore, if pragmatism is our objective, we must attempt to collect our data in the laboratory using a convenient mammalian species,

TABLE II: Some Data on Vitamins Important in Nutrition

Vitamin; Notes on Structure, etc.	Sources	Normal Daily Requirements	Important Functions	Lesions of the Deficiency State	Test(s) for Determination of Deficiency and Notes
Thiamine: (B_1); pyrimidine ring and thiazole. Heat labile and easily oxidized.	Yeast; brans of rice and other cereals; peas; beans; liver; muscle. Poor sources: ~~flour, polished~~ rice, processed cereals, milk.	1 mg (min.) (o. 5 mg/1000 cals.) (age-dependent)	Coenzyme (TPP) for 1. Oxidative decarboxylation 2. Transketolation (inHMP shunt)	Beri-beri Multiple Peripheral neuropathy.	Best is measurement of urinary thiamine excretion per unit time by thiochrome test (blue fluorescence). Normal: 100 mcg./24hr. or more, but must know age and weight of individual, because thiamine excretion falls with age. Very sensitive method (0.05 mcg/100 ml).
Riboflavin; (B_2); 3-ring structure with ribotyl group. Heat stable.	Meats, liver, germ and brans of cereals, yeast. Exposure to light may destroy it: also pasteurization and U.V. irradiation.	1.5-2.0 mg.	Coenzyme (FAD;FMN) for flavoproteins, dehydrogenase enzymes involved in hydrogen transfer for cellular respiration. Function as hydrogen acceptors in this biologic oxidation system.	Most prevalent avitaminosis in U.S.A. Causes cheilosis, seborrhea, glossitis; photophobia, circumcorneal vascularization.	General principles similar to those for thiamine. Riboflavin levels in plasma and urine vary widely, depending upon tissue saturation and state of protein metabolism. A brief period of neg. N balance, for ex., mobilizes it from the tissues and actually causes plasma level to increase. Chemical assay based on yellow-green fluorescence of the free, oxidized form. Must use red glassware.
Niacin: nicotinic acid. Pyridine ring. Occurs mostly bound in body, as coenzymes NAD and NADP	Liver, lean meat, yeast, poultry, wheat germ. Poor sources: ~~most fruits~~ and vegetables. Heat stable, but water soluble; thus, may be lost in cooking. Many diets low in quality protein are also low in all "B" vitamins. Corn low in trp., and niacin is bound in corn.	13-21 mg niacin equivalent (50 mg. trp. = 1 mg niacin in man). Pellagra is usually due to a combined deficiency of trp. and niacin.	Coenzyme (NAD;NADP) which function, together with FAD and flavoproteins, in the oxidative reactions described for other basic metabolic reactions such as oxidative decarboxylation fatty acid synthesis' and oxidation, and oxid of org. acids (TCA cycle)	Pellagra. Early symptoms: weakness, lassitude anorexia, indigestion, glossitis especially of tip and margins of tongue. The three "D's" of advanced pellagra are dermatitis, diarrhea and dementia. Lesions cannot be correlated with known functions of niacin.	Urinary excretion of N-methyl nicotinamide, a metabolite. Normal daily excretion over 3 mg, but this test is actually inadequate to assess borderline niacin deficiency. More sensitive is measurement of the urinary N-methyl pyridone with subjects on standard diets of 10 mg. niacin and 1000 mg trp., but this test is more difficult than the above. Perhaps best for population groups is a "loading test". 4-6 hrs. after a test dose of 50 mg. of niacinamide, measure timed urinary excretion of N-methyl nicotinamide. Subjects with incipient pellagra excrete only 5% of test dose; normals excrete 25% or more.
Pyridoxine (B_6) Pyridine ring. In body occurs in the active coenzyme form, as pyridoxal phosphate. Heat labile.	Meats, viscera, yeast, wheat germ, whole wheat. Milling of wheat may reduce the content by 90%. Milk rather low.	1-2 mg recommended by NRCFNB, but actual requirement is probably 0.5mg or less.	As coenzyme for transaminases, deaminases, decarboxylases, desulfurases, and others. Functions in active transport of amino acids; in synthesis of cysteine from homocystine and serine, of serine from glycine, and of delta amino levulinic acid (heme synthesis). Folate metabolism.	Hypochromic anemia, ? dermatitis, irritability and convulsions, esp. in infants. Peripheral neuritis in patients on isoniazid for T.B. The necessity of B_6 in synthesis of gamma amino butyrate (GABA), important in brain metabolism, may help explain convulsions in B_6 deficient infants.	Pyridoxine is necessary in Conversion of tryptophane to niacin in the body. After a tryptophane loading dose, if the subject is B_6-deficient, he is unable to convert the tryptophane to niacin, and instead excretes xanthenuric acid in the urine. (see Vilter, 1953, J. Lab. and Clin. Med. <u>42</u>:335). Any significant increase above normal in the excretion of xanthenuric acid after a tryptophane load is indicative of B_6 deficiency.
Pantothenic Acid Pantoic acid and beta alanine. Part of coenzyme A. Fairly heat stable. Easily hydrolyzed in acid and alkali (peptide bond).	Egg yolk, kidney, liver and yeast are excellent. Lean beef, broccoli, skimmed milk and sweet potatoes are fair sources. Large losses may occur in wheat flour manufacture and cooking of meat, but not of vegetables. However, the vitamin occurs in all foods.	5-10 mg/ kg. of diet. 20 mg when given as the calcium salt. (Calcium pantothenate).	As part of coenzyme A, it combines with acetate to form "active acetate", with basic functions in carbohydrate, fat, and amino acid metabolism. Also forms acetylcholine. "Active succinate" (succinyl coenzyme A) takes part in heme synthesis Acetyl Co-A helps form cholesterol and thus steroid hormones. Protein-bound pantothenic acid in the cell (acyl carrier protein) is a coenzyme required for fatty acid synthesis.	Spontaneous deficiency is very rare, because of the presence of this vitamin in all foods. However, the "burning feet syndrome" was thought due to pantothenic acid deficiency in WWII prisoners in the Philippines and Japan, and in malnourished south Indian populations.	No chemical methods for its determination are available. Most favored assay techniques are microbiological. An improved technique using Lactobacillus casei is popular. (see Clarke, M. F., 1957. Anal. Chem. <u>29</u>:135.)
Folic acid; pteroylglutamic acid. A combination of the pteridine nucleus, PABA, and glutamic acid. Heat labile and easily oxidized.	Liver, kidney, yeast, cauliflour, dark green, leafy vegetables. Fair amount in Milk	0.5-1.0 mg	Major one is one-carbon unit metabolism, such as formyl, hydroxy-methyl, and fornimino group transfers. Purine synthesis, esp. thymine. Cell mitosis: metaphase to anaphase. (Powerful antagonists are aminopterin, and amethopterin).	Nutritional megaloblastic anemias. "Tropical'sprue (Folic acid should not be used to treat pernicious anemia. The hematologic response is not permanent, and the neuropathy is unresponsive).	Methods of choice (1) microbiolo gical assay using L. casei on L. fecalis. (2) "FIGLU" excretion test. In the metabolism of histidine, there is a folate-dependent step where form-iminoglutamate is converted to glutamate. When folate-deficient subjects are given a loading dose of histidine, there is increased urinary excretion of formimino-glutamic acid (figlu).

Vitamin: Notes on Structure, etc.	Sources	Normal Daily Requirements	Important Functions	Lesions of the Deficiency State	Test(s) for Determination of Deficiency and Notes
Biotin. Organic ring compound with side groups.	Egg yolk, liver, yeast, seeds. Meats (muscle) rather low, but biotin is present in all foods.	Unknown, but human experimental subjects of biotin deficiency recovered on 75-300 mcg. m	As coenzyme in carboxylation and decarboxylation reactions. Best known reaction is that of CO_2 transferring agent in conversion of acetyl Co-A to malonyl Co-A (extramitochondrial) synthesis of fat).	Spontaneous deficiency almost unknown. Avidin, a protein in egg white, binds biotin, making it unavailable for intestinal absorption. The human experimental deficiency causes dermatitis, anorexia, and myalgia.	No chemical method. Bioassay using chicks. Microbiological assay using L. arabinosus.
B_{12}; cyanocobalamin "Corrin" ring with atypical nucleotide containing cobalt. Heat stable.	All foods of animal origin, especially liver and kidney. Muscle meats, milk products and eggs are rather rich. Not present in plants.	Only 1 mcg needed to treat pernicious anemia, but man's req't ranges from 0.6 to 2.8 mcg.	1) As coenzyme with FAD and vit. B_{12} in methyl group synthesis from reduced folic acid and serine. 2) In conversion of propionate to succinate, which then enters TCA cycle. 3) In gluconeogenesis, fat metab., synthesis of thymine, and protein synthesis.	Pernicious Anemia. True dietary deficiency very rare. In most cases, there is an intestinal absorptive defect, notably in pernicious anemia, where "intrinsic factor", a gastric mucoprotein, is lacking.	Two types of tests are available: 1) Microbiological assay, using L. leischmanii or Euglena gracilis (see Hutner, et. al. 1949. Proc. Soc. Exp. Biol. and Med. 70:118). 2) the "Shilling test", another "loading test". Give 0.5 mcg. B_{12} labeled with cobalt-60 p.o., with or without intrinsic factor. Without intrinsic factor, 80-90% of the radioactivity appears in feces (with intrinsic factor, only 20-25%). If a flushing parenteral dose of 1mg of "cold" B_{12} is given 1 hr. after the oral dose, the amount of radioactivity flushed out in urine may be used as a measure of B_{12} absorption. (see Shilling. 1953. J. Lab and Clin. Med. 42:946). For a double isotope procedure for rapid estimation of B_{12} absorption, see Katz et. al. 1963. J. Lab. and Clin. Med. 61:266.
A Beta ionone ring with side group. Easily oxidized. Fat soluble. A_1 - retinal; A_2 - 3 dehydro retinal.	Liver, milk butter. Only animal sources contain preformed vit. A. yellow and green vegetables contain carotenes.	5000 I. U. (One I. U. - 0.3 mcg of vit. A alcohol) A_2 is only 40% as potent as A_1.	1. Dim light vision 2. Maintenance of epithelia and mucous membranes 3. In bone formation. 4. In synthesis of mucopolysaccharides (PAPS) 5. ? in adrenal steroid synthesis	1. Nyctalopia 2. Xerophthalmia and keratomalacia 3. Keratinization of epith. tissues and folliculitis 4. Large, soft, spongy bones; failure of skeletal growth.	Carr-Price reaction, in which a blue color is produced when solution of antimony trichloride in chloroform is added. Serum determination is good measure of vit. A status, because serum level does not change after ingestion (not true for serum carotenes). Carotenes not absorbed as well as vit. A, but beta carotene can produce two molecules of vit. A after absorption. In deficiency, dry keratinized epithelial tissues are more susceptible to invasion by infectious organisms.
D A group of sterols, chiefly of animal origin, which vary in potency. D_2 and D_3 most important. Fat soluble	Fish liver, oils, and viscera. Provit. D_2(ergosterol) in yeast. Mammals can synthesize pro-vit. D_3 from cholesterol, which is then activated in skin by U. V. light.	400-800 I. U. Req't very low or nil if there is exposure to sunlight. Ergosterol is activated to D_2 (calciferol) by U. V. in skin.	1. Increases calcium and phosphate absorption from gut. 2. With PTH, increases resorption of bone minerals. 3. Promotes ossification of cartilage 4. Increases serum citrate level 5. Activates alk. phosphatases.	Rickets. In children, bones become soft and bowed, with enlarged ends, and their mineral content decreases. Adult rickets (osteomalacia) is rare.	No good test for practical clinical use. The only test suitable for survey work is measurement of plasma alkaline phosphatase, which increases in Vit. D deficiency before clinical rickets occurs. Must always compare with normal children of same age group. A bioassay and a spectrophotometric assay are available but difficult.
E Alpha tocopherol; chroman ring with side chain. Fat soluble. Easily oxidized.	Milk, eggs, muscle meats, green plants, cereals, wheat germ, corn and soy bean oil.	30 mg of the acetate. (large amounts of unsaturated fat in diet increase requirement because the fats oxidize vit. E).	Not well understood. There is evidence that it acts as a cofactor in the electron transfer system between cytochromes b and c.	May rarely cause anemia in children. (Acts with selenium to prevent liver necrosis in protein-deficient rats).	Thin layer chromatography followed by spectrophotometric assay can be used and is quite sensitive. For measuring vit. E content of foods, a bioassay using pregnant rats is done. The fetuses are resorbed on E-deficient rations.
C Ascorbic Acid; lactone ring. Water soluble. Easily oxidized.	Citrus fruits and most other fruits, most vegetables, esp. salad greens. Easily destroyed by cooking.	70 mg is recommended but probably too high. 10 mg a day prevents scurvy in man.	Biochemical functions still not known. Helps maintain normal intercellular material of cartilage, dentine, and bone. Probably has several other functions.	Scurvy. In children, the joints of growing bones are affected, especially the costochondral junctions. Subperiosteal hemorrhage may cause severe pain. In adults, gum and skin changes.	Chemical test based on dye reduction by the reduced form of ascorbic acid. Has been adapted to the microdetermination of blood ascorbic acid. Only 0.01 ml. of serum needed. However, plasma levels are good indicators of recent intake but poor for tissue concentration. Leukocyte content of vit. C is best assay, but difficult. Best way to assess status is to give an oral dose of ascorbic acid (15 mg/kg body wt.) and measure plasma level 4 hrs. later. Depleted subjects show little or no increase.

extrapolate our results to human populations, and then attempt to test our conclusions on the latter.

Briefly described in Table III are some valuable tests of immunologic capacity. They are outlined and described in the table. The format for the table derives from the basic dual nature of the immune system: humoral and cellular. Each of these major limbs has an afferent and efferent part, and each limb involves cells which are usually morphologically indistinguishable but functionally heterogeneous. The general model, much of which is based upon experiment tal verification, is as follows: I. Lymphoid cells controlled by the bursa of Fabricius or its mammalian equivalent make humoral antibodies, some of which are "protective". The antibodies fall into 5 categories of immunoglobulins : IgG, IgM, IgA, IgD, and **IgG. IgG**, IgM, IgA may be primarily "protective" against disease, depending upon the antigen (infectious agent). In the primary response IgM is mainly produced; in the secondary response, IgG predominates. These responses may be altered differentially by nutrient deprivation, again depending upon the antigenic stimulus. II. Thymic-dependent lymphoid cells participate in CMI, which is responsible for delayed hypersensitivity, allograft rejection, and the graft-versus-host reaction, and which restricts intracellular parasitism (mycobacteria, viruses, fungi and protozoa) and maintains surveillance for neoplastic cells. Usually more than one cell type participates in the process of CMI (cell "cooperation"). Lymphoid cells involved in the immunologic recognition and response to altered or foreign cells appear to release a variety of soluble factors which would tend in nonspecific ways to destroy such abnormal cells (lymphotoxin) and render surrounding tissues resistant to further viral infection (interferon). Additionally, factors may be released which encourage the influx of granulocytes (leucotactic factor) and macrophages (migration inhibitory factor) and recruit lymphoid cells by specific (transfer factor) and nonspecific means (blastogenic factor) to augment the total response. The relationship between nutrient depletion and CMI has only recently begun to be explored. In some human benign and malignant lymphoproliferative diseases probably caused by oncogenic viruses: Burkitt's lymphoma, carcinoma of the posterior nasal space, infectious mononucleosis, and sarcoidosis, CMI is impaired. This impairment of cellular immune defense results in unrestricted proliferation of the virus-infected cells, predisposing to increased production of virus and presentation of its antigens to the antibody-forming apparatus of the host, with consequent increases in humoral antibody production. The end result is a sort of "antibody escape", where the diseased individual manifests CMI poorly, but may have inordinately high titers of antibody against the virus. Another interesting example of this escape mechanism occurs in patients with a mycobacterial disease: the lepromatous form of leprosy. Some of these individuals have poor CMI and high levels of circulating antimycobacterial antibodies. In the case of leprosy, it is likely that the insufficiency of CMI is a congenital defect, allowing Mycobacterium leprae to establish itself inside cells and produce slowly progressive leprosy, where levels of antimycobacterial antibodies have little protective effect against the disease.

The few recent studies attempting to demonstrate alteration of CMI by malnutrition indicate that, for example, protein malnutrition impairs the delayed skin hypersensitivity reaction to tuberculin. Delayed hypersensitivity probably plays a role in resistance to tuberculosis, and most authorities agree that protein deficiency impairs resistance to tuberculosis. On the other hand, some animal studies indicate that protein deficiency enhances cell-mediated immunity with respect to defense against certain intracellular organisms. For example, the proportion of mice surviving infection with swine influenza virus is greater in protein-deficient mice than in normally-fed mice. My own experimental results indicate that moderate protein depletion, while decreasing humoral response in general (mice, rats) may enhance cell-mediated responsiveness, but the level and duration of protein depletion is quite important, as is the age of the animal. Humoral responsiveness is attenuated in states of moderate protein depletion; as the depletion worsens, CMI is affected. I have found that chronic protein depletion (8 % casein diets) in inbred mice leads to a higher survival rate after infection with pseudorabies virus than does a diet normal in protein (27 % casein). The enhancement of CMI in depleted mice of the same inbred strain was further demonstrated by more rapid rejection of skin allografts and the greater in vitro blastoid response to phytohemagglutinin of spleen lymphocytes from depleted mice. Experiments concerning the humoral limb of the immune system gave a different kind of result. Whereas protein deficiency depresses the primary antibody response to sheep erythrocytes, it has no effect on either the primary or secondary agglutinating antibody response to Brucella abortus. Thus, chronic protein deplection in a particular strain of inbred mouse seems to have a disjunctive effect on humoral and cell-mediated immune responses, and a differential effect on the humoral antibody response, depending upon the type of antigen used to elicit it.

TABLE III. ASSAY OF IMMUNOLOGIC COMPETENCE IN MAN

some appropriate tests

HUMORAL IMMUNITY

TEST	Material and Methods	Interpretation and Comments
1. Immunoglobulins G, M, and A assays. (serum conc. of D and E are too low to be easily detected by this method).	Less than 1 ml. of serum. For each immunoglobulin, 0.01 ml of serum is required. Single radial immunodiffusion technique (Mancini).	The Mancini method (Mancini, G., et. al. 1965. Immunochem. 2:235) assumes a linear relationship between the area of antigen-antibody precipitate and the concentration of protein to be measured. The antigen diffuses into agar containing specific antiserum. Standard sera, with known amounts of immunoglobulin in question, must be used for reference. See also Rowe, D.S. 1969. Bull. Wld. Hlth. Org. 40:613. (Many other types of tests are available).
2. Complement assay.	Less than 1 ml. of serum, fresh, or frozen at -70°C.	The many proteins of the complement system usually function, after an antigen-antibody reaction has occurred, to destroy foreign cells, notably bacteria. Complement assays require skill and specialized equipment, and should not be attempted in the field.
3. Immunoglobulin E assay (Reagin; skin-sensitizing antibody).	1. Prausnitz-Kustner (PK) test (*in vivo*; qualitative) 2. The red cell linked antigen-antiglobulin reaction (*in vitro*; see Coombs et. al. 1968. Lancet i, 1115).	These globulins mediate immediate type hypersensitivity, and include the reagins, which have high affinity for skin, leukocytes, and other cells. They cause anaphylactoid reactions in man, which release histamine, serotonin, "slow reacting substance", and kinins. Anaphylactoid reactions may be mild (mucosal surface involvement, as in asthma), or severe (systemic, as in penicillin anaphylaxis). IgE may represent a major defense in both the upper and lower respiratory tracts.
4. Isohemagglutinin assay (for anti-A and anti B titers)	May be done on fresh (fingerstick) or stored blood. Needed are: at least 0.4 ml. of subject's serum, and normal types A_1 and B red cells. Reference anti-A and -B antisera are commercially available.	Four major blood group phenotypes exist in man: A, B, O (H), and AB. These are antigens which are a structural part of the surface membranes of red blood cells. Most persons with type A, B, or O blood haveserum antibodies (isohemagglutinins) against the A or B antigen not part of their own cells. "Natural antibodies", cross-reacting with A or B antigens, cause titers of the A or B isohemagglutinins to vary widely in different individuals.
5. Circulating antibody assay after antigenic stimulation		These tests measure the person's ability to produce antibodies to specific antigens, and range widely in sensitivity depending upon the method used.
A. Typhoid-paratyphoid vaccine or plain tyhoid (S. typhosa, heat-phenol or acetone-killed) vaccine both used.	Commercial vaccine is injected usually subcutaneously, according to manufacturer's directions. Pre-and post-inoculation Salmonella serum antibody titers are determined. (Many bacterial, viral, or other types of antigens may be used, depending upon relevance to prevailing conditions).	These assays are a measure of the ability of the tested subject to respond to antigenic stimulation by producing specific antibodies. They are tests of the immunocompetence of the humoral immune response, without regard to the biological function (if any) of the induced antibodies. Naturally acquired infectious disease almost always results in better immune protection against a reinfection with the same organism than does passive immunization. For example, in the case of S. typhosa, either infection or vaccination causes rises in at least 3 specific IgM antibodies, directed against the H, O, and Vi Salmonella antigens. However, vaccination induces mainly anti-H antibodies, while the natural disease, typhoid fever, may cause greater increases in anti-O and anti-Vi antibodies, generally considered most important prophylactically.
B. Shick test for diphtheria antitoxin. Done either before or after antigenic stimulation.	0.1 ml diphtheria toxin intradermally. Use heat-inactivated toxin as control. For prior immunization, use commercial D-T vaccine.	This is a highly qualitative test to answer the question whether the subject has already, or is able, to produce antitoxin to diphtheria toxin. A positive skin reaction (induration and erythema), beginning at 24-48 hours, and persisting for 7 days, indicates absence of circulating antitoxin (hence, susceptibility to diphtheria). If Shick-positive subjects are then immunized with diphtheria-tetanus vaccine, and remain Shick positive, a very poor humoral antibody-producing capacity is present.

TABLE III

(continued)

CELLULAR IMMUNITY

1. Skin tests A. Delayed hypersensitivity (D. H.) to "natural" antigens	Tuberculin (P. P. D.), histoplasmin, coccidioidin, mumps virus (inactivated), staphylococcal antigen, Candida antigen, and Trichophyton antigen have all been used.	These tests of cell-mediated immunity (CMI) are all similar in that they inquire into the competence of the thymic-dependent limb of the immune response. The presence of induration and/or erythema (depending upon the test) at the site of injection, usually after 48 hours, indicates: (1) that CMI is at least partially operative, and (2) that the tested subject has prior sensitization to the test antigen. Antigens should be chosen with consideration of (2).
B. Contact sensitization to artificial antigens	Dinitrochlorobenzine (DNCB), oxazolone (a synthetic pencillin-like substance), or picrylated albumin.	Similar to the above skin tests in detecting CMI, but probably more potent elicitors of it (see text). Since these chemicals are not antigens which usually occur in nature, the subject is first sensitized to them with a stronger solution than when challenging the patient 2-3 weeks later. Interpretation of all skin tests requires experience and the use of standardized techniques.
2. <u>In vitro</u> tests A. <u>Lymphocyte</u> "transformation" by mitogens. Non-specific (e.g. phytohemagglutinin) or specific (e.g. C. albicans) mitogens may be used.	The subject's blood lymphocytes, obtained by standard methods from heparinized blood; culture media; decomplemented serum; the mitogen to be used for lymphocyte stimulation.	Special laboratory apparatus and conditions are required for these tests; consequently the necessary materials should be assembled in a central place. In the presence of mitogens, small lymphocytes transform into large, metabolically-active (blastoid) cells, and proliferate. These events may be monitored either morphologically (% blast transformation) or biochemically with tritiated thymidine, which is incorporated into the DNA of dividing lymphocytes. The latter technique gives a more quantitative estimate of CMI, which, if impaired, is reflected in decreased transformability of the subject's lymphocytes <u>in vitro</u>.

3. Notes on other tests for CMI: Clinical examination of the patient and of his blood, bone marrow, and lymphoid tissue should yield valuable information on the extent to which his thymic-dependent immune system is active. Other <u>in vitro</u> tests are also available.

In Table III I have listed the immunological tests which are practicable for general use, where the resources of a research laboratory may not always be available. It is evident that measurements within both limbs of the immune system are necessary, and that tests should be selected on the basis of the type of information needed and the prevailing conditions.

Immunoglobulin assays are numerous and vary widely in sensitivity. The most sensitive methods (bactericidal and virus neutralization tests) can detect as little as 10^{-5} mg of antibody nitrogen per ml. of test material. Less sensitive are precipitin tests and immunoelectrophoresis. The radial immunodiffusion method of Mancini is intermediate in sensitivity between these two extremes and has been adapted for field use. Materials are available commercially to permit the expeditious performance of these measurements. For example, using this technique, Egyptian workers found that serum concentrations of immunoglobulins G, M, and A were all below normal in babies with clinical kwashiorkor under the age of 7 months. IgM concentration continued to be very low in kwashiorkor victims as old as 2 years, and 6 months after cure of the disease IgM levels remained low. IgA concentration was low before age 7 months, but in infants 1-2 years old the levels were thrice normal. The pattern for IgG was similar to that of IgA, but IgG concentrations in the older babies were only moderately elevated. (Aref, G. H., et. al; in press). I emphasize this work for two reasons:(1) current immunologic methodology was used to study the breakdown of "natural" immunoglobulins in a malnourished population before and after dietary treatment; and (2) the immunologic implications of the results are important, especially with respect to IgM. For several reasons, IgM may be the most suitable immunoglobulin candidate for combating bacterial infections in the systemic circulation (see Altemeier, W. A., et. al., J. Immun. 103: 924, 1969). The Egyption work indicates that protein-calorie malnutrition in infancy may suppress IgM synthesis for a prolonged period, perhaps permanently. The production of IgM represents, both phylogenetically and ontogenetically, the most primitive and basic humoral antibody response. Clearly, further investigation should be done of the effect of malnutrition on the production of IgM directed toward specific antigenic stimuli in malnourished subjects.

The low IgA levels in early life followed after 1-2 years by elevations three times normal, found in the Egyptian study, also raises an interesting point. **IgA** is synthesized by plasma cells in mucous epithelial surfaces of the respiratory tract and intestine and in nearly all excretory glands. Secreted onto mucous epithelial surfaces and also into the bloodstream, IgA serves as the first line of immunological defense against invasion by microorganisms, and is especially important in preventing gastrointestinal infections and infections of the secretory glands. Infants in poorly-developed areas invariably have high infectious loads of gastrointestinal microorganisms, frequently succumbing to diarrheal disease during the first years of life, especially after weaning. (IgA is present in breast milk, although withdrawal of this source of IgA is not the most likely principal cause of weanling diarrhea. Substitution of post-weaning feeding formulas of far less nutritional value than breast milk is undoubtedly the major cause of these diarrheas). It has recently been found that antigentic stimulation via the gut not only leads to production of IgA in bowel secretions, but also to the selective production of antibodies of IgA type in the circulation. Conceivably, IgA might combine with antigen and make it a less effective stimulator to the IgM-or IgG-producing cells. This mechanism might explain the shift from low to high IgA levels in very young and in older infants, respectively, found in the Egyption study. It also might, in part, account for the continued low levels of circulating IgM in these infants.

This leads to the mention of a subject important in malnourished populations, that of antigenic competition, defined as follows: when two or more antigens are administered to a subject, the response to each individual antigen may be less than the expected response if a single antigen were given by itself. Antigenic competition probably represents a competition for commitment of individual cells to the production of a specific antibody. Individual cells usually produce only one kind of antibody upon stimulation by several antigens. This type of competition for antibody-producing cells by multiple antigens may play a role in the poor immune resistance of malnourished populations, where the infectious loads (exposure to many different pathogens) in early life are enormous. Such children usually have higher than normal IgG levels, as illustrated in the case of the Egyptian study, where infants 1-2 years old had high levels. These antibody levels are high because of the hypersensitized state of the individual, but this does not mean that his humoral defense mechanisms for specific infections are more active. They are probably less able to cope with specific "superinfections" than in the case of normal children. For example, Gambian children repeatedly infected with malaria in early life, have higher gamma globulin levels than their malaria-free counterparts, and much of this gamma globulin is undoubtedly specific immunoglobulin directed toward malaria and other endemic infectious diseases. However, when given tetanus toxoid, an antigen unrelated to the malaria plasmodium, the malaria-infected children responded very poorly in producing tetanus antitoxin (McGregor I. A. and Barr, M. 1962. Trans. Royal Soc. Trop. Med. Hyg. 56:364). This specific example of antigenic competition is doubtless one of the many which are occurring throughout the less developed world, and illustrates the fact that malnourished populations cope immunologically with their infections progressively less efficiently as the infectious load increases.

Concerning tests for CMI (Table III), the ones available can be done *in vivo* (skin tests), or *in vitro* (cellular studies). CMI is much more difficult to quantitate than is humoral immunity. Depending upon antigenic strength, a wide range of cellular immune responsiveness can be detected with skin tests, but only qualitatively. Patients with lepromatous leprosy and depressed CMI often cannot be sensitized to dinitrochlorobenzene, but can be induced to develop skin sensitivity reactions to the more powerful antigen: hemocyanin. Thus, the degree of depression of CMI could probably be semi-quantitatively assessed if a battery of sensitizing antigens of increasing antigenic strength were developed for this purpose, but the investigator would still be faced with many problems of interpretation of results, such as antigenic cross-reactivity, the subjects' pre-exposure to antigens, etc.

IgE, or skin-sensitizing reagins, are immunoglobulins responsible for immediate hypersensitivity reactions (15 minutes to about 6 hours after exposure to the antigen), as opposed to reactants which determine delayed hypersensitivity (D. H.), where IgE does not participate. D. H. manifests itself 24-48 hours or more after antigenic challenge. Some tests for D.H. are outlined in Table III, under the heading cellular immunity. Some cellular studies for CMI require the procurement of stained blood, bone marrow, or solid lymphoid tissues, so that the presence and morphology of the cell types participating in CMI may be assessed. The range of tests available for the *in vitro* study of CMI is expanding rapidly, but because the performance of the more complex tests usually requires special skills and apparatus, strictly field use of them is currently impractical.

At the outset, I implied that world food supply was decreasing with respect to world population. Very recently, I have read that during the past 2-3 years, world food production per capita has increased. These statements, superficially, appear to contradict each other. The real question is probably disguised in the statements themselves, and cannot be answered in such simple terms. The answer has to do with the quality-food imbalances that exist today on our planet, and with the myriad forces which have created that imbalance. In 1964, president Mohammed Ayub Khan made the following statment: "In ten years' time, human beings will eat human beings in Pakistan".

Nutrient deficiency, especially that of protein in the less-developed world, has produced generations of people who are incompetent physically, mentally, and perhaps immunologically. If, indeed, the world's quality foods were more evenly distributed, it would not be necessary, except from a basic standpoint, to study the interactions of malnutrition and immunocompetence, and president Khan, in 3 years' time, would have to "eat his words". In the meantime, it will become necessary to explore, in greater depth, the effects of malnutrition on immunologic performance.

We are now in a position to consider priorities and feasibilities in attempting to assess nutritional status. Our current degree of sophistication in nutritional metabolism and in immunobiology has created new problems, primarily logistic and technical. An analogy with the field of cancer might be drawn. We no longer think of cancer as a single type of deviation from the normal, with a single etiology. We think of cancer as a variety of different diseases of different etiology, and our diagnosis must be quite specific for proper treatment to be applied. In much the same way, we cannot regard malnutrition as a single entity, but as many different diseases leading to many types of dislocations in bodily homeostasis, depending upon the etiology of the disease. More than any other biological abnormality, malnutrition is culture-dependent. Consequently, in order to cope with problems of human malnutrition, we must thoroughly understand the culture in which they develop and then apply our methods of diagnosis and treatment.

THE RELATION OF PARENTS' EDUCATION AND EDUCATIONAL ASPIRATIONS FOR CHILDREN TO FERTILITY IN TAIWAN

Fred Arnold, Population Studies Center, University of Michigan

I. INTRODUCTION

Much of the research behind current efforts to control fertility and slow population growth has focused on the relative success of various methods of contraception and on the administrative efficiency of family planning programs. According to Kingsley Davis, family planners evince an "intense preoccupation with contraceptive technology."[1] Family planners have placed less emphasis on the longer range solutions which require an investigation of the socio-economic determinants of fertility norms and family size desires. In particular, the role of education as a determinant of fertility norms and behavior has not received the attention it deserves. If we are interested in encouraging couples to limit the size of their families, we should look for characteristics of individuals that are both determinants of fertility behavior and subject to social control. Ronald Freedman has suggested that education might be such a "leading" variable.[2] If basic changes in the attitudes and structure of society may be a necessary antecedent to any major reduction in fertility, as some authors have suggested, then education may be able to stimulate such changes. There is need, therefore, for research to determine what role education might play in contributing to significant declines in fertility.

The research that has been done on the education-fertility relationship suggests that educational attainment has a very powerful influence on fertility and family planning behavior in both industrial and nonindustrial countries. Knodel and Tangoantiang have shown that the link between education and fertility was quite strong during the demographic transition in Europe. They found that neither industrialization nor urbanization were consistently related to the fertility declines in Europe between 1860 and 1920. In fact, the only index examined whose movement was found to parallel that of fertility measures, almost without exception, is illiteracy.[3] In a 1963 United Nations study of 125 countries, the correlation between the gross reproduction rate and the literacy rate of women 15 and over was -.76.[4] This is one of the highest coefficients revealed in that study.

While there are indications that the educational differentials in fertility are narrowing in some developed countries,[5] the inverse relationship still persists. Notable exceptions to this negative association exist, but they are found mostly in population subgroups such as Catholics in the United States.[6] Inverse relationships between education and fertility have been found in such diverse countries as the United States, Japan, Trinidad, Malaysia, Sweden, Taiwan, England, the USSR, India, Korea, France, Italy, and Panama. In short, nearly all studies to date have concluded that education is one of the principal factors associated with fertility. Most previous studies have used sample survey or census data to document this negative relationship. However, very rarely have researchers gone beyond simple cross-tabulations between education and fertility, occasionally standardized for age. Leon Tabah recognized this problem at a recent UNESCO Meeting of Experts on "Education and Family Planning." He indicated that, "the tables showing the relationship between fertility and education published by those who have undertaken such studies are not usually a satisfactory basis for a thorough analysis of this relationship."[7] This paper attempts to partially remedy this situation by investigating the link between education and fertility in a multivariate framework.

Clyde Kiser has shown that the correlation between education and fertility is particularly high in the demographic transition period.[8] Since Taiwan is currently experiencing its demographic transition, it is an excellent country for an intensive study of the effects of education on fertility.

The mechanism by which increased educational attainment induces lower family size goals is no doubt complex. Education has social, economic, and demographic effects on an individual. In Taiwan, education will, in general, increase one's exposure to mass media, enhance one's opportunities for a better job and higher income, influence age at marriage and the probability of ever getting married, increase aspirations to own modern consumer goods and induce significant changes in traditional views and attitudes. All of these effects have implications for fertility and family planning behavior.

One attitudinal variable that may be affected by the educational process is the aspirations of parents for their children's education. It has been suggested that more highly educated parents will have higher educational aspirations for their children, and that these parents will tend to limit the number of children they bear, in order to be able to fulfill their higher aspirations.[9] This paper will focus on the relationship between education and fertility in Taiwan and the extent to which educational aspirations for children act as an intervening variable in this relationship. It will also investigate the interaction between the education of the husband and that of the wife as it affects their family size decisions.

An island-wide fertility survey conducted in Taiwan in the summer of 1969 provides the source of data for this study. Interviews were conducted with a cross-section of 2,200 currently married Taiwanese husbands whose wives were of childbearing age (less than 42 years old). The husbands were asked questions on a variety of subjects, ranging from their demographic and socio-economic characteristics to their attitudes and behavior with respect to fertility and contraceptive practices. The wives had been interviewed in a previous survey 20 months earlier, so that detailed fertility histories of the couples merely required updating in the more recent survey.

II. EDUCATIONAL DIFFERENTIALS IN FERTILITY AND FAMILY PLANNING IN TAIWAN

The educational composition of the Taiwanese population has been changing rapidly since the end of the Japanese occupation of the island. The number of primary schools in Taiwan doubled between 1944 and 1969, while the number of secondary schools increased tenfold. The number of universities and junior colleges increased from 5 in 1944 to 74 in 1969. During the same period of time, the

Table 1--Educational Composition of Married Couples, Age 22-42, Taiwan, 1969

Educational Attainment	Husbands	Wives
None	10.6%	36.6%
Primary school, not grad.	10.9	8.9
Primary school grad.	53.7	43.2
Junior high grad.	8.7	6.3
Senior high grad. or higher	14.9	4.7
Other types of schools or NA	1.2	0.3
TOTAL	100.0	100.0
Number of cases	2279	2279

enrollment rate of primary school age children rose from 71.3% to 97.5%. In 1968, the Taiwanese government extended basic elementary education from six years to nine years. The enrollment rate at the public junior middle school level (grades 7-9) increased from 57.6% before this educational reform to 73.7% in 1969.[10]

Table 1 shows the educational composition of respondents and their wives, based on the 1969 Economic Correlates of Fertility Survey in Taiwan. The husbands are, in general, better educated than their wives; however, more than half the wives are primary school graduates, and nearly 2/3 of the women have received some formal education.

The usual inverse relationship between education and fertility is evident in Table 2 for a variety of measures of fertility and both husband's and wife's educational attainment. Women with no formal education have an average of 48% more living children than women who have continued their education beyond junior high school. Similarly, women with no education want, on the average, 36% more children than women in the highest education category. Finally, ideal family size for women in the lowest education group is approximately 1 child greater than for those who went beyond junior high school. Furthermore, more highly educated women marry later in life than those with less education. Thus, educational differentials in fertility might be even greater if we considered all women in their childbearing years rather than only currently married women. For all three fertility measures, the reduction in fertility is most pronounced for women who have gone beyond a primary school education.[11] This threshold is quite important in light of the recent extension of basic education in Taiwan through the junior high school level. Speculation about the effects of the recent educational reform yields some interesting results. If all the women in our sample with less than a junior high school education had the same low level of fertility as actual junior high school graduates of the same age, then the average number of living children in the sample would drop by 19%, from 3.7 children to 3.0 children. Similarly, women over 35 would have 3.7 living children, on the average, instead of 4.7. These calculations are based on the unadjusted means, and, therefore, they do not take into account variables (such as type of area) that are related to both education and fertility, but are not necessarily affected by education. Thus, they should be thought of as an upper limit of the potential reductions in fertility that could result from universal junior high school education.

In the case of husbands, the major drop in fertility does not appear until the senior high

Table 2--Fertility and Family Planning Attitudes and Behavior, by Wife's and Husband's Education

Education	Average number of children*: Living	Wanted	Ideal	Percent ever used contraception	No. of cases
WIFE					
None	4.0	4.5	4.0	48%	822
<Primary	3.8	4.3	3.9	57	212
Primary grad.	3.6	4.2	3.7	58	949
Junior high	3.0	3.6	3.3	72	164
Senior high +	2.7	3.3	3.1	82	114
HUSBAND					
None	4.1	4.7	4.2	37%	246
<Primary	3.7	4.3	3.9	47	256
Primary grad.	3.8	4.4	3.9	54	1143
Junior high	3.7	4.2	3.7	65	275
Sr. high grad.	3.1	3.6	3.4	73	234
College	2.9	3.4	3.1	85	107
TOTAL	3.7	4.2	3.8	56%	2261

*Wanted--no. living at interview + no. additional wanted; survey question was: "Let's see now, you have ___ children. Do you feel that the number you now have is sufficient or would you like to have more?" [IF MORE] "How many additional children would you like to have?"
Ideal--survey question was: "If you were just getting married and could have just the number of children you wanted, how many would you like to have had when your wife is through having children, about age 45?"

school level. Nevertheless, husbands with a junior high school education have, on the average, 11% fewer living children than those with no formal education. The final column in Table 2 shows a strong positive relationship between educational attainment and contraceptive use. Additional data (not shown here) indicate that more highly educated couples not only use contraception more frequently, but also begin use at lower parities than those with less education. Among women who have used contraception, those with more than a primary school education began using after 2.6 births on the average, whereas women with no education began after 4.3 births.

Some additional data confirm the inverse relationship between education and fertility and the direct relationship between education and contraceptive use for every five-year age group in our sample. Thus, these associations are not simply a spurious result of the relationship of the couple's age to both education and fertility.

Table 3 shows how husband's and wife's education interact in fertility and family planning decisions. For each of the three measures of fertility, it can be seen that the influence of wife's education supplements, and in most cases dominates that of the husband. This is consistent with the results of an earlier study of family planning in Taiwan in which Ronald Freedman and John Takeshita found wife's education to be "the most consistently useful measure in our wide range of work with the Taiwan data."[12] It is possible that less tradition-oriented husbands choose better educated wives, and that wife's education is partially a proxy for the orientation of the husband toward modern values (including small family size). However, we do not believe that this phenomenon can

Table 3--Fertility and Family Planning Attitudes and Behavior, by Husband's(H) and Wife's(W) Education

Education	Average number of children* Living	Wanted	Ideal	Percent ever used contraception	No. of cases
H<PRIMARY GRAD.	3.9	4.4	4.0	42%	495
W none	4.0	4.6	4.1	38	336
W<prim. grad.	4.0	4.4	4.0	58	62
W prim. grad.	3.6	4.1	3.8	43	93
W jr. high +	[2.8]	[3.8]	[4.0]	[50]	4
H PRIMARY GRAD.	3.8	4.4	3.9	54%	1132
W none	4.1	4.6	4.0	54	432
W<prim. grad.	3.7	4.2	3.9	55	126
W prim. grad	3.7	4.2	3.8	54	543
W jr. high +	2.8	3.6	3.3	45	31
H JUNIOR HIGH	3.7	4.2	3.7	65%	273
W none	4.0	4.3	3.7	62	45
W<prim. grad.	[3.6]	[4.1]	[3.6]	[72]	17
W prim. grad.	3.7	4.3	3.8	62	173
W jr. high +	3.4	3.8	3.6	83	38
H SR. HIGH +	3.1	3.6	3.3	76%	337
W none	[3.2]	[3.5]	[3.8]	[25]	4
W<prim. grad.	[3.3]	[3.7]	[4.0]	[42]	7
W prim. grad.	3.5	3.9	3.5	75	129
W jr. high +	2.8	3.4	3.2	79	197
ALL COUPLES	3.7	4.2	3.8	56%	2237

*See Table 2. [ ] = base less than 25 cases.

account for all the effect of wife's education on fertility. Therefore, if a reduction in fertility is considered to be a desirable goal in Taiwan, then any educational policy or practice which neglects the education of females may well be misdirected.

Husband's education, on the other hand, is more important than wife's for determining the couple's use of contraception. Contraceptive use is positively related to husband's education for each level of educational attainment of the wife. However, there is no consistent positive relationship between contraceptive use and wife's education, after controlling for husband's education. This finding appears at first to be somewhat anomalous and we can only speculate as to its cause. Perhaps the husband has the predominant influence on the decision to use contraception, whereas the wife determines the effectiveness with which it is used and, thereby, ultimate fertility. This is quite plausible, since the most popular contraceptive methods in Taiwan (the loop, Ota ring, and pill) are female-oriented. However, these interrelationships require further study before we can reach any firm conclusions.

III. EDUCATIONAL ASPIRATIONS FOR CHILDREN

One possible mechanism by which increased education becomes translated into smaller family size goals is through the aspirations of the parents for their children's education. More highly educated parents have higher educational aspirations for their children. The realization of these high educational goals may require couples to limit the size of their families. A further reduction in fertility is implied by this mechanism, although it is considerably more long range. If parents with high educational aspirations for their children are able to fulfill these aspirations, then their children, in turn, may be expected to have low fertility when they reach childbearing age.

Table 4--Percent of Respondents Who Plan to Send Their Children to College, by Parents' Education

Husband's education	Wife's education: None	< Pr. grad.	Pr. grad.	Jr. high+	Total
			SONS		
< Primary grad.	40	44	53	[75]	43%
Primary grad.	53	61	74	87	65
Junior high	76	[82]	86	97	86
Sr. high grad. +	[80]	[100]	97	98	97
TOTAL	48%	59%	76%	96%	67%
			DAUGHTERS		
< Primary grad.	22	28	33	[50]	25%
Primary grad.	31	38	53	63	43
Junior high	44	[47]	70	82	66
Sr. high grad. +	[75]	[86]	85	97	92
TOTAL	27%	36%	57%	89%	48%

Survey question was: "We'd like to know what plans you have for the education of your children. What is the highest school which you expect your sons (daughters) to attend?"
[ ] = less than 25 cases.

Although the Economic Correlates of Fertility Survey interviewed husbands and asked them for their aspirations for their children's education, Table 4 shows quite clearly that both husband's and wife's education have a strong influence on these aspirations. Educational plans for both sons and daughters are strongly affected by the education of both parents. Even in a culture that is so highly educated-oriented, it is striking that 2/3 of the respondents expressed a desire to have their sons attend college and nearly half want their daughters to attend college. However, it is quite likely that some of these aspirations are more "realistic" and more certain of fulfillment than others. For this reason, the survey included a question concerning the respondent's knowledge of the costs of sending a child to college. Presumably, if a respondent had thought about the costs of educating his children and could give an estimate of educational costs, then his plans for sending his children to college may be considered more "realistic."* In fact, only 25% of all respondents wanted their children to attend college _and_ evinced some knowledge of costs.

The percent of respondents who had such "realistic plans" to send their children to college is shown in Table 5. Husband's educational attainment and total family income are both important determinants of educational aspirations. Data not shown here reveal that income is a somewhat more important determinent of educational aspirations for daughters than sons. In fact, 44% of low income respondents have higher educational aspirations for their sons than for their daughters, while only 15% of high income respondents favor their sons' education.

In order to determine the extent to which educational aspirations for children have an independent effect on fertility and family planning behavior, a multivariate analysis is necessary, and this will be undertaken in the next section. However, some indication of the influence of educational aspirations can be seen in Table 6. Respondents were asked what they considered to be the main advantages of having a small family and

Table 5--Percent of Respondents Who Both Plan to Send Their Children to College and Have Some Knowledge of the Costs of College Education, by Husband's Education and Total Family Income

Husband's education	Income, 000's NT$ per year <12	12-24	24-48	48+	Total
< Primary grad.	3	10	10	[15]	8%
Primary grad.	9	12	25	34	18
Junior high	[25]	30	38	40	36
Sr. high grad. +	[100]	53	63	68	65
TOTAL	8%	15%	33%	49%	25%

Survey questions for knowledge were: "Do you know or have you ever tried to find out how much it costs to send a child to college?" and "How much do you think it will be, considering all costs?"

the main disadvantages of having a large family. Those who spontaneously mentioned that it is easier to educate children in a small family (or harder in a large family) had lower fertility than others by each of the three fertility measures and were more likely to be using contraception at the time of the interview. Moreover, these differentials in fertility appear for every level of husband's education and the differentials in contraceptive use for all husbands with less than a senior high school education. Thus, it seems that if children's education is a salient factor to a couple, they will tend to limit the size of their family and use contraception more readily. It may be argued that some of the causality runs in the opposite direction, i.e., couples with smaller families are better able to educate their children and, therefore, they have higher aspirations for their education. However, the fact that the negative relationship between educational aspirations and fertility appears for measures of desired and expected fertility as well as actual fertility helps determine the direction of causality.

IV. MULTIPLE CLASSIFICATION ANALYSIS

In this section, the links between education of parents, educational aspirations for children and fertility and family planning are clarified by means of multiple classification analysis (MCA),[13] an extension of multiple regression using dummy variables. This technique is useful for examining the effect of each of several predictors simultaneously on a dependent variable when the effects of other predictors are held constant. MCA makes no special assumptions about the linearity or ordering of the categories of the independent variables and is equipped to handle independent variables on a nominal scale. The dependent variable is required to be either dichotomous or on an interval scale. The MCA program yields a mean value of the dependent variable for each category of the independent variables and an adjusted mean, controlling simultaneously for all the other independent variables considered.

In Table 7, we have used a dichotomous dependent variable to study the determinants of educational aspirations for children. The first column shows the unadjusted class means of percent of respondents having "realistic" college plans for their children. There is a very strong positive relationship between this variable and either husband's or wife's education. We want to know how much of this relationship is a reflection of other variables such as family income, which are

Table 6--Fertility and Family Planning Behavior and Attitudes in Relation to Salience of Children's Education

[M, mentioned children's educ. spontaneously; NM, did not]

Husband's education	Avg. no. of children: Living M	Living NM	Ideal M	Ideal NM	Wanted M	Wanted NM	Percent currently using contr. M	NM
None	4.0	4.2	3.9	4.2	4.3	4.7	34%	30%
< Primary grad.	3.6	3.8	3.7	3.9	4.1	4.3	47	32
Primary grad.	3.6	3.9	3.7	3.9	4.1	4.4	46	43
Jr. high grad.	3.4	3.6	3.5	3.8	4.0	4.3	53	50
Sr. high grad.	3.1	3.2	3.2	3.5	3.2	3.8	62	62
College	2.7	3.2	3.1	3.2	2.9	3.6	76	78
ALL HUSBANDS	3.4	3.8	3.5	3.9	3.9	4.4	52	44

related to both education and educational aspirations. This can be seen in the remainder of the columns in Table 7, which contain the "adjusted means." These figures show the net relation between education and educational aspirations after various combinations of other independent variables have been held constant. The differentials in educational aspirations by parents' education are diminished somewhat when controlling for total family income and the type of area in which the respondent lives. They are reduced slightly more when controlling for the educational attainment of the other spouse as well. However, wife's age and number of living children have practically no net effect on educational aspirations. Parents' educational attainment, particularly that of the husband, retains a strong positive net effect after controlling for all the other independent variables considered.

A multiple classification analysis of the relation between education and educational aspirations as determinants of fertility is shown in Table 8. In this table, the number of living children is related to education of the parents and their plans for children's education. Differentials by husband's education are diminished when controlling for the effects of either wife's education or educational aspirations for sons. In fact, the inverse relationship disappears when the effects of wife's education and educational aspirations for sons and daughters are taken into account (next to the last column). This finding lends support to the hypothesis that educational aspirations for children act as an intervening variable between parents' educational attainment and their fertility. Differentials by wife's education are also reduced when other independent variables are controlled for. However, the net inverse relationship between number of living children and wife's education persists in every case.

The great importance of parents' education as determinants of fertility and family planning is further demonstrated in Table 9, which shows the values of the partial beta coefficients for predictors in 5 different MCAs. These betas indicate the relative importance of the various predictors in their joint explanation of the dependent variable.† In nearly every case, duration of marriage emerges as the principal determinant of fertility and family planning. While the largest part of the variance in the dependent variables is explained by

Table 7--Relation between Socioeconomic Characteristics and Educational Aspirations for Children. Percent Planning to Send Children to College and Stating Knowledge of Costs, by Husband's Education(H), Wife's Education(W), and Total Family Income(I)

	No. of cases	Un-adj. mean	Mean adj. for type of area and--		
H'S EDUCATION			I	I,W	I,W,G,N
< Primary grad.	483	8%	12%	15%	15%
Primary grad.	1096	18	19	20	20
Jr. high	266	36	32	31	32
Sr. high grad. +	332	65	58	52	53
W'S EDUCATION			I	I,H	I,H,G,N
None	792	10%	15%	20%	20%
< Primary grad.	207	14	17	20	21
Primary grad.	914	30	28	27	28
Jr. high +	264	61	53	37	37
INCOME, NT$			H	H,W	H,W,G,N
<12,000	399	8%	17%	18%	18%
12,000-23,999	697	15	20	21	21
24,000-47,999	720	33	29	29	29
48,000 +	361	49	36	34	34
GRAND MEAN = 25%					

G = Wife's age, N = No. of living children.

demographic factors, socio-economic variables are also quite important. The betas for combined educational attainment of the parents are much higher than the betas for any other socio-economic variable included in the analysis. Thus, parents' education seems to be a major determinant of decisions about fertility and family planning in Taiwan. Wife's education appears to be decidedly more important than husband's education in its influence on fertility, whereas husband's education is considerably more important than wife's as a determinant of contraceptive use.

The net influence of educational aspirations for children on the fertility variables, while considerably smaller than the influence of parents' education, is at least as large as the effects of income per adult or type of area, variables which have conventionally been considered to be important determinants of fertility. Moreover, in this analysis we have used a fairly crude measure of educational aspirations for children. Perhaps educational aspirations would have shown up as a stronger determinant of fertility if we had an improved measure for it. Future surveys that include educational aspirations would do well to expand the questioning on this topic.

V. SUMMARY AND CONCLUDING REMARKS

Parents' educational attainment has a very important bearing on fertility and family planning in Taiwan. More highly educated couples have a smaller number of living children and lower family size goals than couples with less education. Moreover, the higher the education of the parents, the more receptive they are to the use of contraceptives to limit the size of their families. Thus, education facilititates the process of economic development not only by making the populace more literate and enhancing worker productivity, but also by inducing reductions in fertility. Therefore, if slower population growth is considered to be a desirable goal in Taiwan, increased government expenditures on education should be favorable

Table 8--Multiple Classification Analysis of Number of Living Children

	Un-adj.	Means adjusted for--							
H's Education	mean	WE	AS	DM	IN	AR	EP	A	B
< Prim. grad.	3.9	3.7	3.8	3.8	3.9	3.9	3.8	3.6	3.7
Prim. grad.	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.7
Jr. high	3.7	3.8	3.8	3.8	3.7	3.7	3.7	3.9	3.8
Sr. high +	3.1	3.5	3.2	3.3	3.2	3.1	3.1	3.6	3.6
W's Education		HE							
None	4.1	4.0	4.0	3.9	4.0	4.0	4.0	4.0	3.8
< Prim. grad.	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8
Prim. grad.	3.6	3.6	3.7	3.7	3.7	3.7	3.7	3.7	3.7
> Prim. grad	2.9	3.0	3.0	3.3	3.0	2.9	2.9	3.1	3.4
Ed. Plans and Kn. of Costs		HE	WE						C
< Sr. high	4.2	4.1	4.0						3.9
Sr.-no costs	3.8	3.7	3.7						3.8
Sr.-costs	4.2	4.2	4.2						3.8
Col.-no costs	3.5	3.5	3.5						3.7
Col.-costs	3.5	3.7	3.8						3.6
GR. MEAN = 3.7									

HE = Husband's education, WE = Wife's education
AS = Educational aspirations for sons
IN = Income per adult, AR = Type of area
DM = Duration of marriage
EP = Educational plans and knowledge of costs
HN = Husband's employment status
A = Spouse's ed., ed. asp. sons, ed. asp. dtrs.
B = Spouse's ed., AS, IN, AR, DM, HN
C = HE, WE, IN, AR, DM, HN

to a continuing fertility decline in that country. Moreover, since the educational attainment of the wife seems to be more decisive than that of the husband as a determinant of actual and desired family size, educational policies should not neglect the education of women. We expect that the educational reform instituted in 1968 will significantly affect the ultimate fertility of those now in school. Further efforts in the educational sphere should provide even greater dividends and shoud be well-received by parents who have such high aspirations for their children's education.

Educational aspirations for children seem to have a role as an intervening variable between parents' education and fertility. Parents with higher educational attainment have higher aspirations for their children's education. They also appear to have more "realistic" plans to give their children a college education since they are more aware of the costs of such an education. These high aspirations, in turn, induce them to limit the size of their families because of the high expected costs of their children's education. Although this is the primary role of educational aspirations, they also exert some influence on fertility independently of parents' education. Thus, the possibility of independently stimulating educational aspirations and cost consciousness with respect to children's education should not be overlooked.

Future studies in this area will examine the importance of other intervening variables such as age at marriage, income, occupation, and attitudinal factors. Hopefully, we will be able to sort out the strongest paths and, thus, gain a better understanding of the link between education and fertility.

Table 9--Betas for Multiple Classification Analyses of Fertility and Family Planning--Five MCAs with Various Combinations of Independent Variables for Each of Four Dependent Variables

Dep. Vars.	HE	WE	DM	AS	EP	HW	AR	HN	IN	R^2
No. living children										
(1)	.06	.20	---	---	---	---	---	---	---	.05
(2)	---	---	---	.12	---	.18	---	---	---	.06
(3)	---	---	---	---	.15	.22	---	---	---	.06
(4)	.05	.08	.64	---	.05	---	.03	.03	.03	.47
(5)	---	---	.64	---	.05	.11	.03	.02	.03	.47
No. wanted children										
(1)	.09	.20	---	---	---	---	---	---	---	.07
(2)	---	---	---	.10	---	.21	---	---	---	.06
(3)	---	---	---	---	.09	.23	---	---	---	.06
(4)	.07	.11	.40	---	.06	---	.04	.02	.06	.23
(5)	---	---	.40	---	.06	.15	.04	.02	.06	.22
No. ideal children										
(1)	.12	.18	---	---	---	---	---	---	---	.07
(2)	---	---	---	.05	---	.26	---	---	---	.07
(3)	---	---	---	---	.07	.25	---	---	---	.07
(4)	.08	.14	.16	---	.07	---	.03	.05	.07	.10
(5)	---	---	.16	---	.07	.20	.03	.06	.08	.10
% Ever used contracept.										
(1)	.19	.06	---	---	---	---	---	---	---	.05
(2)	---	---	---	.06	---	.21	---	---	---	.05
(3)	---	---	---	---	.12	.19	---	---	---	.06
(4)	.15	.08	.36	---	.09	---	.02	.05	.12	.20
(5)	---	---	.35	---	.09	.19	.02	.05	.13	.19

HW = Husband's and wife's education combined. For other abbreviations, see Table 8.

ACKNOWLEDGEMENT: Research was supported by the Population Council and the Population Studies Center, University of Michigan. The author would like to thank Eva Mueller and Ronald Freedman for many helpful suggestions.

NOTES

*For Taiwanese couples, the options for sending their children to college are varied, including local or away from home colleges, public or private colleges, military colleges, and even study abroad. The costs of higher education, then, depend on the type of institution visualized by the respondent. Consequently, no unique figure can be construed as the "correct" cost of a college education and all estimates were accepted.

†These betas should not be confused with the beta coefficients in multiple regressions. In the context of multiple classification analysis, beta is defined to be a positive value. The direction of the relationship is shown by the deviations of the adjusted mean from the grand mean. Although the beta statistic should be used with some caution, it can be useful as a measure of the relative importance of the predictors when it is looked at in conjunction with the adjusted means of the predictors.

1. Kingsley Davis, "Population Policy: Will Current Programs Succeed?", Science, Vol. 158, Nov. 10, 1967, p. 734.

2. Ronald Freedman, "Norms for Family Size in Underdeveloped Areas," in David Heer, ed., Readings on Population.

3. Unpublished research by Knodel and Tangoatiang, cited in Freedman, op. cit., p. 168.

4. United Nations, Population Bulletin No. 7, 1963, with Special Reference to Conditions and Trends of Fertility in the World.

5. Clyde Kiser, Wilson Grabill, and Arthur Campbell, Trends and Variations in Fertility in the United States, Harvard University Press, Cambridge, 1968, p. 151.

6. For differentials in actual and expected fertility among Catholics in the U. S., see C. F. Westoff, R. G. Potter, P. C. Sagi, and E. G. Mishler, Family Growth in Metropolitan America, Princeton, Princeton University Press, 1961, Ch. 13, and Ronald Freedman, David Goldberg, and Doris Slesinger, "Current Fertility Expectations of Married Couples in the U. S.", Population Index, Vol. 32, 1966, pp. 181-197.

7. Léon Tabah, "Relationship between Education and Fertility," United Nations Educational, Scientific, and Cultural Organization, Meeting of Experts on "Education and Family Planning," Paris, 20-24 April 1970.

8. Clyde Kiser, "Educational Differentials in Fertility in Relation to the Demographic Transition," Congress of the International Union for the Scientific Study of Population, London, Aug. 1969.

9. See T. Paul Schultz, A Family Planning Hypothesis: Some Empirical Evidence from Puerto Rico, The Rand Corporation, Santa Monica, Calif., Dec. 1967, p. 8. Also see Ronald Freedman and Lolagene Coombs, "Economic Considerations in Family Growth Decisions," Population Studies, Vol. 20, No. 2, Nov. 1966.

10. China, Taiwan Provincial Government, Essentials of the Taiwan Provincial Administration, 1970, p. 3.

11. It is interesting to note that in India the threshold at which number of children ever born per woman begins to drop sharply is also at the junior high school level. See Murari Majumdar, "Some Findings from Family Planning Data of the National Sample Survey," p. 403 in International Union for the Scientific Study of Population, Contributed Papers, Sydney Conference, 1967.

12. Ronald Freedman and John Y. Takeshita, Family Planning in Taiwan: An Experiment in Social Change, Princeton, Princeton University Press, 1969, p. 79n.

13. For a comprehensive discussion of MCA, see Frank Andrews, James Morgan, and John Sonquist, Multiple Classification Analysis, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, 1967.

WHITE-NONWHITE INCOME DIFFERENTIALS AS A MEASURE OF DISCRIMINATION*

Bruce Bosworth, St. John's University

Introduction and Purpose

This study is concerned with an analysis of income differentials of white and nonwhite males, and the relationship of such differentials to thier education and age. When education and age are equal for whites and nonwhites, income differentials that exist are an indication of racial discrimination against nonwhites.

The empirical investigation is based on data from the 1940, 1950, and 1960 Censuses. The first section comprises the estimation of white and nonwhite incomes based on a multiplicative regression model. This model looks at income as a joint effect of education and age. In the second section a discrimination measure that goes from zero to one (from no discrimination to complete discrimination) is used to analyze the estimated income data in order to observe changes in discrimination over time. In addition, cohorts are utilized to observe the effect of discrimination with regard to age groups moving through time.

Three questions are to be answered. The first being, "how have the levels of discrimination against nonwhite males changed within each year being considered as their education and age increase on the same basis as white males?" The second question being, "how have the levels of discrimination against nonwhite males changed between the years being considered as their education and age increase on the same basis as white males?" The third question being, "how have the levels of discrimination changed with regard to the education of specific groups of men as these groups (or cohorts) move through time?"

The Model

As a basis for determining whether different racial groups having the same levels of education and age have different incomes or earnings a regression model was utilized. This model assumes that the effects of education and age on income are a joint effect or an "interaction."(1) Such a model is analogous to a production function. The model is multiplicative and can be written as follows:

$$(1) \qquad I_{ij} = a\, E_j^{\,b}\, A_j^{\,c}\, e_{ij}$$

where i denotes an education classification; j denotes some age group; I denotes the education of individuals; A denotes the age of individuals; e denotes the stochastic or random term that includes other factors.

The Data

The multiplicative model was fitted to the following cross sectional data:(2) white and Negro median earnings of males 25 to 64 years old, by years of school completed for 1939; white and nonwhite income of males 25 to 64 years old, by years of school completed for 1949; and white and nonwhite median earnings of males 25 to 64 years old, by years of school completed for 1959. Functions were estimated for white and nonwhite males for the United States for the years 1939, 1949, and 1959. (3) Since the source data is presented by intervals for the cross classifications of years of school completed and age in years, it was necessary to adjust these intervals to arrive at a single value for years of school completed and age in years.(4)

The Estimation Results

For white and nonwhite males in 1939, 1949, and 1959 equation (1) was estimated in the following form:

$$(2) \qquad \log I_{ij} = \log a + b \log E_j + c \log A_i + \log e_{ij}.$$

The estimating equations for each group in each year are presented in Table 1. The use of the multiplicative model given by (1) to estimate white and nonwhite incomes as function of years of schooling and age worked very well. In all cases over eighty percent of the variation in incomes was explained.

From the estimating equations in Table 1, Appendix Tables I, II, and III were obtained. These tables give the estimated incomes for white and nonwhite males by years of school completed and age for 1939, 1949, and 1959. The income values in these tables will be analyzed in a following section with the objective of measuring changes in discrimination against nonwhites with regard to their education and age.

*This research was supported by a grant from the Business Research Institute, St. John's University, College of Business Administration.

TABLE 1

ESTIMATING EQUATIONS FOR INCOME OF WHITE AND NONWHITE MALES FOR 1939, 1949, AND 1959

Year and Race		Constant	Regression Coefficients Years of School	Age	R^2
1939	White	1.7220	.7139 (.0389)	.4355 (.0822)	.918
	Nonwhite	1.9505	.4840 (.0266)	.2721 (.0556)	.917
1949	White	2.5342	.4641 (.0349)	.3112 (.0721)	.859
	Nonwhite	2.7165	.4375 (.0299)	.0948 (.0617)	.871
1959	White	2.9952	.4540 (.0465)	.1784 (.0919)	.825
	Nonwhite	2.9277	.4724 (.0327)	.1002 (.0646)	.910

Note: Standard errors are in parentheses.

Measuring Discrimination

One method of measuring discrimination that is suggested by Gary S. Becker is through the use of the "market discrimination coefficient" or MDC.(5) This measure of discrimination is defined as

$$(3)\quad MDC = \frac{Y(W)}{Y(N)} - \frac{Y_o(W)}{Y_o(N)}$$

where Y(N) and Y(W) represent the actual incomes of N and W, $Y_o(N)$ and $Y_o(W)$ represent the income of N and W without discrimination. If it is assumed that W and N are perfect substitutes with perfect competition in the market place, then $Y_o(W) = Y_o(N)$ and (3) reduces to

$$MDC = \frac{Y(W)}{Y(N)} - 1 = \frac{Y(W) - Y(N)}{Y(N)}$$

In this form the MDC represents the percentage difference between the incomes of W and N with respect to the income of N. If it is assumed that Y(N) will be less than, or equal to Y(W), then MDC values will range upward from zero. The further away from zero the MDC gets the greater the discrimination against N. Because the scale of the MDC is from zero upward, a problem of interpretation can then arise. Values greater than one (1.50, 5.00, 9.00, etc.) lack clear meaning as to the level of discrimination against N. Such values would only suggest that there may exist high levels of discrimination against N.

To provide a clearer indication of the discrimination faced by N a modified measure of the MDC was developed. This modified measure has a scale going from zero to one. That is from no discrimination to complete discrimination. Such a scale is obtained by comparing income differences between W and N against W, rather than against N as is the case with the MDC. So that if it is assumed that W and N are perfect substitutes in the market place, then

$$DM = \frac{Y(W) - Y(N)}{Y(W)} = 1 - \frac{Y(N)}{Y(W)}$$

where DM is the modification of the MDC, and is called the "Discrimination Measure." This Discrimination Measure will be used to analyze the estimated income data presented in Appendix Tables I-III. The income of white males is represented by Y(W), and the income of nonwhite males is represented by Y(N).

Analysis of Income Differences

Discrimination measures were calculated for the white and nonwhite estimated incomes given in Appendix Tables I-III. These calculated DM values will help to provide answers to the three questions stated earlier. Table 2 shows the DM values calulated for 1939, 1949, and 1959, by years of school completed and age.

In 1939 as education and age increased the DM values increased. Within each age group as education increased from elementary to college the DM's increased. The DM's for the lowest education class (2.5 years of school completed) suggest that incomes for nonwhites,

TABLE 2

DISCRIMINATION MEASURES FOR 1939, 1949, AND 1959 BASED ON ESTIMATED MEDIAN INCOME

Years of School Completed	Age in Years				
	27	32	40	50	60
1939					
Elementary					
2.5	.200	.222	.251	.276	.297
5.5	.333	.350	.373	.397	.415
7.5	.378	.395	.416	.438	.454
High School					
9.5	.412	.428	.448	.468	.483
12	.441	.457	.477	.496	.510
College					
14	.416	.482	.495	.513	.527
17	.485	.499	.517	.534	.548
1949					
Elementary					
2.5	.273	.298	.332	.336	.402
6	.289	.315	.347	.378	.402
8	.294	.320	.352	.383	.406
High School					
10	.229	.324	.356	.386	.410
12	.302	.327	.359	.389	.413
College					
14	.305	.330	.362	.392	.415
17	.308	.333	.365	.395	.418
1959		Age in Years			
		30	40	50	60
Elementary					
3.5		.329	.344	.355	.364
8		.318	.333	.346	.354
High School					
10		.316	.331	.342	.352
12		.313	.329	.340	.350
College					
14		.311	.327	.338	.348
17		.309	.324	.336	.345

27 to 60 years old were about 20 to 30 percent less than that of white males having the same education and age. The highest education class(17 years of school completed) had DM values suggesting that nonwhite males 27 to 60 years old with a college education had incomes from 49 to 55 percent less than white males with the same education and age. In 1939, the less educated younger nonwhite males faced less discrimination, as shown by lower DM values, than the older better educated nonwhite males.

For 1949 the same overall pattern in DM values as exhibited in 1939 is apparent in Table 2. That is, DM values increasing as education and age increase. For nonwhite males in the lowest education class (2.5 years of school completed) who are 27 to 60 years old, the DM values indicate their incomes were 27 to 40 percent less than the incomes of white males having the same education and age. At the college level, nonwhite males with 17 years of schooling who are 27 to 60 years old had incomes 31 to 41 percent less than that of white males having the same education and age. As was the case in 1939, in 1949 the older bettter educated nonwhite male faced more discrimination than the younger less educated nonwhite male.

In 1959 the patterns of the increasing DM values as education and age increase previously found in 1949 and 1939 are not found in the values presented in Table 2. For 1959 it can be seen that as education increases for each age the DM values decline for the first time. But, as age increases within each education classification the DM values increase as was the case in 1939 and 1949. In 1959 nonwhite males in the lowest education class(3.5 years of school completed) who are 30 to 60 years old, had DM values indicating that their incomes were 33 to 36 percent less than the incomes of white males having the same education and age. Nonwhite males with a college education who are 30 to 60 years old had incomes as indicated by the DM values that were 31 to 35 percent less than that of white males having the same education and age. Older better educated nonwhite males did not face more discrimination than younger less educated nonwhite males as was the case in 1939 and 1949.

With regard to the question of changes in discrimination between years, the DM values in Table 2 when followed from year to year show a decline over time in each education and age classification. To obtain an overall view of the changes in the levels of discrimination against nonwhites having the same education and age as whites that have occurred over time, summary statistics for the values in Table 2 were calculated. Table 3 presents for 1939, 1949, amd 1959, means, standard deviation, and coefficients of variation for the DM values in Table 2.

TABLE 3

SUMMARY STATISTICS FOR DM VALUES; 1939, 1949, AND 1959

Year	Mean	Standard Deviation	Coefficient of Variation
1939	.425	.090	.212
1949	.352	.043	.122
1959	.336	.015	.045

All of the statistics in Table 3 indicate a decline in the DM values over time. In 1939 the DM values had a mean and standard deviation of .425 and .090 respectively. By 1959 the mean DM had dropped to .336 with a standard deviation of .015. The coefficients of variation show the variation about the mean DM for each year on a relative basis. This measure of relative dispersion shows a decline from 21 percent in 1939 to 4.5 percent in 1959. These declining summary statistics suggest that over time discrimination against nonwhite males having the same education and age as white males has been declining. Furthermore, these measures also suggest that the discrimination against nonwhite males is not only declining, but is also becoming more uniform over the education and age classifications shown in Table 2.

With regard to the question of changes in discrimination against the same group of men over time, a cohort analysis was used to examine changes in discrimination against nonwhite males over time when they have the same education and age as white males. This type of analysis compares the same group of men at different points in time rather than comparing a different group of men having the same age at different points in time. Men born in the period 1905-1914 comprise a cohort that in 1939 will consist of men 30-34 years old; in 1949 35-44 years old; and in 1959 45-54 years old. Cohorts for 1905-1914 and 1895-1905 are used in conjunction with the discrimination measures shown in Table 2.

Table 4 shows the discrimination measures for these two cohorts. For the men comprising the 1905-1914 cohort the DM values decline as this group moves through the time period 1939-1959. In 1939, at age 32 the DM values increased as education increased. At age 40 in 1949 the DM values begin to show uniformity over the years of schooling. By age 50 in 1959 it appears that as education increases the DM values decline. The DM values for the cohort of 1895-1904 show the same pattern as the 1905-1914 cohort. That is, increasing DM values as education increases in early years, and then declining DM values as education increases in later years.

The DM Values in Table 4 were plotted for each cohort and appear as Figure 1. From this figure the increase in DM values as education increases for younger men (32 and 40) can be seen. As these men get older (40 and 50) the DM values begin to level off as education increases. Finally, the DM values for men 50 and 60 years old in 1959 show declines as education increases. Both the values in Table 4 and the plots in Figure 1 suggest that over the period 1939 to 1959, there was a decline in the discrimination against nonwhite males as their education and age increased on the same basis as white males.

Summary and Conclusions

This study set out to analyze white and nonwhite income differentials as a means of measuring changes in discrimination against nonwhite males having the same education and age as white males. Use was made of a discrimination measure going from zero (indicating no discrimination) to one (indicating complete discrimination). This measure describes relative differences in incomes given white and nonwhite males have the same education (years of schooling) and the same age. Although the study was limited to discrimination against nonwhite males with regard to their education and age, the results obtained do suggest that for the United States as a whole there have been declines in the discrimination against nonwhite males having the same education and age as white males between the years 1939 and 1959.(6)

TABLE 4

DISCRIMINATION MEASURES FOR COHORTS OF 1905-1914, AND 1895-1904, BY SELECTED AGES AND YEARS OF SCHOOL COMPLETED

	Cohort of 1905-1914			Cohort of 1895-1904		
Years of School Completed	32 years old in 1939	40 years old in 1949	50 years old in 1959	40 years old in 1939	50 years old in 1949	60 years old in 1959
8	.395	.352	.346	.416	.383	.354
10	.428	.356	.342	.448	.386	.352
12	.457	.359	.340	.477	.389	.350
14	.482	.362	.338	.495	.392	.348
17	.499	.365	.336	.517	.395	.345

Note: DM values are from Table 2.

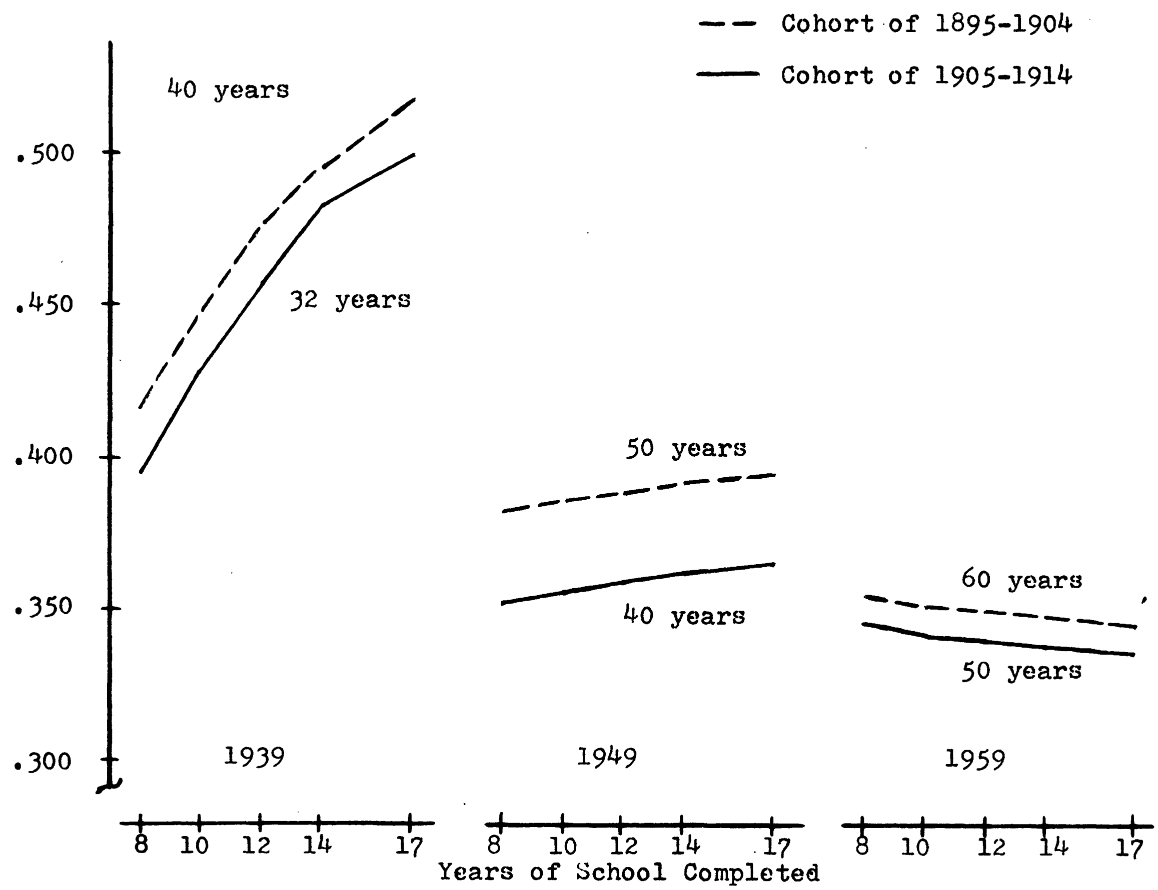


FIGURE 1

PLOTS OF DISCRIMINATION MEASURES OF THE COHORTS OF 1895-1904 AND 1905-1914

APPENDIX TABLES I-III. ESTIMATED INCOME OF WHITE AND NONWHITE MALES, SELECTED AGES, BY YEARS OF SCHOOL COMPLETED FOR 1939, 1949, AND 1959

Years of School Completed	Age in Years									
	27	32*	40	50	60	27	32*	40	50	60
	White Income					Nonwhite Income				
1939										
Elementary										
2.5	426	459	506	557	603	341	357	379	403	424
5.5	748	805	887	978	1059	499	523	556	590	620
7.5	933	1005	1107	1220	1321	580	608	646	686	721
High School										
9.5	1105	1190	1311	1445	1564	650	681	724	769	808
12	1305	1405	1549	1707	1848	729	763	810	861	905
College										
14	1457	1569	1729	1906	2063	785	822	873	928	975
17	1674	1802	1986	2189	2370	862	903	959	1019	1071
1949										
Elementary										
2.5	1460	1539	1650	1768	1915	1062	1080	1103	1126	1146
6	2191	2310	2477	2655	2810	1558	1583	1617	1652	1681
8	2504	2640	2830	3034	3211	1767	1796	1834	1873	1906
High School										
10	2778	2929	3139	3365	3561	1948	1980	2022	2066	2102
12	3023	3187	3416	3662	3877	2110	2144	2190	2237	2276
College										
14	3247	3424	3670	3933	4163	2257	2294	2343	2393	2435
17	3553	3746	4016	4304	4556	2457	2497	2551	2605	2651
1959										
Elementary										
3.5		3204	3373	3510	3626		2151	2214	2264	2306
8		4664	4909	5119	5278		3179	3272	3346	3407
High School										
10		5161	5433	5654	5840		3532	3636	3718	3786
12		5607	5902	6142	6345		3850	3963	4052	4127
College										
14		6013	6330	6587	6805		4141	4262	4358	4439
17		6567	6913	7194	7432		4539	4671	4777	4865

*In 1959 age in years begins at 30 years.

Footnotes

(1) Hill, T. P. "An Analysis of the Distribution of Wages and Salaries In Great Britain," Econometrica, Vol. 27, No. 3, July, 1959.

(2) The 1939 data was derived from the U. S. Bureau of the Census, Sixteenth Census of the United States: 1940, "Education: Educational Attainment by Economic Characterisitc and Marital Status," Tables 29 and 31; the 1949 data was derived from the U. S. Bureau of the Census, United States Census of Population: 1950, Series P-E, No. 5B, "Education," Tables 12 and 13; the 1959 data was derived from U. S. Bureau of the Census, United States Census of Population: 1960, Series PC(2)-7B, "Occupation by Earnings and Education," Table 1.

(3) The 1939 data used is divided into "white" and "Negro" categories. The "Negro" category is not used in 1949 and 1959. Rather a "nonwhite" category is used in those years. The difference between "nonwhite" and "Negro" populations is slight, since 92 percent of the nonwhite population is Negro, and statistics of the Negro (nonwhite) population generally reflect conditions of the nonwhite (Negro) population.

(4) The years of schooling intervals were adjusted by using the midpoint of the interval for closed end intervals. For open end intervals, 4 years of schooling or more, 17 years was the value used. The age intervals were adjusted by using the midpoint of the interval, or in some cases the midpoint was rounded to the nearest whole year.

(5) Becker, Gary S., The Economics of Discrimination,(The University of Chicago Press, Chicago, 1959), p. 14.

(6) Additional analyses were carried out using other models that were linear and multiplicative in form along with data that covered the period up to 1967; these analyses also suggested declines in discrimination against nonwhite males having the same education and age as white males. See my unpublished doctoral dissertation, "A Quantitative Analysis of White-Nonwhite Income Differentials as Related to education and Age," New York University, Graduate School of Business Administration, 1971.

THE CONTINUITY CORRECTION IN THE NORMAL APPROXIMATION

Gary L. Brager, Baltimore County Public Schools

There arise experimental situations in educational research where a binomial model is assumed and inferential statistics based upon normal curve theory are employed to test hypotheses (e.g., Z test for the mean). One example is student guessing behavior on multiple choice tests (See for example, Ebel,) where the probability of obtaining a correct answer by guessing is $\frac{1}{K}$, K = number of choices per item.

As is well known, the use of the normal curve to evaluate probability levels of Bernoulli sequences, or more specifically cumulative Bernoulli sequences, depends upon the asymptotic convergence of the binomial distribution function to that of the normal distribution function (e. g., Fisz, 1963, or almost any standard text on mathematical statistics or probability). What is also well appreciated is that for finite samples, the utilization of normal integrals to evaluate binomial situations involves some error. Many studies have investigated and catalogued facets of the error related to such variables as sample size and population proportion, among others. (See, for example, Mosteller, 1961; Jeffreys, 1961; Raff, 1956; Mood and Graybill, 1963; Dayton, 1964; Plackett, 1964; Munroe, 1951.)

One of the more widely used techniques for making the normal integral more valid for binomial experiments is the Yates-type continuity correction. Typically, a constant of one-half is subtracted from the absolute value of the difference between the sample value and the assumed population mean in forming the Z statistic.

Thus, instead of the analytically derived $Z = \frac{X - np}{(npq)^{\frac{1}{2}}}$, we have $Z_c = \frac{|X - np| - \frac{1}{2}}{(npq)^{\frac{1}{2}}}$.

This "correction" factor is usually attributed to Yates and dates to 1921 in course outlines of the Department of Applied Statistics at University College, London (Plackett, 1964).

It is the contention of the present paper that the purpose and operation of the Yates-type correction in the normal approximation to the binomial is still misunderstood, particularly among educational statisticians. If one samples the standard texts in educational research and statistics (e.g., Fryer, 1966; Cornell, 1956; Walker and Lev, 1953), one comes away with the fact that the purpose of the correction for continuity is to effect better fitting normal curves for the binomial model.

Todhunter (1931) reports Yates as first applying that connotation to integrals of $f(Z_c)$ compared to $f(Z)$. Lindgreen (1964) employs an intuitive approach to demonstrate that using Z_c provides less error in the approximation than using Z. Richardson (1944) concurs in this interpretation, while Cochran (1953) and Dayton (1964) devote papers to finding other subtractors than one-half to provide better fitting normal approximations.

In recent years, several authors, (e.g., Montel and Greenhouse, 1968) have demonstrated, by using computer-based procedures, that the continuity correction may not provide better fitting theoretical curves at all, but will tend to reduce the percent of false rejections of the null hypothesis.

The present paper investigated this relationship of using the normal approximation to test binomial hypotheses but considering an additional factor of importance in educational research which has not been explored previously by applied statisticians.

Typically, the research setting in which the binomial model is assumed (e.g., test-taking behavior) calls for the formulation of the non-directional hypothesis. This concept is related to the removing of a-priori bias between two opposing methods. It is the contention of some educational statisticians (e.g., Raths, 1964) that research hypotheses in educational experiments should almost always be non directional.

In the binomial situation, if we are given N and π and wish to calculate $P(X \leq Xo)$, where Xo is a given number of defined successful outcomes in a Bernoulli sequence, a non-directional hypothesis would be defined as follows:

Directional probability refers to the probability of all successes at least as extreme from np as Xo and in the same algebraic direction from np as X. Non-directional probability includes the directional total plus the probability of all successes at least as extreme from np as (X^1) and in the same algebraic direction from np as (X^1), where (X^1) is the nearest integer value to the X^1 whose Z score is the additive inverse of the Z score of X.

Thus, non-directional hypotheses yield probabilities of sample occurrence greater than or equal to directionally calculated probabilities. Very few studies of the continuity correction and the error in using normal integrals have considered the case of non-directional binomial hypothesis. Those that have (e.g., Mosteller, 1961), have assumed that symmetry would negate studying the inverse sample results. However, the binomial distribution is not always symmetrical about its mean value, and one cannot assume the error function will simply be proportioned on both sides np.

Thus, the present paper investigated the relative error distribution in using the standard unit normal curve for binomial sequences. Computer generated samples were employed to calculate the relative error for directional and non-directional hypotheses and for corrected and uncorrected Z-statistics.

Since the skewness of the binomial distribution appears in the literature to be of paramount importance in assessing the degree of convergence between the normal and binomial, the present study controlled for skewness.

For a given value of skewness, binomial probabilities were calculated for various n and p combinations. It can be shown (Fisz, 1963) that for the binomial $SK = \frac{q - p}{(npq)^{\frac{1}{2}}}$

Since skewness is not bounded and since the number of usable samples generated varies with the skewness, and further, since practical study had shown few instances where SK>1, one was chosen as an upper bound for SK. The number of usable samples for study is limited by SK, since we can show that $p = \frac{1}{2}\left(1 + \left(\frac{ny^2}{4+ny^2}\right)^{\frac{1}{2}}\right)$, $y = SK$

Thus, for large values of SK, p will be very small and the associated Z statistic will be too large for meaningful comparisons. The skewness was incremented by 0.02 from zero through one, and the program generated the relative error distribution, summary statistics and the percent of rejections of null hypotheses by the normal approximation when the binomial probabilities were greater than .05 or .01, respectively. This procedure was replicated for the directional and non-directional hypotheses and uncorrected and corrected Z statistics.

There were little data to support the belief, held in educational research, that Zc unqualifyingly produces significantly better fitting normal curves to the binomial model. On the other hand, the Zc's were associated with somewhat lower mean relative errors, but not for all values of skewness.

In fact, for directional hypotheses, small binomial skewness resulted in better fitting curves for uncorrected than for corrected Z statistics. As the skewness increased Zc did yield normal curves with less relative error. It is interesting to note that the corrective factor of Zc was typically greater for non-directional than for directional hypotheses in lowering the mean value of the relative errors, and that as the skewness increased Zc's corrective power also increased. Further, even the uncorrected non-directional curves had less relative error than the directional counterparts. The following table lists the mean relative errors for selected values of binomial skewness.

Moreover, the distributions of relative error were all positively skewed and platykurtic. As the binomial skewness increased, the relative error distributions became less positively skewed, a result of the general tendency of the error to increase as the binomial skewness increased. The non-directional errors were the most positively skewed. The distributions of relative error were non-normal in form.

The second characteristic of the present investigation dealt with the percent of Type I errors under the four modes of the Z statistic. The data presented show clearly that employing the Yates-type correction factor results in a substantial reduction of Type I errors in using normal integrals to evaluate binomial hypotheses over the uncorrected Z statistics. This is an expected result for directional hypotheses, but is not an obvious result for the non-directional case. In fact, as the following table shows, the reduction in Type I errors is the largest for non-directional corrected hypotheses. Additionally, for some values of binomial skewness, there were no Type I errors made when using the non-directional Zc.

Of interest is to note that, while in general the corrected Z statistics had less percent of Type I errors than the uncorrected Z scores, such percent for the most part increases as the binomial skewness increases. At the .01 level, the percent of Type I errors is generally larger than at the .05 levels across all four modes of Z statistics.

Further, the disparity between the percent of Type I decision errors for corrected and uncorrected Z statistics is not as great at the .01 level than at the .05 levels.

It is also important to note that even when the binomial skewness is zero and the directional and non-directional error distributions are identical, the percent of Type I errors are unequal. This again emphasizes the need to investigate the non-directional case for binomial hypotheses.

The present paper investigated the distributions of relative error in using normal integrals to assess probabilities for binomial hypotheses under four modes of the standardized Z statistic: directional uncorrected, directional corrected, non-directional uncorrected, non-directional corrected. Computer based procedures generated relative error distributions, summary statistics and Type I error tables for each mode of Z statistic while controlling for the binomial skewness. Skewness was incremented from 0.00 to 1.00 by steps of 0.02, and n and p combinations for each skewness value, n = 2(100)(2).

In summary, then, the following conclusions seem tenable:

1. In general, when normal integrals are used for binomial hypotheses, the mean relative errors increase as the binomial skewness increases.

2. Non-directional Z statistics yield better fitting normal curves relative to the binomial probabilities than do directional Z statistics.

3. In general, corrected Z statistics are associated with somewhat less mean relative errors than uncorrected Z statistics. However, for binomial skewness near zero, directional uncorrected Z's have less mean error than directional corrected Z's.

4. The percent of Type I decision errors made when normal probabilities are used generally increases as the binomial skewness increases.

5. Corrected Z's are associated with lower percents of Type I errors than uncorrected Z's.

6. Non-directional Z's are associated with lower percents of Type I errors than directional Z's.

7. The percent of Type I decision errors made using normal integrals is generally inversely related to the alpha level of the null hypothesis.

It would appear that when using normal integrals to evaluate binomial hypotheses, the skewness of the binomial affects both the relative error in the calculation of probabilities and the percent of error in rejecting the null hypothesis. The Yates-type continuity correction combined with a non-directional hypothesis minimizes both the relative error and the error in decision compared to the other modes.

It would seem that for nearly symmetrical binomials ($SK<.10$), a "better fit" is obtained with the directional uncorrected than the directional corrected. However, even in this case, one lessens the percent of Type I decision errors by using Zc.

Thus, the data of the present study seem to support the contention that the primary effect of the continuity correction in the normal approximation to the binomial is to provide a more conservative test of hypothesis rather than to obtain better fitting normal integrals.

<u>REFERENCES</u>

Cochran, W.G. Sampling techniques. New York: John Wiley and Sons, Inc., 1953.

Cornell, F.G. The essentials of educational statistics. New York: John Wiley and Sons, Inc., 1956.

Dayton, C.M. An empirical evaluation of small expected values in 2 x 2 contingency tables on the chi-square test of significance. Unpublished Ph. D. dissertation, University of Maryland, 1968.

Fisz, M. Probability theory and mathematical statistics. New York: John Wiley and Sons, Inc., 1963.

Fryer, H.C. Concepts and methods of experimental statistics. Boston: Allyn and Bacon, Inc., 1966.

Hoel, P.G. Introduction to mathematical statistics. New York: John Wiley and Sons, Inc., 1954.

Jeffreys, H. Theory of probability. Oxford at the Clarendon: University of Cambridge Press, 1961.

Lindgreen, B.W. and McElrath, G.W. Introduction to probability and statistics. New York: The Macmillan Company, 1959.

Mantel, N. and Greenhouse, S.W. What is the continuity correction? The American Statistician. 1968, 22, 27-30.

Mood, A.M. and Graybill, F.A. Introduction to the theory of statistics. New York: McGraw-Hill Book Company, Inc. 1963.

Mosteller, F. <u>et al</u>. Probability and statistics. reading: Addison-Wesley Publishing Company, Inc., 1961.

Munroe, M.E. Theory of probability. New York: McGraw-Hill Book Company, Inc., 1951.

Plackett, R.L. The continuity correction in 2 x 2 tables. Biometrika. 1964, 51, 327-337.

Raff, M. On approximating the point binomial. Journal of the American Statistical Association. 1956, 51, 293-303.

Raths, J. The null hypotheses model for decision making. Mimeographed course notes. University of Maryland, 1964.

Richardson, C.H. An introduction to statistical analysis. Washington: War Department, 1941.

Thiele, J. Normal approximation to the binomial distribution fails to account for skewness. Biometrika, 1931, 2, 283.

Todhunter, I. A history of the mathematical theory of probability; from the time of Pascal to that of LaPlace. New York: G.E. Stechert and Company, 1931.

Walker, H.M. and Lev, J. Statistical inference. New York: Holt, Rinehart and Winston, 1953.

Table 1. Mean Relative Errors

Binomial Skewness	Directional Uncorrected	Directional Corrected	Non-Directional Uncorrected	Non-Directional Corrected
0.00	.235	.095	.235	.095
.02	1.232	2.115	.291	.118
.04	2.367	5.623	.325	.132
.06	1.577	3.876	.338	.141
.08	.688	1.267	.352	.147
.10	.489	.659	.362	.164
.12	.446	.444	.374	.172
.14	.434	.342	.381	.191
.16	.455	.306	.391	.199
.18	.463	.285	.399	.214
.20	.465	.277	.408	.226
.22	.464	.278	.411	.223
.24	.474	.314	.421	.235
.26	.481	.301	.421	.249
.30	.497	.292	.428	.255
.34	.522	.300	.447	.271
.38	.538	.299	.447	.314
.42	.550	.300	.468	.306
.46	.553	.289	.464	.295
.50	.570	.305	.473	.363
.60	.587	.312	.499	.282
.70	.600	.318	.473	.305
.80	.618	.337	.523	.390
.90	.636	.372	.550	.363
1.00	.602	.327	.478	.304

Table 2. Percent of Type I Decision Errors at .05 and .01 Levels

Binomial Skewness	Directional Uncorrected		Directional Correctional		Non-Directional Uncorrected		Non-Directional Corrected	
	.05	.01	.05	.01	.05	.01	.05	.01
0.00	3.40	2.96	0.00	0.00	3.70	2.81	0.00	0.00
0.02	4.08	3.06	0.53	0.98	3.82	3.65	0.27	0.35
.04	3.56	3.95	0.61	0.91	4.02	3.92	0.30	0.30
.06	5.00	4.88	1.03	2.18	4.13	4.25	0.34	0.34
.08	5.80	5.56	1.03	2.32	5.20	4.84	0.39	0.51
.10	6.72	6.06	1.13	2.82	6.46	4.61	0.42	0.70
.12	6.40	6.68	1.07	3.83	4.13	5.83	0.61	0.00
.14	7.00	6.70	1.48	3.95	5.78	4.11	0.16	0.00
.16	7.90	7.58	1.58	4.74	3.39	4.19	0.18	0.00
.18	7.82	7.82	0.74	4.28	6.29	3.74	1.49	0.18
.20	7.13	8.74	1.76	4.31	6.24	4.46	0.00	0.59
.22	6.38	9.19	1.24	5.59	6.19	5.63	0.00	0.62
.24	8.72	10.08	1.72	5.36	2.13	5.43	0.00	1.72
.26	8.32	9.74	3.39	4.97	3.04	5.48	0.00	0.68
.30	13.97	11.35	2.20	4.96	2.18	5.68	0.00	2.45
.40	15.08	10.42	3.49	4.74	3.55	6.21	0.00	0.75
.50	10.27	10.96	3.09	8.77	3.20	8.68	5.41	1.29
.60	6.43	12.14	1.08	8.38	1.43	8.33	0.00	2.71
.70	6.81	14.99	0.63	7.57	1.64	8.45	0.00	5.99
.80	11.32	15.60	11.19	14.08	10.40	12.84	0.00	5.42
.90	16.67	17.01	1.26	5.04	2.43	16.67	0.00	1.26
1.00	17.89	20.33	6.12	21.43	17.75	18.29	0.00	20.92

THE EFFECTS OF "IMPROVED" METHODS OF COLLECTING INCOME STATISTICS ON NON-RESPONSE RATES--1969 CPS EXPERIMENTS

John J. Chai, Syracuse University

I. Introduction

The interest in and use of CPS (Current Population Survey) income statistics have increased in the recent past as the study of socio-economic problems such as poverty or underemployment among different socio-economic classes and geographic areas has been found increasingly necessary.

However, the accuracy of the CPS income statistics has been the matter of serious concern among the users of these statistics because of high nonresponse rates 1/ (14.0% in 1965, 14.8% in 1966, 21.9% in 1967, and 17.2% in 1968) and underreporting of some income items such as the income earned from sources other than wages and salary. The nonresponse rates on income statistics are considerably higher than the ones on the labor force statistics (Approximately 6%) which also are collected by the CPS sample.

Research projects to explore possible ways to improve the quality of the income data were begun at the Census Bureau at a modest level in 1968 and at more extensive level in 1969 and is continuing in 1970.

The purpose of this paper is to present a global analysis of the results of the 1969 CPS experiments to improve the quality of income statistics, which were conducted in March and April of 1969. Particularly, this paper focusses its attention on the effects of various "improved" CPS income collection methods (see the following section for the description of the methods tested) on the nonresponse (NA) rates.

A. Brief Description of the Programs to Improve Quality of Income Statistics.

The 1969 CPS projects 2/ on improvement of quality of income statistics are divided into two parts. The first consists of improvements in the "field procedures"--(1) increased amount of interviewer training and changing the time of training, (2) expansion of field office editing procedures, (3) review and expansion of clerical coding procedures, (4) modification of questionnaire content and design, and (5) extension of the interview period.

The second consists of experimental study of "improved" CPS income statistics collection methods.

The first part of the project was designed to be implemented without further testing the effectiveness of each of the five improvements and the second part was designed to test experimentally the effectiveness of the various collection methods.

B. Brief Description of the 1969 CPS Experimental Design

B.1 Collection Methods Tested: The following are modifications or additions to the previous CPS procedures tested experimentally in an effort to examine whether these new methods may improve the quality of income statistics.

(1) Advance letter before the regular interview (A.L.). An advance letter solicits the cooperation of respondents and it provides an advance income form in which respondents are asked to record their answers prior to the regular interviews.

There are a number of causes of the nonresponse rates--sloppiness in leaving blank entries that could be zero, unavailability of information because the wrong household member is being interviewed, refusal, etc. The advance letter might make it possible for more persons in the household to participate in providing response and it would give them more opportunity to consult records, and thus, reducing some of the elements of the causes of nonresponses.

(2) Change of the time of collecting income data from March to April (I_A).

(3) Income (I) and work experience (WE) questions are asked in a single interview. Previously income questions were asked in March and work experience in February and April. The regular CPS procedure is that the income statistics are collected in March and work experience (labor force characteristics) in February; for the two new rotation groups in March (i.e., 1st and 5th month-in-sample), work experience is collected in April. The second procedure (collection of income statistics close to the income return deadline) above was tested to see whether more accurate (less NA rates and less underreporting of income) could be obtained and the third procedure was tested to see whether work experience questions would jog the respondents memory so that more accurate income is reported.

B.2 Assignment of the Collection Methods to be Tested to CPS Subsamples

There are eight independent rotation groups in CPS sample--each representing a random sample of approximately 6000 households. 3/ The assignment of these methods to be tested is presented in Table 1.

B.3 Experimental Group Composition

For the purpose of comparing different collection methods, the experimental groups are formed as shown by Table 2. It is noted from Tables 1 and 2 that each experimental group is treated with a different income collection procedure. For example, Experimental Group C for 1969 includes rotation groups which were enumerated for income items in the month of April.

B.4 Analysis of the Experimental Results

The analysis presented in this paper is based on the cross classification of nonresponse rates by four types of income ("wage and salary," "nonfarm self-employment income," "farm self-employment income," and "income other than earnings") and by three person characteristics for all persons 14+ ("with total income," "with no income of specified type" and "with income of specified type.") The method of analysis used consists of comparisons of non-response rates among the various experimental groups (see Table 2) for which different income collection procedures are applied. In order to assess the "significance" of the differences of nonresponse rates, the standard errors for the estimated nonresponse rates are estimated. 4/

II. Evaluation of the NA Rates on Income Items by Experimental Groups

This study explores possible answers to the four basic questions (see Table 3) and the results are summarized briefly below.

(1) The 1969 March CPS NA rates in most cases are not lower than the 1968 March CPS and in most of the cases studied the 1969 March CPS rates are significantly higher than the 1968 rates in a statistical sense.

(2) The March 1969 CPS with advance letter did not produce better results than the March 1968 CPS; but the former seems to be a better procedure than the March 1969 CPS procedure without advance letter.

(3) The collection of income information in April 1969 does not seem to be a better procedure than (a) the 1968 March CPS procedure, (b) the 1969 CPS with advance letter, and (c) the 1969 March CPS without advance letter. And this seems to be true for all the categories included in the study.

(4) The 1969 April CPS with work experience and income combined seems to be no better than any of the 1968 CPS procedures, but better than the 1969 April CPS with income only. Of course, this does not mean that the procedure "work experience and income combined" itself is not better than other CPS procedures included in the experiments. In other words, if the procedure "work experience and income combined" had been used in March CPS rather than in April CPS, the results might have indicated the NA rates for the procedure to be smaller than those for other CPS procedure tested in the experiments. In fact, "work experience and income combined" procedure has been tested in March 1970 and the preliminary results indicate that the NA rates are lower for this procedure than the ones for other procedures.

FOOTNOTES

1/The nonresponse rate for a given sample in this paper is defined as the ratio of the number of sample elements (households) with one or more income items allocated over the total number of elements (households) in the sample after allocation.

2/More detailed accounts of the projects are given by the internal memorandum (U.S.Bureau of the Census) from D.B. Levine, et.al. to C.F. Taeuber and M.H. Hansen dated September 11, 1968.

3/See U.S. Bureau of the Census, THE CURRENT POPULATION SURVEY--A REPORT ON METHODOLOGY, Technical Paper No.7, U.S. Government Printing Office, Washington,D.C.

4/The estimates of the standard errors for the estimated nonresponse rates (Tables C and D) may be obtained from the author upon request.

5/Due to the limitation of space, no supporting data, except Table A, are included in this report.However, these data are available from the author upon request.

* Views expressed in this paper are those of the author and not necessarily those of the Bureau of the Census. The author gratefully acknowledges the assistance and comments given by Joseph Waksberg, Barbara Boyes, M. Ono, and other members of the Bureau of the Census.

Table 1. Assignment of the CPS Income Statistics Collection Methods				
CPS Sample	Rotation Group	February * Month 1	March Month 2	April Month 3
A23	8	(8)WE		
A24	1	(7)WE	(8)AL&I	
	2	(6)WE	(7)I	(8)LF
	3	(5)WE	(6)I	(7)LF
	4		(5)LF	(6)I&WE
	5			(5)LF
A25	4	(4)WE		
	5	(3)WE	(4)I	
	6	(2)WE	(3)LF	(4)I
	7	(1)WE	(2)I	(3)LF
	8		(1)I	(2)WE
	1			(1)LF

*This is the regular CPS Procedures, i.e., no changes in the CPS procedure in February.

This table shows the CPS sample for 1969 only.
WE = Work experience
AL = Advance letter
I = Income
LF = Labor force
() The numbers in parentheses indicate month of enumeration; e.g., (8) denotes the 8th or the last month of enumeration.

Table 2. Experimental Group Composition			
Exp. Group	Collection Methods Applied	Rotation Group and Month-in-Sample*	
		1969	1968
A	All with income data collected in March	1(8), 2(7),3(6),5(4) 7(2), and 8(1)	1(4),3(2),4(1),5(8) 6(7), and 7(6)
B	March: income data, no advance letter	2(7),3(6),5(4),7(2) and 8(1)	1(4),3(2),4(1),6(7) and 7(6)
C	All with income data collected in April	4(6), and 6(4)	1(4), and 7(6)
D	March: Advance letter	1(8)	5(8)
E	April: Income and work experience	4(6)**	7(6)
F	April: Income only	6(4)**	1(4)

*Month-in-samples are given in parentheses.
**These indicate the month-in-samples in April, 1969.

Table 3. Questions and Experimental Groups Compared	
Questions	Experimental Group Compared*
1) Did the 1969 Procedures produce better results than the 1968 CPS Procedures? ('Better' in the sense that the 1969 NA rates are lower than the 1968 NA rates.)	A_{69} and A_{68}
2) Did the 1969 March CPS with the advance letter for collecting income information produce better results than	
(a) the 1968 March CPS?	D_{69} and D_{68}
(b) the 1969 March CPS without the letter?	B_{69} and D_{69}, B_{68} and D_{68}
3) Is the collection of income data in the month of April better than the collection of income data,	
(a) in March 1968?	C_{69} and C_{68}
(b) in March 1969 with the advance letter?	C_{69} and D_{69}, C_{68} and D_{68}
(c) in March 1969 without the advance letter?	C_{69} and B_{69}, C_{69} and B_{68}
4) Is the combination of work experience and income as a single interview a better procedure for collecting income data than	
(a) the 1968 March CPS procedure?	E_{69} and E_{68}
(b) the 1969 March CPS procedure with the advance letter?	E_{69} and D_{69}, E_{68} and D_{68}
(c) the 1969 March CPS without the letter?	E_{69} and B_{69}, E_{68} and B_{68}
(d) the 1969 April CPS in which only income items were asked?	E_{69} and F_{69}, E_{68} and F_{68}

*Subscripts indicate year in which a particular experimental group was enumerated.

Table A--Nonresponse Rate in Percent by Type of Income and Rotation Groups for all Persons 14+: 1968 and 1969 CPS

Type of Income		Rotation Group 1/ All Rotations	1(8,4) AL	2(7,3)	3(6,2)	4(6,1)WI	5(4,8)	6(4,7)IA	7(2,6)	8(1,5)
Wage and Salary Income:										
Total persons 14+	1969	7.98	6.98	8.21	8.27	8.62	8.81	10.18	7.74	5.07
	1968	6.10	6.43	7.22	6.68	4.64	6.61	6.46	6.21	5.02
	1969 less 1968	1.88*	.55	.99	1.59	3.98	2.20	3.72	1.53	.05
With no income of specified type	1969	2.20	1.98	2.40	2.45	2.10	2.55	2.47	2.33	1.36
	1968	1.02	.91	1.52	1.22	.66	1.05	1.17	.82	.82
	1969 less 1968	1.88*	1.07	.88	1.23	1.44	1.50	1.30	1.51	.54
With income of specified type	1969	11.87	10.25	12.09	12.09	13.21	12.94	15.62	11.34	7.55
	1968	9.59	10.19	11.03	10.27	7.32	9.83	10.20	9.95	7.89
	1969 less 1968	2.28*	.06	1.06	1.82	5.89	3.11	5.42	1.39	-.34
Mean income after allocation	1969	$5,101	$5,247	$5,517	$4,770	$5,209	$4,826	$4,933	$4,987	$5,563
	1968	$4,629	$4,703	$4,602	$4,137	$4,702	$4,874	$4,743	$4,565	$4,790
	1969 less 1968	$ 472	$ 544	$ 915	$ 633	$ 507	$ - 48	$ 190	$ 422	$ 773
Non-farm Self Employment Income:										
Total persons 14+	1969	5.45	4.96	5.63	5.90	5.96	6.36	6.17	5.19	3.44
	1968	3.48	3.68	3.94	3.48	2.76	3.48	3.85	3.72	2.91
	1969 less 1968	1.97*	1.28	1.69	2.42	3.20	2.88	2.32	1.47	.53
With no income of specified type	1969	4.58	4.24	4.68	4.99	5.15	5.45	5.20	4.24	2.67
	1968	2.70	2.86	3.18	2.64	1.98	2.77	3.00	2.95	2.19
	1969 less 1968	1.88*	1.38	1.50	2.35	3.17	2.68	2.20	1.29	.48
With income of specified type	1969	21.51	18.22	23.06	22.48	21.45	23.00	24.22	22.11	17.54
	1968	18.22	19.07	18.34	18.11	16.16	17.73	20.79	19.36	16.61
	1969 less 1968	3.29*	- .85	4.72	4.37	5.29	5.27	3.43	2.75	.93
Mean income after allocation	1969	$7,929	$5,319	$8,326	$4,734	$7,395	$7,591	$6,926	$15,144	$7,250
	1968	$6,008	$6,420	$6,388	$6,187	$7,290	$7,097	$5,062	$ 4,755	$4,893
	1969 less 1968	$1,921	-$1,101	$1,938	-$1,453	$ 105	$ 494	$1,864	$10,389	$2,357

Table A--Nonresponse Rate in Percent by Type of Income Rotation Groups for all Persons 14+: 1968 and 1969 CPS (continued)

Type of Income		Rotation Group 1/ All Rotations	1(8,4) AL	2(7,3)	3(6,2)	4(6,1)WI	5(4,8)	6(4,7)IA	7(2,6)	8(1,5)
Farm Self Employment Income:										
Total persons 14+	1969	5.08	4.66	5.19	5.59	5.54	5.99	5.82	4.86	3.02
	1968	2.95	3.02	3.61	2.92	2.27	2.95	3.28	3.27	2.32
	1969 less 1968	2.13*	1.64	1.58	2.67	3.27	3.04	2.54	1.59	.70
With no income of specified type	1969	4.96	4.60	5.08	5.42	5.41	5.88	5.60	4.75	2.91
	1968	2.79	2.90	3.40	2.73	2.10	2.88	3.06	3.03	2.24
	1969 less 1968	2.17*	1.70	1.68	2.69	3.31	3.00	2.54	1.72	.67
With income of specified type	1969	10.11	7.44	9.91	12.74	10.45	10.14	13.19	9.29	7.07
	1968	9.06	8.69	11.95	10.41	7.76	6.03	10.67	12.36	5.05
	1969 less 1968	1.05	-1.25	-2.04	2.33	2.69	4.11	2.52	-3.07	2.02
Mean income after allocation	1969	$2,366	$2,216	$2,780	$2,682	$2,957	$1,013	$2,298	$2,687	$2,261
	1968	$3,208	$ 961	$2,980	$5,718	$2,370	$4,397	$4,428	$1,991	$2,329
	1969 less 1968	-$ 842	$1,255	-$ 200	-$3,036	$ 587	-$3,384	-$2,130	$ 696	-$ 68
Income other than Earnings:										
Total persons 14+	1969	10.22	8.96	10.80	10.17	12.18	11.33	12.30	9.22	6.83
	1968	7.44	7.80	8.65	6.82	5.47	8.13	8.59	7.70	6.33
	1969 less 1968	2.78*	1.16	2.15	3.35	6.71	3.20	3.71	1.52	.50
With no income of specified type	1969	7.60	7.02	8.05	8.07	8.56	8.89	8.26	7.03	4.73
	1968	6.17	6.36	7.67	5.75	4.27	6.64	7.21	6.38	4.97
	1969 less 1968	1.43*	.66	.38	2.32	4.29	2.25	1.05	.65	- 24
With income of specified type	1969	15.10	12.37	15.90	14.06	19.59	15.99	20.34	13.21	10.34
	1968	10.17	10.96	10.82	9.13	7.79	11.54	11.68	10.71	9.08
	1969 less 1968	4.93*	1.41	5.08	4.93	11.80	4.45	8.66	2.50	1.26
Mean income after allocation	1969	$1,588	$1,547	$1,460	$1,560	$1,600	$1,762	$1,719	$1,371	$1,626
	1968	$1,511	$1,390	$1,394	$1,610	$1,371	$1,497	$1,534	$1,598	$1,703
	1969 less 1968	$ 77	$ 157	$ 66	-$ 50	$ 229	$ 265	$ 185	-$ 227	-$ 77

*The difference (1969 NA rate less 1968 NA rate) for "all rotations combined" is statistically significant at 2 sigma level.

1/ The numbers in the parentheses indicate month-in-sample. The first numbers are the rotation groups that are enumerated in 1969 and the second numbers are those enumerated in 1968.

A--Advance letter; WI--Work experience and Income in April. IA--Income in April.

UNBIASED ESTIMATES FOR STRATIFIED SUBSAMPLE DESIGN

Robert S. Cochran, Univ. Wyo. - Texas A & M Univ.

Harold F. Huddleston, U. S. D. A.

The sampling design to be discussed in this paper is an unusual but apparently effective way of combining several different ways of sampling and constructing estimates. The discussion will center around the design, estimation procedures, and some empirical results.

A. The Sampling Plan

The first step in the sampling operation is the selection of a single stage cluster sample. In application, the clusters are actually segments in the usual USDA sense. Each cluster, or segment, is divided into tracts. The tracts found in the sample of segments are then classified into various strata depending upon some characteristic of the tract.

The second stage of the sampling procedure is to draw a stratified sample of tracts from the group of tracts found in the segments sampled in the first stage. If the original sample of tracts had been drawn at random from the entire population, stratified, and then a stratified sample of tracts drawn at the second stage, the design would be classical double sampling for the stratification problem. If the tracts selected at the second stage had been selected at random from within the segments, the design would be a classical two-stage cluster sample. The actual plan is a compromise between these two operations.

B. The Estimation Formulas

Notationally, a sample of m segments is drawn at random from a population of M segments. In the sample of m segments there are found n_h tracts belonging to the h^{th} stratum. The second part of the sample calls for v_h tracts to be selected at random from the h^{th} stratum. There will be found v_{hs} tracts at the intersection of the h^{th} stratum and the s^{th} segment. Associated with each unit found in this intersection will be a characteristic, a_{hst}. These characteristics may be combined to give the following population and sample quantities:

$$a_{hs} = \sum_t^{v_{hs}} a_{hst} \qquad a = \sum_h^H a_h = \sum_s^m a_s$$

$$a_h = \sum_s^m a_{hs} \qquad A_s = \sum_h^H \sum_t^{n_h} a_{hst}$$

$$a_s = \sum_h^H a_{hs} \qquad A = \sum_s^m A_s$$

$$\hat{A}_1 = \frac{M}{m}\sum_s^m A_s = \frac{M}{m} A$$

$$\hat{A}_2 = \frac{M}{m} A = \frac{M}{m}\sum_h^H \hat{A}_h = \frac{M}{m}\sum_s^m \hat{A}_s$$

$\hat{A}_1$ is the estimate of the population total if there is a complete enumeration of all tracts in the sampled segments. $\hat{A}_2$ is based on the stratified sample of tracts.

Spelling out $\hat{A}_2$ in more detail we have

$$\hat{A}_2 = \frac{M}{m}\sum_s^m \hat{A}_s = \frac{M}{m}\sum_s^m \sum_h^H \frac{n_h}{v_h} \sum_t^{v_{hs}} a_{hst} \tag{1}$$

In this form $\hat{A}_s$ is seen to be a "domain" estimate of the s^{th} segment total based upon the stratified sample that cuts across all segments.

When $\hat{A}_2$ is written as

$$\hat{A}_2 = \frac{M}{m}\hat{A} = \frac{M}{m}\sum_h^H \hat{A}_h$$

its unbiasedness is apparent since $\hat{A}_h$ is an unbiased estimate of A_h and thus $\hat{A}$ is an unbiased estimate of A.

The variance of $\hat{A}_2$ is easily obtained by applying the principle of conditional expectation. As such

$$V(\hat{A}_2) = V_1 E_2(\hat{A}_2) + E_1 V_2(\hat{A}_2) \tag{2}$$

$$= \frac{M^2}{2}\left(1 - \frac{m}{M}\right)\sum_s^M \frac{\left(A_s - \frac{A}{M}\right)^2}{M - 1}$$

$$+ \frac{M}{m}\left[\sum_s^M V(\hat{A}_s) + \frac{m-1}{M-1}\sum_{s\neq s'} \mathrm{Cov}(\hat{A}_s\hat{A}_{s'})\right]$$

If the sample of tracts had been drawn following traditional two-stage sampling methods, the term involving covariances would have been zero, but since the sample was drawn following strata boundaries, the covariance of the domain estimates must be present. Yates (1960) indicates that these covariances will be negative so this stratified scheme will be better than the traditional two-stage cluster operation.

The derivation of the unbiased estimate of this expression is obtained by splitting it into its components, estimating each of them, and then recombining the component estimates.

For the "between" cluster component,

$$\frac{M^2}{m}\left(1 - \frac{m}{M}\right)\frac{\sum_s^M\left(A_s - \frac{A}{M}\right)^2}{M - 1},$$

the estimate is easily shown to be

$$\frac{M^2}{m}\left(1 - \frac{m}{M}\right)\left[\frac{\frac{M}{m}\sum_s^m\left(\hat{A}_s - \frac{\sum_s^m \hat{A}_s}{m}\right)^2 - \frac{M}{m}\sum_s^m \nu(\hat{A}_s) + \frac{\nu(\hat{A}_2)}{M}}{M - 1}\right] \tag{3}$$

For the "within" term,

$$E_1V_2(\hat{A}_2) = E_1\{\frac{M^2}{m^2} V_2(\sum_s^m \hat{A}_s)\}$$

$$= \frac{M}{m}\left[\sum_s^M V(\hat{A}_s) + \frac{m-1}{M-1} \sum_{s\neq s'} Cov(\hat{A}_s\hat{A}_{s'})\right]$$

the estimate is simply

$$\frac{M^2}{m^2} \nu(\sum_s^m \hat{A}_s). \tag{4}$$

Thus

$$\nu(\hat{A}_2) = \frac{M^2}{m}(1 - \frac{m}{M}) \frac{\sum_s^m \left(\hat{A}_s - \frac{\sum_s^m \hat{A}_s}{m}\right)^2}{m - 1}$$

$$+ \frac{M}{m}\sum_s^m \nu(\hat{A}_s) + \frac{M(M-1)}{m(m-1)}\left[\nu(\sum_s^m \hat{A}_s) - \sum_s^m \nu(\hat{A}_s)\right] \tag{5}$$

Since

$$\sum_s^m \hat{A}_s = \hat{A} = \sum_h^H \frac{n_h}{v_h} \sum_s^m \sum_s^{v_{hs}} a_{hst}$$

is a stratified estimate of A, $(\sum_s^m \hat{A}_s)$ can be computed by

$$\nu(\sum_s^m \hat{A}_s) = \sum_h^H \frac{n_h^2}{v_h}(1 - \frac{v_h}{n_h})$$

$$\cdot \frac{\sum_s^m \sum_t^{v_{hs}} \left(a_{hst} - \frac{\sum_s^m \sum_t^{v_{hs}} a_{hst}}{v_h}\right)^2}{v_h - 1} \tag{6}$$

Since $\hat{A}_s$ is a "domain" type estimate for the s^{th} segment when the stratified sample of tracts cuts across all segments, the quantity $\nu(\hat{A}_s)$ can be written as

$$\nu(\hat{A}_s) = \sum_h^H \frac{n_h^2}{v_h}(1 - \frac{v_h}{n_h})$$

$$\cdot \frac{\sum_t^{v_{hs}} \left(a_{hst} - \frac{\sum_t^{v_{hs}} a_{hst}}{v_h}\right)^2}{v_h - 1}$$

When the number of tracts in the intersection of the h^{th} stratum and the s^{th} segment, n_{hs}, is known, these formulas can be altered to take this into account.

Some additional algebra can be used to give a more compact form for $\nu(\hat{A}_2)$ as

$$\nu(\hat{A}_2) = \frac{M^2}{m}(1 - \frac{m}{M}) \frac{\sum_s^m \left(\hat{A}_s - \frac{\sum_s^m \hat{A}_s}{m}\right)^2}{m - 1}$$

$$+ \frac{M^2}{m^2}\left[\sum_h^H \frac{n_h^2}{v_h}(1 - \frac{v_h}{n_h})\frac{(-\sum_{s\neq s'} a_{hs}a_{hs'})}{v_h(v_h - 1)}\right]^2$$

$$+ \frac{M}{m}\sum_s^m \nu(\hat{A}_s) .$$

C. Some Empirical Results

An area sample survey was conducted in Ohio in June and December, 1969, using this sampling procedure. The purpose of the survey was to estimate the inventory of major livestock characteristics, spring planted crops, farm labor and operator characteristics, and the harvest of fall planted crops. The area sample survey data was collected on a subunit or tract basis for each of the segments. A tract being the land under the control of an operator. The tracts were then regrouped into nine strata based on previously established criteria for a survey which would have different emphasis then the initial June Survey. However, a number of survey characteristics were common to both surveys. The data for six of these characteristics are given in the following tables. In table 1, column 1, are the estimates referred to as $\hat{A}_1$. These are the usual single stage cluster sample estimates for the data obtained in June. In column 2 are the estimates of the June totals based upon the stratified subsample and computed using the $\hat{A}_2$ formula. In column 3 are the estimates of the December totals based upon the stratified subsample and also computed using the $\hat{A}_2$ formula. The December data has been obtained on the same units as the column 2 June estimates.

In table 2 the estimates of the variances of the estimates in table 1, column 2 and column 3 are presented in total as well as broken into three components. The first component is computed between estimates of segment totals. The second component is the usual within segment component. The third component reflects the covariance of estimates of segment totals. The relative size of these components follows the usual pattern of most of the variance estimates being obtained from the between component.

Table 1

Estimates of Characteristic Totals

Characteristic	Based on Complete June Survey	Based on Sample from June Survey	Based on December Data - June Sample
	(000)	(000)	(000)
1	2,530	3,141	3,290
2	2,252	2,891	3,236
3	1,017	1,703	1,232
4	11,847	6,824	7,411
5	1,012	1,288	1,304
6	45	46	95
M	107,858		
m	350		
Σn_h	2,377		
Σv_h		487	487

Table 2

Estimates of Variance Components

Characteristic	Between		Within		Covariance		Total	
	June	December	June	December	June	December	June	December
	(000,000)	(000,000)	(000,000)	(000,000)	(000,000)	(000,000)	(000,000)	(000,000)
1	245,123	272,023	175	240	-2,282	-1,921	243,016	270,342
2	105,499	141,990	79	114	-2,186	-2,092	103,392	140,012
3	193,538	86,396	389	174	- 482	- 506	193,445	86,064
4	17,077,493	13,526,947	33,185	24,748	-2,659	-4,367	17,108,019	13,547,328
5	13,414	17,577	20	31	- 429	- 491	13,005	17,117
6	655	1,298	1	2	0	- 4	656	1,296

D. Reference

Yates, F. (1960). Sampling methods for censuses and surveys. Charles Griffin and Co., London, third edition.

METHODOLOGY IN A STUDY OF PHYSICIANS' ATTITUDES TOWARD NURSE-MIDWIVES

Seth B. Goldsmith, Tulane University and Monroe Lerner, The Johns Hopkins University

Traditional solutions to the problem of the physician "shortage" in the delivery of health services have usually involved increasing the annual output, and hence the supply, of physicians (1). However, difficulties and limitations inherent in this approach have demonstrated the desirability of seeking alternative and/or supplementary solutions to this problem. One such alternative, currently the subject of wide interest, is to extend the productivity of practicing physicians by encouraging them to use other professionals and/or para-professionals, e.g., nurse-midwives, to provide medical care under the physician's professional supervision.

The feasibility of this solution, at least as far as the involvement of nurse-midwives in maternity care is concerned, indeed its effectiveness and efficiency, has been generally accepted, in settings other than the United States, e.g., in some Western European countries. Thus, in England and in Sweden, the evidence has indicated that nurse-midwives, at least in the settings characteristic of those countries, can deliver a high quality of maternity care, perhaps as high as that of physicians (2,3,4). This may also be true of other professional and/or para-professional personnel, but hard and fast evidence unfortunately is lacking, at least up to this point.

In considering the feasibility of this alternative solution, it seems clear that a major difficulty is likely to be the unwillingness of professionals to delegate tasks for which they must assume legal and professional responsibility. However, a major assumption of the present research is that the degree of this unwillingness is likely to be conditional, i.e., it is likely to depend upon the circumstances including, for example, their perception of the nature of the task itself and the extent to which, in their opinion, it can be performed by another professional or para-professional person. The objective of the present study was to determine the extent to which a professional group--in this instance, obstetricians--would be willing to permit a para-professional group--in this instance, nurse-midwives--to perform specific aspects of their jobs.

The development of this study, i.e., its movement from the stage of problem formulation to the stages of research design and methods, necessarily involved decisions which, it is thought, might be of interest to students in this field, especially those doing research on the use of para-professional personnel. The present discussion reports the bases for some of the major decisions made in this study, e.g., those involved in selecting the specific professional relationship to serve as the object of study; the selection of the specific attitudes within that relationship which were most pertinent to the present study; the selection of the most appropriate study setting; the selection of the study population; questionnaire design; and selection of a data-gathering strategy.

THE APPROPRIATE PROFESSIONAL RELATIONSHIP

Four physician-nonphysician health personnel relationships were initially considered: the anesthesiologist/nurse-anesthetist; the pediatrician/pediatric-associate; the internist (or general practitioner)/physician's assistant; and, the obstetrician/nurse-midwife. These are the four relationships which have been the focus of perhaps the most interest during the current discussions about the necessity for extending the productivity of physicians through the use of paramedical personnel.

From among these four, the relationship to be studied should provide the optimum mix of the following factors: a) it should readily permit generalization to the entire universe of physician/paramedical relationships; b) it should help to define the potential outer limits of the paramedical roles, i.e., the maximum degree to which the physician is likely to delegate authority and responsibility for the delivery of services to any paramedical person; c) it should be a clearly defined and structured role, so that physicians might be able, as objectively as possible, to appraise the paramedical person's ability to deliver a satisfactory level of care; and, d) the medical specialty concerned should consist of a series of routine, easily analyzed procedures which would be relatively amenable to study.

On the basis of these criteria, the obstetrician/nurse-midwife relationship was selected. The study of other paramedics such as physician's assistants and pediatric associates was rejected because the potential roles of these paramedics are extremely vague, a situation caused by the lack of common structure in their educational programs and role expectations.

The study of nurse-anesthetists presented quite the opposite problem; their role is too clearly defined. Asking anesthesiologists to define the already well-established role of the nurse-anesthetist would not in any real measure be determining the extent to which a professional group would be willing to permit a new professional group to perform aspects of their jobs. Nurse-midwives, in contrast to the other three, have had a well-defined and highly respected role in many other countries. Considering their high level of education and training, a definition of the extent of the nurse-midwives' role might well provide the utmost that could be expected in terms of the limits of paramedical authority and responsibility.

THE SPECIFIC ATTITUDE

Selection of the obstetrician/nurse-midwife relationship as the focus of the research, however, imposed certain constraints on the study.

For one thing, it dictated that the study measure attitudes toward a hypothetical possibility as opposed to the measurement of behavior in a real situation. This limitation occurs because very few of the real situations of the type required were available, i.e., very few nurse-midwives are available for employment, and the possibility of finding obstetricians who had had an opportunity to interact directly with nurse-midwives was, therefore, remote (5).

The specific attitude to be studied was the physician's opinion about the ability of nurse-midwives to perform specific technical functions. Since technical competence represents one of the foundations of professionalism and professional relationships, it seemed reasonable to study this dimension of professionalism as opposed, for example, to such other aspects of obstetrical practice as acceptability of a paramedical person to the patient, monetary or financial reward or penalty to the obstetrician concomitant upon the introduction of a paramedical person, etc. Discovering and defining the professional consensus, if any, about the ability of the nurse-midwife to perform specific technical functions should help to define the limits of authority and responsibility that practicing professionals might be likely to delegate to a paramedical person in the delivery of health services.

Finally, the question was: How were researchers to discover and define this professional consensus about the ability of the nurse-midwife to perform specific technical functions? The issue was: Should obstetricians be asked, "Do you think nurse-midwives should be able to ...?" or "Would you allow a nurse-midwife to ...?" If asked this first alternative question, it was reasoned, the obstetrician would merely be making an <u>independent professional judgment</u> about a nurse-midwife's probable abilities, but, in the second alternative, the obstetrician would have to make a decision which might be threatening to him, i.e., he would have to state whether he would, <u>under unclear circumstances</u> delegate professional authority and responsibility to a paramedical person. The first alternative wording of the question was deemed to be clearly preferable and was adopted.

THE STUDY SETTING

The choice of a study setting required a decision among the two major alternatives of using a national or local setting, and, if local, whether it should be a region, a state, e.g., Maryland, an urban, or a rural area. The decision was to use the State of Maryland as the study setting.

It is obviously easier to generalize from the findings based on a national sample. However, the expense of using a national sample--the cost of obtaining a list of the study population and gathering the data--was considered too great for this study. Additionally, the use of a national sample in this study would have lost the special advantages inherent in certain settings, such as physicians having had some previous experience with nurse-midwives. The final factor leading to the rejection of the national study setting was the inability to obtain a national medical or obstetrical association endorsement for the study. This lack of endorsement, it was reasoned, would be a serious handicap in the gathering of data for the study.

While the decision not to use a national or regional sample obviously imposed some limitations on one's ability to generalize from the findings of this study the advantages of doing research in the State of Maryland were considered an acceptable trade-off. The primary advantage of this study setting lay in the fact that for the past 25 years nurse-midwives have worked in the various health departments and hospitals throughout the State. Two of Baltimore's leading hospitals employ a large number of nurse-midwives and one of the two medical schools in the State is closely affiliated with a nurse-midwifery training program. Since this research was also interested in the effect of a number of background and experience factors on physicians' attitudes toward nurse-midwives, the fact that many physicians in Maryland had had previous contact with nurse-midwives was considered important.

From a logistical viewpoint, Maryland offered two advantages. First, the State Medical Society was interested in the study and was willing to endorse it. Such interest and endorsement were thought to be extremely useful in terms of gaining cooperation and increasing responses to the questionnaire. Second, the researchers were located in Maryland so that the possibility of telephone and personal follow-up of non-respondents became both convenient and monetarily feasible.

SELECTION OF THE STUDY POPULATION

It was decided that the universe studied should be comprised of obstetricians and general practitioners with a secondary specialty of obstetrics who were born prior to 1937. The choice of whether or not to have an age cutoff was the issue in making this decision. To get at the homogeneous group of physicians who could have had experiences with nurse-midwives or paramedics during internship, residency, or military service meant that a 32 year old cutoff point was required. Clearly, many physicians of that age might not have had these experiences; however, they would have had the opportunity for such experiences. Setting the age limit higher than 32 would not allow for the sampling of the largest possible number of physicians who could be used to test hypotheses relating to the association between age and experience variables and attitudes.

With this single age exclusion, all other physicians, regardless of type of practice, e.g., solo or group, or the nature of their principal employer, e.g., government, medical school or self-employed, were considered part of the universe. Selection was not related to these factors because it was reasonably expected that all physicians practicing obstetrics had some attitudes toward the professional functioning of nurse-midwives. Further, although the importance of opinion leaders was recognized, no attempts were made in selecting the population to evaluate the extent of each physician's influence within medical circles and the community at large. Such evaluation was considered too time consuming and somewhat beyond the bounds of this study.

Having decided what the universe would be, the question was where and how could a list of the study population be obtained. Several possibilities were considered. One way would have been to compile a list of physicians' names from their signatures on birth certificates in the State. This was rejected as being too time consuming and, more importantly, not representative. The problem inherent in this alternative is that the signer of the document may not have been the actual deliverer of the infant or even the physician who cared for the mother during the prenatal or postpartum phases.

Such other potential sources of a physician list as the two State obstetrical societies, medical societies, and the State general practitioner association were investigated. It was found, however, that their information was difficult to obtain and essentially incomplete. The American Medical Association's directory was rejected as well because four to five years had elapsed since the collection of its data. The Physicians' Record Service of the American Medical Association presented the most reasonable alternative. For a fee of slightly under $100 the Association provided a list of all the obstetricians and general practitioners with a secondary specialty of obstetrics in the State. Their information also provided data about the physician's background such as medical school attended, date of graduation, year of birth, type of practice, and employer. Finally and perhaps most importantly, this data was current --two to four weeks old--an advantage not found in the other alternatives.

QUESTIONNAIRE DESIGN

The writing of the questions for the attitudinal questionnaire initially required a definition of the tasks performed by obstetricians in the course of delivering normal maternity care. One way of defining these tasks would have been to ask a number of obstetricians to enumerate the functions they perform in providing normal maternity care. This self-analysis was rejected as being too time consuming for both the physicians and the researchers.

An alternative process of analyzing maternity care by outlining standard obstetrical textbooks was chosen. While this means of defining the tasks might be considered by some to be too theoretical, it had the advantage of providing uniform information inexpensively. It was reasoned that the gap between the theoretical and the practical could be bridged at the time of the pretesting of the questionnaire. From the various textbooks, 150 statements were collected which described the normal obstetrical process. This number was reduced to 91 by editing and eliminating statements that appeared to overlap.

At this point, it was necessary to decide between the alternatives of devising an a priori Likert or Thurstone type scale or a post factum scale by using the Guttman model or factor analysis. Two complications were foreseen in the use of a priori scaling techniques. First, there would be the problem of using judges from an already small universe and thus depleting the universe of possible study subjects. Second, there would be the pragmatic consideration of the insurmountable logistical problems inherent in getting a large number of physicians to evaluate a number of statements. The use of factor analysis was chosen both because of its relative logistical simplicity and its value in testing the theoretical framework upon which the study was built. The final attitude scale was a result of the factor analysis, but a hypothesized scale existed prior to the creation of the new scale. Correlating this hypothesized scale and the devised scale was a possibility present in post factum but not in a priori scaling. A close correlation of the hypothesized and devised scale, as was the case, provides significant support to the theories underlying the hypothesized scale.

Having decided to do post factum scaling, the remaining 91 statements were then analyzed for clarity and functional relevance by the nurse-midwives at The Johns Hopkins University. Statements which clearly reflected the functions of the nurse-midwife, or those which were in doubt, were retained. As a result of this elimination process, 42 statements remained. These were then reviewed by the obstetrical consultants to the study who suggested that the tasks performed by physicians in providing normal maternity services could be grouped into five major categories: (1) diagnostic judgment; (2) prenatal care; (3) normal delivery; (4) postpartum care and family planning services; and, (5) operative medical care. These categories, it was reasoned, reflected not only the major functions performed during normal obstetrical care, but the continuum of technical and professional expertise required to perform these functions. This led to the hypothesis that obstetricians would perceive nurse-midwives as being most able to perform tasks requiring the least degree of technical and professional competence. In order to test the validity of this scale and, at the same time not influence the respondents, no attempt was made to reflect this continuum or categorization in the design of the actual questionnaire. The questionnaire asked the physicians to answer the 42 questions about the ability of the nurse-midwife to perform the obstetrical care functions. Responses were made along a five point scale of strongly agree to strongly disagree.

A pretest to check the clarity of the questionnaire, the procedure for its administration, and the validity of the instrument was carried out in another state. The specific hospital selected for pretesting had a large obstetrical service, employed a nurse-midwife, and was remote from the study site. Additionally, the chief of obstetrics and the administrators were interested enough in the study to make it possible to interview many of the respondents.

The alternative of using a part of the non-selected sample in the State of Maryland was considered an easier and less costly way of pretesting. However, the difficulty in interviewing physicians scattered throughout the State, along with the probable contamination of the sample made this alternative unacceptable.

DATA GATHERING

A review of several studies concerned with alternative data gathering strategies revealed that no single method is sacrosanct (6,7,8). Therefore, the realistic factors of time and personal bias governed the decision to use mailed questionnaires, rather than personal interviews. The time and financial constraints are perhaps best explained by the fact that there were 210 physicians in the sample, the land area of the State of Maryland is almost 10,000 square miles, and there was only one investigator to do all of the interviewing. Almost of equal importance in making the decision was a sensitivity about the introduction of the investigator's bias into the personal interview. This sensitivity was, in part, a result of the interviewer's experiences during the pretest where, at times, he was forced to leave the role of "detached observer" to play the role of "defender", or "advocate", of nurse-midwifery.

Because the mailed questionnaire and the strategy to increase responses resulted in an unusually high response rate (93.6 per cent), the data gathering stategy will be briefly described. A number of different techniques may have contributed to the success of this effort, but, since none of these variables were controlled, it is virtually impossible to state their individual significance.

The underlying strategy in data gathering was to approach the physician in a personalized manner, thereby, stimulating his interest and desire to cooperate in the study. All letters were personally typed and addressed on an automatic typewriter and then hand signed. The personal approach also was emphasized in the questionnaire on which the physician's name, medical school, and year of graduation were filled in beforehand.

Several days prior to the mailing of the first letter and questionnaire, the president of the State Medical Society sent each physician in the sample a letter asking for his cooperation in the study. The questionnaires and cover letters were then stamped by hand and mailed special delivery. Non-respondents received a follow-up special delivery letter and questionnaire 15 days later, followed by another special delivery letter to non-respondents 20 days thereafter. By the 40th day of the data collection effort, more than 90 per cent of the sample had responded at a total cost per response of approximately $1.42.

* * *

This report has traced the development of a medical care research study from formulation of the problem through a number of critical decision stages to a final research design. These critical decisions involved alternatives that eventually led to the selection of the obstetrician/nurse-midwife relationship as the appropriate professional relationship to study, the selection of the physician's attitude toward the technical ability of the nurse-midwife as the specific relationship that was the most relevant to study, the selection of Maryland as the study setting, and the use of all obstetricians and general practitioners 32 years of age or older as the study population. This paper also reviewed the alternatives considered prior to the designing of the questionnaire and, finally, it described the strategy used in collecting the data. It is hoped that this presentation will be of interest to those responsible for conducting research on the use of para-professionals, as well as, to collaborators in such research.

FOOTNOTES

(1) Excellent reviews of previous health manpower studies are found in Rashi Fein's The Doctor Shortage: An Economic Analysis (Washington: The Brookings Institution, 1967), pp. 6-13 and the Report of the National Advisory Commission on Health Manpower, Volume II, (Washington: Government Printing Office, 1967), pp. 265-277.

(2) U. S. Department of Health, Education, and Welfare, International Comparison of Perinatal and Infant Mortality: The United States and Six West European Countries, (Washington: Government Printing Office, 1967).

(3) H. Arthure, et.al., Report of Confidential Enquiries Into Maternal Deaths in England and Wales, 1964-1966, (London: Her Majesty's Stationery Office, 1969).

(4) L. Runnerstrom, "The Effectiveness of Nurse-Midwifery In a Supervised Hospital Environment", Bulletin of the American College of Nurse-Midwifery, XIV:2 (May 1969), pp. 40-52.

(5) Nurse-Midwives--U.S.A. Descriptive Data. New York: American College of Nurse-Midwifery, 1968.

(6) J. Hochstim, "A Critical Comparison of Three Strategies of Collecting Data From

Households", Journal of the American Statistical Association, 62 (September 1967), pp. 976-989.

(7) W. L. Slocum, et.al., "Increasing Response to Questionnaires and Structured Interviews", American Sociological Review, 21:2 (April 1956), pp. 221-225.

(8) A. M. Burgess and J. T. Tierney, "Bias Due to Non-Response in a Mail Survey of Rhode Island Physicians' Smoking Habits - 1968", New England Journal of Medicine, 282:16 (April 16, 1970), p. 908.

PROJECTIONS OF CONSUMER INCOME AND EXPENDITURES FOR 1985

Roger A. Herriot and Herman P. Miller*
U.S. Bureau of the Census

Introduction

This paper summarizes some recent work being done at the Bureau of the Census. Our concern here is with the development of graphical techniques for projections of consumer income and expenditures. In addition, an attempt is made to separate the contributions of population growth and economic growth (increase in average family income) on the expected increase in expenditures for major categories of consumption between now and 1985.

Development of Control Totals and "Corrected" Size Distribution

Due to the income underreporting in the CPS, the money income of families and unrelated individuals was adjusted to control totals using data supplied by the Office of Business Economics (OBE). Beginning with the OBE's "personal income total," we subtracted out those types of "personal income" which are not included in the Census's (BC) "money income" and added those sources which are included in Bureau of the Census "money income" but are excluded from "personal income." The major items subtracted from "personal income" are all types of imputed income, income of non-profit institutions, medicare payments to beneficiaries, and various types of lump sum payments. The major items which were added to "personal income" are employees contributions to social insurance and periodic payments for life insurance.

After computing control totals for each source of income, these income source control totals were then distributed to the income intervals using the preadjustment percentages. For example: if the BC $8,000 to $8,999 income interval had 5.6 percent of the total wage and salary income before adjustment, this interval is then given 5.6 percent of the OBE wage and salary control aggregate. Within each income interval, we then added together the aggregate amount for each source to arrive at a total money income aggregate for each income interval. These "new" total money income aggregates were then divided by the number of families and unrelated individuals in the intervals before any adjustments were made. This yields a set of new interval means which were then plotted directly above the preadjustment interval means using lognormal graph paper. The "corrected" income distribution based on OBE control totals is obtained by connecting the new interval means and reading from the graph, the number of families and unrelated individuals within each income interval. While this type of "raking" adjustment, which assumes an equiproportional underreporting of income, leaves much to be desired, it was found to yield a distribution very close to that developed by Dr. Budd at OBE for 1964.[1]

The Projection of the 1985 Aggregate Income

To project aggregate total money income for 1985, we used the following model:

$$TMY = GNP^*$$

$$GNP^* = \frac{GNP^*}{MH} \times \frac{MH}{E} \times \frac{E}{LF} \times LF$$

$$LF = \left(\frac{LF_{mi}}{P_{mi}}\right)\left(P_{mi}\right) + \left(\frac{LF_{fi}}{P_{fi}}\right)\left(P_{fi}\right)$$

Where

GNP^*	=	Constant dollar GNP with average productivity increase inputed to government workers.
MH	=	Manhours worked per year
E	=	Employed labor force
LF	=	Total labor force
LF_{mi}	=	Male labor force in the i th age group
LF_{fi}	=	Female labor force in the i th age group
P_{mi}	=	Male population in the i th age group
P_{fi}	=	Female population in the i th age group

Below are presented the assumed relationships between the variables for 1985 along with their 1968 values. In addition we have presented the magnitude of the variables for 1968 and 1985 along with the implicit average annual rate of change for each variable for this period. We used the "C" population series and 3.0% increase in output per manhour which was applied to government workers as well as those in the private economy in order to maintain a reasonably constant relationship between constant dollar output and constant dollar money income. The average labor force participation rates for males and females were estimated using the weighted average of the estimated labor force participation rates for each age group for each sex.

* The views presented here are not necessarily those of the Census Bureau or any other governmental agency.

	Description		1968	1985
$\frac{TMY}{GNP^*}$	= Money income to output ratio........................		.72	.72
$\frac{GNP^*}{MH}$	= Output per man-hour @ 3.0% average annual increase...		\$5.54	\$ 9.16
$\frac{MH}{E}$	= Average hours worked per year per worker..............		1,966	1,851.2
$\frac{E}{LF}$	= Employment rate..................................		.966	.960
$\frac{LF_m}{P_m}$	= Male average labor force participation rate...........		.813	.804
$\frac{LF_f}{P_f}$	= Female average labor force participation rate.........		.411	.433
TMY	= Total Money Income...........	Billions in 1968 \$.. % change—av. an. rate..	622.3 —	1,263.4 4.3
MH	= Total Man-Hours Worked.......	Billions.. % change—av. an. rate..	156.2 —	191.4 1.2
E	= Employed Labor Force.........	Thousands.. % change—av. an. rate..	79,455 —	103,382 1.6
LF_m	= Male Labor Force (Age 16+)...	Thousands.. % change—av. an. rate..	53,030 —	68,051 1.5
LF_f	= Female Labor Force (Age 16+).	Thousands.. % change—av. an. rate..	29,242 —	39,639 1.8
P_m	= Male Population (Age 16+)....	Thousands.. % change—av. an. rate..	65,238 —	84,692 1.6
P_f	= Female Population (Age 16+)..	Thousands.. % change—av. an. rate..	71,117 —	91,464 1.5

Projection of the 1985-Income Size Distribution

The projected income size distribution, shown in table 1, is based on the assumption that in the next 15 years, the Lorenz Curve (that is the cumulative percent distribution of families and of income) will be constant. On a lognormal graph this assumption results in a parallel upward shift of the cumulative frequency distribution of families receiving less than a certain level of income.

The procedure used to project the size distribution is as follows: we first calculated the 1985 aggregate income assuming no population growth using the 1968 number of families and unrelated individuals and the projected 1985 mean income. This 1985 aggregate is then distributed to the income intervals using the 1968 percentages. Thus if the \$4,000 to \$5,000 income interval had 2.8 percent of the total income in 1968, this interval is then given 2.8 percent of the estimated 1985 aggregate. These interval aggregates are then divided by the number of families and unrelated individuals in that interval in 1968. This yields a set of new interval means which are then plotted directly above the 1968 interval means on lognormal graph. When these new means are connected using a French curve, we have the projected 1985 distribution. From this new distribution we can then read off the percentage of all families and unrelated individuals which have income below any particular level of income. By applying these family and unrelated individuals percentages against the projected number of families and unrelated individuals in 1985, we can calculate the estimated number of families and unrelated individuals in each income interval.

Based on our assumptions, real incomes would grow by over 100 percent during the next 17 years. In 1968 the money income of families and unrelated individuals totaled about 629 billion dollars. It is expected to reach about \$1.3 trillion by 1985. The average (mean) income of families and unrelated individuals is expected to increase from about \$9,800 in 1968 to about \$14,700 in 1985 measured in constant purchasing power. At present about 17 percent of all families and unrelated individuals receive incomes over \$15,000 and account for about 39 percent of the income. By 1985 about 44 percent of families and unrelated individuals are expected to have incomes above \$15,000 at constant 1968 prices, and will account for over 3/4 of the total income. Moreover, because of the combined impact of both income and population growth, the total amount of purchasing power, in constant dollars, at this upper income level will be over four times as great as in 1968.

Projection of Consumer Expenditures in 1985

Given the above configuration of purchasing power in 1985, our next concern is to examine how it will be spent. We have projected the major types of consumer expenditure using the procedures outlined below.

We adjusted each category of consumer expenditures for price changes and calculated, for a number of the postwar years, each category's proportion of total personal consumption expenditures in constant dollars. For each type of consumer expenditure, we then plotted the proportions on semi-log paper and by fitting a linear trend line, we calculated an average rate of change in its proportion of total consumer expenditures. For example, the proportions of total consumer expenditure which goes for food has been declining during the postwar period. We fitted a linear trend line to this series and projected the proportion to be spent on food in 1985 (relative to total consumption expenditures). When this was done for each major type of expenditures, the sum of these percentages for 1985 added to more than 100 percent. The percentages were then proportionately reduced so that their sum would equal 100 percent. The net result was a slight reduction in longrun rate of increase of those expenditure types which have been increasing, and a slight increase in the longrun rate of change of those consumer types which have been falling. Estimates were made for 1980 using this procedure, and the estimated percentage for each expenditure type came to within one half of one percent of the Bureau of Labor Statistics (BLS) 1980 personal consumption forecasts.

For the 1985 expenditure projections, shown in table 2, we adjusted OBE's Personal Consumption Expenditure (PCE) so that they reflected, as much as possible, only the money expenditure of families and unrelated individuals.

Consequently, we excluded the expenditures of nonprofit foundations, net imputed rent on owner-occupied nonfarm dwellings, imputed proprietors income from owner-occupied farm dwellings, space rental values of institutional buildings, food and fuel produced and consumed on farms, services furnished without payment by financial intermediaries, and employees food, lodging, and clothing furnished by employers. By subtracting the above items from their respective expenditure type, we obtained an estimate of the money expenditure of families and unrelated individuals. It should be noted that an imputed amount covering capital consumption allowances, taxes and interest expenses on owner-occupied dwellings are included in our figure for housing expenditure.

We also developed a technique which is useful in allocating the proportion of the increase in each type of consumer expenditure that can be attributed to population growth and an overall measure including factors which result in increasing affluence. This procedure rests on the following two identities:

$$x_o = \left(\frac{x_o}{P_o}\right)\left(P_o\right)$$

and

$$x_1 = \left(\frac{x_1}{P_1}\right)\left(P_1\right)$$

where

x_o = Total expenditure on x in 1968

x_1 = Total expenditure on x in 1985

P_o = Number of families and unrelated individuals in 1968

P_1 = Number of families and unrelated individuals in 1985

The change in expenditure on x between 1968 and 1985 is then equal to:

$$x_1 - x_o = \left(\frac{x_1}{p_1}\,p_1\right) - \left(\frac{x_o}{p_o}\,p_o\right)$$

By adding the following terms:

$$p_o\frac{x_1}{p_1} - p_o\frac{x_1}{p_1} + p_1\frac{x_o}{p_o} - p_1\frac{x_o}{p_o} + p_o\frac{x_o}{p_o} - p_o\frac{x_o}{p_o}$$

and factoring the result, we obtain:

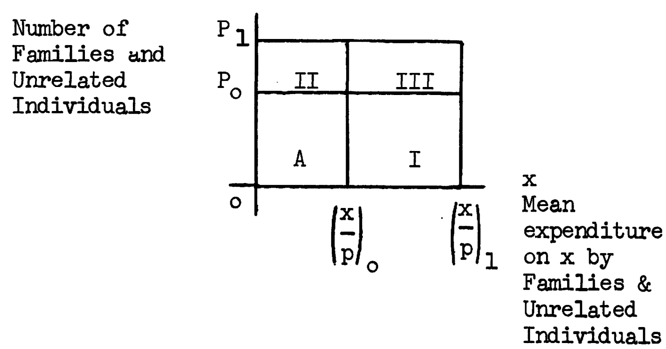
$$\left(x_1 - x_o\right) =$$

$$\left(p_o\right)\left(\frac{x_1}{p_1} - \frac{x_o}{p_o}\right) + \left(\frac{x_o}{p_o}\right)\left(p_1 - p_o\right) + \left(p_1 - p_o\right)\left(\frac{x_1}{p_1} - \frac{x_o}{p_o}\right)$$

Thus:

$$\Delta x = p_o\,\Delta\left(\frac{x}{p}\right) + \left(\frac{x}{p}\right)_o\,\Delta p + \Delta p\,\Delta\left(\frac{x}{p}\right)$$

The first term shows the "pure" income effect (i.e., the change in average expenditure on x times a constant population); the second term shows the "pure" population effect (i.e., the change in population times a constant average consumption of x); and the final term shows the combined effect of the change in population times the change in average consumption of x. The following box diagram is useful in showing the magnitude of the above effects.



On the "x" axis we measure the average expenditure on x, and on the "p" axis we measure the number of families and unrelated individuals. The area of box A then measures the aggregate expenditure on x for period zero $p_o\left(\frac{x}{p}\right)_o$. The area of box I shows the increase in expenditure for x attributable to the "pure" income effect, box II shows the increase effect of the increases in the income and population factors. For our analysis we partitioned box III into separate population and income effects. The rule we used to allocate box III was to prorate it according to the relative magnitude of the "pure" population effect (box II) and the "pure" income effect (box I). Thus if the area of box I is twice as large as box II, then 2/3 of box III is attributed to income factors and 1/3 is attributed to population factors.

Using the above procedure we have estimated the proportion of the projected increases in income and major categories of consumer expenditure, which can be attributed to population and to income factors. The estimates are shown in table 2. About 60-percent of the increase in aggregate income can be attributed to increasing affluence and the remaining 40-percent of the increase is attributable to population growth. Increasing affluence is also the more important factor in the projected increase in all the major categories of consumer expenditure except for food and clothing. Of the remaining categories, increases in income account for almost three-fourths of the increase for housing and recreation and from 51 to 61-percent of the increase for the remaining categories.

Table 1.--DISTRIBUTION OF FAMILIES AND UNRELATED INDIVIDUALS AND AGGREGATE INCOME, BY INCOME LEVELS: 1968 AND 1985

(All figures in 1968 dollars)

Income levels	1968 1/	1985 estimate
FAMILIES AND UNRELATED INDIVIDUALS		
Number (millions)	64.3	85.9
Percent distribution		
Total	100.0	100.0
Under $3,000	14.5	8.6
$3,000 to $4,999	12.7	8.1
$5,000 to $9,999	32.8	21.4
$10,000 to $14,999	23.0	21.9
$15,000 to $24,999	13.1	26.8
$25,000 and over	3.9	13.2
Median Income	$8,470	$12,600
AGGREGATE INCOME		
Amount (billions)	$629	$1263.4
Percent distribution		
Total	100.0	100.0
Under $3,000	2.5	1.0
$3,000 to $4,999	5.1	2.2
$5,000 to $9,999	25.0	11.0
$10,000 to $14,999	28.2	17.9
$15,000 to $24,999	24.3	33.1
$25,000 and over	14.9	34.8
Mean Income	$9,779	$14,700

1/ Amounts adjusted to compare with OBE control totals.

Table 2.--AGGREGATE CURRENT CONSUMPTION EXPENDITURES OF ALL FAMILIES AND UNRELATED INDIVIDUALS: 1968 AND 1985

(In billions of 1968-dollars except percents)

Income levels	1968	1985	Percent increase	
			Attributable to population change	Attributable to income change
Total aggregate income....	$628.9	$1,263.4	40.0	60.0
Taxes, Savings.................	117.1	277.9	30.4	69.6
Expenditures for current consumption.....................	511.8	985.5	43.1	56.9
Food, beverage, tobacco....	124.7	183.3	77.2	22.8
Housing (shelter)..........	49.1	128.1	26.1	73.9
Household operation and furnishings...............	75.9	141.9	45.4	54.6
Clothing and clothing materials.................	55.5	95.6	53.3	46.7
Personal and medical care..	47.7	96.6	39.2	60.8
Transportation.............	72.2	130.1	49.0	51.0
Recreation, education, contributions, and other.....	86.7	209.9	29.2	70.8

1/ Edward C. Budd and Daniel B. Radner, "The OBE Size Distribution Series: Methods and Tenative Results for 1964," American Economic Review, May 1969, Vol. LIX, No. 2, pp. 435-449.

AMERICAN IMMIGRATION IN THE SIXTIES

Richard Irwin and Robert Warren, Bureau of the Census*

Introduction

Immigration added about 3.9 million persons to the United States population between the 1960 and 1970 Censuses. This is a moderate increase over the 3.0 million figure for the preceding ten years. The relative importance of immigration as a component of population change increased significantly as decreasing fertility rates through most of the period reduced the level of natural increase. By the end of the decade, immigration was contributing about 20 percent of total population growth as compared to only 11 percent in 1960.

The relative stability of the level of immigration during the past twenty years contrasts strongly with the large fluctuations which occurred during the first half of this century. During the ten years before 1914 immigration averaged about one million each year but dropped very sharply during World War I. The restrictive immigration laws passed after World War I, the great depression during the thirties, and World War II have acted to keep immigration far below the levels reached before 1914. In the absence of changes in our immigration laws, net immigration in the future should continue at about 400,000 per year.

The 3.9 million immigration figure cited above is a net value composed of five major components. These components and the figures for the decade 1960 to 1970 are as follows:

Net alien immigration[1]	3,125,000
Net arrivals from Puerto Rico	152,000
Net arrivals of civilian citizens	446,000
Conditional entrants[2]	380,000
Emigration	250,000
Net civilian immigration, total	3,854,000

The estimates of immigration given above are for civilians only. Movements of the Armed Forces to and from posts overseas do not appear as a component of immigration, because the population coverage of the national estimates under consideration here includes the Armed Forces overseas. The Armed Forces overseas increased by about 400,000 between April 1, 1960 and April 1, 1970.

Net alien immigration, the largest component of net civilian immigration, consists of aliens who have been lawfully admitted for permanent residence under the provisions of the Immigration and Naturalization Act, as adjusted by the Census Bureau for use in developing national population estimates. In addition to net alien immigration, the movement of resident aliens and citizens to and from the United States must be taken into account. Estimates are developed of the net movement of Puerto Ricans to the United States, the net movement of government employees and their dependents from posts overseas, and for the emigration of aliens and citizens. Finally, conditional entrants, as defined by the Census Bureau, are added to national population totals at the time of entry.

Net alien immigration

Net alien immigration added about 3,125,000 persons to the United States population between the 1960 and 1970 Censuses. The figures by year for net alien immigration and net civilian immigration are as follows:

Period	Net alien immigration	Net civilian immigration
July 1, 1969 to March 31, 1970	250,000	351,000
1968-1969	347,000	383,000
1967-1968	355,000	420,000
1966-1967	331,000	429,000
1965-1966	323,000	425,000
1964-1965	297,000	323,000
1963-1964	292,000	341,000
1962-1963	306,000	356,000
1961-1962	284,000	365,000
1960-1961	271,000	385,000
April 1, 1960 to June 30, 1960	69,000	76,000
Total	3,125,000	3,854,000

The Immigration Act of 1965 was a major change in national immigration policy. The national origins system, established by the Act of 1924, assigned to each foreign country of the Eastern Hemisphere an annual quota based on the national origins of the United States population in 1920. As a result, large quotas were assigned

[1]Consists of immigrant aliens as reported by the Immigration and Naturalization Service as adjusted by the Bureau of the Census for use in developing national population estimates. The categories of immigration listed are described in detail in later sections. They correspond to those in use by the Bureau of the Census and do not necessarily agree with the terminology used by the Immigration and Naturalization Service.

[2]Conditional entrants, as defined by the Bureau of the Census, include aliens who are permitted to enter the United States at the discretion of the Attorney General without an immigration visa.

*National Population Estimates and Projections Branch, Population Division. The views expressed are those of the authors, and do not necessarily reflect the views of the Bureau of the Census.

to countries of northern and western Europe, which had contributed most heavily to immigration during the settlement of America. The Act of 1965 abolished this system and substituted a limitation of 20,000 per country with an overall limitation of 170,000 for the Eastern Hemisphere. At the same time, a stronger preference was given to close relatives of United States citizens and resident aliens. Under the old law 50 percent of the quota was reserved for close relatives; under the new Act, 74 percent of the available numbers are so reserved. Another important change was the imposition of a limitation for the first time on immigration from the Western Hemisphere. The preference system does not apply to the 120,000 immigrants allotted to independent countries of the Western Hemisphere; immigration from these countries is essentially on a first-come, first-serve basis with no individual country limitation.

The Immigration Act of 1965 resulted in increased immigration in the latter part of the past decade. The regular provisions of the Act became fully operative on July 1, 1968, and in fiscal years 1968 and 1969 net alien immigration was 355,000 and 347,000, respectively. For the first half of the decade, the average was only 290,000.

The geographic distribution of immigrants by place of origin also changed as a result of provisions of the new act. Under the national origins quota system, northern Europe enjoyed a highly favored position in regard to the availability of immigrant visas, but the Act of October 3, 1965 placed each country in the Eastern Hemisphere on an equal basis in competition for visas. According to figures published by the Immigration and Naturalization Service, the proportion of immigrants from the countries of Southern Europe and Asia increased considerably in 1969 compared to the first half of the decade, as shown by the following table:

Continent of origin	In thousands: 1968-1969	In thousands: Average 1961-65	Percent of total: 1968-1969	Percent of total: 1961-1965
Total	359[a]	290	100.0	100.0
Europe	114	106	31.8	36.6
Southern[b]	70	27	19.6	9.5
Other	44	79	12.2	27.2
Asia	73	21	20.3	7.2
North America	138	129	38.5	44.4
Other	33	34	9.1	11.7

[a]Includes aliens adjusted to permanent resident status.

[b]Southern Europe includes Albania, Bulgaria, Greece, Italy, Portugal, Spain, Turkey, and Yugoslavia.

The shift in geographic origin also resulted in a change in the racial composition of immigrants in the 1960's compared to the previous decade. Estimates of net civilian immigration by race between 1950 and 1970 are as follows:

Race	Number (thousands): 1950-60	Number (thousands): 1960-70	Percent of total: 1950-60	Percent of total: 1960-70
Total	2,975	3,854	100.0	100.0
White	2,842	3,330	95.5	86.4
Negro and other races	132	524	4.4	13.6

The estimates of net alien immigration are based on data collected by the Immigration and Naturalization Service, and do not include any estimate of unrecorded immigration. Akers (1) discussed the problem of illegal entries across the Mexican and Canadian borders, and found no conclusive evidence about the possible volume of unrecorded immigration. During fiscal year 1969, of the 283,557 violators of the immigration laws who were located by the Immigration and Naturalization Service, 167,174 entered illegally at other than ports of inspection (2). Almost 97 percent of these were illegal entries of Mexican aliens across the border between the United States and Mexico. No estimates have been made of the number of aliens who entered illegally and were not apprehended later.

Alien immigration as defined by the Immigration and Nationality Act includes only aliens who intend to establish permanent residence in the United States. However, one sizeable group of alien immigrants has traditionally been exempt from the residency requirement of the Act. This group, called commuters, is composed of citizens of Canada and Mexico who have been admitted for permanent residence in the United States but choose to reside in Canada or Mexico and commute each day to places of employment in the United States (3). To become a commuter an alien must first obtain lawful immigrant status as prescribed by the Immigration and Nationality Act. Upon admission, the alien receives an alien registration receipt card, form I-151, commonly known as a "green card." The green card is renewable each year and can be used for admission to the United States for work each day. To retain his commuter status the alien must continue to be permanently employed in the United States. The crucial difference between commuters and other alien immigrants is that commuters are not required to establish permanent residence in the United States. Those commuters who never become residents of the United States should not be included in estimates of immigration for making national population estimates. However, no separate statistics are regularly collected on aliens who plan to become commuters. The Immigration and Naturalization Service takes periodic sample counts of the number of commuters. The latest sample counts, made in January 1967, produced the following data on commuters from Mexico and Canada:

Mexico	42,641
Canada	10,688
Total	53,329

These figures cannot be used to estimate the number of commuters included in the net alien immigration figures between 1960 and 1970 because there is no way of determining the number of commuters who were absent from work on the day of the check, the number of commuters who later gave up their commuter status to move to the United States, or the number of commuters who remain permanent residents of Mexico or Canada. A further difficulty is the absence of data showing the year immigration status was obtained.

Net arrivals from Puerto Rico

Since natives of Puerto Rico are citizens of the United States, no official action is required when a Puerto Rican moves from the island to the mainland or vice versa. In preparing estimates of the population of the United States, it is necessary to account for the net movement between the United States and Puerto Rico. Passenger statistics to and from Puerto Rico by sea and air are published monthly by the Puerto Rican Planning Board. The net movement indicated by these total passenger figures is assumed to be "net arrivals from Puerto Rico" to the United States. The net movement for the intercensal period by year was as follows:

July 1, 1969 to March 31, 1970	28,000
1968-1969	-7,000
1967-1968	19,000
1966-1967	34,000
1965-1966	30,000
1964-1965	11,000
1963-1964	4,000
1962-1963	5,000
1961-1962	11,000
1960-1961	14,000
April 1, 1960 to June 30, 1960	4,000
Total	152,000

The total for the decade is considerably lower than the total of 456,000 for the 1950-60 decade, but during the past ten years there does not appear to be a consistent trend.

These figures must be regarded as approximations, as there are problems involved in interpreting the data. The passenger statistics include all persons arriving and departing, and tourists far outnumber migrants permanently changing their residence. Presumably the arrival and departure of tourists cancels out, leaving a net figure which accurately reflects migration, but the gross movement is so large that a small bias in collecting the data could significantly affect the net value. In addition, while this net amount is assumed to represent immigration into the United States, the data include all movements to and from Puerto Rico. Although traffic directly between the United States and Puerto Rico dominates these statistics, significant passenger movements between Puerto Rico and other parts of the world are included. Probably the great bulk of this traffic represents tourist movement, but to some degree, Puerto Ricans changing residence must also be involved.

Further problems relate to the seasonal nature of tourist movement. Heavy in-migration occurs in the winter months followed by heavy out-migration. Fortunately, during the summer months the volume of movement is low so that comparisons between fiscal year totals are relatively unaffected by this factor.

The estimate of Puerto Rican immigration will be compared with the 1970 Census results for Puerto Rico and the United States, but the detailed statistics needed for this analysis will not be available until late in 1971.

Net arrivals of civilian citizens

This category of immigration takes into account only those citizens overseas employed by the Federal Government, their dependents, and the dependents of the Armed Forces overseas. On April 1, 1960 about 42,000 U.S. citizens were employed by the Federal Government in countries throughout the world. Dependents of these civilian Federal employees and dependents of Armed Forces personnel numbered about 550,000 for an overseas total of 593,000. By April 1, 1970, the total number had decreased to 445,000, a net change of 148,000.

Births and deaths to these citizens overseas must be taken into account before deriving an estimate of net migration. Death statistics are not available, and are not taken into account in the estimating procedure. The number is small since the average age of this group is low. Births on the other hand are a significant factor. Available birth statistics indicate about 298,000 births for the decade. These births, in combination with the decrease of 148,000 in civilians overseas, indicate a net immigration into the United States of 446,000 for the 10 year period. This is 166,000, or 59 percent, greater than the total of 280,000 recorded in the 1950's.

Annual figures for the 60's were as follows:

July 1, 1969 to March 31, 1970	69,000
1968-1969	34,000
1967-1968	29,000
1966-1967	43,000
1965-1966	62,000
1964-1965	34,000
1963-1964	48,000
1962-1963	51,000
1961-1962	14,000
1960-1961	53,000
April 1, 1960 to June 30, 1960	9,000
Total	446,000

The figure of 69,000 for the last nine months is based on an estimated decrease in the number of civilian citizens overseas. This estimate does not take account of data collected by the 1970 Census.

The estimating procedure ignores the possibility that employees leaving Federal employment overseas may remain overseas, and that citizens accepting federal employment may already be overseas. A further problem relates to the completeness of the statistics on births overseas. Only births occurring in Department of Defense hospitals overseas are taken into account. Incomplete figures show about 2,000 births to U.S. citizens in private hospitals overseas each year. The actual number is not known. If the estimate of births overseas were increased to take this error into account, the estimate of net arrivals of civilian citizens, and therefore net civilian immigration, would be increased by the same amount.

The incompleteness of the birth statistics results in an understatement of net immigration to the United States. An error in the opposite direction is caused by ignoring deaths to citizens overseas. Since complete statistics on births and deaths are not available, and since the numbers involved are apparently quite small, no adjustment has been introduced for these factors.

Conditional entrants

Cuban refugees, Chinese refugees from Hong Kong, and escapees from communist dominated countries of Eastern Europe added about 380,000 to intercensal immigration in the 1960's. Immigrants in these three categories are grouped as conditional entrants following the usage of the Bureau of the Census, although they have also been referred to as refugees, parolees, exiles, and escapees. Conditional entrants are aliens without official immigration status who are permitted to enter the United States at the discretion of the Attorney General. Although the admissions are on a temporary basis, one of the conditions of entry is that political conditions do not permit the return of conditional entrants to their former countries. The Law of November 2, 1966 permitted conditional entrants who had been continually present in the country for two years to adjust their status to permanent resident aliens. Conditional entrants are added to national population totals in the year of arrival rather than the year that official immigration status is obtained. Therefore, when statistics from the Immigration and Naturalization Service are used for making national population estimates, aliens who have adjusted status during the previous year are subtracted to avoid a double count.

Cubans fleeing the Castro regime after 1959 accounted for about 365,000 of the estimated 380,000 conditional entrants who entered the United States in the 1960's. Chinese refugees from Hong Kong and refugees from communist countries of Eastern Europe added about 15,000 to the total. Annual figures for conditional entrants during the decade were as follows:

July 1, 1969 to March 31, 1970	33,000
1968-1969	42,000
1967-1968	45,000
1966-1967	46,000
1965-1966	34,000
1964-1965	5,000
1963-1964	19,000
1962-1963	16,000
1961-1962	75,000
1960-1961	64,000
April 1, 1960 to June 30, 1960	-
Total	380,000

When Fidel Castro came to power in Cuba in January 1959, members of the Batista government fled to Florida by air and small craft. Within a few months, increasing numbers of Cubans disenchanted with the new government began to arrive on commercial airlines and ships. Many Cubans entered the United States as visitors, students, or nonimmigrant aliens and remained in the Miami area in the hope of returning to Cuba soon. However, relations between the U.S. and Cuba deteriorated rapidly, and diplomatic relations were broken on January 30, 1960. Immigrants from Cuba poured into the United States by any means possible including commerical airlines, small boats, and commandeered aircraft. The exodus increased to nearly 2,000 per week, and more than 150,000 Cubans entered the United States between July 1960 and October 1962.

What had been a flood of refugees was reduced to a trickle after President Kennedy announced the quarantine of Cuba on October 22, 1962. Commercial transportation between the U.S. and Cuba ceased and a few privately owned small boats became the only means of escape. Between December 1962 and the first half of 1963, the American Red Cross was allowed to transport 7,800 refugees to the United States including 1,100 prisoners of war from the "Bay of Pigs" invasion and many of the prisoners' relatives.

Only a small number of refugees were able to reach the United States in the three years after October 1962. However, in October 1965, President Johnson announced that Cubans seeking refuge in the United States would be welcomed, and the subsequent negotiations led to the establishment of a regularly scheduled airlift from Cuba to the United States. Between November 1965 and March 1970, about 189,000 Cubans arrived on the airlift. Since the inaugural flight on November 6, 1965, between 3,000 and 4,000 Cubans have flown to the United States each month, and use of the airlift has shown no signs of diminishing.

Emigration

No direct measures of the emigration of United States residents are currently available. Indirect evidence suggests that about 250,000 residents, both citizens and resident aliens, left the country permanently between 1960 and 1970. This estimate is based on data presented in statistical publications of various foreign countries and on data recently developed by the Social Security Administration showing the migration of OASDHI beneficiaries to and from the United States. The overseas movement of government employees, their dependents, and dependents of the Armed Forces overseas have been taken into account in the section on the net arrivals of civilian citizens. Immigration estimates presented in this section are independent of the four categories presented above.

More than half of the estimated 250,000 emigrants were United States citizens; the exact proportion could not be determined. Data on emigration of resident aliens, that is, the permanent return movement of immigrants lawfully admitted for permanent residence, have not been collected by the Immigration and Naturalization Service since 1957. The data provided by the Social Security Administration show the in- and out-migration of both citizens and aliens, most of whom are over age 65. The statistics for individual countries suggest that the movement is primarily by U.S. citizens, although not all countries provide a classification of migrants by citizenship.

Statistics obtained from Canadian officials indicate that approximately 158,000 U.S. residents, mostly citizens, have been admitted for permanent residence in Canada during the last 10 years. These figures include only persons who have convinced Canadian immigration officials that they plan to reside permanently in Canada. Students, visitors, and workers admitted temporarily are not included in the Canadian figures. The proportion of U.S. emigrants who actually remain permanently in Canada cannot be determined because there are no restrictions on the return of U.S. citizens to the United States, and no official action is required by either country in the event of such return. On the other hand, the same is true of the 432,000 residents of Canada who have entered the U.S. as immigrants during the same period so that to some extent, the two errors offset one another.

Data on the movement of U.S. residents to Israel, Australia, Italy, and Mexico during the last 10 years are less complete than those for Canada. According to figures provided by the Jewish Agency in New York, the estimated net movement of U.S. residents to Israel from 1960 to 1969 was about 15,000. More than half of this number emigrated to Israel after 1967. The number of U.S. residents who moved permanently to Australia is estimated to be about 12,000 (4). The amount of permanent movement to Australia from 1966 to 1969 was twice the amount from 1960 to 1963. The number of Italians who have returned to Italy after having previously migrated to the United States during the last 10 years is estimated to be about 5,000 according to Italian Embassy personnel in Washington. Statistics for Mexico indicate that about 1,000 U.S. residents move to that country each year, but these data may be seriously incomplete (5).

On the basis of statistics on the migration of OASDHI beneficiaries between the United States and all foreign countries during 1968 and 1969 and other data provided by the Social Security Administration, the net emigration of OASDHI beneficiaries was estimated to be about 75,000 for the decade. The majority were over 65 years of age. A number of these OASDHI beneficiaries have migrated to the individual countries discussed above. Therefore, to avoid duplication the estimated number of OASDHI beneficiaries emigrating to each of these countries was subtracted from the total movement to each of the countries mentioned above.

Based on the procedure outlined above, U.S. emigration for the intercensal period is estimated to be roughly 250,000. Estimates by year are as follows:

July 1, 1969 to March 31, 1970	28,000
1968-1969	33,000
1967-1968	29,000
1966-1967	26,000
1965-1966	24,000
1964-1965	23,000
1963-1964	23,000
1962-1963	22,000
1961-1962	20,000
1960-1961	18,000
April 1, 1960 to June 30, 1960	4,000
Total	250,000

The rather fragmentary data that are available for the individual countries indicate that the tempo of emigration increased considerably after 1965. The observed increase may be partly a function of unreliable data, and if real, may prove to be a temporary phenomenon. Much more complete and accurate data would be needed in order to measure the total emigration of U.S. citizens and resident aliens and to evaluate the causes and possible consequences of this movement.

Limitations of the estimates

The estimated net civilian immigration of 3.9 million for the decade should be regarded as an approximation rather than a precisely measured quantity. The largest category, net alien immigration, is the result of a careful registration and tabulation procedure. There are greater possibilities for error in the smaller categories, such as net arrivals from Puerto Rico.

Detailed tabulations from the 1970 Census will provide a powerful tool for strengthening the estimates of immigration in all categories, but better statistics on a current basis are needed in several fields. Data are needed showing the number of commuters who never become United States residents. If this is impossible to obtain, there should be, as a minimum, a careful periodic count of persons in this category along with figures showing the turnover involved. Better current data are needed for the civilian population overseas. Complete reporting of births to citizens overseas is also needed, and more frequent and systematic tabulations of federally affiliated civilians overseas are required. Finally, little is known about the emigration of citizens and aliens and about illegal immigration.

It is not possible to set limits to the possible variation of the immigration estimates from the "true" situation. Some of the missing data series tend to cause a high estimate of immigration, while some tend in the opposite direction. In each category the estimates have been put at the level which seems the most reasonable, given the data available. Perhaps the greatest need at present is for better information on emigration since this category seems to offer the largest potential for error.

References

(1) Akers, Donald S., Immigration Data and National Population Estimates for the United States, Demography, Volume 4, No. 1, 1967, pp. 262 to 272.

(2) Department of Justice, Annual Report of the Immigration and Naturalization Service, 1969, page 12.

(3) Commuters; Historical Background, Legal Challenges, and Issues, Report of the Select Commission on Western Hemisphere Immigration, January 1968, page 101.

(4) Commonwealth Bureau of Census and Statistics, Yearbook of Australia, annual publications, 1963 - 1969.

(5) Secretaria de Industria y Comercio, Anuario Estadístico de los Estados Unidos Mexicanos, 1969.

Academic and Socio-Economic Factors Related to Entrance and Retention at Two- and Four-Year Colleges in the Late 1960's

A. J. Jaffe and Walter Adams,*
Bureau of Applied Social Research
Columbia University

Introduction

Open Enrollment Plans and Programs

As we enter the 1970's, controversy mounts over open enrollment in American colleges, and especially so for selective 4-year ones. Most of the "target" students rank low on test distributions, and include few who followed college preparatory programs in high school. Earlier attempts to enroll such students in selective colleges were usually limited to relatively small groups. Today, however, pressures have increased for extended, or even universal, access to college.

One critical question concerns eventuations for less able students entering selective colleges, or even unselective ones. Considerable recent information is available on college entrance, type of college entered, and duration of attendance for students with various academic and socio-economic characteristics, including the types of students open enrollment programs are addressed to. Such hard data on survival rates for various groups of students at different types of colleges afford insights into probable outcomes of open enrollment programs, minus special assistance to the relevant students. From such data one could roughly gauge how essential such assistance would be (assuming its effectiveness) to assure the less promising entrant a reasonable chance for academic survival.

The Study Data

In Fall, 1965, the Census Bureau interviewed a national sample of high school seniors as part of its Current Population Survey. Information was obtained on post-high school plans, plus personal and background information about the student. The seniors were followed up in Fall 1968 to determine proportions entering college, proportions attending 2- and 4-year colleges, and proportions still enrolled in Fall 1968. By that date a Fall 1966 college entrant would normally be in the junior year.

It was also possible to determine relationships between post-high school behavior and personal and background variables. In this paper we discuss such relationships for the following: (a) the student's high school curriculum; (b) his estimate of his brightness relative to high school classmates; (c) average high school grade; (d) the college entrant's estimate of his brightness relative to college classmates; (e) average college grade; (f) family income; (g) occupation of household head; and (h) father's educational attainment.

The Analysis

We may ask a number of questions for the various student groups. First, is there any one variable with an especially strong relationship to post-high school behavior? If so, with what sorts of behavior is it most associated (college entrance *per se*, type of college entered, or continuation in college)? Second, which group of variables is most significantly related to post-high school behavior, the academic or the socio-economic ones? Are relationships about equally strong for all variables comprising each of these groups, or are there considerable differences between variables within each group? Next, can we determine whether it is the economic or the social component of socio-economic status that is the stronger determinant of post-high school behavior?

Overall, can we infer the likely outcome for less academically promising and/or lower socio-economic students who might enter college under open enrollment? Here we would base our inferences on one of two assumptions. First, large numbers of students who would not have qualified in former years might be admitted to formerly selective 4-year colleges. The alternative assumption is that admission standards might be maintained at most 4-year colleges, but that increased numbers of less promising high school graduates would be encouraged and assisted to enter 2-year colleges.

Our main emphasis is on college dropout, since it is the survival rate that would largely determine success or failure of open enrollment. Nevertheless, we shall also consider college entrance, since it determines the cast of players who struggle for survival on the academic stage. We are further interested in variables which not only predict the likelihood of college entrance, but also the likelihood of survival at college. Should there be such variables, open enrollment might dissipate the relationship to entrance by simple fiat, but fail to weaken the relationship to dropout or continuation--assuming no radical changes were made in the colleges themselves, such as introduction of less demanding curricula. We now turn to the findings, and refer the reader to appended Tables 1 and 2.

---

*This article reports upon findings from a larger study in progress (A. J. Jaffe, Director), supported by the U. S. Office of Education. The interviews were conducted by the U. S. Census Bureau and the tabulations and analysis were made at the Bureau of Applied Social Research.

Findings

The High School Curriculum

Of the eight student characteristics in our analysis, the high school curriculum is of overriding significance, not only for entering college, but also for type of college entered (2-year or 4-year) and for survival at each type of school.

College entrance. By Fall 1968 about 53% of the high school graduates had entered college, one-third enrolling initially in 2-year and two-thirds in 4-year ones. A few students transferred from one type of college to another. About three-quarters of the students ended up in 4-year colleges, or attended such schools only. In this analysis we classify the transfers by the type of colleges they finally reached, since these are the schools most relevant to dropout or continuation.

Overall, about 17 in 20 college preparatory students entered college, but only about 5 in 20 students who had followed other programs. About 17 in 20 students at senior colleges were college preparatory students. Nevertheless, this was the case for over 10 in 20 students in 2-year schools. In sum, college preparatory students predominate among all college entrants. If open enrollment were to eliminate the high school program as a criterion for college entrance we might well expect significantly larger proportions of non-college preparatory entrants. We may well ask about the probable effect on college dropout.

College dropout. Overall, about a third of all the college entrants had dropped from college by Fall 1968. Additional students will drop, of course, subsequent to Fall 1968. About 5 in 20 preparatory entrants had dropped, whereas this was the case for 12 in 20 entrants who had followed other programs. Dropout was nearly three times as frequent at 2-year as at 4-year colleges, respectively representing 13 in 20 versus less than 5 in 20 students. But it is also true that at 2-year and 4-year colleges alike the high school program was a strong determinant of dropout. Under 4 in 20 orelege preparatory 4-year students dropped, compared with nearly 9 in 20 non-preparatory ones. About 11 in 20 college preparatory 2-year students dropped, compared with 15 in 20 non-preparatory ones.

Implications of the above findings for open enrollment. If substantially larger proportions of non-preparatory students were to enroll in 4-year colleges under open enrollment we would expect nearly half of these students to drop within several years, basing our expected rate upon the observed one. The rate might be even higher if non-preparatory students who now enroll in 2-year colleges were to select 4-year ones instead, once the admissions criteria were relaxed. We would not expect them to fare as well at academically more demanding schools. In short,we would expect any large increase in non-preparatory 4-year enrollments to raise the dropout rate sharply, with whatever effects on the academic climates of the colleges. For the dropouts, benefits from a brief stay would seem extremely doubtful.

If substantially larger proportions of non-preparatory students were to enter 2-year colleges under open enrollment, we would simply expect about 3 in 4 to drop without transferring or completing the 2-year program. The net effect would be to raise the current 64% dropout rate at these schools to appreciably more than 2 in 3 of all entrants.

Other Variables and Junior College Dropout

We found the very high dropout rate for non-college preparatory students at 2-year colleges rather surprising. Presumably these colleges are designed to salvage this type of student. We therefore asked whether non-preparatory 2-year dropouts possess other unfavorable academic characteristics which might lead to dropout--or does dropout relate to financial or other components of socio-economic status? Very briefly,none of the seven other student characteristics tabulated, whether academic or socio-economic, had a statistically significant relationship to 2-year dropout. In contract, the high school curriculum was associated with 2-year college dropout at the .01 (chi square) probability level--a very strong relationship, given the small size of the sample. We infer, then, that high school curricular choice, though considerably related to ability and class, nevertheless reflects underlying personal attributes and predispositions, quite apart from income, class, and ability.

Grade and Academic Self-image Relative to Classmates in High School and College

Overall, grades in high school and college and the student's estimate of his brightness relative to high school and college classmates are strongly related to all aspects of post-high school behavior--college entrance, type of college entered, and continuation at college--with the exception of continuation at 2-year colleges. The student with poor marks at a 2-year college, for example, is about as likely to remain enrolled as a student with good marks.

College entrance. The grade and self-image variables are far less related to college entrance, however, than is the high school curriculum. College preparatory students are about three and a half times as likely to enter college as non-preparatory ones. Students who think they are brighter than average in high school are about twice as likely as the pessimists to enter college. Students with better high school grades are about a third again as likely to enter college as students with poorer grades.

Type of college entered and 4-year college dropout. For type of college attended,however,and for the 4-year college continuation rate, average high school grade and academic self-image relative to high school classmates present roughly equal degrees of association.

Implications of the above findings. Many of the students for whom open enrollment programs are designed receive poor grades in high school. Most such students would have low self-images as well. Of all college entrants in the sample, 8 in 10 with better grades had better self-images and 7 in 10 with poorer grades had poorer self-images. Since self-image is far more related to college entrance than grade, it would appear that liberalized college access for students with poorer grades might yield only modest increases in proportions entering. Major increases would probably stem from efforts to raise the confidence level of academically poorer high school graduates, whether by high school counselling, encouragement from colleges to apply, media emphasis on the availability of college, or other measures.

For those enrolled in a 4-year college both variables are about equally related to continuation. This would seem reasonable, since the college student is under pressure to earn reasonably adequate marks. Open enrollment, accordingly, though it might induce students with poor marks and/or poor self-images to enter college, might well find that large proportions of such students would drop. We argue as follows:

About half of the students had consistent grades and self-images in high school and college. Of this half, those with poorer grades and/or self-images were particularly likely to drop from 4-year schools. For the other half of the entrants, grades and/or self-images changed between high school and college, and the strong tendency was for depreciation for both variables--but especially so for self-image. Depreciation was strongly associated with dropout for both variables. Of all grade changes 63% were unfavorable, but of all self-image changes 86% were unfavorable. Presumably a student faces stronger competition, by and large, in college than in high school. Though some might manage to maintain acceptable grades by working harder, it would seem plausible that academic self-esteem would nevertheless suffer.

At 2-year colleges the lack of relevance of either high school or college grades or self-images to dropout is itself not too relevant, given the extent of dropout at these schools, for whatever reasons.

Socio-economic Variables

College entrance. For college entrance all three socio-economic variables have strong and roughly equal relationships, but this relationship is not nearly as strong as for the high school curriculum, or even academic self-image in high school. For example, the difference in college entrance between students with family incomes of $7,500 and over and those with under $7,500 is 25 percentage points, whereas it is 38 percentage points between students with better and poorer self-images in high school, and 60 percentage points between college preparatory and non-preparatory students. Clearly the academic variables are more significant than the economic one for college entrance.

Further findings for family income. Apart from college entrance, family income has no apparent relationship to type of college attended, and no statistically significant relationship to retention at either 2- or 4-year colleges. It would seem that after the initial sorting of entrants and non-entrants by family income, most of those who do enter are able to finance college, and dropout or continuation are determined by non-financial factors. Apparently, then, continuation of low income college entrants, the likely entrants under open enrollment, would not be seriously threatened by inability to meet expenses, nor would increased student aid assure such continuation.

Occupation of head of the student's household and educational attainment of the student's father. It is important to note, however, that non-financial aspects of socio-economic status not only have statistically significant, and indeed strong, relationships to college entrance rates, but also to the dropout rate at 4-year colleges.

The occupation of the head of the student's household predicts college entrance about as well as family income. Roughly 7 in 10 white collar high school graduates entered college, whereas this was the case for only 4 in 10 graduates from blue collar backgrounds. For both occupation of head and educational attainment of father there are statistically significant relationships both to type of college attended and to continuation or dropout at 4-year colleges, in sharp contrast to the absence of such relationships for family income.

In sum, financial and non-financial indicators of socio-economic status are about equally and positively related to entering college, but only the non-financial elements have a significant positive relationship to attending a 4-year rather than a 2-year school, and to continuation in a senior college.

Implications of the above findings. Very simply, it would seem that financial liberalization of access to college might lead to some increase in proportions enrolling, though we doubt if the increase would be very large. For the less affluent entrant, however, increased student aid would not be likely to reduce the dropout rate appreciably. It would seem that it is the non-financial aspect of socio-economic status, including the educational tradition handed on from father to son, that most relates to continuation or dropout. To what extent open enrollment programs could develop effective means of countering the relationship between lower social class and a high incidence of college dropout we do not know, but it is nevertheless clear that the problem is a more complex one than if lack of money were the principal socio-economic determinant. Once again, assuring continuation in college appears appreciably more difficult than inducing more high school graduates to enter.

Policy Implications

Educators appear to have considerable awareness of the pivotal role of curricular choice for college entrance and for the type of college attended, but less awareness of its relationship to continuation or dropout from college. Nor do they seem to be aware of our inferential finding that it is not only the curriculum *per se* that determines post-high school behavior, but also, and perhaps more significantly, less understood and enduring social and psychological correlates of the curricular decision in the student's early teens.

It appears that advocates of limited access to senior colleges, but open enrollment at junior ones, are overly optimistic in their presumed belief that large proportions of non-college preparatory type students would survive at 2-year schools. If these students were admitted to more rigorous 4-year colleges they would be even less likely to survive. Furthermore, since at 2-year colleges academic performance and academic self-image bear virtually no relationship to retention or dropout, there is no particular reason to believe that supplementary or remedial academic assistance would appreciably reduce the dropout rate. There is more reason to believe that dropout might be reduced at 4-year colleges by academic assistance, since 4-year college retention appears to have a significant relationship to academic variables other than the high school curriculum.

In sum, we suggest that advocates of open enrollment, whether for 2-year or 4-year colleges, or both, must face the no doubt unpleasant possibility that the college careers of many of the target students would be brief. If open enrollment were to proceed rapidly without a full appraisal of this possibility, and without effective measures to assure continuation,disappointment and frustration consequent to widespread dropout might outweigh benefits for the students who would survive.

We suggest the advisability of open enrollment programs of limited size for the years immediately ahead, with extensive evaluation an essential feature of their design, and with options retained to modify such pilot programs, or even abandon them, should they fail to realize their objectives. Once widely adopted, an educational innovation such as open enrollment would be difficult to reverse, irrespective of its outcome for students and their colleges. Once conferred, it is politically awkward to withdraw egalitarian measures, which tend to be viewed almost immediately as inalienable rights.

We are not primarily concerned about more traditional open enrollment programs, such as the system which has pertained to California public higher education for many years. In that state all high school graduates have access to some sort of higher education, but the type of institution a student is eligible to enter depends upon his academic performance in high school. Rigorous academic requirements at the state university branches, and moderately rigorous ones at the state colleges, preserve the academic quality of these institutions, and assure that only students reasonably likely to survive enter them. Less able and/or less well prepared students enter public 2-year colleges, and if they do well have the option of transferring to senior ones. Even if the 2-year entrant drops after a couple of semesters, there is the strong possibility that he may benefit, occupationally and otherwise, from a program more commensurate with his academic capacity. There is considerable evidence that even limited exposure to 2-year terminal vocational courses represents an advantage for entry into a desirable job.

What we are most concerned about is the more recent version of open enrollment, whose advocates oppose the maintenance of elite, or even moderately selective, public colleges or universities. Irrespective of academic performance,it is claimed that all high school graduates should have access to the public college of first choice, however rigorous the program and unlikely the entrant's survival. For open enrollment of this sort, or approximations thereto, we foresee the waste, frustration, and demoralization consequent to frequent dropout, or else (and this we feel may be the more likely possibility) adulteration of the academic offering so as to accomodate the "high risk" student. If the latter were to occur, the damage would be sustained by the better students seeking rigorous training at formerly elite institutions, and by society in general, since it would lose the full measure of the able student's potential services. It is such open enrollment in all colleges, 4-year as well as 2-year, and academically rigorous as well as less rigorous, which we feel should be approached with extreme caution and subjected to scrupulous evaluation.

We conclude that the egalitarian impulse alone, however admirable, is insufficient justification for radical change in admission to higher education, and that open enrollment should stand or fall on the basis of demonstrable effects upon college and students.

Table 1. Post-high school behavior of 1966 high school graduates, as of Fall 1968, by academic characteristics of graduates

Post high-school behavior

Academic characteristics	All HS graduates No.	%	All coll entrants %	All coll non-entrants %	All coll entrants No.	%	4-yr coll* %	2-yr coll** %	All coll entrants No.	%	Contd %	Dropped %	All 4-yr entrants* No.	%	Contd %	Dropped %	All 2-yr entrants** No.	%	Contd %	Dropped %
High school curriculum:																				
College prep.	599	100	84	16	501	100	81	19	501	100	74	26	405	100	81	19	96	100	45	55
All other	637	100	24	76	153	100	50	50	153	100	41	59	77	100	57	43	76	100	25	75
All curr.	1236	100	53	47	654	100	74	26	654	100	67	33	482	100	77	23	172	100	36	64
Chi square		p =<.001				p =<.001				p =<.001				p =<.001				p =<.01		
Relative brightness in HS:																				
Above average	505	100	75	25	380	100	83	17	380	100	76	24	317	100	84	16	63	100	38	62
Av. or below	741	100	37	63	277	100	61	39	273	100	55	45	169	100	66	34	108	100	35	65
All levels	1246	100	53	47	657	100	74	26	653	100	67	33	486	100	77	23	171	100	36	64
Chi square		p =<.001				p =<.001				p =<.001				p =<.001			Not significant, no clear direction			
Average grade in HS:																				
A or B	546	100	62	38	340	100	87	13	340	100	77	23	296	100	84	16	44	100	34	66
Less than B	701	100	46	54	319	100	60	40	319	100	55	45	191	100	68	32	128	100	37	63
All grades	1247	100	53	47	659	100	74	26	659	100	67	33	487	100	77	23	172	100	36	64
Chi square		p =<.001				p =<.001				p =<.001				p =<.001			Not significant, no clear direction			
Relative brightness in college:																				
Above average					224	100	81	19	224	100	80	20	182	100	89	11	42	100	40	60
Av. or below	Not applicable				431	100	70	30	431	100	60	40	303	100	71	29	128	100	35	65
All levels					655	100	74	26	655	100	67	33	485	100	78	22	170	100	36	64
Chi square						p =<.01				p =<.001				p =<.001			Not significant, no clear direction			
Average grade in college:																				
A or B					284	100	81	19	284	100	80	20	229	100	89	11	55	100	40	60
Less than B	Not applicable				362	100	70	30	362	100	58	42	252	100	68	32	110	100	36	64
All grades					646	100	74	26	646	100	68	32	481	100	78	22	165	100	38	62
Chi square						p =<.01				p =<.001				p =<.001			Not significant, no clear direction			

*Includes students who attended a 4-year college only and those who transferred to a 4-year college from a 2-year one.
**Includes students who attended a 2-year college only and those who transferred to a 2-year college from a 4-year one.

Table 2. Post-high school behavior of 1966 high school graduates, as of Fall 1968, by socio-economic characteristics of graduates

Post-high school behavior

Socio-economic characteristics	All HS graduates No.	%	All coll entrants %	All coll non-entrants %	All coll entrants No.	%	4-yr coll* %	2-yr coll** %	All coll entrants No.	%	Contd %	Dropped %	All 4-yr entrants* No.	%	Contd %	Dropped %	All 2-yr entrants** No.	%	Contd %	Dropped %
Respondent's family income:																				
$7500 & over	547	100	65	35	358	100	75	25	358	100	69	31	268	100	79	21	90	100	39	61
Under $7500	615	100	40	60	246	100	73	27	246	100	60	40	179	100	73	27	67	100	25	75
All incomes	1162	100	52	48	604	100	74	26	604	100	66	34	447	100	75	25	157	100	33	67
Chi square	p =<.001				Not significant,no clear direction				p =<.05				Not significant, in expected direction				Almost .05, in expected direction			
Occupation of head of household:																				
White collar	495	100	69	31	340	100	81	19	340	100	75	25	274	100	83	17	66	100	42	58
All other	746	100	42	58	314	100	66	34	314	100	57	43	208	100	70	30	106	100	32	68
All occupations	1241	100	53	47	654	100	74	26	654	100	67	33	482	100	77	23	172	100	36	64
Chi square	p =<.001				p =<.001				p =<.001				p =<.001				Not significant, in expected direction			
Father's education:																				
College grad.	147	100	86	14	125	100	86	14	125	100	82	18	108	100	86	14	17	100	59	41
Some college	134	100	72	28	97	100	69	31	97	100	63	37	67	100	72	28	30	100	43	57
H.S.grad.	338	100	59	41	199	100	74	26	199	100	64	36	147	100	76	24	52	100	29	71
Less than H.S.grad.	507	100	35	65	178	100	66	34	178	100	56	44	118	100	71	29	60	100	28	72
All levels	1126	100	53	47	599	100	74	26	599	100	65	35	440	100	77	23	159	100	35	65
Chi square	p =<.001				p =<.001				p =<.001				p =<.05				Not significant, in expected direction			

*Includes students who attended a 4-year college only and those who transferred to a 4-year college from a 2-year one.
**Includes students who attended a 2-year college only and those who transferred to a 2-year college from a 4-year one.

PROJECTIONS OF 1969 INCOME SIZE DISTRIBUTION FOR FAMILIES AND UNRELATED INDIVIDUALS COMBINED FOR STATES AND SELECTED SMSA's

Joseph J. Knott and Mitsuo Ono, U.S. Bureau of the Census*

Introduction

The demand for information on levels and distribution of income of demographic units in "smaller" geographic areas, such as Standard Metropolitan Statistical Areas (SMSA) and counties, has increased significantly in the past few years. Although the 1970 Population Census will provide such data for income year 1969, there is still a need for current information each year after the census. The purpose of this paper is to outline an estimation procedure that can be used to develop projected income size distribution data for these areas for consumer units (families and unrelated individuals combined). Projections of 1969 income size distributions for states and selected SMSA's were computed by this estimation method. Also included are some 1967 data for analytical purposes.

The data presented herein are considered as experimental information and do not represent official estimates. We plan, at a later time, to compare these projected data against official estimates and to analyze the differences.

This document is divided into four parts. The first part outlines the simple projection procedure used to obtain data on income size distributions covering income year 1969 for states and selected SMSA's as shown in table 1. The second part analyzes these findings. The third part compares these estimates with aggregate income obtained from various sources. The last part briefly presents a summary and direction for further research.

Derivation of Projected Income Size Distribution Data

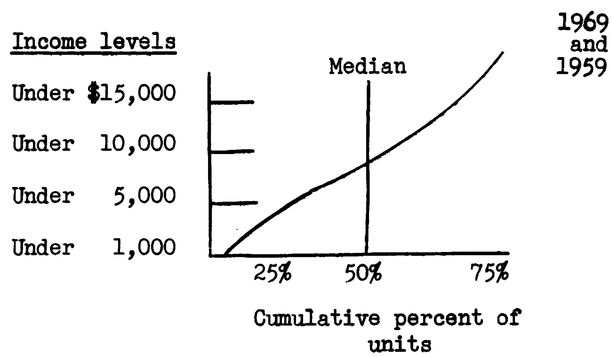
Table 1 presents projections of median income and income distribution for all states and a few selected SMSA's for 1969. These estimates were based upon a "naive" projection procedure, but it appears to give reasonable results. The key idea behind this procedure is that any cumulative income (lognormal) distribution can be described by two parameters, i.e., the median value (or the "positional" parameter) and the overall variance of the distribution (or the "shape of curve" parameter).1/ Thus, any change in a distribution over two points in time can be classified under the following:

	Parameters	
Case	"Position"	"Shape of curve"
1	Same	Same
2	Same	Change
3	Change	Same
4	Change	Change

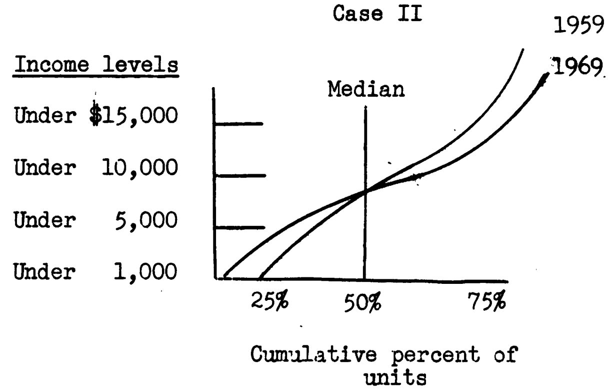
In turn, each of these parameters can be made a function of certain socioeconomic variables at the microeconomic level, e.g., the unemployment rate, occupational and educational mix of family heads, the propensity to work of wives and other family members, the age mix of employed family heads, etc. 2/ Also, for a given area, variables can be regrouped under the "internal" and "external" effects. 3/ These effects are still under study and are to be the subject of a future paper.

In Case I as shown below the median income level and the relative shapes of the income distribution curve are assumed to remain constant between the two points in time, e.g., 1959 and 1969. This type of stability can result from compensating positive and negative factors. In Case II, the medium income level remains constant but the distribution of income changes. In this example, the 1969 distribution is assumed to be more equal than the 1959 distribution. The decrease in the overall "slope" of the 1969 distribution compared with the 1959 distribution reflects a smaller variance of the 1969 distribution than the 1959 distribution. In Case III, the "shape" or variance of the distribution remains the same over the period but the median income level increases between the two points in time. (Case I may be considered a Case III with zero growth.) In Case IV, both the median income level and the overall "slope" had increased over the two points in time.

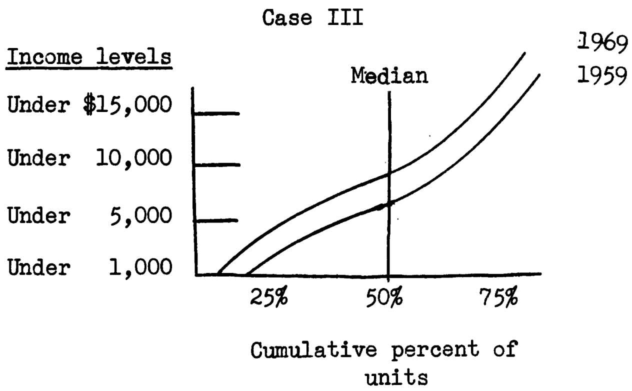
Case I

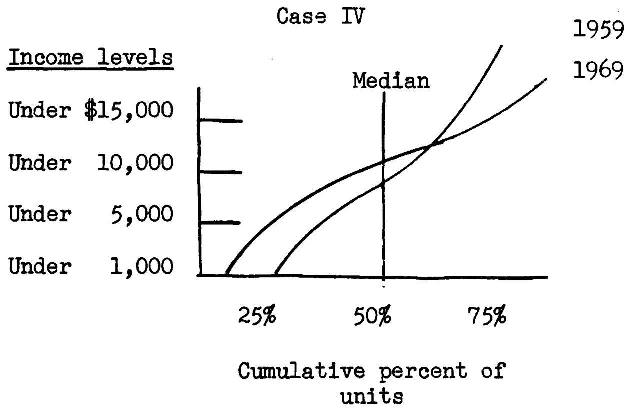


Case II



* The views expressed herein are not necessarily those of the U.S. Bureau of the Census. We wish to acknowledge suggestions by Dr. Murray S. Weitzman, Chief of the Economic Statistics Programs, Population Division, and staff members of the Consumer Income Statistics Branch.





Two basic sources of statistical data were used to analyze income distributions over time. These were (1) income tabulations obtained from the Current Population Survey conducted by the Bureau of the Census and (2) statistical tabulations of adjusted gross income data from the Internal Revenue Service.

Findings from these analyses show that for smaller population areas, e.g., counties, Case IV is the more typical one. For larger population areas, such as for states and for metropolitan areas not experiencing large population structural shifts, changes in income size distribution tend to follow Case III where the "shape" of the income distribution curve remains fairly constant but median income levels increase over time. This suggests that projections of income distributions for many areas can be made using one parameter (changes in median family income) instead of two. Under these conditions, the problem resolves itself in finding the most "efficient" carry-forward of median family income and assuming that the "shape" of the curve remained fairly constant over time.

If it is assumed that the "shape" of the income size distribution itself does not change, what is needed then is some rate of increase which is assumed to be constant over the entire distribution. One method of computing this rate of increase involves the following formula:

$$\text{Projection factor} = \frac{\dfrac{PI_{69}}{TR_{70}}}{\dfrac{PI_{59}}{TR_{60}}} \times 100$$

PI_t = OBE Personal Income for year t,

TR_t = Total Resident Population at time t.

The above rate can then be applied to income size class limits resulting in a projected distribution.4/ The conventional income size distribution classes can then be obtained by interpolation, assuming a linear distribution of units in each income class interval.

The assumptions involved in this simple projection procedure are: (1) The income level of all units change at the same rate, (2) neutrality of the internal and external effects on the shape of the distribution, and (3) the rate of change in income level of all units is equal to the rate of change in per capita personal income. This procedure was used to develop the data shown in table 1.

Selected Analysis of Findings

In this second part, we attempt to evaluate the projections presented in table 1 by comparing them against national and regional data obtained from the Current Population Survey (CPS). Shown on Chart I are income distributions of consumer units (families and unrelated individuals) obtained from the 1960 census, 1960 CPS (both covering income year 1959), the March 1970 CPS, and the projected 1970 census income distribution for the United States (both covering income year 1969).

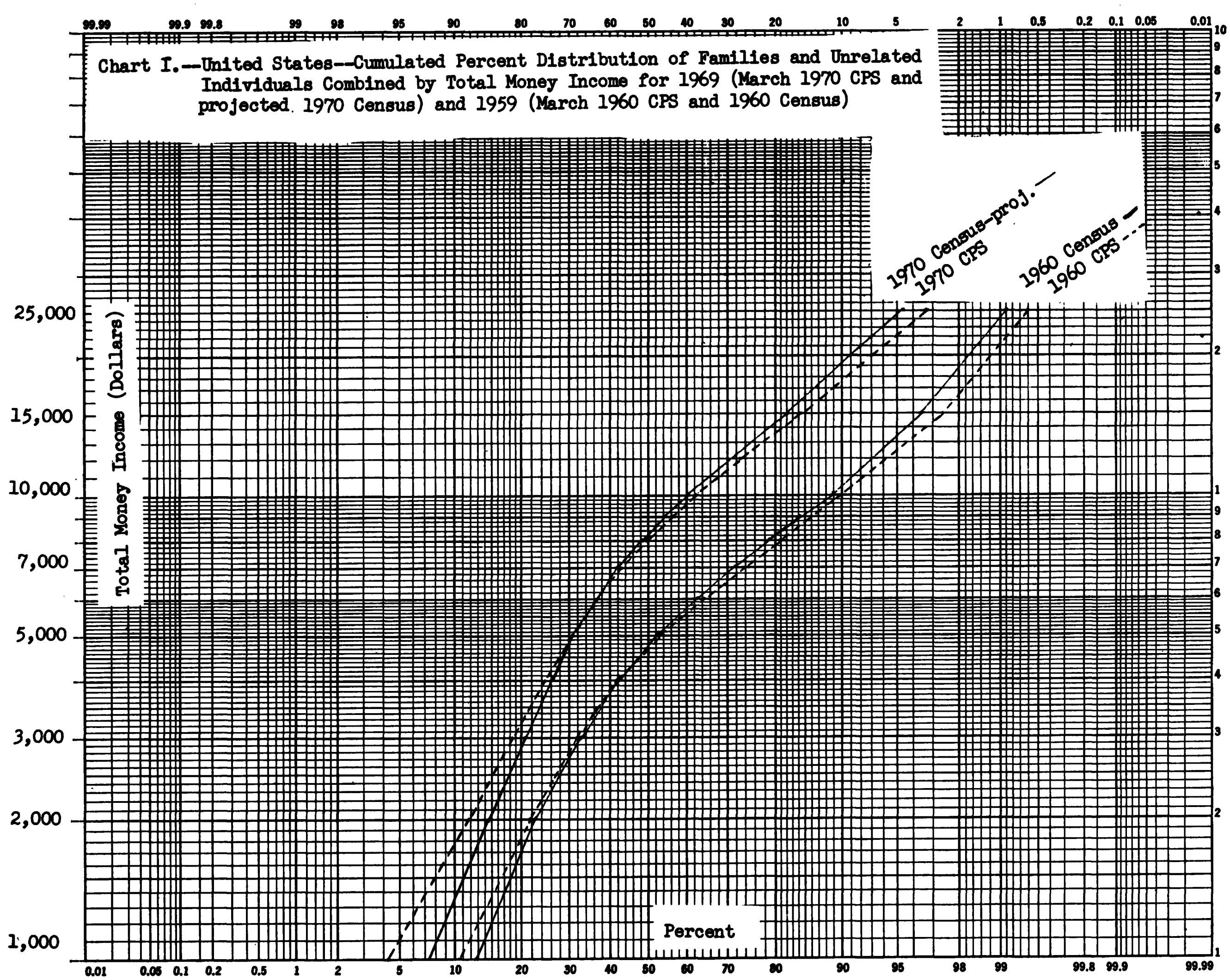
In table A below, the March 1970 CPS shows a smaller percent of families and unrelated individuals with income under $3,000 and shows a greater percent between $3,000 - $15,000 than the projected 1970 census data. Data indicate that the projected data tend to overstate, somewhat, the percentage of units at the extremes.

In summary, the projections appear to be reasonable as compared with CPS data for the United States. Also, projected data are found to be consistent with regional CPS income size distribution data. We consider the income data obtained from the CPS to be a good approximation of what the actual census will show.

Table A.--UNITED STATES--FAMILIES AND UNRELATED INDIVIDUALS COMBINED BY TOTAL MONEY INCOME FOR 1969 (MARCH 1970 CPS AND PROJECTED 1970 CENSUS) AND 1959 (MARCH 1960 CPS AND 1960 CENSUS)

Income class	1969			1959		
	March 1970 CPS	Projected 1970 Census	Difference (CPS-Census)	March 1960 CPS	1960 Census	Difference (CPS-Census)
Total................	100.0	100.0		100.0	100.0	
Under $3,000...............	18.3	20.4	-2.1	31.5	32.5	-1.0
$3,000 - $4,999............	12.2	10.9	1.3	21.1	19.7	1.4
$5,000 - $6,999............	12.4	10.9	1.5	21.1	19.6	1.5
$7,000 - $9,999............	19.2	18.1	1.1	16.0	16.2	-.2
$10,000 - $14,999..........	22.0	21.9	0.1	7.6	8.4	-.8
$15,000 and over...........	15.6	17.9	-2.3	2.6	3.7	-1.1
Median................	$8,170	$8,317	-$147	$4,759	$4,791	-$32

Chart I.--United States--Cumulated Percent Distribution of Families and Unrelated Individuals Combined by Total Money Income for 1969 (March 1970 CPS and projected 1970 Census) and 1959 (March 1960 CPS and 1960 Census)



Reconciliation of Aggregate Income Obtained From Income Distribution Data

In the third part, we compare aggregate total money income computed from the 1960 census and projected 1967 and 1969 data with two independent sources of aggregate income: Adjusted gross income (AGI) and personal income. 5/ If the projected income size distributions for each state are reasonable approximations of the actual distribution, then the 1969 ratio of census aggregate **total money income** to the independent aggregate income source should be close to the 1959 ratio. This type of analysis can be used to identify areas for which the simple projection procedure used in this paper would not be appropriate and alternate projection methods would be necessary.

Table B below shows differences in the ratios of aggregate total money income to adjusted gross income (AGI). For 46 of the states, the absolute value of the difference is less than .10.

Table B.--DISTRIBUTION OF THE DIFFERENCES IN THE RATIOS OF AGGREGATE TOTAL MONEY INCOME TO AGGREGATE TOTAL ADJUSTED GROSS INCOME FROM 1959 TO 1967 AND 1969 INCOME FOR THE 50 STATES AND THE DISTRICT OF COLUMBIA 1/

Difference in ratios	1959 to 1967		1959 to 1969	
	Number	Percent	Number	Percent
Total........	51	100.0	51	100.0
Less than .02.....	16	31.4	16	31.4
.02 to .05........	23	45.0	19	37.3
.05 to .10........	7	13.7	7	13.7
.10 and over......	5	9.8	9	17.6

1/ The 1969 adjusted gross income data were estimated by increasing the 1968 AGI by the average annual increase from 1959 to 1968.
Source: Unpublished tabulation.

The difference in the ratios of aggregate total money income to personal income between 1959 and 1969 are presented below in table C. Overall, data show that the 1969 ratio of census money income to total personal income remains fairly similar to the 1959 ratio.

Table C.--DISTRIBUTION OF THE DIFFERENCES IN THE RATIOS OF AGGREGATE TOTAL MONEY INCOME TO PERSONAL INCOME BETWEEN 1959 AND 1969 FOR THE 50 STATES AND THE DISTRICT OF COLUMBIA

Difference in ratios	Number	Percent
Total.............	51	100.0
Less than .01...........	16	31.4
.01 to .02..............	12	23.5
.02 to .03..............	10	19.6
.03 to .05..............	8	15.7
.05 and over............	5	9.8

Source: Unpublished tabulation.

Summary and Direction for Future Study

Empirical projections of income size distribution for small areas involve analysis of the complex interaction of many institutional changes. In trying to probe for an empirical model by which these projections can be made, we classified changes in the form of income size distributions into four basic types. These models cover changes essentially in two parameters: (1) The median income level (the positional parameter) and (2) the "shape" of the curve (the variance parameter). Empirical evidence, however, shows that among the four models, only two are typically found. Thus, empirical data for large areas show that, as a rule only the positional parameter tends to change while the "shape of the curve" parameter remains fairly constant. Using this finding, projections of income distribution were developed and tested against independent sources. This comparison showed that, overall, the projections appear reasonable. However, the ultimate test is to compare them against actual census results. This will be done at a later date. Also, more work is planned to determine what changes in socioeconomic variables are associated with changes in the two parameters noted above.

FOOTNOTES

1/ A theoretical model can be shown simply as follows:

Income distribution = f(M,V)
Where M = Median income
V = Variance

2/ For example, see "State Differentials in Income Concentration" by Ahmad Al-Samarrie and Herman P. Miller in The American Economic Review, March 1967.

3/ The "internal" effect relates to changes in the income distribution resulting from the income upgrading or downgrading of the population within an area, assuming no changes in income distribution due to migrants. The "external" effect relates to changes in income distribution due to migrants only. These two effects interact with each other in generating different types of income size distribution curves for small areas at different points in time.

4/ Other ways of projecting median family income are:

a. Imputing the growth rate of median income of a region to its subareas.
b. Imputing a growth rate of median income based upon the average growth rate over some past period.

See technical appendix for the projection procedure used. For a graphic projection technique, see "A Graphic Technique for Projecting Family Income Size Distribution" by Mitsuo Ono, Proceedings of Social Statistics Section, American Statistical Association, 1969.

5/ The comparison of the aggregate total money income with personal income is not strictly independent for this particular set of projections because we used per capita personal income to derive the projection factor for these projections.

Table 1.--PROJECTED 1/ 1969 INCOME SIZE DISTRIBUTION FOR FAMILIES AND UNRELATED INDIVIDUALS

Region, State (including D.C.), and SMSA	Base 2/ (000)	Total	Under $3,000	$3,000 to $4,999	$5,000 to $6,999	$7,000 to $9,999	$10,000 to $14,999	$15,000 and over	Median 3/ (dollars)	Mean 4/ (dollars)
United States	65,431	100.0	20.4	10.9	10.9	18.1	21.9	17.9	8,317	10,072
Northeast	16,198	100.0	16.5	9.4	10.9	19.8	23.9	19.6	9,062	10,753
North Central	18,209	100.0	19.0	9.9	10.1	18.6	23.6	18.6	8,823	10,285
South	19,314	100.0	25.5	13.2	12.0	16.1	18.5	14.8	6,878	9,198
West	11,821	100.0	18.8	10.4	10.9	18.7	23.0	18.3	8,625	10,115
Northeast										
Maine	323	100.0	21.9	13.5	14.7	20.6	19.3	9.8	6,962	8,150
New Hampshire	245	100.0	19.2	10.7	13.1	21.7	22.5	12.8	7,963	9,066
Vermont	145	100.0	22.5	11.6	12.4	18.9	21.2	13.5	7,557	9,207
Massachusetts	1,880	100.0	16.7	9.4	10.7	19.8	24.7	18.8	9,068	10,579
Rhode Island	317	100.0	19.9	11.0	12.2	20.2	22.6	14.2	8,059	9,415
Connecticut	985	100.0	13.3	7.3	9.0	19.1	26.9	24.4	10,254	12,113
New York	6,217	100.0	16.2	9.4	10.8	18.9	23.4	21.1	9,181	11,152
New Jersey	2,318	100.0	14.1	8.0	9.8	19.3	25.9	22.8	9,822	11,471
Pennsylvania	3,768	100.0	17.7	9.8	11.6	21.5	23.1	16.3	8,542	9,927
North Central										
Ohio	3,371	100.0	17.3	8.7	9.7	19.8	25.6	19.0	9,241	10,474
Indiana	1,658	100.0	18.3	9.4	10.4	19.3	24.3	18.3	8,902	10,266
Illinois	3,688	100.0	17.2	8.7	9.6	17.9	24.5	22.2	9,480	11,133
Michigan	2,734	100.0	16.6	8.4	8.3	18.5	25.9	22.4	9,756	11,299
Wisconsin	1,391	100.0	18.7	10.1	10.7	20.3	24.4	15.7	8,610	9,682
Minnesota	1,208	100.0	20.3	11.0	10.9	18.0	22.9	16.8	8,313	9,914
Iowa	925	100.0	22.3	12.0	11.6	18.1	21.3	14.9	7,704	9,441
Missouri	1,594	100.0	25.4	12.7	11.7	18.3	19.1	12.9	7,029	8,583
North Dakota	188	100.0	21.2	12.7	12.2	18.0	20.2	15.6	7,613	9,575
South Dakota	210	100.0	24.0	12.8	11.4	15.9	20.0	15.9	7,320	9,440
Nebraska	491	100.0	21.0	13.0	12.8	18.3	20.3	14.5	7,479	9,377
Kansas	751	100.0	21.1	11.6	11.5	18.7	21.4	15.9	7,965	9,731
South										
Delaware	177	100.0	19.9	11.1	12.7	20.4	20.4	15.5	7,902	9,888
Maryland	1,225	100.0	16.4	8.8	9.8	17.4	23.6	24.0	9,605	11,734
District of Columbia	306	100.0	19.3	10.9	12.5	19.1	18.2	19.9	8,059	10,947
Virginia	1,444	100.0	22.3	12.1	11.3	15.9	19.2	19.1	7,772	10,349
West Virginia	513	100.0	26.5	12.4	11.3	18.0	19.6	12.0	6,941	8,373
North Carolina	1,480	100.0	25.3	13.4	12.4	16.4	18.2	14.5	6,837	8,988
South Carolina	725	100.0	27.9	12.9	11.7	15.1	17.9	14.5	6,557	8,718
Georgia	1,350	100.0	24.3	13.6	12.1	15.7	18.2	16.2	7,000	9,501
Florida	2,334	100.0	23.8	14.8	13.9	17.2	17.4	12.9	6,631	8,763
Kentucky	974	100.0	28.3	13.6	11.6	15.8	17.7	13.0	6,371	8,546
Tennessee	1,176	100.0	27.4	13.7	12.3	15.7	17.7	13.3	6,429	8,655
Alabama	984	100.0	28.5	13.8	11.9	15.7	17.4	12.6	6,263	8,350
Mississippi	621	100.0	36.6	15.3	11.3	13.0	13.5	10.3	4,722	7,179
Arkansas	591	100.0	32.4	16.3	12.8	14.5	14.0	10.0	5,179	7,417
Louisiana	1,061	100.0	27.8	14.5	12.1	15.5	17.2	13.0	6,257	8,485
Oklahoma	853	100.0	26.5	13.6	12.0	17.0	18.0	12.8	6,644	8,601
Texas	3,500	100.0	24.3	12.8	12.0	16.8	19.2	14.9	7,153	9,195
West										
Montana	228	100.0	22.5	13.0	14.0	21.1	18.9	10.4	7,059	8,259
Idaho	217	100.0	20.5	13.0	14.3	22.0	19.8	10.6	7,297	8,437
Wyoming	107	100.0	18.0	11.6	13.2	21.9	22.1	13.1	7,984	9,231
Colorado	732	100.0	20.4	12.1	12.6	19.9	21.0	13.9	7,733	9,105
New Mexico	292	100.0	23.0	12.6	13.0	18.9	19.2	13.3	7,212	8,750
Arizona	551	100.0	21.1	11.3	11.3	18.7	21.3	16.3	8,008	9,668
Utah	308	100.0	18.2	9.6	12.1	23.6	23.1	13.4	8,325	9,302
Nevada	176	100.0	17.7	10.7	12.1	19.7	22.8	17.1	8,453	10,054
Washington	1,157	100.0	18.5	10.1	9.8	19.1	24.4	18.0	8,891	10,107
Oregon	704	100.0	19.7	10.5	11.3	21.2	23.2	14.2	8,248	9,372
California	7,002	100.0	18.2	10.1	10.4	17.9	23.5	19.9	8,926	10,511
Alaska	112	100.0	22.1	12.6	8.9	13.1	17.9	25.4	8,466	11,149
Hawaii	235	100.0	16.0	11.2	11.2	15.9	19.8	25.9	9,157	12,911
SMSA's										
New York - SCA 5/	5,520	100.0	14.5	8.9	10.7	18.9	24.6	22.5	9,558	11,161
Chicago - SCA 6/	1,752	100.0	14.7	7.8	9.3	17.8	26.7	23.7	10,007	10,439
Los Angeles - Long Beach	2,402	100.0	15.8	9.1	9.7	17.3	24.8	23.1	9,662	11,191
Boston, Massachusetts	933	100.0	16.3	8.8	9.5	17.7	25.1	22.8	9,662	11,157
Baltimore, Maryland	620	100.0	16.8	9.1	10.8	19.6	24.6	19.1	9,071	10,340
Washington, D.C.	945	100.0	14.8	8.6	10.4	17.1	22.6	26.6	9,845	11,966
Detroit, Michigan	1,267	100.0	13.4	7.2	6.8	15.2	28.5	28.8	11,283	12,619
Omaha, Nebraska - Iowa	179	100.0	17.5	10.4	11.8	21.6	23.4	15.2	8,445	9,500
Columbia, South Carolina	98	100.0	31.3	13.6	12.5	14.9	15.8	11.8	5,800	7,604
Knoxville, Tennessee	121	100.0	23.0	12.4	12.4	18.3	20.9	12.9	7,345	8,545
Nashville, Tennessee	152	100.0	23.1	12.0	11.9	16.6	20.6	15.8	7,540	9,081
Houston, Texas	542	100.0	17.7	10.8	12.4	19.6	22.8	16.7	8,384	10,675
Denver, Colorado	422	100.0	19.2	10.6	12.2	20.5	22.9	14.6	8,178	9,897

1/ The projection procedure was to multiply the income class limits by a projection factor to obtain a projected income distribution. The traditional income class limits were then obtained by interpolation assuming a linear distribution of units in the projected income class.

2/ The 1970 number of families and unrelated individuals was estimated by assuming a proportional increase with total resident population from 1960 to 1970.

3/ Computed on the basis of $1,000 intervals under $10,000.

4/ Computed by assuming midpoint to be the mean for each income class below $15,000; the mean of the $15,000 and over income interval was estimated assuming a Pareto relationship.

5/ New York - Northeastern New Jersey Standard Consolidated Area.

6/ Chicago - Northwestern Indiana Standard Consolidated Area.

APPENDIX

General Procedure for Projecting an Income Size Distribution for 1959 to 1969 1/

Step 1.--Obtain a benchmark income size distribution for geographic area. Table a gives the income size distribution for families and unrelated individuals by total money income for Maine in 1959.

Step 2.--Accumulate the distribution. Table b shows the accumulated distribution for Maine.

Step 3. Obtain a projection factor by one of the methods discussed in the paper. The projection factor used in the paper for Maine from 1959 to 1969 is 1.731.

Step 4.--Multiply the income class limits by the projection factor.

The projected distribution (table c) is obtained by multiplying the class limit in table b by the projection factor. For example:

$1,000 x 1.731 = $1,731
$2,000 x 1.731 = $3,462
:
$15,000 x 1,731 = $25,965
$25,000 x 1,731 = $43,275

Step 5.--Obtain the "conventional" income class limits ($1,000, $2,000, ... $8,000, etc.) by assuming the units are linearly distributed within the projected income class.

$$\text{Under } 1{,}000\text{:}\quad 0 + \frac{1{,}000-0}{1{,}731-0} \times (41{,}872-0) = 24{,}189$$

$$\text{Under } 2{,}000\text{:}\quad 41{,}872 + \frac{2{,}000-1{,}731}{3{,}462-1{,}731} \times (80{,}094-41{,}872) = 47{,}812$$

:

$$\text{Under } 25{,}000\text{:}\quad 299{,}233 + \frac{25{,}000-17{,}310}{25{,}965-17{,}310} \times (313{,}148-299{,}233) = 311{,}596$$

Step 6.--Disaccumulate the projected income size distribution either as an absolute or a percent. Table e gives the projected 1969 income size distribution for families and unrelated individuals for Maine.

1/ The computer program can be obtained by writing to Joseph Knott, Population Division, U.S. Bureau of the Census, Washington, D.C. 20233

Table a.--FAMILIES AND UNRELATED INDIVIDUALS IN MAINE FOR 1959

Income	Number
Total...........	318,316
Under $1,000..........	41,872
1,000 - 1,999.........	38,222
2,000 - 2,999.........	37,131
3,000 - 3,999.........	41,100
4,000 - 4,999.........	39,420
5,000 - 5,999.........	35,947
6,000 - 6,999.........	26,749
7,000 - 7,999.........	18,366
8,000 - 8,999.........	12,552
9,000 - 9,999.........	7,874
10,000-14,999.........	13,915
15,000-24,999.........	3,823
25,000 and over.......	1,345

Table b.--ACCUMULATED

Income	Number
Under $1,000........	41,872
Under 2,000........	80,094
Under 3,000........	117,225
Under 4,000........	158,325
Under 5,000........	197,745
Under 6,000........	233,692
Under 7,000........	260,441
Under 8,000........	278,807
Under 9,000........	291,359
Under 10,000........	299,233
Under 15,000........	313,148
Under 25,000........	316,971
Total..........	318,316

Table c.--PROJECTED 1969 ACCUMULATED INCOME SIZE DISTRIBUTION

Income	Number
Under $1,731...........	41,872
Under 3,462...........	80,094
Under 5,193...........	117,225
Under 6,924...........	158,325
Under 8,655...........	197,745
Under 10,386...........	233,692
Under 12,117...........	260,441
Under 13,848...........	278,807
Under 15,579...........	291,359
Under 17,310...........	299,233
Under 25,965...........	313,148
Under 43,275...........	316,971
Total............	318,316

Table d.--1969 INCOME SIZE DISTRIBUTION (CONVENTIONAL CLASS LIMITS)

Income	Number	Percent
Under $1,000....	24,189	7.6
Under 2,000....	47,812	15.0
Under 3,000....	69,893	22.0
Under 4,000....	91,634	28.8
Under 5,000....	113,085	35.5
Under 6,000....	136,386	42.8
Under 7,000....	160,056	50.3
Under 8,000....	182,829	57.4
Under 9,000....	204,909	64.4
Under 10,000....	225,676	70.9
Under 15,000....	287,160	90.2
Under 25,000....	311,597	97.9
Total......	318,316	100.0

Table e.--PROJECTED PERCENTAGE INCOME SIZE DISTRIBUTION FOR 1969

Income	Percent
Total.....	100.0
Under $1,000........	7.6
1,000 - 1,999.......	7.4
2,000 - 2,999.......	7.0
3,000 - 3,999.......	6.8
4,000 - 4,999.......	6.7
5,000 - 5,999.......	7.3
6,000 - 6,999.......	7.5
7,000 - 7,999.......	7.1
8,000 - 8,999.......	7.0
9,000 - 9,999.......	6.5
10,000-14,999.......	19.3
15,000-24,999.......	7.7
25,000 and over.....	2.1

THE 1971 CANADIAN CENSUS TRACT PROGRAM

George J.V. Kokich, Dominion Bureau of Statistics

Introduction

In the Canadian census, Census Tracts are defined as "permanent small census statistical areas in the larger urban communities, and are as homogeneous as possible in terms of economic status and social living conditions".(1) With this definition in mind, the purpose of this paper is to demonstrate the continuing need for Census Tract information in Canada, even though new Bureau Programs will soon provide a wide range of diversified data for urban areas.

The 1971 Canadian Census Tract Program is one of several Programs within the Dominion Bureau of Statistics designed to provide users with detailed census information for different statistical areas. More specifically, this particular Program deals with such information on an intra-urban level for the larger urban communities in Canada. The Program is set up as a means of dividing eligible cities into small geographic areas having permanent boundaries. As noted in The Canadian Census Tract Manual,

> "The purpose of this Program is to primarily provide a general-purpose geographical basis for the collection of local statistics, as well as meet the wide variety and increasing number of requests to those requiring such information. The scope of this Program is to prepare and publish census statistics, maps, area descriptions, and related census information for every eligible center in Canada."(2)

Some of the other 1971 Bureau Programs that are designed to provide users with census information for urban areas, are the Census Metropolitan Area Program, the Enumeration Area Program and the Geocoding Program. Each of these Programs as the Census Tract Program is developed in a different way, designed for different purposes, yet are complementary since each of these Programs satisfy different user needs, at different statistical area levels.

The scope of the present paper is to focus on the preparation and development of the 1971 Canadian Census Tract Program by presenting a summary of intercensal research and planning concerning the development of the Program; the role and co-operation of the local Census Tract committee in the development of a Census Tract proposal; the nature of the 1971 Program; and, a conclusion regarding the Program's significance, as an alternate source of urban information from the 1971 Census.

Intercensal Research and Planning

Following the 1966 quinquennial census of Canada, the Census Division's Geography Section established a Census Tract Unit for the first time, in order to carry out the necessary intercensal research and planning for the Canadian Census Tract Program in 1971 as well as in future censuses. The decision to create such a Unit, was prompted by two considerations: the expanding coverage of the Canadian Census Tract Program from an original two centers in 1941 to a total of twenty-eight eligible centers in 1966, as well as the increasing requests to the Bureau for detailed intra-urban information from the Canadian census.

As conceived, the specific role of this new Unit is to carry out social science research on Canada's Census Tract Program; to plan, develop and implement the 1971 Program; and, to promote this Program among various departments, agencies and urban research-orientated groups and individuals, particularly on a local basis. In 1968 the Census Tract Unit began to carry out this role by reviewing the relevant literature; by examining previous Canadian Census Tract Programs; by comparing the earlier Programs to similar intra-urban Programs; and, by consulting other Bureau and non-Bureau professionals who were familiar with Census Tract data or related information for their comments and suggestions.

As a result of this work to date, the Unit has prepared and developed its Program, and documented guidelines in a new publication entitled The Canadian Census Tract Manual,(3) similar to the American Census Tract Manual.(4) Summarized in the manual are the nature of the Canadian Census Tract Program, the role of the local Census Tract committee and the method of preparing a suitable 1971 Census Tract proposal to the Census Division's Geography Section. In previous censuses, relevant memos and detailed verbal instructions were provided to Canadian local Census Tract committees regarding the preparation of a suitable Census Tract proposal to the Bureau. A new periodic memo entitled "The Canadian Census Tract Memo", similar in idea to the "Census Tract Memo" prepared by the US Bureau of the Census, has also been initiated in order to present a summary report on the general progress and activity of all local Census Tract committees across Canada. The first issue of this Memo(5) has already been distributed.

Following the 1971 Census, the Unit will continue to carry out its original role. In particular, there will be a focus upon additional social science research on Canada's Census Tract Program, such as preparing and publishing some cross-sectional and historical socio-economic studies of Canadian cities according to city size groups. A wider promotion of the Program will also be planned in detail. In this regard, the Unit will be pleased to assist any committee, group or individual regarding Census Tract and/or related intra-urban information.

The Role and Co-operation of the 1971 Canadian Local Census Tract Committee

A very important factor in planning the preparation and promotion of the 1971 Canadian Census Tract Program was the role of the local Census Tract committee. The role of the local committee in Canada is important because of the very nature of the Program (described in the next part of this paper), and the detailed knowledge of local conditions by the committee members. As in the past, the committees were made up of a large variety of professionals who jointly participated with the Bureau's representatives to prepare a suitable Census Tract proposal. As a general rule, the Mayor, city clerk or in most cases the city planner organized such committees at the request of the Bureau's Dominion Statistician, by inviting planners, assessors, educators, real estate specialists, transportation experts, etc., to represent the city at a local Census Tract meeting.

Once the committees were organized, meetings were arranged across Canada during the latter part of 1969 and early 1970. Prior to such meetings, copies of The Canadian Census Tract Manual and of selected 1966 Census Tract bulletins(6) were sent to each committee, as the bases for discussion. Only three urban centers, Winnipeg, Sarnia and Peterborough did not require a meeting with the Bureau's representatives regarding the 1971 Program because they did not consider it necessary to make any significant changes to their 1966 Program.

During each meeting, the Bureau's representatives explained the nature of the 1971 Canadian Census Tract Program, defined the role of the local Census Tract committee, and emphasized that the success or failure of any Census Tract Program would depend on the initiative and enthusiasm of the local committee in carrying out its role within the Program. This role was defined as follows:

(i) "to understand the nature and objectives of the Canadian Census Tract Program;

(ii) to assist the Census Division in the preparation and implementation of a Census Tract Program on a local basis;

(iii) to promote locally, the use of Census Tracts as the basis of data collection by the various local municipal departments and/or agencies;

(iv) to provide the Census Division's Geography Section with relevant information on a continuing basis, regarding the Census Tract boundaries; and,

(v) to keep in touch with the Census Division's Geography Section regarding the success of the Program in your city."(7)

Most of the local committees accepted this role, and agreed to co-operate with the Bureau. As a result, most of the committees have submitted a 1971 Census Tract proposal, and in a few cases, also a 1971 Enumeration Area(8) proposal to the Bureau. Although it was encouraging to receive some complete and well prepared proposals from the more enthusiastic local committees interested in the availability of the 1971 Census information, most of the local committees did not follow all the instructions found in The Canadian Census Tract Manual regarding the contents of a suitable Census Tract proposal. For example, many committees did not send a current Census Tract street index as requested, a description of the boundaries of each proposed Census Tract, or the population count for each proposed tract. Such omissions delayed the review and finalization of the Census Tract proposals well into 1970.

The reason for such omissions appears to have been the wide range of interests among committee members. Many committee members, particularly those in the metropolitan areas for example, also expressed an interest in the alternate Programs, especially the Geocoding Program. Similarly, some committee members representing the smaller cities also expressed an interest in the Enumeration Area Program. In both cases, the amount of interest in alternate Programs varied depending on each committee member. Only a few committee members indicated a lack of interest in any Bureau Program.

The Nature of the 1971 Canadian Census Tract Program

Previous to 1971, only a general description of Census Tracts was available in the various Canadian census bulletins. More detailed information however, was available upon request from the Census Division's Geography Section. In the 1971 Census, a new policy has been adopted in this regard with the decision to publish both the 1971 Canadian Census Tract definition and criteria in all Canadian Census Tract bulletins. It is also noteworthy to point out that the Canadian and American Census Tract definition and criteria are very similar.

Three sets of criteria(9) indicate the requirements for the 1971 Census Tracts defined earlier in this paper. First of all, in order to have a Program, a city must meet the criterion of eligibility. This criterion states that "each city having a minimum of 50,000 persons at the previous quinquennial or decennial census is eligible for a Census Tract Program".

If a city was eligible for a Census Tract Program, then either the criteria for the delineation of Census Tracts or the revision of Census Tracts was applied, depending on which set of criteria were appropriate. The following criteria were used for delineating new Census Tracts:

" (i) a population between 2,500 and 8,000 except for Census Tracts in the Central Business District or Institutional Census Tracts which may have a lower population;

(ii) an area that is as homogeneous as possible in terms of economic status and social living conditions;

(iii) boundaries that follow permanent and easily recognized lines on the ground;

(iv) as much as possible, a compact shape."

Similarly, the following criteria, if necessary, were used to revise established Census Tracts:

" (i) of primary importance are the permanency of the Census Tract boundaries and the population criteria above, in this order;

(ii) of secondary importance is the shape of the Census Tract(s) under review;

(iii) of least importance is the homogeneity criterion."

A total of twenty-nine urban centers which should include approximately 60% of Canada's population in 1971, will have a Census Tract Program. In addition to those cities that had a 1966 Program, the cities of Guelph, Thunder Bay and Sault Ste. Marie will be added to the 1971 Program. As a result of recent changes in municipal organization in the Province of Ontario, the following cities have been combined in the 1971 Census: Kitchener and Guelph, as well as St. Catharines and Niagara Falls. Also beginning in 1971, the Census Tract Program for Canada's capital city will be identified as "Ottawa-Hull", instead of Ottawa as in the past.

As in the previous Census Tract Programs, a large variety of both published and unpublished intra-urban information will be available from the Bureau, for each participating city included in the 1971 Program. This variety will include detailed census statistics on population, housing, retail trade and service trades; a series of reference maps and diagrams portraying the 1971 Census Tracts with or without the component enumeration areas and city blocks respectively; and, area descriptions of each Census Tract and/or enumeration area, indicating the exact boundaries of each statistical area. Conversion tables indicating the relationship between the 1971 Census Tracts and earlier Programs are also being prepared for distribution. The availability of these tables will be noted in future issues of "The Canadian Census Tract Memo".

Significant modifications to the previous numbering system have also been introduced. This new numbering system which was in a large part based upon the US Bureau of the Census' proposed revisions in their Census Tract Numbering system as found in the "Census Tract Memo No. 4",(10) will be a nationally uniform numeric system consisting of 5 digits at the most. The first three digits will represent the tract number, while the two suffix digits that can appear in later censuses will indicate tract splits. Essentially, this new revised approach facilitates the identification of the Census Tracts in the central city (e.g. Montreal), as well as the identification of the Census Tracts (if there are any) in the peripheral municipalities juxtaposition to the central city (e.g. Laval, Longueuil, etc. juxtaposition to Montreal), according to the location of these tracts in a county(s) or census division(s). A more detailed technical paper which elaborates upon the 1971 Census Tract numbering system is currently being prepared. When available for distribution, a note will appear in "The Canadian Census Tract Memo".

In each urban centre that was eligible for a 1971 Census Tract program, the local committees were strongly encouraged to coordinate and match local municipal statistical area information systems with the Bureau's Census Tract Program. The above innovation was generally accepted, and wherever adopted will provide a beneficial exchange of data sources between federal and municipal departments, as well as a more efficient use of the Census Tract Program on a local basis for intra-urban analysis and research.

The committees were also strongly encouraged to emphasize the major advantage of Census Tracts, namely, permanent boundaries in order to ensure the historical comparability of intra-urban information for Canadian cities. The principle was also accepted and used by the majority of the local committees. In a few cases, boundary changes were proposed and accepted, but the number of acceptable changes were kept to an absolute minimum. Changes to the previous 1966 Census Tract boundaries were permitted in exceptionally difficult or special cases such as hard to recognize former municipal limits, or the virtual disappearance of a street (previously used as a Census Tract boundary) replaced by an expressway.

Conclusion

Historically, the Census Tract Program was significant because it was the only centralized and nationally uniform Program in Canada providing cross-sectional and longitudinal socio-economic information on a permanent intra-urban level. In 1971, even though other Bureau Programs will also provide census information for Canadian cities, the Census Tract Program will continue to be a significant Program for many users.

As such, the Census Tract Program will provide detailed information for small permanent areas. By comparison, the other Programs will generally differ in the following regards: the Census Metropolitan Area Program will also provide detailed information and at two levels, first the component municipalities and secondly, the urbanized core and surrounding fringe; the Enumeration Area Program will provide only broad census data categories (such as population data, housing data, etc.) either on summary computer tapes or print-outs; and, the Geocoding Program will provide either broad or detailed census data depending on the nature of the requests for user-specified areas.(11)

In view of the preceding, the original significance of the Census Tract Program will remain in 1971, even though alternate data sources will be available, since the other Bureau Programs are designed to meet different types of requests. As a result of the availability of alternate Bureau Programs, users will be able to choose the most appropriate Program to suit their needs. Compared to the other 1971 Programs for example, Census Tracts will be particularly suitable in analysing the new journey-to-work data, internal migration trends in the cities, as well as historical trends of various demographic and housing phenomena.

As developed to date, the Census Tract Program is simple to understand, easy to use, answers general purpose needs, and is inexpensive. Users should consider all Bureau Programs that are available, before making one or more choices.

Footnotes

(1) Kokich, 1969 ... George J.V. The Canadian Census Tract Manual. Canada. Dominion Bureau of Statistics, Census Division, Geography Section, 10 pages. Ottawa, July, 1969, p. 6.
(2) Ibid., p. 3
(3) Ibid., copies are available in English or French, upon request from the Census Tract Unit of the Census Division's Geography Section.
(4) Bureau of the Census, 1966 ... United States. Department of Commerce. Bureau of the Census. Census Tract Manual (Fifth Edition). 83 pages. Washington, D.C., January, 1966.
(5) DBS, 1970 ... Canada. Dominion Bureau of Statistics, Census Division. "The Canadian Census Tract Memo" No. 1, 1 page. Ottawa, December, 1970. Copies of "The Canadian Census Tract Memo" are available in English or French upon request, from the Census Tract Unit of the Census Division's Geography Section.
(6) The 1966 Census Tract bulletins provide final 1966 Census population and housing totals by Census Tracts, comparative 1961 figures by Census Tracts, and map(s) indicating the Census Tract boundaries, for each eligible city.
(7) Kokich, 1969, pp. 3-4.
(8) In the Canadian Census, Census Tracts are generally larger or occasionally equal, but never smaller than an Enumeration Area. Normally, Enumeration Areas are component parts of a Census Tract. In rare cases however, one Enumeration Area may equal one Census Tract, as for example, a Census Tract demarcating a large hospital or institution.
(9) Ibid., p. 6.
(10) Bureau of the Census, 1963 ... United States. Department of Commerce. Bureau of the Census. "Census Tract Memo No. 4" pp. 1-5. Washington, D.C., August 9, 1963.
(11) More detailed information on any of these Programs is available upon request from the Census Division's Geography Section.

ESTIMATION OF TRANSITION PROBABILITIES FROM SAMPLE SURVEY DATA AND USES

J.C. Koop and M.P. Singh
Dominion Bureau of Statistics
Ottawa

1. Introduction

Consider a universe where the units are grouped into a number of class intervals or strata according to the variate values of an important characteristic of each of these units. Hereafter the variate value of such a characteristic will be termed the stratification variable. Due to changes in the stratification variable each group of units assigned to a particular stratum (out of a total of L) at a point or epoch of time before a survey is now distributed in a particular manner over L+1 strata, hereafter to be termed transitional strata, at the time of the survey, the (L+1)th stratum consisting of units having one or more special characteristics, e.g. going out of business in the context of a farm economy. In this paper we shall define the special characteristic for those units found in each of the (L+1)th strata as one of having zero variate values for all variables of interest to the survey.

At a point or epoch of time before a survey let N_h be the number of units in the hth stratum $(h = 1, 2, \ldots, L)$. This is known to the sample surveyor. At the time of the survey the stratification variable of some of the units of the hth stratum changes, and N_{hj} out of N_h units are found in the j^{th} transitional stratum $(j = 1, 2, \ldots, L+1)$; the sample surveyor has evidence of this change only from his sample. Let x_{hjk} be a variate value of interest of the k^{th} unit of the j^{th} transitional stratum of the h^{th} original stratum $(h = 1, 2, \ldots, L;$ $j = 1, 2, \ldots, L+1;\ k = 1, 2, \ldots, N_{hj})$. The problem is to estimate

$$T = \sum_{h=1}^{L} \sum_{j=1}^{L+1} \sum_{k=1}^{N_{hj}} x_{hjk}, \tag{1}$$

on the basis of a stratified sample of $n_1, n_2, \ldots, n_L$ units drawn by simple random sampling without replacement from each respective original stratum.

$$\text{If } \bar{\bar{x}}_{hj} = \sum_{k=1}^{N_{hj}} x_{hjk}/N_{hj}, \text{ for all } (h,j),$$

is the mean of the j^{th} transitional stratum of the h^{th} original stratum, then (1) may be rewritten as

$$T = \sum_{h=1}^{L} \sum_{j=1}^{L} N_{hj}\, \bar{\bar{x}}_{hj}, \tag{1.1}$$

since $\bar{\bar{x}}_{hj} = 0$ for $h = 1, 2, \ldots, L$ and $j = L+1$. Further with respect to the original stratification if $\bar{\bar{x}}_h = \sum_{j=1}^{L} N_{hj}\, \bar{\bar{x}}_{hj}/N_h$, then (1.1) may also be rewritten as

$$T = \sum_{h=1}^{L} N_h\, \bar{\bar{x}}_h. \tag{1.2}$$

2. Estimation Problems

The stratified sample shows that n_{hj} out of n_h units, drawn from the h^{th} original stratum, are found in the j^{th} transitional stratum in regard to which

$$0 \le n_{hj} \le n_h, \quad \text{for all } (h,j). \tag{2}$$

The universe total T may be estimated
(i) by standard theory, effectively ignoring the information provided by (2) regarding the transition of units from stratum to stratum, or
(ii) by taking into account these transitions.

Estimation by standard theory. In this case the usual estimator of T is

$$\hat{T} = \sum_{h=1}^{L} N_h\, \bar{x}_h, \tag{3}$$

where $\bar{x}_h$ is the mean of the sample units in stratum h. Alternatively in (1.1) if unbiased estimates of N_{hj} and $\bar{\bar{x}}_{hj}$ are substituted, viz.

$$\hat{N}_{hj} = \frac{n_{hj}}{n_h} N_h, \tag{4}$$

and

$$\bar{x}_{hj} = \sum_{k=1}^{n_{hj}} x_{hjk}/n_{hj}, \tag{5}$$

for all relevant (h,j), then it is not difficult to see that the resulting estimator does not involve the n_{hj}'s and is identical to (3). The variance of this estimator is

$$V(\hat{T}) = \sum_{h=1}^{L} N_h^2 \frac{S_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right), \tag{6}$$

where

$$S_h^2 = \sum_{j=1}^{L+1} \sum_{k=1}^{N_{hj}} (x_{hjk} - \bar{\bar{x}}_h)^2/(N_h - 1) \quad \text{for all h.}$$

Estimation taking into account transition probabilities. The probability of transition of a unit in stratum h to the j^{th} transitional stratum is $p_{hj} = N_{hj}/N_h$, and on the basis of sample data an unbiased estimate of this transition probability is $\hat{p}_{hj} = n_{hj}/n_h$ for all (h,j). Thus we have a matrix, with L+1 rows and L columns, of the estimated transition probabilities

$$(\hat{p}_{hj}). \tag{7}$$

This matrix multiplied by the column vector of the number of units in the L strata at the point of epoch of time before the survey, i.e., $\{N_1, N_2, \ldots, N_L\}$, yields a vector of L+1 unbiased estimates of the number of units in the L+1 strata at the time of the survey, the number in the j^{th} stratum being

$$\hat{N}_{.j} = \sum_{h=1}^{L} \hat{p}_{hj} N_h, \tag{8}$$

which is simply the sum over h of the estimate given by (4). By classical theory the variance of $\hat{N}_{.j}$, for all j, is given by

$$V(\hat{N}_{.j}) = \sum_{h=1}^{L} N_h^2 \frac{p_{hj}(1-p_{hj})}{n_h} \left(\frac{N_h-n_h}{N_h-1}\right). \tag{9.1}$$

Also

$$\mathrm{Cov}(\hat{N}_{.j}, \hat{N}_{.j'}) = -\sum_h N_h^2 p_{hj}p_{hj'} \left(\frac{N_h-n_h}{N_h-1}\right)/n_h \quad \text{for all } j \neq j'. \tag{9.2}$$

On the basis of the estimated stratum numbers given by (8), a biased but consistent estimator of T is

$$T' = \sum_{j=1}^{L} \hat{N}_{.j} \left(\sum_{h=1}^{L} n_{hj} \bar{x}_{hj} / \sum_{h=1}^{L} n_{hj} \right). \tag{10}$$

The loss in efficiency of the original stratification is to some extent regained by restratification through pooling each set of L transitional strata relating to the same range of x-values; this is the justification for constructing an alternative estimator T' given by (10).

We shall derive briefly an expression for the bias of T'. Substituting the expression for $\hat{N}_{.j}$ in (10), and denoting the random variables specified in (2) by $\underset{\sim}{n}_{hj}$ we find

$$E(T') = E\{E(T'|\underset{\sim}{n}_{hj})\}$$

$$= E\left\{ \sum_{j=1}^{L} \left(\sum_{h=1}^{L} \frac{N_h}{n_h} n_{hj} \right) \left(\frac{\sum_{h=1}^{L} n_{hj} \bar{\bar{x}}_{hj}}{\sum_{h=1}^{L} n_{hj}} \right) \right\}$$

$$= \sum_{j=1}^{L} \sum_{h=1}^{L} \frac{N_h \bar{\bar{x}}_{hj}}{n_h} E\left(\frac{n_{hj}^2}{\sum_{h=1}^{L} n_{hj}}\right)$$

$$+ \sum_{j=1}^{L} \sum_{h \neq h'}^{L} \frac{N_h \bar{\bar{x}}_{h'j}}{n_h} E\left(\frac{n_{hj} n_{h'j}}{\sum_{h=1}^{L} n_{hj}}\right). \tag{11}$$

In order to evaluate the expectations of ratios of functions of random variables found in (11) we need the following result due to one of us. If X and Y are pairs of random variables with X assuming all values but not zero, then

$$E\left(\frac{Y}{X}\right) = \frac{E(Y)}{E(X)} - \frac{\mathrm{Cov}(Y,X)}{E^2(X)} + \frac{1}{E^2(X)} E\left\{\frac{Y}{X} (X-E(X))^2\right\}. \tag{12}$$

With this result, we find

$$E\left(\frac{n_{hj}^2}{\sum_h n_{hj}}\right) = \frac{E(n_{hj}^2)}{E(\sum_h n_{hj})} - \frac{\mathrm{Cov}(n_{hj}^2, \sum_h n_{hj})}{E^2(\sum_h n_{hj})} + \frac{1}{E^2(\sum_h n_{hj})} E\left\{\frac{n_{hj}^2}{\sum_h n_{hj}} \left(\sum_h (n_{hj}-E(n_{hj}))\right)^2\right\}, \tag{13}$$

and

$$E\left(\frac{n_{hj}n_{h'j}}{\sum_h n_{hj}}\right) = \frac{E(n_{hj}n_{h'j})}{E(\sum_h n_{hj})} - \frac{\mathrm{Cov}(n_{hj}n_{h'j}, \sum_h n_{hj})}{E^2(\sum_h n_{hj})} + \frac{1}{E^2(\sum_h n_{hj})} E\left\{\frac{n_{hj}n_{h'j}}{\sum_h n_{hj}} \left(\sum_h (n_{hj}-E(n_{hj}))\right)^2\right\}. \tag{14}$$

Further details of the derivation of E(T') are found in the appendix. Considering only the leading terms of (13) and (14) we find

$$E(T') = T + \frac{1}{n} \Big[\sum_{j=1}^{L} \Big\{ \sum_{h>h'} N_h N_{h'} p_{hj} p_{h'j} (\bar{\bar{x}}_{hj}-\bar{\bar{x}}_{h'j}) \left(\frac{n_h}{N_h} - \frac{n_{h'}}{N_{h'}}\right) / \left(\sum_{h=1}^{L} \alpha_h p_{hj}\right)$$

$$+ \sum_{h=1}^{L} N_{hj} \bar{\bar{x}}_{hj} (1-p_{hj}) \left(\frac{N_h-n_h}{N_h-1}\right) / \left(\sum_{h=1}^{L} \alpha_h p_{hj}\right) \Big\} \Big], \tag{15}$$

where $\alpha_h = n_h/n$ is the proportion of the sample allocated to stratum h. Now $\bar{\bar{x}}_{hj}$ is the mean of all the N_{hj} units of the h^{th} original stratum which have moved to the j^{th} transitional stratum. Similarly $\bar{\bar{x}}_{h'j}$ is the mean of all the $N_{h'j}$ units of the h'^{th} original stratum which have moved to the j^{th} transitional stratum. Since by this procedure of restratification the range of x-values in each of the j^{th} transitional strata is controlled, the differences $\bar{\bar{x}}_{hj}-\bar{\bar{x}}_{h'j}$, for all h>h', could not be considerable. Further it is unlikely that the signs of such differences are the same unless the class interval of the stratification variable is wide. Indeed (15) suggests that the bias, $E(T')-T$, can be reduced by narrowing the width of the transitional strata so as to make the $\bar{\bar{x}}$'s for each j as equal as possible, in which situation this bias will be negligible when n is large.

The first steps in the derivation of the variance of T' are as follows:

$$V(T') = V\{E(T'|\underset{\sim}{n}_{hj})\} + E\, V(T'|\underset{\sim}{n}_{hj})\}$$

$$= V\{ \sum_{j=1}^{L} \left(\sum_{h=1}^{L} \frac{N_h}{n_h} n_{hj} \right) \left(\frac{\sum_{h=1}^{L} n_{hj} \bar{\bar{x}}_{hj}}{\sum_h n_{hj}} \right) \}$$

$$+ E\{ \sum_{j=1}^{L} \left(\sum_{h=1}^{L} \frac{N_h}{n_h} n_{hj} \right)^2 \frac{1}{(\sum_h n_{hj})^2} \cdot \sum_{h=1}^{L} n_{hj}^2 \frac{S_{hj}^2}{n_{hj}} \left(1- \frac{n_{hj}}{N_{hj}}\right) \}, \qquad (16)$$

where $S_{hj}^2 = \sum_{k=1}^{N_{hj}} (x_{hjk}-\bar{\bar{x}}_{hj})^2/(N_{hj}-1)$ for all (h,j). The first part of the variance function V{ } may be ascribed to movement of some of the units of the original strata to their respective transitional strata resulting in the $\bar{\bar{x}}_{hj}$'s. This variance function involves the products of the estimates of the transition probabilities n_{hj}/n_h and the corresponding transitional stratum weights $n_{hj}/\sum_h n_{hj}$, and also involves the relevant cross-product terms. It can be shown that the variances and covariances of all these product and cross-product terms are of order $1/n_h^2$ for all h so that there is control on this part of the variance through restratification. The second part of the variance function is due to the variation of the x's in the transitional strata. We again remark that this source of variation can be controlled by narrowing the width of the transitional strata.

The variance of the unbiased estimator given by (6) can be rewritten as

$$V(\hat{T}) = \sum_{h=1}^{L} N_h^2 \left(1- \frac{n_h}{N_h}\right) \frac{1}{n_h} \sum_{j=1}^{L} \left(\frac{N_{hj}}{N_h-1}\right)(\bar{\bar{x}}_{hj}-\bar{\bar{x}}_h)^2$$

$$+ \sum_{h=1}^{L} \frac{N_h^2}{n_h} \left(1- \frac{n_h}{N_h}\right)\left(\frac{N_{h,L+1}}{N_h-1}\right) \bar{\bar{x}}_h^2$$

$$+ \sum_{h=1}^{L} N_h^2 \left(1- \frac{n_h}{N_h}\right) \frac{1}{n_h} \sum_{j=1}^{L} \left(\frac{N_{hj}-1}{N_h}\right) S_{hj}^2 . \qquad (17)$$

The expression speaks for itself. There are reasons to believe that $V(\hat{T})$ can be greater than $V(T')$, because of the middle term in (17).

The variance expression for T' is developed to this point just to gain an insight into the features commented upon in the foregoing account. An exact expression for it will involve all the transition probabilities, and is not useful for estimation purposes because of its complexity.

The problem of variance estimation can be resolved by redesigning the survey so as to have two or more independent stratified samples, each yielding an independent estimate of T. When the total sample consists of two independent replicates, then the variance of the mean $\bar{T}=(T_1'+T_2')/2$ is given by

$$v(\bar{T}) = (T_1'-T_2')^2/4, \qquad (18)$$

despite the complexity of (16).

3. Remarks

The problem has been considered in the context of one-stage stratified sampling partly because it is simple, the strata corresponding to the "states" in the theory of Markov chains, and partly because the problem was originally seen in relation to this sampling design. The problem of estimation can be considered almost along the same lines for more ramified sampling designs.

Acknowledgement. We are grateful to Dr. I.P. Fellegi for some critical remarks which have improved the presentation of this paper. To our colleagues in the Agriculture Division of this Bureau, we are also grateful for many enlightening discussions on the practical problems of estimation in livestock and poultry sample surveys, which have provided the impulse for the estimation problem considered in this paper.

Reference

Koop, J.C. On the derivation of expected value and variance of ratios without the use of infinite series expansions. To appear in Metrika.

Appendix

Derivation of E(T'). For evaluation of terms involved in (13) and (14) we note the following.

$$E(n_{hj}) = n_h p_{hj}, \quad E(n_{hj}^2) = n_h^2 p_{hj}^2 + n_h p_{hj} q_{hj} f_h ,$$

$$E(n_{hj} n_{h'j}) = n_h n_{h'} p_{hj} p_{h'j},$$

$$\text{where } q_{hj}=1-p_{hj}, \quad f_h = \left(1- \frac{n_h-1}{N_h-1}\right),$$

$$Cov(n_{hj}^2, \sum_{h=1}^{L} n_{hj}) = E(n_{hj}^3) - E(n_{hj}^2)\, E(n_{hj})$$

$$\doteqdot n_h^2 p_{hj}^2 (3-q_{hj} f_j) - n_h p_{hj} (2-3q_{hj} f_h),$$

and

$$\mathrm{Cov}(n_{hj}n_{h'j}, \sum_{h=1}^{L} n_{hj}) = \mathrm{Cov}(n_{hj}n_{h'j}, n_{hj})$$

$$+ \mathrm{Cov}(n_{hj}n_{h'j}, n_{h'j})$$

$$= E(n_{h'j})\, V(n_{hj}) + E(n_{hj})\, V(n_{h'j})$$

$$= n_h n_{h'} p_{hj} p_{h'j}\, (f_h q_{hj} + f_{h'} q_{h'j}).$$

Further denoting $E\{\frac{n_{hj}^2}{\sum_h n_{hj}} (\sum_{h=1}^{L} (n_{hj}-n_h p_{hj}))^2\}$

by $R_{hj}^{(1)}$ and $E\{\frac{n_{hj}n_{h'j}}{\sum_h n_{hj}} (\sum_{h=1}^{L} (n_{hj}-n_h p_{hj}))^2\}$

by $R_{hj}^{(2)}$, we get from (11), (13) and (14)

$$E(T') = \frac{1}{n} \sum_{j=1}^{L} \{ \sum_{h=1}^{L} N_h \bar{\bar{x}}_{hj} \frac{n_h p_{hj}^2}{\sum_h \alpha_h p_{hj}}$$

$$+ \sum_{h \neq h'} N_h \bar{\bar{x}}_{h'j} \frac{n_{h'} p_{h'j} p_{hj}}{\sum_h \alpha_h p_{hj}}\}$$

$$+ \frac{1}{n} \sum_{j=1}^{L} \sum_{h=1}^{L} N_h \bar{\bar{x}}_{hj} \frac{p_{hj} q_{hj} f_h}{\sum_h \alpha_h p_{hj}}$$

$$- \frac{1}{n^2} \sum_{j=1}^{L} \sum_{h=1}^{L} \frac{N_h \bar{\bar{x}}_{hj}}{(\sum_h \alpha_h p_{hj})^2} \{ n_h p_{hj}^2 (3-q_{hj} f_h)$$

$$- p_{hj}(2-3q_{hj} f_j) - \frac{R_{hj}^{(1)}}{n_h}\}$$

$$- \frac{1}{n^2} \sum_{j=1}^{L} \sum_{h \neq h'}^{L} \frac{N_h \bar{\bar{x}}_{h'j}}{(\sum_h \alpha_h p_{hj})^2} \cdot$$

$$\{ n_h p_{h'j} p_{hj} (f_h q_{hj} + f_{h'} q_{h'j}) - \frac{R_{hj}^{(2)}}{n_h}\},$$

where $\alpha_h = n_h/n$ is the proportion of sample allocated to stratum h. Ignoring the contributions of third and fourth term for large sample size and adding and subtracting

$$\sum_{h \neq h'}^{L} (N_h \bar{\bar{x}}_{hj}\, p_{hj})(n_{h'} p_{h'j}) / \sum_{h=1}^{L} n_h p_{hj}$$

to the first term we can write

$$E(T') = \sum_{j=1}^{L} \sum_{h=1}^{L} N_{hj}\, \bar{\bar{x}}_{hj}$$

$$+ \frac{1}{n} \sum_{j=1}^{L} \{ \sum_{h=h'}^{L} N_{hj} (\bar{\bar{x}}_{h'j} - \bar{\bar{x}}_{hj}) \frac{n_{h'} p_{h'j}}{\sum_{h=1}^{L} \alpha_h p_{hj}}$$

$$+ \sum_{h=1}^{L} N_{hj} \bar{\bar{x}}_{hj} \frac{q_{hj} f_j}{\sum_{h=1}^{L} \alpha_h p_{hj}}\},$$

which can also be expressed in the form given in (15).

RACE-SEX DISCRIMINATION IN INTERNAL NET MIGRATION

G. K. Kripalani*, Western Michigan University

INTRODUCTION

Investigations relating to factors affecting internal migration of people from one area to another can contribute to better understanding of why people move and help to improve judgments about future population adjustments. Existing methods for predicting an area's population are essentially of the nature of extrapolations of past trends, and since they are not based on analysis of changes in factors influencing net migration, they do not permit much insight into causal factors underlying population change.

Why do people move internally from one part of the United States to another? Several widely divergent motives may underlie the migration behavioral pattern of the people of an area. Better wages or more generally, more favorable economic opportunities, present or potential, represent one major group of factors influencing migration decisions. Another major group of causes stems from "non-economic socio-cultural environment" of the areas of origin of migrants and their anticipated evaluation of corresponding elements of "non-economic socio-cultural environment" in the areas of prospective in-migration. Migration decisions are also affected by costs, information, existence of programs of assistance and kindred factors.

A theory of labor migration which regards relative wage ratio (or income differential) as the sole primary determinant of net migration is considered too simple and too inadequate to be useful in theoretical formulation or in empirical investigations. The general remarks in the preceding paragraphs suggest that neither relative wage ratio nor even some of the major relative economic opportunity factors may completely explain internal net migration behavior. For some of the major independent variables, adequately valid data series may not be available; besides, the nature of some particular variable or variables may be such as to preclude its measurability or observability. For example, valid reasons are advanced that the relative wage ratio should relate to marginal workers confronting migration choice and should not be the ratio of average wages.

The principal premise that underlies this study is that there are at least a few major independent variables affecting net migration and that some of these are non-measurable or nonobservable, and that valid data series for such variables do not exist for use in empirical investigations. The method of analyses used is therefore designed to recognize and take into account this problem of nonobservability of some of the major explanatory variables. It is further recognized that net migration behavior patterns vary between the races, between the sexes and between age groups within each race-sex category. Consequently, there is need for stratification of an area's population into reasonable small homogeneous age, sex and race groups.

It is hypothesized that factors influencing internal net migration decisions of an age-sex-race group are of three categories:

1. Time-related (or sub-area-related) relative opportunity factors. These factors are functions of time in time-series analyses or of a sub-area in cross-section analyses. These factors are the same for all age groups within a race-sex category. These relative opportunity factors are represented by an omnibus variable Z_t which is an index representing all relevant time-related relative opportunity factors. It is assumed that Z_t, which is the independent variable, is nonobservable.

2. Age-related relative opportunity factors. These factors do not vary over time in time series analyses (or over sub-areas in cross-section analyses) but vary between age groups within a race-sex category. Such age-related relative opportunity factors are denoted by a nonobservable index m_i where i refers to age group.

3. Category-related relative opportunity factors. These factors do not vary over time in time series analyses (or over sub-areas in cross-section analyses) and between age groups within a race-sex category. But these factors vary between race-sex categories. These forces are denoted by a_j where j refers to race-sex category.

The results reported in this paper are with reference to the third component of internal net migration which stems from forces which are constant over the time span of a time series study (or over sub-areas of a cross-section study) and over age groups within a race-sex category, but which vary between the four race-sex categories namely white males (WM), white females (WF), nonwhite males (NM), and nonwhite females (NF).

For the purpose of this study, we may define the "race-sex discrimination" index of an area as the race-sex related component of internal net migration of that area (component a_j). It is, however, recognized that the subset S_1 of elements of a socio-cultural environment S giving rise to what is called "race-sex discrimination" may consist of two types of elements--subset S_{11} consisting of elements which are the same for all age groups within a race-sex category (component a_j) and a subset S_{12} consisting of elements which vary between age groups within a race-sex category (component m_{ij}). The latter component may reasonably be thought of, in given situations of being the result of "race-sex discrimination" and should appropriately be attributed to it.**

Model & The Method of Estimation

The model (for each race-sex category) is:

$$Y_{it}/E_{it} = \alpha Z_{it}{}^{\beta i} \quad \varepsilon it \qquad (1)$$

where $Z_{it} = m_i Z_t$ and β_i is the elasticity of response. Taking logarithms, the linear form becomes:

$$Y'_{it} = \alpha' + \beta_i m_i' + \beta_i Z_t' + \varepsilon'_{it} \qquad (2)$$

where prime quantities represent natural logarithms of corresponding unprimed quantities. The notation is:

E_{it} - Population of age group i exposed to risk of net migration during time interval t which would be in the area in the absence of any net migration.

M_{it} - Net migration of age group i during time interval t. M_{it} is positive when there is net in-migration and is negative when there is net out-migration.

$Y_{it} = 1 + M_{it}/E_{it}$ = Survival rate against net migration where $(M_{it} + E_{it})$ is the quantity of supply of labor, and E_{it} is the supply shifter.

Z_t - Non-observable independent variable representing the average index of "relative opportunity".

β_i - age group i's response coefficient (elasticity) to index Z_t.

m_i - index of age-related opportunity factors

α - represents effect of category-related factors (called "race-sex discrimination effects" in this study)

ε_{it} - disturbance terms

The non-linear iterative least squares estimation procedure developed by Johnston and Tolley [1] in their study "Supply of Farm Operators" was used to estimate values of model parameters and the nonobservable variable Z. The basic properties of this model were, however, crucially different in some respects from the properties of Johnston-Tolley model and consequently necessary modifications were introduced in evaluation procedures. It is not proposed to deal with the estimation problems in the report.

In practical language the model separates net migration into three components: category effect 'autonomous' component α which is the same for all age groups within a race-sex category. This component would reflect the amount of net migration that would occur if $Z_t=1$ and $m_i=1$, that is if net migration induced by time-related and age-related factors is zero. It is this component which in this report is said to reflect "race-sex discrimination effect". There are two induced effects, one representing response to time-related omnibus independent variable Z_t and the other to age-related factors m_i. The real difficulty comes in the interpretation of the significance of the forces represented by m_i. Some of the forces underlying m_i may stem from those elements of the "socio-cultural environmental" complex as may be said to represent "race-sex discrimination", while it may legitimately be argued that some of these age-related factors stem from the fact that the assumption of a common index of relative opportunity facing all age-groups is unrealistic and that the index of relative opportunity is a function of both t and i. In such a situation, Z_t would represent an average index of relative opportunity and a part of m_i would represent departures of the omnibus variable for the age group from the average Z_t for the category. Under these conditions, it would be necessary to identify the two subsets of the elements underlying m_i; those that relate to race-sex discrimination and those that reflect the situation that the index of relative opportunity is both age-related and time-related.

The "race-sex discrimination" index of a socio-cultural environment of an area may be viewed as a measure of the net effect of factors other than age-time related factors. Viewed thus a comparative analyses of α's may enable us to answer questions such as:

(1) Are females "potentially" more migratory than males when the influences of age-time related factors are eliminated or equalized out or are Southern nonwhite males potentially more migratory than the Southern nonwhite females?

(2) Does the socio-cultural environment of a state discriminate against females or against nonwhites?

The significance of the positive or negative sign of α' may be clearly understood. Since total internal net migration of a color-sex category for the nation as a whole must be zero, it is easy to see that for each race-sex category:

$$\sum_s \alpha'_{js} W_{js} = 0 \quad (j = WM, WF, NM, NF)$$

where α'_{js} equals race-sex discrimination index of category j in state s, and W_{js} equals proportion of category j population in state s (as proportion of the total category population in the nation).

Consequently, index α' is an index of relative "discrimination", in relation to the average for the nation which is zero. A positive α' does not signify that "discrimination" however defined, is absent in that state; it only signifies that "discrimination", if any in this state, is less than the average for the nation as a whole.

Empirical Results

1. For white males, α' was positive for the MSEA's of most States except for Georgia and

Tennessee. This means that the socio-cultural environment of these MSEA's is less favorable to white males than the average for the nation. For white females, a' for MSEA's of Illinois and Washington States were negative. The results for nonwhites of both sexes for New England and Middle Atlantic States is interesting. a' for nonwhite males and nonwhite females are negative for Massachusetts, New York and Pennsylvania. The implication is that the socio-cultural environment of the MSEA's of these States discriminates against nonwhites of both sexes and that the nonwhites would be migrating out of these MSEA's <u>if</u> the relative economic and noneconomic opportunity factors in these States were only as favorable as the average for the nation as a whole. Similarly a' for nonwhite males and nonwhite females for Florida is negative showing that larger numbers of nonwhites of both sexes would indeed be moving into Florida if the discrimination environment against nonwhites improved in the State. Similar remarks apply to nonwhite females for the MSEA's in the States of Louisiana and Texas.

2. Strictly speaking, comparisons of a'between sexes or races for a given state are not always valid. However, the iterations resulted in an underlying pattern of Z's. If the Z's between States are, in fact, similar, or if the Z's between two race or sex categories within a state are about similar, then the comparisons of a' are valid and such comparisons could serve as basis for a reasonably broad and general interpretation.

Some interesting results were:

(a) metropolitan state economic areas (MSEA) are relatively more favorable to males than to females of both races. a' for white males was greater than a' for white females in 10 states out of 15; a' for nonwhite males was greater than a' for nonwhite females in 11 states out of 13.

(b) inter-racial comparisons did not provide definite evidence as to whether MSEA are more favorable to members of one race rather than another. a' for white males was greater than a' for nonwhite males in 5 states out of 8 and was lower in 3 states; a' for white females was higher than a' for nonwhite females in 4 states out of 7 and was lower in 3.

(c) A study of inter-racial comparisons by region, however, reveals some interesting results. In the Southern States, a' for whites was greater than a' for nonwhites in 5 states out of 6. In California, on the other hand a' for nonwhites was higher than a' for whites of both sexes.

PRINCIPAL REFERENCES

1. Johnston, W. E. and Tolley, G. S., 1968. The supply of farm operators. Econometrica, Vol. 36, No. 2 (April, 1968), p. 365-382.

2. Lyttkens, Ejnar, 1965. On the fixed-point property of Wold's iterative estimation method for principal components, - a paper presented at the International Symposium on Multivariate Analysis at University of Dayton, Dayton, Ohio June 14-19, 1965.

3. Wold, Herman, 1965. Nonlinear Estimation by Iterative Least Squares; unpublished paper, Institute of Statistics, University of Uppsala, Sweden.

4. U. S. Department of Agriculture, 1965. States, counties, economic areas and metropolitan areas, Volume I. Population-Migration Report; Net Migration of the Population, 1950-60, by Age, Sex, and Color. U. S. Government Printing Office, Washington, D. C.

Footnote

**I am indebted to my colleague Louis Junker for suggesting that the variable m_i may represent and capture age-related factors of "race-sex discrimination" and that "race-sex discrimination" index need not necessarily be totally described by component aj.

Acknowledgment

*Main work on this research was done while the author worked at the N.C. State University on a National Science Foundation Project, "Area Population Adjustments in Relation to Economic Activity". The author gratefully acknowledges the help he received from C.E. Bishop, G.S. Tolley, C.H. Hamilton, and F.E. McVay. Results reported in Tables 1 and 2 are contained in author's report to NSF. However, work on the interpretation and significance of a has been done during author's stay at Western Michigan University.

Table 1 . Estimates of category effect, MSEA in net in-migration data analyses

Region/state	White Male (2)	White Female (3)	Nonwhite Male (4)	Nonwhite Female (5)	$H_{1.} \overset{D}{=}$ (2)-(3)	$H_{2.} \overset{D}{=}$ (4)-(5)	$H_{.1} \overset{D}{=}$ (2)-(4)	$H_{.2} \overset{D}{=}$ (3)-(5)
New England					(0, 1)	(0, 1)		
Massachusetts			-.0141	-.0094		0		
Middle Atlantic					(1, 0)	(3, 0)	(1, 0)	(1, 0)
New York			-.0050	-.0139		1		
New Jersey	.1894	.0535	.0562	.0417	1	1	1	1
Pennsylvania			-.0037	-.0060		1		
East North Central					(0, 1)	(4, 0)	(1, 1)	(0, 2)
Ohio	.0336		.0124	.0123		1	1	
Indiana	.0083	.0152	.0170	.0166	0	1	0	0
Illinois		-.0105	.0095	-.0007		1		0
Michigan			.0505	.0194		1		
Wisconsin	.0102							
West North Central					(1, 0)	(1, 0)	(1, 0)	
Minnesota	.0307	.0283	.0157	-.0135	1	1	1	1
South Atlantic and D.C.					(2, 2)	(0, 1)	(1, 0)	(1, 0)
District of Columbia		.0050						
Virginia	.0248	.0196			1			
North Carolina	.0100	.1309			0			
Georgia	-.0382	.0263			0			
Florida	.0325	.0145	-.0072	-.0027	1	0	1	1
East South Central					(1, 0)			
Tennessee	-.0160							
Alabama	.0369	.0191			1			
West South Central					(2, 1)	(2, 0)	(1, 1)	(2, 0)
Louisiana	.0181	.0094	.0296	-.0348	1	1	0	1
Oklahoma	.0337	.0386			0			
Texas	.0422	.0365	.0093	-.0064	1	1	1	1
Mountain					(1, 1)			
Colorado	.0544	.0280			1			
New Mexico	.0095	.0127			0			
Pacific					(2, 0)	(1, 0)	(0, 1)	(0, 1)
Washington	1.7798	-.0151			1			
California	.0163	.0061	.0409	.0207	1	1	0	0

A MODEL FOR INDIVIDUAL-GROUP COMPARISONS

Kinley Larntz and Stephen E. Fienberg, The University of Chicago

1. INTRODUCTION

In problem solving situations, it has been suggested that the superiority of groups over individuals is due simply to the fact that groups consist of several individuals. Here, the Lorge and Solomon [1955] approach to such situations is reexamined using the method of maximum likelihood. Extensions to trichotomous response situations are also presented, and the resulting models are applied to data gathered by Staub [1970]. A more detailed analysis of this set of data, including the partitioning of the likelihood ratio goodness-of-fit statistic, is given in Fienberg and Larntz [1971].

2. THE LORGE-SOLOMON MODEL

For problem solving situations, one criterion for comparing group and individual performance is the difference between the proportion of individuals and the proportion of groups sucessful in the solution of a particular problem. Shaw [1932] compared individuals and groups of size four in just this manner. A group of 41 students was randomly divided into two parts, one part consisting of 5 ad-hoc like-sex four-member groups. The 21 individuals and 5 groups then attempted to solve each of three well-known mathematical puzzles. The data are presented in Table 1.

Lorge and Solomon [1955] suggested that the following hypothesis might provide an adequate description of this data:

> H_0: "Group superiority is a function only of the ability of one or more of its members to solve the problem without taking account of the interpersonal rejection and acceptance of suggestions among its members."

A model for the data, implied by H_0, is described below.

Assume that there are N_I individuals each of which has probability p of not solving a particular problem, and that the number of individuals, n_I, who do not solve the problem is binomially distributed. Similarly, assume that there are N_G groups each of which has probability g of not solving the problem, and that the number of groups, n_G, who do not solve the problem is also binomially distributed.

Under H_0, the probability of a group solution is equal to the probability of the group containing one or more members who can solve the problem, i.e,

$$(2.1) \qquad H_0\colon\ 1-g = 1-p^k,$$

for groups of size k. Thus g is expressible as a function of p under H_0, and the maximum likelihood estimator of p under H_0 is found to be a solution of

$$(2.2) \qquad 0 = (n_I+kn_G)(1-p^k) - (N_I-n_I)(p^k+p^{k-1}+\cdots+p) - k(N_G-n_G)p^k.$$

It can be shown that, under H_0, the Pearson chi-square statistic

$$(2.3) \qquad X^2 = \Sigma\frac{(O-E)^2}{E},$$

and the likelihood ratio chi-square statistic

$$(2.4) \qquad G^2 = 2\,\Sigma\, O \log \frac{O}{E},$$

This research was supported in part by Research No. NSF GS-2818 from the Division of Social Sciences of the National Science Foundation to the Department of Statistics, The University of Chicago.

TABLE 1

Data from Shaw [1932]

Problem I

	I	G_4
NS	18	2
S	3	3
Totals	21	5

Problem II

	I	G_4
NS	21	2
S	0	3
Totals	21	5

Problem III

	I	G_4
NS	19	3
S	2	2
Totals	21	5

TABLE 2

Expected Values and Goodness-of-fit Statistics for the Shaw Data.

Problem I

	I	G_4
NS	17.48	2.40
S	3.52	2.60
Totals	21.00	5.00

$G^2 = 0.226$ ($\alpha = .64$)

$X^2 = 0.306$ ($\alpha = .58$)

Problem II

	I	G_4
NS	19.26	3.54
S	1.74	1.46
Totals	21.00	5.00

$G^2 = 5.663$ ($\alpha = .017$)

$X^2 = 4.823$ ($\alpha = .028$)

Problem III

	I	G_4
NS	18.77	3.19
S	2.23	1.81
Totals	21.00	5.00

$G^2 = 0.059$ ($\alpha = .81$)

$X^2 = 0.070$ ($\alpha = .72$)

are each asymptotically distributed as a X^2 variable with 1 degree of freedom, where O is the observed value, E the expected value based on the maximem likelihood estimate, and the summation is over all four cells of the 2x2 table. The test statistics X^2 and G^2 are goodness-of-fit statistics used to test the fit of H_0 against unrestricted alternatives.

For the Shaw data in Table 1 the maximum likelihood estimates of p are 0.832, 0.917, and 0.894 for Problems I, II, and III respectively. Table 2 contains the expected values under H_0 for each of the three problems along with the values of the goodness-of-fit statistics G^2 and X^2.

3. EXTENSION OF MODEL TO TRICHOTOMIES

The model discussed in Section 2 can be extended to the case of a trichotomy with comparison of individuals and groups of size two. The extension was suggested by data gathered by Staub [1970] on the development of helping patterns in children. In the Staub experiment, kindergarten, first, second, fourth, and sixth grade children, alone or in same-sex pairs, heard sounds of another child's distress from an adjoining room. One of three responses was recorded for each individual or pair. The three responses were 1) NO HELP, 2) VOLUNTEER HELP, and 3) ACTIVE HELP. An excerpt of the recorded data are given in Table 4 under the "OBSERVED" heading.

Looking at the observations for each grade-sex classification separately, the observed table for any particular grade-sex classification is of the form given by Table 3. There are n_1 individuals who gave no help, n_4 pairs that gave no help, etc.

TABLE 3

Observed Table for a Classification

	Ind	Pairs
NH	n_1	n_4
VH	n_2	n_5
AH	n_3	n_6
	N_I	N_G

Assume that there are N_I individuals and the probability of observing (n_1,n_2,n_3), where $n_1+n_2+n_3 = N_I$, is governed by a multinomial distribution with probabilities (p_1,p_2,p_3), where $p_1+p_2+p_3 = 1$. For pairs, assume that there are N_G pairs and the probability of observing (n_4,n_5,n_6), where $n_4+n_5+n_6 = N_G$, is also multinomial, but with probabilities (g_1,g_2,g_3), where $g_1+g_2+g_3 = 1$.

Assuming that a pair consists of two randomly chosen individuals and that the manner in which pairs act is dominated by the more helpful individual, the probabilities for pairs become

$$(3.1)\qquad \begin{aligned} H_0:\ g_1 &= p_1^2 \\ g_2 &= p_2^2 + 2p_1p_2 \\ g_3 &= p_3^2 + 2p_1p_3+2p_2p_3. \end{aligned}$$

Under hypothesis H_0, the parameters (p_1,p_2,p_3) can be estimated by maximum likelihood and the statistics (2.3) and (2.4) can be used to judge the fit of the model. Solving the maximum likelihood equations yields the estimates

$$(3.2)\qquad \begin{aligned} \hat{p}_1 &= \frac{-n_3 + \sqrt{n_3^2 + 4ac}}{2a} \\ \hat{p}_2 &= r\hat{p}_1 \\ \hat{p}_3 &= 1 - (1+r)\hat{p}_1 \end{aligned}$$

TABLE 4

Observed and Expected Values and Goodness-of-fit Statistics for the Staub Data

		Observed I	Observed P	Expected I	Expected P	X^2	G^2
Kindergarten Boys	NH	7	3	5.75	4.14	2.41	3.54
	VH	0	3	1.20	1.91		
	AH	1	2	1.05	1.95		
		8	8	8	8		
Kindergarten Girls	NH	6	5	6.19	4.79	0.83	1.14
	VH	2	2	1.47	2.54		
	AH	0	1	0.34	0.67		
		8	8	8	8		
4th Grade Boys	NH	6	2	4.96	2.69	3.41	4.94
	VH	2	1	1.42	1.77		
	AH	0	4	1.62	2.54		
		8	7	8	7		
4th Grade Girls	NH	2	6	5.38	3.16	11.19	11.25
	VH	1	0	0.38	0.47		
	AH	5	1	2.24	3.37		
		8	7	8	7		
6th Grade Boys	NH	9	5	8.86	5.20	2.40	3.12
	VH	0	2	0.89	1.09		
	AH	2	1	1.25	1.71		
		11	8	11	8		
6th Grade Girls	NH	7	6	6.96	6.06	0.78	1.10
	VH	0	1	0.36	0.64		
	AH	1	1	0.68	1.30		
		8	8	8	8		

where

$$(3.3) \qquad r = \frac{n_2-2n_1-4n_4+S}{2(n_1+2n_4)} ,$$

$$(3.4) \quad S = \sqrt{(2n_1+4n_4-n_2)^2+8(n_2+n_5)(n_1+2n_4)}$$

$$(3.5) \quad a = (1+r)[(n_1+2n_4)(1+r) + (n_3+2n_6) + \frac{2n_5}{2+r}(1+r)],$$

and

$$(3.6) \qquad c = n_1+2n_4 + \frac{2n_5}{2+r} .$$

In Table 4, under the expected heading, the fitted values for the excerpt of the Staub data are given. The X^2 and G^2 goodness-of-fit statistics are also included. Both sets of statistics when compared with a chi-square variable on 2 d.f. indicate that H_0 fits the observed data well, except for fourth grade girls. In fact, fourth grade girls helped more as individuals than they did in pairs.

To examine the effect of grade and sex on the applicability of H_0, one can make use of methods for partitioning likelihood ratio test statistics. Details are given in Fienberg and Larntz [1971].

4. SUMMARY

In the comparison of group versus individual behavior of various sorts, a model of interest is the one which postulates that differences in response can be accounted for simply by the fact that groups consist of several individuals. This model has been applied to two different sets of data in this paper, using the method of maximum likelihood and the standard chi-square goodness-of-fit criteria. Complete details for the analysis of these data are given in Fienberg and Larntz [1971], where alternative models are presented and the small sample behavior of the likelihood ratio goodness-of-fit statistic is discussed.

REFERENCES

[1] Fienberg, S. E. and Larntz, K., Some Models for Individual-Group Comparisons and Group Behavior. To appear in Psychometrika, (1971).

[2] Lorge, I. and Solomon, H., Two models of group behavior in the solution of Eureka-type problems. Psychometrika, 20 (1955), 139-148.

[3] Shaw, M. E., Comparison of individuals and small groups in the rational solution of complex problems. American Journal of Psychology, 44 (1932), 491-504.

[4] Staub, E., A child in distress: the influence of age and number of witnesses on children's attempts to help. Journal of Personality and Social Psychology, 14 (1970), 130-140.

ASPECTS OF INTER-INSTITUTIONAL MOBILITY PATTERNS OF CHEMISTS IN HIGHER EDUCATION

Eun Sul Lee, The University of Texas School of Public Health
Charles H. Proctor, North Carolina State University

1. INTRODUCTION

This study is concerned with mobility of chemists among colleges and universities in the academic training phase as well as in the post-doctoral phase of their professional careers. Patterns of mobility are examined in terms of geographic and social dimensions of the academic stratification system, namely, geographic location and prestige structure of universities.

In their studies of the academic community, sociologists have emphasized its hierarchical structure [5,15]. The hierarchy is usually expressed in terms of institutional "quality" or "prestige". It has been maintained that the mobility of scholars in the academic community follows more the endogamous pattern than the stepladder pattern [7]. On the other hand, it has been common practice for demographers to emphasize the geographic location in their studies of population mobility. Many migration models which use the gravity concept have dealt with mathematical formulations of the relationship between migration and distance [17,19]. However, few studies have attempted to study the degree to which a combination of social and geographic views can account for observed patterns of mobility in the academic community. It is the purpose of this paper to examine these two aspects of mobility pattern by applying a statistical model which has been recently used in the analysis of various types of transaction flow data [2,9,16].

Mobility is defined in this study as a change of institutional affiliation. Three types of mobility emerge upon linking such points of career development as baccalaureate graduation, doctorate graduation, and employment: (1) Baccalaureate-to-doctorate mobility, (2) Doctorate-to-employment mobility, and (3) post-doctoral job mobility.

2. DATA AND SAMPLING PROCEDURE

The data for the study are obtained from the American Chemical Society's Directory of Graduate Research which has been published every two years since 1955.

The population to be studied consists of all faculty members affiliated with chemistry departments offering graduate degree programs in institutions of higher education in the United States. The study sample was selected from seven successive editions of directories, from 1955 to 1967, using a single-stage cluster sampling procedure. The alphabetical listing of the latest directory was first subdivided into 80 clusters and a random sample of 20 clusters were selected. Each cluster is a span of alphabetical listing with unique starting and ending points. These 20 spans of alphabetical listing are canvassed in each edition of directories. The sample, then, consists of all chemists included in these 20 spans in the alphabetical listing in at least one edition of the directories. Thus, included in the sample are those who left the profession as well as those who newly joined the profession during the study period. A total of 1,128 unduplicated individual names were selected following through the seven successive editions of the directory. Of the 1,128 names 244 appear in all seven successive directories and 215 appear only one time.

In order to obtain information about the mobility of chemists, sociometric-type matrices describing the flow of scholars among the 86 institutions were formed for the three types of mobility mentioned above. The 86 institutions were selected based on a criterion that each of them employed at least 15 faculty members and produced 5 or more doctorates in the year 1966-67.

3. METHOD OF ANALYSIS

In a mobility matrix, the diagonal cells represent stayers who did not change their institutional affiliation and off-diagonal cells show movers who changed their institutional affiliation during the period of study. In order not to misinterpret the data, we must use a model that actually describes the phenomena under investigation or is a sufficiently close approximation to it. Blumen and others [3,8] demonstrated that the separation of movers and stayers improved the fitness of their Markov model in their study of the movement of workers among various industrial catagories. Goodman [11] also showed that a "quasi-perfect mobility" model which deals with movers only separating out stayers fits to the actual data much better than the "perfect mobility" model which describes the whole mobility table including stayers. Therefore, in this study an attempt was made to distinguish the movers and the stayers and treated them separately.

The 86 x 86 mobility matrices excluding the diagonal cells were analyzed by examining the departures of the observed frequencies from the expected frequencies of movements. The expected frequencies are derived from a "quasi-perfect mobility" model which assigns zero frequencies along the main diagonal of the matrix and, subject to this constraint, calculates expected values on the assumption of no association between the institution of origin and the institution of destination.

This statistical model was first developed by Savage and Deutsch to analyze international trade data and an iterative technique was used in the maximum likelihood estimation procedure [1,16]. Subsequently Goodman presented a generalization of the model with an alternative iterative procedure [9, 10, 12]. A similar model dealing with incomplete two-dimensional contingency tables was also presented by Bishop and Fienberg [2]. Recently Wagner [18] presented more elaborate discussion on the maximum likelihood estimate for contingency tables with zero diagonal. It is to be noted in this connection that this type of problem can also be handled by a general noniterative procedure presented by Grizzle and others[13].

Because of the small number of cases upon which the analysis was based, it would be inappropriate to interpret and generalize the results based on larger tables of the order of 86 x 86. The analysis of mobility patterns are based on pooled frequencies, observed and expected, from the larger tables according to a combination of regional and prestige level groupings of institutions. Based on the U.S. Census Regions, five regional groupings are used: New England, Middle Atlantic, Midwest, South and West. Four levels of prestige groupings are used based on the American Council on Education study on quality rating of graduate faculty as reported by Carter [6]. By using these regional and prestige groupings, four basic components of mobility are identified: (1) Intra-regional and same prestige level mobility; (2) Intra-regional and different prestige level mobility; (3) Inter-regional and same prestige level mobility; and (4) Inter-regional and different prestige level mobility.

In addition to these four components based on the movers, there is the fifth component representing the stayers. Analysis of the stayer component was done using a different model independently of the analysis of the movers. The expected frequencies for the stayers in the diagonal cells were computed based on the "perfect mobility" model which uses the same procedure in ordinary complete two-dimensional contingency table analysis.

4. FINDINGS

4.1 Baccalaureate-to-Doctorate Mobility

Table 1 presents data on the baccalaureate-to-doctorate mobility among the 86 institutions. The table shows the four components of mobility with the fifth component below the dividing line and each component is further divided into subcomponents. The figure at the first column corresponding to each component represents the percent of the movers with respect to that component. The second column shows the percentage of the movers which would be expected for each component based on the quasi-perfect mobility model mentioned above. Since the first four components are computed on the same base, these four add up to 100 percent. The fifth component is based on a different base, namely, the total including movers and stayers. The third column in the table presents a crude measure of the extent to which the observed values for each component deviate from those which would be expected. The data are presented in the same manner throughout this paper.

It is shown in Table 1 that 35 percent (addition of first two components) of movers moved within the same region. This percentage is much higher than would be expected. It is also shown that 60 percent (addition of the first and third components) of movers moved to institutions of the same prestige level as their baccalaureate institutions, and this figure is slightly higher than the expected percentage. Within the same region, horizontal mobility with respect to prestige level far exceeds expected mobility, while vertical mobility is slightly higher than expected mobility. The stayer component shows that more than 30 percent of the total (movers plus stayers) stayed at the same institution where they received the baccalaureate, and this figure is much higher than would be expected. Thus, the selective tendencies displayed here suggest that the academic stratification system in the baccalaureate-to-doctorate mobility is better represented as a set of regional hierarchies rather than a strict prestige hierarchy.

Table 1. Baccalaureate-to-Doctorate Mobility Patterns of Chemists

Component of Mobility	Observed Percentage	Expected Percentage	Difference (Ob-Exp)
Intra-regional			
(I) Horizontal	21.7%	11.1%	+10.6%
Top level	(19.2)	(10.5)	(+8.7)
Lower levels	(2.5)	(0.6)	(+1.9)
(II) Vertical	13.0	8.3	+4.7
Upward	(8.6)	(5.8)	(+2.8)
Downward	(4.4)	(2.5)	(+1.9)
Inter-regional			
(III) Horizontal	38.0	43.3	-5.3
Top level	(36.7)	(41.7)	(-5.0)
Lower levels	(1.3)	(1.6)	(-0.3)
(IV) Vertical	27.3	37.2	-9.9
Upward	(21.7)	(27.8)	(-6.1)
Downward	(5.6)	(9.4)	(-3.8)
(V) Stayers	30.4	3.1	+27.3
Top level	(22.8)	(2.9)	(+19.9)
Lower levels	(7.6)	(0.2)	(+7.4)
Number of movers		360	
Number of Stayers		157	
Total		517	

The general patterns revealed above may not characterize each region and each prestige level considered separately. In order to investigate this possibility, each region and each prestige level is separately considered. The data in Table 2 reveal that the general tendencies observed above exist in every region, although the sizes of the differences between observed and expected percentages vary from one region to another. For example, the first row shows that the intraregional, horizontal mobility tendency is weakest in the South and the Middle Atlantic, but they will have more mobility than expected in this respect. The intraregional, vertical mobility shown in the second row also exceeds the expected mobility with the exception of New England where it is balanced. On the other hand, the percentage differences for interregional mobility are all negative in sign, with the lone exception of the South. But the plus sign appears only in horizontal mobility. In other words, the South is sending its baccalaureate graduates to institutions of the same prestige level in other regions for their doctorate slightly more often than would be expected, while interregional, vertical mobility is still far less than the expected mobility. The tendency to remain at the same institution is strongest in the

Table 2. Analysis of Regional Differences in Baccalaureate-to- Doctorate Mobility Patterns

Component of Mobility	New England Ob	New England Exp	New England Dif	Middle Atlantic Ob	Middle Atlantic Exp	Middle Atlantic Dif	Midwest Ob	Midwest Exp	Midwest Dif	South Ob	South Exp	South Dif	West Ob	West Exp	West Dif
Intra-regional															
(I) Horizontal	24.5	12.5	+12.0	10.5	5.0	+5.5	32.1	19.4	+12.7	3.4	1.4	+2.0	30.9	11.8	+19.1
(II) Vertical	2.0	2.0	0.0	11.8	6.8	+5.0	19.3	14.3	+5.0	10.4	5.1	+5.3	14.7	8.7	+6.0
Inter-regional															
(III) Horizontal	63.3	63.9	-0.6	40.8	43.8	-3.0	32.1	40.1	-8.0	22.4	21.4	+1.0	39.7	51.9	-12.2
(IV) Vertical	10.2	21.6	-11.4	36.9	44.4	-7.5	16.5	26.2	-9.7	63.8	72.1	-8.3	14.7	27.6	-12.9
(V) Stayers	33.8	7.8	+26.0	31.5	1.8	+29.7	34.3	2.8	+31.5	8.7	0.2	+8.5	26.9	4.3	+22.6
Number of movers	49			76			109			58			68		
Nuhber of stayers	25			35			57			15			25		
Total	74			111			166			73			93		

Table 3. Analysis of Prestige Level Differences in Baccalaureate-to-Doctorate Mobility Patterns

Component of Mobility	Level 1 (High) Ob	Level 1 (High) Exp	Level 1 (High) Dif	Level 2 Ob	Level 2 Exp	Level 2 Dif	Level 3 Ob	Level 3 Exp	Level 3 Dif	Level 4 (Low) Ob	Level 4 (Low) Exp	Level 4 (Low) Dif
Intra-regional												
(I) Horizontal	30.4	16.7	+13.7	11.3	1.9	+9.4	4.1	0.8	+3.3	3.2	1.3	+1.9
(II) Vertical	5.7	3.5	+2.2	28.3	20.4	+7.9	30.6	18.2	+12.4	12.9	7.7	+5.2
Inter-regional												
(III) Horizontal	58.1	66.5	-8.4	5.7	7.0	-1.3	2.0	2.8	-0.8	3.2	1.3	+1.9
(IV) Vertical	5.8	13.3	-7.5	54.7	70.7	-16.0	63.3	78.2	-14.9	80.7	89.7	-9.0
(V) Stayers	34.2	4.3	+29.9	29.3	0.8	+28.5	21.0	0.3	+20.7	11.4	0.8	+10.6
Number of movers		227			53			49			31	
Number of stayers		118			22			13			4	
Total		345			75			62			35	

Table 4. Analysis of Trends in Baccalaureate-to-Doctorate Mobility Patterns

Component of Mobility	Old Cohort* Observed	Old Cohort* Expected	Old Cohort* Difference	Young Cohort** Observed	Young Cohort** Expected	Young Cohort** Difference
Intra-regional						
(I) Horizontal	24.4%	12.2%	+12.2%	19.6%	10.3%	+9.3
Top level	(23.7)	(11.5)	(+12.2)	(15.7)	(9.8)	(+5.9)
Lower levels	(0.7)	(0.7)	(0.0)	(3.9)	(0.5)	(+3.4)
(II) Vertical	12.8	9.0	+3.8	13.2	7.8	+5.4
Upward	(10.3)	(7.1)	(+3.2)	(7.4)	(4.9)	(+2.5)
Downward	(2.5)	(1.9)	(+0.6)	(5.8)	(2.9)	(+2.9)
Inter-regional						
(III) Horizontal	34.6	43.6	-9.0	40.7	43.1	-2.4
Top level	(32.7)	(42.3)	(-9.6)	(39.7)	(41.2)	(-1.5)
Lower levels	(1.9)	(1.3)	(+0.6)	(1.0)	(1.9)	(-0.9)
(IV) Vertical	28.2	35.2	-7.0	26.5	38.8	-12.3
Upward	(22.4)	(28.2)	(-5.8)	(21.1)	(27.5)	(-6.4)
Downward	(5.8)	(7.0)	(-1.2)	(5.4)	(11.3)	(-5.9)
(V) Stayers	39.3	3.1	+36.2	21.5	3.0	+18.5
Top level	(30.4)	(2.9)	(+27.5)	(15.4)	(2.8)	(+12.6)
Lower levels	(8.9)	(0.2)	(+8.7)	(6.1)	(0.2)	(+5.9)
Number of movers		156			204	
Nubmer of stayers		101			56	
Total		257			260	

*Received bachelor's degree before 1950 **Received bachelor's degree in 1950 or after

Midwest where the "Big Ten" schools are located, and it is weakest in the South which has few prestigious doctorate producing institutions.

The data presented in Table 3 indicate that, at each prestige level of baccalaureate institutions, the movement to doctoral institutions tend to be oriented toward both the same region and the same prestige level. These tendencies are consistent with the ones observed above. The patterns of mobility at the lowest prestige level are similar to those observed for the South in Table 2. Examining the third row, it appears that institutions in the lower prestige levels tend to send their baccalaureate graduates to the same prestige level but in different regions. This interregional, horizontal mobility at the lowest level is, in fact, slightly more than would be expected but this is not the case at higher prestige levels. The rate of remaining in the same region and prestige level tends to be lower at lower prestige levels. The rate of remaining at the same institution also becomes lower as the prestige level decreases.

In order to investigate possible trends in the baccalaureate-to-doctorate mobility patterns, the young and old scholars are analyzed separately. The results of this investigation are presented in Table 4. The data reveal that both the young and the old cohort do not deviate from the general mobility pattern oriented toward the same region and the same prestige level. There is some indication that the regionalistic orientation is gradually decreasing. The intraregional and horizonal tendencies are somewhat weaker for the young cohort. The tendency of remaining at the same institution is less profound for the young cohort than for the old cohort.

4.2 Doctorate-to-Employment Mobility

An analysis of mobility patterns from the institution of doctorate training to the institution of employment is presented in Table 5. Of those who were not inbred into their doctoral institutions, 14 percent obtained positions in the same region and the same prestige level as their doctoral institutions. This observed percentage is somewhat higher than the expected percentage. Twenty-one percent of them moved to institutions of different prestige levels in the same region, and it is also higher than the expected percentage. Here again the tendency toward intraregional mobility is apparent and it is consistent with the tendency observed in the mobility from the institution of baccalaureate training to the institution of doctorate training. Examining the first and the third components together, it is found that the horizontal mobility with respect to the prestige level accounts for 39 percent of the total movers. This observed percentage is only slightly higher than the expected percentage. Thus, regionalistic tendencies appear to be stronger than the selective prestige level tendencies.

Another feature of the data presented in Table 5 appears in the second and the fourth components. Although the intraregional, vertical mobility as a whole exceeds the expected mobility, the upward mobility (decomposition of the intraregional, vertical mobility) shows the negative deviation from

Table 5. Doctorate-to-Employment Mobility Patterns of Chemists

Component of Mobility	Observed Percentage	Expected Percentage	Difference (Ob-Exp)
Intra-regional			
(I) Horizontal	13.6%	8.8%	+4.8%
Top level	(12.8)	(8.0)	(+4.8)
Lower levels	(0.8)	(0.8)	(0.0)
(II) Vertical	20.8	12.3	+8.5
Upward	(1.5)	(1.9)	(-0.4)
Downward	(19.3)	(10.4)	(+8.9)
Inter-regional			
(III) Horizontal	25.1	27.2	-2.1
Top level	(23.0)	(24.8)	(-1.8)
Lower levels	(2.1)	(2.4)	(-0.3)
(IV) Vertical	40.5	51.7	-11.2
Upward	(3.2)	(6.2)	(-3.0)
Downward	(37.3)	(45.5)	(-8.2)
(V) Stayers	10.3	1.3	+9.0
Top level	(6.2)	(1.1)	(+5.1)
Lower levels	(4.1)	(0.2)	(+3.9)
Number of movers		617	
Number of stayers		71	
		688	

the expected mobility. This tendency toward downward mobility may be considered as normal, since new doctorates tend to move down from the prestige level of their doctoral institutions during their early careers [5, P. 181]. In fact, 57 percent, examining the first column, of the movement was downward mobility, while 39 percent remained at the same prestige level and only four percent moved upward. The fifth component in Table 5 indicates institutional inbreeding patterns. One out of ten scholars held positions at the same institution where he received his doctorate. This rate of institutional inbreeding is much higher than would be expected.

An analysis of regional differences in the doctorate to employment mobility is presented in Table 6. First of all, a comparison of the second and the third components of each region indicates that selective regional tendencies are stronger than selective prestige level tendencies for each region. In all five regions, the observed percentage of the movers who remained in that region but moved to a different prestige level exceeds the expected percentage in this component. On the other hand, the observed percentage of movers who moved to a different region but remained in the same prestige level is lower than the expected percentage in all regions with two exceptions - New England and the West. Although the general patterns are similar in each region, there emerge some regional peculiarities. For example, the South retained 67 percent of its doctorate graduates who are not inbred into their doctoral institutions, while New England retained only 21 percent in this respect. The South also had the highest rate of inbreeding (18%) and the lowest rate of inbreeding is found in New England. Thus, the data suggest that the regions with predominantly lower prestige institutions tend to have higher inbreeding rates and stronger intraregional tendencies.

Table 6. Analysis of Regional Differences in Doctorate-to-Employment Mobility Patterns

Component of Mobility	New England Ob	New England Exp	New England Dif	Middle Atlantic Ob	Middle Atlantic Exp	Middle Atlantic Dif	Midwest Ob	Midwest Exp	Midwest Dif	South Ob	South Exp	South Dif	West Ob	West Exp	West Dif
Intra-regional															
(I) Horizontal	13.3	5.6	+7.7	6.5	4.6	+1.9	18.1	14.2	+3.9	10.9	4.4	+6.5	10.9	5.0	+5.9
(II) Vertical	7.5	4.1	+3.4	16.3	9.9	+6.4	23.3	16.9	+6.4	56.5	20.4	+36.1	18.2	9.5	+8.7
Inter-regional															
(III) Horizontal	35.8	33.4	+2.4	27.2	30.1	-2.9	18.5	21.9	-3.4	8.7	23.5	-14.8	33.6	31.8	+1.8
(IV) Vertical	43.4	56.9	-13.5	50.0	55.4	-5.4	40.1	47.0	-6.9	23.9	51.7	-27.8	37.3	53.7	-16.4
(V) Stayers	5.5	1.0	+4.5	16.4	1.0	+15.4	8.1	1.4	+6.7	17.8	0.9	+16.9	11.3	1.5	+9.8
Number of movers		120			92			249			46			110	
Number of stayers		7			18			22			10			14	
Total		127			110			271			56			124	

Table 7. Analysis of Prestige Level Differences in Doctorate-to-Employment Mobility

Component of Mobility	Level 1 (High) Ob	Level 1 (High) Exp	Level 1 (High) Dif	Level 2 Ob	Level 2 Exp	Level 2 Dif	Levels 3 & 4 (Low) Ob	Levels 3 & 4 (Low) Exp	Levels 3 & 4 (Low) Dif
Intra-regional									
(I) Horizontal	15.3	9.5	+5.8	1.5	4.6	-3.1	11.4	5.1	+6.3
(II) Vertical	18.5	11.2	+7.3	27.7	18.5	+9.2	40.0	20.0	+20.0
Inter-regional									
(III) Horizontal	27.5	29.6	-2.1	16.9	15.4	+1.5	5.7	12.0	-6.3
(IV) Vertical	38.7	49.7	-11.0	53.9	61.5	-7.6	42.9	62.9	-20.0
(V) Stayers	7.7	1.4	+6.3	17.7	0.6	+17.1	28.8	0.6	+28.2
Number of movers		517			65			35	
Number of stayers		43			14			14	
Total		560			79			49	

Table 8. Analysis of Trends in Doctorate-to-Employment Mobility Patterns

Component of Mobility	Old Cohort* Observed	Old Cohort* Expected	Old Cohort* Difference	Young Cohort** Observed	Young Cohort** Expected	Young Cohort** Difference
Intra-regional						
(I) Horizontal	16.1	10.0	+6.1	11.1	7.5	+3.6
Top level	(15.5)	(9.0)	(+6.5)	(10.1)	(6.8)	(+3.3)
Lower levels	(0.6)	(1.0)	(-0.4)	(1.0)	(0.7)	(+0.3)
(II) Vertical	22.3	12.9	+9.4	19.2	11.7	+7.5
Top level	(4.2)	(1.9)	(+2.3)	(1.6)	(1.9)	(-0.3)
Lower levels	(18.1)	(11.0)	(+7.1)	(17.6)	(9.8)	(+7.8)
Inter-regional						
(III) Horizontal	22.9	28.1	-5.2	27.4	26.4	+1.0
Top level	(21.6)	(26.1)	(-4.5)	(24.4)	(23.4)	(+1.0)
Lower levels	(1.3)	(2.0)	(-0.7)	(3.0)	(3.0)	(0.0)
(IV) Vertical	38.7	49.0	-10.3	42.3	54.4	-12.1
Upward	(3.5)	(5.1)	(-1.6)	(2.9)	(7.2)	(-4.3)
Downward	(35.2)	(43.9)	(-8.7)	(39.4)	(47.2)	(-7.8)
(V) Stayers	15.3	1.9	+13.4	4.7	0.6	+4.1
Top level	(10.1)	(1.8)	(+8.3)	(1.9)	(0.3)	(+1.6)
Lower levels	(5.2)	(0.1)	(+5.1)	(2.8)	(0.3)	(+2.5)
Number of movers		310			307	
Number of stayers		56			15	
Total		366			322	

*Received the Doctorate before 1955 **Received the Doctorate in 1955 or after

Prestige level differences in mobility patterns of doctorates who obtained positions in another institution of higher education are examined in Table 7. Stronger intraregional than intraprestige level tendencies are consistently displayed in each prestige level. Apart from these general tendencies the data show that institutions in the lower prestige levels tend to retain their graduates more in their own institutions and send more to institutions in the same region than institutions in the higher prestige levels. The rate of inbreeding is eight percent at the highest prestige level, while the rate is more than three times higher (29%) at the lowest level.

An Analysis of mobility patterns for the young and old cohorts is given in Table 8. The results show that the general tendencies toward intraregional and intraprestige level mobility are consistent in both young and old cohorts. Although the differences between the corresponding percentages for the two cohorts on each component are quite small, there is some indication that the tendency toward intraregional mobility is more prominent for the old cohort than for the young. The rate of inbreeding is much higher for the old cohort (15%) than for the young (5%). While there is more downward mobility than upward mobility in both cohorts, the old cohort shows more upward mobility than the young cohort. These slight differences between the young and old cohorts may reflect the development of postdoctoral careers of scholars. At earlier stages of their careers, scholars are more willing to accept jobs at lower prestige institutions than their doctoral institutions and more willing to move to other regions. As their professional experiences increase, they tend to move upward to more prestigious institutions and many of them return to their Alma Mater.

4.3 Postdoctoral Job Mobility

In order to obtain information about postdoctoral job mobility, six origin-destination matrices were formulated for each of the six biennial periods. Although it is possible to analyze each of these matrices separately, all six matrices were combined due to insufficient number of cases in each matrix. Thus the fifth component is not considered in this section. The total number of moves identified among the 86 institutions during the entire 12 year period adds up to only 83.

Analysis of job mobility patterns during the 1955-1967 period is presented in Table 9. The positive differences in the first two components indicate excess intraregional mobility and the positive sign in the first and third component reflect excess horizontal mobility with respect to prestige level. About the same magnitude of differences between the observed and expected percentage in the second and the third components suggest that selective prestige level tendencies are as strong as selective regional tendencies in the job mobility.

Regional differences in the job mobility patterns are analyzed in Table 10. Although the small number of cases in each region do not allow any detailed observations, selective prestige level tendencies seems to be stronger than regionalistic tendencies in all regions except for the Middle Atlantic. Those who were employed in the Middle Atlantic seem to prefer to move to a different prestige level in that region rather than moving to the same prestige level in other regions.

Table 9. Postdoctoral Job Mobility Patterns

Component of Mobility	Observed Percentage	Expected Percentage	Difference (Ob-Exp)
Intra-regional			
(I) Horizontal	8.4%	7.5%	+0.9%
Top level	(4.8)	(5.5)	(-0.7)
Lower levels	(3.6)	(2.0)	(+1.6)
(II) Vertical	21.7	13.9	+7.8
Upward	(4.8)	(4.5)	(+0.3)
Downward	(16.9)	(9.4)	(+7.5)
Inter-regional			
(III) Horizontal	33.7	26.1	+7.6
Top level	(28.9)	(20.7)	(+8.2)
Lower levels	(4.8)	(5.4)	(-0.6)
(IV) Vertical	36.2	52.5	-16.3
Upward	(9.6)	(18.3)	(-8.7)
Downward	(26.6)	(34.2)	(-7.6)
Number of moves		83	

Prestige level differences are considered in Table 11. Institutions in lower prestige levels are more strongly oriented toward intraregional mobility than institutions at the highest level. Scholars in higher prestige level institutions tend to move to institutions in the same prestige level in other regions.

An attempt is made in Table 12 to investigate any possible trends in job mobility by carrying out separate analysis for the 1955-1961 and 1961-1967 periods. The corresponding percentages for the two periods are quite similar. Thus, there seems to be no marked change in the job mobility patterns.

5. SUMMARY AND CONCLUSIONS

The results reported in Tables 1 through 12 suggest the following conclusions about interinstitutional mobility patterns of chemists at various stages of their careers:

(1). The mobility from baccalaureate to doctorate training was characterized by stronger tendencies toward regionalism than toward prestige level homogeneity. It appeared that for their doctoral study, students tended to stay at the same institution or move to institutions of the same prestige level in the same region. Consequently, interregional and interprestige level mobilities were less than would be expected based on the quasi-perfect mobility model. The regionalistic tendencies seem to be stronger for the institutions at the lower prestige level.

(2). The mobility from the doctoral institution to the institutuion of employment was also oriented toward stronger regionalistic tendencies than selective prestige level tendencies. Thus, the academic stratification system in doctorate-to-employment mobility can be said to be a set of regional hierarchies rather than a rigid prestige hierarchy. It was noted that downward mobility was more common, especially at earlier stages of postdoctoral careers. Institutions at the lower prestige level had a relatively higher rate of inbreeding and a stronger regionalistic orientation.

Table 10. Analysis of Regional Differences in Postdoctoral Job Mobility Patterns

Component of Mobility	New England Ob	New England Exp	New England Dif	Middle Atlantic Ob	Middle Atlantic Exp	Middle Atlantic Dif	Midwest Ob	Midwest Exp	Midwest Dif	South Ob	South Exp	South Dif	West Ob	West Exp	West Dif
Intra-regional															
(I) Horizontal	9.1	7.3	+1.8	5.6	3.9	+1.7	14.8	11.9	+2.9	0.0	4.3	-4.3	7.7	9.2	-1.5
(II) Vertical	18.2	0.9	+17.3	33.3	10.6	+22.7	14.8	22.2	-7.4	21.4	14.3	+7.1	23.1	11.5	+11.6
Inter-regional															
(III) Horizontal	54.5	33.6	+20.9	16.7	38.3	-21.6	33.3	21.1	+12.2	28.6	20.0	+8.6	46.1	30.8	+15.3
(IV) Vertical	18.2	58.2	-40.0	44.4	47.2	-2.8	37.1	44.8	-7.7	50.0	61.4	-11.4	23.1	48.5	-25.4
Number of moves		11			18			27			14			13	

Table 11. Analysis of Prestige Level Differences in Postdoctoral Job Mobility Patterns

Component of Mobility	Level 1 (High) Ob	Level 1 (High) Exp	Level 1 (High) Dif	Lower levels Ob	Lower levels Exp	Lower levels Dif
Intra-regional						
(I) Horizontal	7.8	9.0	-1.8	9.4	5.0	+4.4
(II) Vertical	17.6	12.0	+5.6	28.1	16.5	+11.6
Inter-regional						
(III) Horizontal	47.1	33.9	+13.2	12.5	13.8	-1.3
(IV) Vertical	27.5	45.1	-17.6	50.0	64.7	-14.7
Number of moves		51			32	

Table 12. Analysis of Trends in Postdoctoral Job Mobility Patterns

Component of Mobility	1955-61 Period Ob	1955-61 Period Exp	1955-61 Period Dif	1961-67 Period Ob	1961-67 Period Exp	1961-67 Period Dif
Intra-regional						
(I) Horizontal	9.5	6.6	+2.9	7.7	7.0	+0.7
(II) Vertical	27.6	18.9	+8.7	19.2	13.0	+6.2
Inter-regional						
(III) Horizontal	23.8	21.4	+2.4	39.5	32.8	+6.7
(IV) Vertical	39.1	53.1	-14.0	33.6	47.2	-13.6
Number of moves		29			54	

(3). The postdoctoral job mobility patterns were characterized by about equally strong tendencies of regionalism and prestige homogeneity. The regionalistic tendencies were relatively stronger for institutions at the lower prestige levels and the tendencies toward homogeneous prestige were more operative for institutions at the higher prestige levels. Thus, scholars in the institutions at the higher levels tended to move to institutions of the same prestige level in other regions, while those in the institutions at a lower prestige level tended to move within that region.

A fairly strong tendency toward intra-regional mobility has been reported in some studies of mobility of scientific manpower [14]. From an economist's viewpoint Brown [4] describes the academic market place as a series of submarkets partially isolated from each other by geography and other characteristics. These views are consistent with the findings of this study. Although mobility exerts a significant influence in equalizing the distribution of talent, both quantitatively and qualitatively, there exists a consistent tendency toward regionalistic stratification of those doctoral trained chemists who are employed in the academic setting. Implications of these regionalistic tendencies in institutional mobility in various phases of career development of scholars would be the persistence of regional inequalities in the quality of higher education. Regional tendencies would impose restrictions on the development of institutions at the lower spectrum of quality and newly emerging universities. Regions where higher education is comparatively less effective are likely to remain so if migration is allowed to be the primary equilibrium force.

As to the question whether the regionalistic tendency might be decreasing, this study did not provide any substantial evidence. Further efforts to investigate trends in these patterns over time would be relevant. Finally, it should be pointed out that the small number of cases has imposed certain limitations to the findings in this study. It is hoped that the findings of this study will stimulate a larger study along these lines.

REFERENCES

[1] Alker, Haywood R., Jr., "An IBM 709 Program for the Gross Analysis of Transaction Flows", Behavioral Science, 7 (1962), 498-499.

[2] Bishop, Yvonne M., and Stephen E. Fienberg, "Incomplete Two-Dimensional Contingency Table", Biometrics, 25 (1969), 119-128.

[3] Blumen, Isadore, Marvin Kogan, and Philip J. McCarthy, The Industrial Mobility of Labor as a Probability Process, Cornell Studies of Industrial and Labor Relations, Vol. 4, Ithaca, N.Y.: The New York State School of Industrial and Labor Relations, Cornell University, 1966.

[4] Brown, David G., The Mobile Professors, Washington, D.C.: American Council on Education, 1967.

[5] Caplow, Theodore and Reece McGee, The Academic Marketplace, New York: Basic Books, Inc., 1958.

[6] Carter, Allan M., An Assessment of Quality in Graduate Education, Washington, D. C.: American Council on Education, 1966.

[7] Ferriss, Abbott L., "A hypothesis on institutional mobility of teachers in higher education", College and University, 42 (1966), 13-28.

[8] Goodman, Leo A., "Statistical methods for the mover-stayer model", Journal of the American Statistical Association, 56 (1961), 841-868.

[9] Goodman, Leo A., "Statistical methods for the preliminary analysis of Transaction flows" Econometrica, 31 (1963), 197-208.

[10] Goodman, Leo A., "A short computer program for the analysis of transaction flows", Behavioral Science, 9 (1964), 176-186.

[11] Goodman, Leo A., "On the statistical analysis of mobility tables", American Journal of Sociology, 70 (1965), 564-585.

[12] Goodman, Leo A., "The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries", Journal of the American Statistical Association, 63 (1968), 1091-1131.

[13] Grizzle, James E., C. Frank Starmer, and Gary G. Kock, "Analysis of Categorical Data by Linear Models", Biometrics, 25 (1969), 489-504.

[14] Hargens, Lowell L., "Patterns of mobility of new Ph.D.'s among American academic institutions", Sociology of Education, 42 (1969) 18-37.

[15] Reisman, David, Constraint and Variety in American Education, Garden City, New York: Doubleday and Company, Inc. 1957.

[16] Savage, I. Richard and Karl W. Deutsch, "A statistical model of the gross analysis of transaction flows", Econometrica, 28 (1960), 551-572.

[17] Stouffer, Samuel A., "Intervening opportunities and competing migrants", Journal of Regional Science, 2 (1960), 1-26.

[18] Wagner, S.S., "The Maximum-Likelihood Estimator for Contingency Tables with Zero Diagonal", Journal of American Statistical Association, 65 (1970), 1363-1383.

[19] Zipf, George K, "The P_1P_2/D hypothesis: on the intercity movement of persons", American Sociological Review, 11 (1946), 677-686.

THE COMPLETENESS OF ROUTINELY COLLECTED RECORDS IN MEDICAL CARE RESEARCH

Monroe Lerner and Zili Amsel

The Johns Hopkins University

(This work was supported by grants [8 R01 HS00110 and 8 T01 HS00012] from the National Center for Health Services Research and Development and [5 D04 AH 00076] from the National Institutes of Health, U.S. Department of Health, Education and Welfare.)

Although household interview surveys have generally been the preferred mechanism for data collection in social science research, some question has recently been raised about the further usefulness of this mechanism. This comes about partly because at least some populations, perhaps especially residents of inner cities and university communities, may have begun to define themselves as "over-interviewed", i.e., they appear to feel that they have become too vulnerable as target populations for anybody at all who wants to do a survey on any matter of his own choosing, however irrelevant and/or inconsequential it may be to the population interviewed. As a consequence, both the resistance of potential survey respondents, and non-response rates, appear to be rising.

But in addition to this apparent increase in resistance, question arises also because of the many problems inherent in household interviews, perhaps too well-known to merit detailed repetition here. In brief, bias is inherent in the interview situation itself, it is believed, since the interview constitutes essentially a small-group interaction characterized by social-psychological dynamics, in which the characteristics of both respondent and interviewer play a role. This may be particularly true if the interview takes place in the respondent's home, with other family members present. The household interview, especially under these circumstances, becomes an "obtrusive" measure. 1

But also, it is a very expensive form of obtrusive measure, largely because of the high costs of travel, locating the respondent in his own home, the all-too-frequent necessity for numerous callbacks in order to achieve a high response rate, etc. Mail and telephone surveys, often advocated as alternative data-collection mechanisms, also raise many questions, not the least of which is the relatively low response rate of the former and the potential hazards of abuse inherent in the latter.

Conceived of for social research purposes either as an alternative or as a supplement to household interviews, routine records collected in response to the administrative needs of ongoing agencies may serve as a useful data-collection mechanism. Clearly, since social research and administration serve different purposes, information is likely to be collected in response to each need; however, the information collected should not be entirely different because, at least in some degree, these needs may overlap.

Also, some items of information are presumably useful for both types of purposes, especially because "usefulness" can be considered as a matter of degree rather than as an absolute. As one example, items relating to the socio-economic status of a user of health services (e.g., a hospital patient) may be maximally useful for social research purposes, but only minimally useful for administrative purposes; however, their usefulness for the latter is likely to be higher than zero. Many other examples of overlap of need can be cited, and in either direction, i.e., items maximally useful for social research purposes and only minimally so for administration, and items maximally useful for purposes of administration and only minimally so for social research. The questions are, how useful for social research purposes are routine records in general, and how to devise instruments for the routine collection of administrative and social-research data that will provide an optimum mix of the information required for these two purposes?

Routine data collection for administrative purposes appears to have at least one presumed major advantage over the usual household interview in that the purpose of the former is less likely to be subject to question by the respondent than the purpose of the latter. The administrative needs of an agency, especially one providing a service, are more likely to be defined as legitimate by the clients of that agency than is the case for the social research, or even the market research, needs served by household interviews. Because these needs are likely to be defined as legitimate, it seems reasonable to suppose that there would be less resistance on the part of the respondent to providing the requested information, and to providing it as accurately and as completely as possible. However, this is a proposition which should be tested.

Routine data collection for administrative purposes, like household interviews, involves the use of forms which may be self-administered or administered by interviewers. In either case, but especially in the latter, the process is subject to many of the same social-psychological problems as those which characterize the household interview, except for "resistance" by the respondent. But also, routine data collection may have these additional advantages: 1) the information is usually collected by highly experienced interviewers who know the subject matter thoroughly, and 2) the information is obtained at much less cost. But a problem of interest to the researcher is this: How complete is the information obtained in this process, i.e., with what degree of completeness is the information called for by routine forms actually obtained?

In the present study, the authors attempted to measure completeness on a small number of hospital inpatient admission forms filled out under ordinary operating conditions in a large municipal hospital. Hospital personnel completing the forms, and patients and others providing the data, had no knowledge that a study was being, or would be, conducted and that the forms might receive anything other than routine treatment. Thus, as nearly as possible, this study represents "unobtrusive" observation of behavior in a "natural" situation.

I. STUDY BACKGROUND AND SETTING

The present study was carried out as one "by-product" of a survey of inpatient utilization. During one week in May 1968, 221 inpatient admissions took place in the ordinary course of this hospital's operations, with a fairly elaborate admitting form filled out for each patient. The patient population served by this hospital consists primarily of three types: neighborhood residents, largely white ethnic; nonwhite maternity patients from elsewhere in the city; and a much smaller number of patients admitted for specialized treatment. Patient care in this hospital is provided by a full-time house staff; unlike voluntary hospitals, no outside physicians have staff privileges within it.

Admission to the inpatient wards of this hospital may occur by any of the following three methods:

1. From the emergency room. On the physician's decision to admit the patient, and if the patient is able to provide the required information and to take the time to do so (or if it can be provided by persons accompanying him), the emergency room's admitting officer fills out the admitting form immediately in the emergency room and the admitting physician signs it. If the patient is admitted to the inpatient ward immediately, however (and no one else is able to provide the required information), an admitting officer subsequently visits the patient in the ward, fills out the admitting form and returns with it to the emergency room where the admitting physician signs it. Finally, if the patient is unable to provide the required information, and no one else who could provide it is present, the admitting officer is permitted to search the patient for identification; under these circumstances, the form may be filled out either in the emergency room or in the ward.

2. On a pre-planned basis, either through the emergency room, an outpatient clinic, or from elsewhere in the hospital. The "preadmission" data contained on the admitting form is filled out during the patient's pre-admission visit by an admitting officer either in the emergency room, in the outpatient clinic, or in the hospital's admitting room. The form is retained at the hospital's information desk. When the patient subsequently arrives to be admitted, he obtains the form from the information desk, and an admitting officer completes it.

3. Directly from an outpatient clinic. For clinic patients seriously ill enough to be admitted directly from the clinic, preadmission data are obtained by the clinic staff. An admitting officer obtains the remainder of the data in the clinic prior to the patient's admission to the ward.

The admitting officers in this hospital do not specialize by type of admission; any admitting officer can, and does, service any type of admission. All admitting officers must be college graduates or have the equivalent in years of clerical experience. Satisfactory performance on a civil service examination is a prerequisite for the job, and on-the-job training is provided to the newcomer by experienced admitting officers.

II. THE ADMITTING FORM

The admitting form used in this institution is quite an elaborate document; it is intended to obtain a relatively large volume of information, some of it fairly detailed and/or complex. The elaborateness of the form results from the hospital's need for a large volume of detailed information on some patients, although this same amount of information is not needed for all patients; however, the same form is used for all. The admitting officers probably understand the nature of the hospital's differential needs, but the differences are not always clear to the outsider. This is because the form itself does not make the differences clear, as could be done, for example, by having explicit instructions on the form to the individual filling it out that some questions are to be asked only of specified categories of patients or only if preceding questions result in designated responses. It is also not initially clear to the outsider, because it is not explicitly stated on the form, nor is it always observed in practice, that in some instances two or three items are intended as mutually exclusive alternative responses to a single, often implicit, question.

The form is filled out in duplicate on some items, but in sextuplicate on others. It has two sides, but only the first side was used in the present study (see Appendix I, which shows the first side). The backside of the page was excluded because it requests highly specialized information applicable only to a very small number of patients -- i.e., items of information requested under these headings: in case of accident; if occupational injury; dead on arrival; animal bites; and communicable diseases. It also contains some "business office" information items, and space for a "plate imprintation". (Each patient served by the hospital, inpatient and outpatient alike, receives an addressograph plate containing identifying information plus a date of issue.)

III. RESULTS OF THE STUDY

Table I shows the completed-response rates, by major category, for some selected items, arranged in order of degree of completeness. An item on each form was classified as "completed" when a "usable" response was obtained, even if the information was not as detailed as requested or the answer was "no" or "none". That is, items were considered as incomplete only when they had been left blank. For a surprisingly large number of items the completed-response rate was 100 percent. These are clearly the items which provide information required <u>on each patient</u> as a matter of necessity by the hospital for its administrative purposes. However, the completed-response rate was less than 100 percent for a number of the items shown on the table, with 61 percent for "how brought to hospital" the lowest among those listed.

However, it should be emphasized that it was frequently not possible from the data at hand to tell whether no entry for an item meant that the answer was none; that the question was irrelevant -- i.e., that it was not applicable to, or was inappropriate to, the specific patient or the circumstances of his admission; or simply that the admitting clerk had failed to obtain, or enter if obtained, the requested information. Thus the name of the referring or family physician was entered on only 62 percent of the forms, but this may mean that the other 38 percent did not have a referring or family physician, that the question was not applicable,* or that the information was either not asked for or not entered.

This points to what is perhaps one of the fundamental difficulties in working with administrative records for social research purposes. That is, the essential aspect of the hospital's concern was not with -- a typical social research question -- whether or not the patient has a family physician who might have referred him to the hospital. What it was concerned with was the physician's name if the patient did have one, so that he might be contacted by the hospital staff physicians if this was indicated. This concern of the hospital explains the form of the question. (Not "were you referred here by a physician?" or "do you have a family physician?", but rather "what is the name of your referring or family physician?"). Nevertheless, it leaves the social researcher in the dark as to the precise meaning of "no entry" for this item.

For purposes of analysis, seven items were selected from among those with completed-response rates of less than 100 percent. The items and rates are: occupation, 93; father's name, 90; length of residence, 86; previous admission, 85; name of spouse, 82; previous address, 75; and referring or family physician, 62 percent. (For four of these items, the completed-response rate was computed from a base of 221 forms, while for the other three it was computed from a smaller base.) The completed-response rate for the 7 items combined was 82 percent, and the question was, how account for their lesser degree of completeness? Special cross-tabulations were run for these items by: the demographic characteristics of the patients (color-sex and age); the pay status of the patient; the direct source of the information to the admitting clerk (i.e., patient or a surrogate as informant); the source of referral for impatient admission (route of the patient within the hospital to the admitting office); whether the information had been secured at an interview prior to the admission or at the admission itself; and whether the admission was for a maternity or other type of case.*

The special cross-tabulations

Relatively little variation was found by demographic characteristics of the patient. The completed-response rate was almost identical for each sex (83 percent for males and 82 percent for females). It was higher for white patients than for nonwhite patients (84 percent against 81 percent). Considering each of the four color-sex groups among patients separately, white males had the highest completed-response rate (86 percent) and non-white males the lowest (80 percent). None of these differences were of the magnitude that might have been expected.

By age, not much in the way of a consistent pattern was evident. Completed-response rates were highest for patients aged 70-79 (88 percent) and 30-39 (87 percent). They were lowest for the eight patients aged 80 and over (62 percent), but except for this group, the completed-response rates for all of the age-groups were fairly close to the mean,

*This would have been the case, for example, for inpatient maternity cases referred by a Health Department Clinic.

*Because of space limitations, the present discussion is limited only to the data from the cross-tabulations by demographic characteristics of the patient and by the direct source of the information.

with the lowest being the 40-49 year age-group (79 percent).

Some variation in the completed-response rates was evident in accordance with the direct source of the information provided to the admitting clerk, i.e., whether by the patient himself (which occurred in 130 of the 221 cases, or 59 percent of the total), or by a surrogate for him. For present purposes these surrogates were divided into 3 major categories -- a member of the patient's immediate family, i.e., mother, father, wife, husband, son, or daughter; an "other informant", often a friend, other relative, ambulance driver or policeman, etc.; and an "unknown" informant.

A member of the patient's immediate family provided information for 66 of the 221 patients, or 30 percent of the total. The largest category among these immediate family members consisted of the mother of the patient (25 of the 66), the wife (11), and the husband (10). The father, son, or daughter of the patient provided information to the admitting clerk in smaller numbers of cases (8, 4, and 8, respectively). The category of "other informant" accounted for 19 of the 221 cases, about 9 percent of the total, while an "unknown informant", i.e., identity of the informant not indicated on the form, provided the information for 6 cases, about 3 percent of the total.

Somewhat surprisingly, the completed-response rate was lower when the patient was the informant (82 percent, exactly at the mean for the entire experience), than when the informant was an immediate family member (86 percent). Among the immediate family members, the completed-response rate was highest for wife of the patient (95 percent), and lowest for daughter (81 percent) and mother (82 percent). Although lower than for an immediate family member, the completed-response rate was relatively high for the category of "other informant" (83 percent). The category of "unknown informant", as would be expected, resulted in the lowest completed-response rate, only 53 percent. It is possible that this might have been the failure of the admitting clerk, in not noting the identity of the informant, also to fail to record other information about the patient.

IV. CONCLUSIONS

Social researchers intending to work with administrative records are likely, on the basis of the present study, to find high completed-response rates for most of the items contained in these records. However, investigators will have to be fully familiar with the administrative needs served by the records and with the settings in which the information is obtained if the records are to be maximally useful for social research purposes.

REFERENCES

1 Webb, Eugene J. et.al., Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand-McNally, 1966.

[redacted] ADMISSION OUT PATIENT [redacted]

NAME—LAST FIRST MIDDLE OR MAIDEN

ARRIVAL DATE TIME A.M. P.M.

STREET ADDRESS HOW LONG

DISPOSITION DATE TIME A.M. P.M.

OCCUPATION SOC. SECURITY NO.

CITY OR COUNTY ZIP CODE CODE TELEPHONE

BIRTHPLACE CITIZEN YES NO

BIRTH DATE AGE-YRS. RACE SEX M F RELIGION MARITAL STATUS PLATE

SPOUSE NAME LIVING DEAD

FATHER'S NAME BIRTHPLACE MOTHER'S MAIDEN NAME BIRTHPLACE

NEXT OF KIN OR RESPONSIBLE PERSON RELATIONSHIP ADDRESS ZIP CODE TELEPHONE

RELATIVE (NAME) ADDRESS (OTHER THAN ABOVE) ZIP CODE HOW LONG TELEPHONE

HOW BROUGHT TO HOSPITAL BROUGHT BY WHOM ADDRESS TELEPHONE

AMBULANCE NO. AND CO. PATIENT'S COMPLAINT: REFERRING PHYSICIAN OR FAMILY PHYSICIAN

POLICE DISTRICT AND NO. ADDRESS TELEPHONE

PROVISIONAL DIAGNOSIS AND TREATMENT RENDERED:

ASSIGNED TO ADMITTED FROM

ADMIT. TO ACUTE ☐ CHRONIC ☐

ROOM #______ BED #______

ADMITTED OR TREATED BY

(1) DR.

(2) DR.

(3) DR.

PREVIOUS ADDRESS (RESIDENCE MUST COVER 1 FULL YEAR) ZIP CODE HOW LONG

NO. OF DEPENDENTS IN FAMILY ADULTS CHILDREN

SOURCE OF INCOME

PATIENT EMPLOYER ADDRESS EARN PER $ SOCIAL SECURITY NO.

OTHER EMPLOYER ADDRESS EARN PER $ SOCIAL SECURITY NO.

OTHER EMPLOYER ADDRESS EARN PER $ SOCIAL SECURITY NO.

INSURANCE GROUP ☐ INDIVID. ☐ CARRIER OR UNION SUBSCRIBER

VETERAN YES ☐ NO ☐

BLUE CROSS ☐ BLUE SHIELD ☐ STATE SUBSCRIBER MEMBERSHIP NO.

SERVICE CONNECTED DISABILITY YES ☐ -C # NO ☐

CHECK ONE IF APPLICABLE COMPLETE NECESSARY FORMS → TITLE X YES ☐ NO ☐ NO. TITLE XVIII YES ☐ NO ☐ A - B NO. TITLE XIX YES ☐ NO ☐ NO.

ACUTE ADMISSION	RATES	CHRONIC ADMISSION—NURSING
$ PER DIEM	OASI	☐ $
$ DELIVERY	OTHER	☐ $
$ ALLOWANCE	PTS. FUND	☐ $
$ PROFESSIONAL FEE	FULL VENDOR	☐ $
DESCRIPTION	PARTIAL VENDOR	☐ $
	TOTAL	$

PREVIOUS ADMISSION TO [redacted] YES ☐ DATE NO ☐ ACTIVE DPW YES ☐ NO ☐

INFORMATION GIVEN BY:

PRE-ADMISSION BY: DATE

ADMISSION BY: DATE

INCOME: ☐ COUNTY ☐ NON-RES. ☐ INS. ☐ W.C. ☐

OTHER ☐ SPECIFY

RATE CHANGED TO $ PER DIEM ☐ DELIVERY ☐ REASON BY DATE

Knowing that it is to be relied on by [redacted] for determining hospital charges and for other purposes, I/we hereby certify that the information contained above and on the reverse hereof is true and correct to the best of my/our knowledge and belief. In consideration of the admission and/or treatment of the above patient I/we agree to pay to [redacted] the applicable hospital charges and professional fees each week as accrued and to pay the balance in full, or assign to the Hospitals insurance benefits and submit necessary claim papers to cover the amount due, or to obtain a State-Aid certificate by the time of discharge.

Date________________ Seal ________________ Seal

GUARANTOR

TABLE I. COMPLETED-RESPONSE RATES FOR SELECTED ITEMS BY ITEM

Item	Base	Completed-Response Rate
Name	221	100
Address	221	100
Birthdate	221	100
Age	221	100
Sex	221	100
Race	221	100
Patient's Complaint	221	100
Provisional Diagnosis and Treatment Rendered	221	100
Wing to Which Assigned	221	100
Source of Payment*	221	100
Source of Income*	221	100
Arrival and/or Disposition Date	221	100
Arrival and/or Disposition Time	221	100
Name of Interviewer (Preadmission and/or Admission)	221	100
Religion	221	99
Marital Status	221	98
Birthplace	221	95
Number of Dependents in Family	221	95
Occupation**	150	93
Father's Name	221	90
Length of Residence	221	86
Previous Admission	221	85
Admitted From	221	83
Name of Spouse	125	82
Previous Address	53	75
Referring or Family Physician	221	62
How Brought to Hospital	221	61

* This question is not specifically asked. However, all forms contain this information somewhere on the form.

** This information is given either for the patient or for some member of the patient's family.

COST-BENEFIT ANALYSIS OF NARCOTIC ADDICTION TREATMENT PROGRAMS
with Special Reference to Age

Irving Leveson,[1] New York City Planning Commission

Introduction

Efforts to deal with consequences of poverty, race, and other social problems have resulted in a proliferation of treatment and rehabilitation programs for criminals, addicts, and the mentally ill. These have included skills training, professional and nonprofessional psychotherapy, drug therapy, and incarceration and surveillance. Many such efforts are experimental or pilot projects. To date there has been little in the way of systematic efforts to evaluate, either publicly or privately, the success of these programs.

The amount of information in the public domain is growing, and public agencies providing funds are beginning to insist that evaluations take place. To attain comparability among such evaluations requires agreement on methodologies of general applicability. A "simplified" methodology which permits comparisons among divergent addiction rehabilitation and treatment methods is developed and illustrated here. It is believed that the techniques have applicability to many other programs.

A critical feature of rehabilitation efforts is that they tend to deal with very different groups. We would like to have cost-benefit information for each treatment applied to each population type. In lieu of this, we find it necessary to assume each treatment to be most appropriate to a particular population type. We examine costs and benefits for each group and then adjust for differences between groups in ease of treatment and in the value of benefits from successful treatment. In a related approach, the efficient choice of client group for a given program is considered.

There has been great concern over the growth in the drug problem among youth and its far-reaching implications for institutions such as schools and the military. There has been great concern over the lack of a consensus on any treatment method to effectively deal with the young. At the same time, there has been a proliferation of treatment programs attempting to deal with the young which require assessment. These efforts are critically bound up with discovery of the basic causes of addiction. Rather than develop a more general methodology for comparing programs dealing with divergent population groups, we focus here on the role of age in the success in treatment programs.

The Underlying Cost-Benefit Model

The benefit to society of having one less addict[2] depends on the number of years that an addict would have remained addicted without assistance. Because of a finite life expectancy and the possibility that many persons "mature out" of addiction as they grow older, the number of years of addiction remaining is greater for younger addicts. The number of remaining years of addiction is represented by ℓ.

The benefits of a reduction of a man-year of addiction (b_{ij}) are represented as a function of age (i) as well as the number of years (j) of addiction remaining.[3] Future benefits are discounted by an appropriate rate (r). The full expression for the benefits of $N = \sum_i^a N_i$ persons withdrawing from addiction is given by summarizing over individuals with different ages and numbers of years of addiction remaining:

$$B = \sum_{i=1}^{a} \sum_{j=1}^{\ell} \frac{N_{ij}\, \ell_j\, b_{ij}}{(1+r)^j} . \quad (1)$$

Equation 1 can be greatly simplified. First, we assume that the number of years of addiction remaining is solely a function of age. Therefore we can write

$$B = \sum_{j=1}^{\ell} \frac{N_j\, \ell_j\, b_j}{(1+r)^j} . \quad (2)$$

Now we examine the assumptions implicitly required to use simpler measures of success than the full expression in Equation 2. Most studies ignore differences among age groups in benefits per addict successfully treated. The effectiveness measure becomes

$$E = \sum_{j=1}^{\ell} \frac{N_j\, \ell_j}{(1+r)^j} , \quad (3)$$

where B = bE. The most commonly used measure of success is the proportion (σ) of persons admitted to treatment (A) who successfully withdraw. The reduction in the number of addicts is $N = A\,\sigma$. If we compared different programs treating the same types of people, b_1 would equal b_2 and ℓ_1 would equal ℓ_2, so that the probability of successful withdrawal would be sufficient to gauge success. Otherwise, use of this measure is equivalent to assumming that

$$\sum_{j=1}^{\ell} \frac{\ell_j}{(1+r)^j} = 1 \quad (4)$$

as well as ignoring age differences in benefits per man-year averted. It gauges success by the number of addicts successfully withdrawing without regard to length of time they would have been addicted. In doing so, it overemphasizes very immediate gains.

Reducing the average number of addicts over a number of years requires measuring success by the number of man-years of addiction averted, or

$M = N \ell$. This can be derived from Equation 3 as

$$M = \sum_{j=1}^{\ell} N_j \ \ell_j \qquad (5)$$

where $r = o$ and $B = b\,M$. This measure is particularly simple to use. However, it overemphasizes benefits which are deferred in time. If we wish to take into account that we value benefits today more than those which will not be received for many years, it is necessary to discount future benefits by an appropriate discount rate (r). The measure which does this is the discounted number of man-years of addiction averted, designated as E as defined in Equation 3.

Next we consider cost effectiveness. Where ℓ is used, a discounted ℓ could easily be substituted. Let individuals going through a program either be treated successfully at a cost c_s or unsuccessfully at a cost c_u. The average cost (c_t) of treating an addict in a program having a proportion successfully treated of σ is

$$c_t = \sigma\, c_s + (1-\sigma)\, c_u \,.^5 \qquad (6)$$

The cost of treating A addicts is A c_t, so that the cost per successfully treated is

$\frac{A\, c_t}{\sigma A} = \frac{c_t}{\sigma}$. Designating the cost per successfully treated addict as c,

$$c = c_s + \left(\frac{1-\sigma}{\sigma}\right) c_u \,. \qquad (7)$$

As the proportion successfully treated increases, cost per success falls because the costs of relatively fewer unsuccessful cases are included.

The ratio of costs of addicts admitted to treatment to man-years of addiction averted (cost-effectiveness ratio) is

$$\frac{c_t\, A}{M} = \frac{\sigma\, c_s + (1-\sigma)\, c_u}{\sigma\, \ell} \,. \qquad (8)$$

Assume that programs with a larger ℓ have a lower σ. If across programs σ is proportional to ℓ, the ratio of the cost-effectiveness ratios is the same as the ratio of the costs per person treated. If σ varies more than ℓ, the costs per person treated would have to be lower in the programs with the higher ℓ for them to compare favorably with the others. If c_u is small relative to c_s or σ is large

$$\frac{c_t\, A}{M} \simeq \frac{c_s}{\ell} \,. \qquad (9)$$

Empirical Evidence on Age Differences

We have examined the evidence on differences among age groups in the benefits of successful treatment and the chances for successful treatment. The analysis indicates (a) that differences in the prevalence of opiate use among age groups reflects some sort of maturation (and/or mortality) since without maturation they imply implausible patterns in the number of users over time, (b) derives the average number of years of addiction remaining by ethnic group based on age patterns under alternative assumptions about past rates of growth of addiction, and (c) compares the higher retention rates of the New York City Methadone Program for the youngest addicts with the lower derived maturation rates and suggests that the young may do the same or worse if "natural" maturation accounts for part of the dropping out of older persons. (This material is available upon request.)

A Cursory Look at Major Programs

Virtually any program will pay in terms of benefits to the nonaddict population. Even if we incarcerate persons at a cost of over \$8,000 per year for several years, if crime of \$10,000 to \$20,000 per person per year is averted,[6] the program can be justified. However, this calculation does not consider that, even with a high payoff, funds for addiction services may be limited, and it does not take into account the undesirability of involuntary incarceration. While it may be clear that the payoff to addiction programs in general is very high, there is still a major issue as to what proportion of the addict population can be dealt with by each approach.

As of the end of 1969, there were about 11,000 New York City addicts in treatment programs. In addition, there were about 5,000 in pretrial detention[7] and others serving prison terms or on parole. Nearly all treatment funding comes from the New York State Narcotic Addiction Control Commission which has a current operating budget of \$38 million per year and a \$200 million capital construction program. The three largest programs for the treatment of drug addicts in New York City are the program of the New York State Narcotic Addiction Control Commission, the Methadone Maintenance Program, and the Phoenix House Program of New York City's Addiction Services Agency.

The State program combines voluntary or involuntary incarceration with therapeutic and rehabilitation services for up to five years, followed by an aftercare phase involving surveillance. The program, depending strongly on court and parental coercion, has suffered abscondance rates of over one-third per year and high rates of recidivism. However, it has been successful in moving persons from an institutional setting into aftercare. Furthermore, it is the only one of the three largest programs to have given primary attention to younger addicts. One-third of the persons in the program are under age 20 and two-thirds are under age 25. The availability of rehabilitation services under the program appears to be increasing. At the end of 1969 there were 5,000 addicts in the program, with a large expansion planned as new facilities being constructed become available.

The Methadone Program uses the drug Methadone, itself addicting but providing no "high," to block the effects of opiates. After a few weeks of withdrawal, the drug is administered orally on an outpatient basis for an indefinite period of time. There are about 2,000 addicts currently under treatment, and a doubling of efforts is under way. This program tends to concentrate on older addicts as indicated by comparisons with the Narcotics Register for September 1969:

	Methadone Program	Narcotics Register
Less than 25	10%	33%
25 - 34	51	41
35 and over	39	16

The Methadone Program is the only one of the three to make public regular evaluation reports. The evidence of high retention rates, improvement in employment experiences, and reduced criminality have led to a belief that this is the most effective method currently available.

The Phoenix House program relies on encounter techniques involving confrontation of addicts with each other and with ex-addicts. After two years of institutionalization, a process of gradual re-entry into the community begins. Phoenix House has generally served about a thousand persons, but only a small fraction have re-entered the community through the program.

Scant information is available with which to compare the effectiveness of major program alternatives, even using a simplified framework devised to minimize data requirements.

The Methadone program reported about a .7 success rate for persons on drug maintenance for 36 months. Lack of success in that program reflects both voluntary and involuntary termination stemming from factors such as criminal behavior and alcoholism.

The State program reported that after its first 21 months of operation, 44 percent of the 1,893 persons returned to the community had not resumed drug use. Thirty percent were returned to a rehabilitation center after it was discovered that they were reusing narcotics, and 26 percent could no longer be located (warrants were issued against them).[8] In view of the fact that the program had only 188 persons on aftercare after its first year and 836 after 21 months, the average length of time in the community would have to have been substantially less than one year. In view of the short time, the program's experience is discouraging.

A study of a roughly similar program, that of the California Rehabilitation Center at Corona, found that 35 percent were in good standing (or had been removed from the program while in good standing) one year after release. After three years the figure had fallen 19 percent.[9] There is nothing in the experience of the New York State program to date to indicate it is doing any better than this.

If we take seriously the public mandate to prepare addicts for entry into the community, the fact that only a handful of Phoenix House residents have actually done so after their period of institutionalization must be treated as if the proportion dealt with successfully is negligible. This program could only be considered successful if society were willing to place a high value on withdrawal in a setting of permanent institutionalization.

We are now in a position to compare programs on the assumption that they continue to treat the same types of people as they now treat with the same degree of success. If we compare the benefits of the Methadone program designated by subscript 1 relative to the State program designated by subscript 2, we have

$$\frac{B_1}{B_2} = \frac{b_1}{b_2} \cdot \frac{\sigma_1}{\sigma_2} \cdot \frac{\ell_1}{\ell_2} \quad . \qquad (10)$$

The number of persons admitted to each program is assumed to be the same. Assuming the California experience to be applicable to New York State, $\frac{\sigma_1}{\sigma_2}$ equals about .7/.2 or 3 1/2.

The mean ages and corresponding years of addiction remaining for the three major programs are as follows:

	Age	Approximate Years Remaining
Narcotic Addiction Commission	25	13
Addiction Services Agency	27	11
Methadone Maintenance	32	8

The age distribution of the Methadone program implies that the cost of the program must be about 2/3 of the others to compensate for the difference in years of addiction remaining, if all had equally high success rates. If the years of addiction remaining were discounted, this difference would be modified somewhat. The value to society of treating the kinds of persons treated by the State program may be higher than those treated with Methadone. There is some evidence that older addicts commit fewer crimes. Younger addicts can be expected to remain on drugs longer, with greater adverse effects on health. Furthermore, treatment of younger addicts may have greater benefits of preventing other persons from being induced to use drugs. However, there would have to be a ratio of b_1 to b_2 as low as about 3/7 in order for the State program to have equally large benefits.

When one considers costs, the problem of program choice becomes more difficult. In the past, the State program required over $8,000 in operating costs alone per man-year of incarceration, and currently employs nearly one person for every addict in residential or aftercare activities. Completion of the program probably requires an average of about $15,000 per person. For a comparable cost, the Methadone program could provide detoxification and drug maintenance for more than a decade. After that time, it may be possible for maturation effects to be substituted for drug maintenance.

While there may not be a large difference between the State and the Methadone programs in the cost per completed program, those programs within the <u>lowest</u> proportion of persons treated successfully will tend to have the lowest cost per person admitted to treatment. This occurs because the cost per person treated (c_t) is an average of the cost per person treated successfully (c_s) and the cost of persons who do not successfully complete the program, some of whom have partial treatment (c_u). (See Equation 6.)

Higher costs per person admitted to treatment in a more successful program will tend to offset part of the gains from a more successful treatment. The later stages in treatment at which persons drop out the larger will c_u be relative to c_s, and the less the cost per person admitted to treatment will vary with the probability of success.

When the cost per admission is considered, the weight of evidence is probably still heavily in favor of Methadone, since in the State program there is probably only a modest difference in cost between those who successfully withdraw and those who do not. However, this bears closer scrutiny, since success may depend on the amount of rehabilitation provided and the State has been expanding the amount of real rehabilitation service provided. Furthermore, some deduction of welfare of the addict must be made for the involuntary nature of the State program and the dependence on Methadone.

On the surface, the Addiction Services Agency program is by far the cheapest, estimated by the agency to cost $7,500 for the entire 2 1/4 year process. While in theory it could have an added benefit beyond the time of maturation out of drug use because it is directed to improvement in many aspects of social adjustment, this does not occur without re-entry.10

<u>Choice of Age Group within a Program</u>

Selecting age groups for treatment so as to maximize the number of man-years of addiction averted is equivalent to minimizing the number of addicts. If the success of treatment were the same at each age, we would concentrate on the youngest addicts who have the greatest number of years of addiction remaining. On the other hand, if the number of years of addiction remaining were the same at each age and success of treatment were greater for older persons, we would concentrate on the oldest addicts. (We are defining success net of any normal maturation tendencies.) Here we consider the optimal age when variations in both years of addiction remaining and success of treatment are taken into account.

The number of man-years of addiction averted for a program is given by

$$M = A \sigma \ell . \qquad (11)$$

For simplicity, we assume that the average number of years of addiction remaining in a program can be represented as a proportion (p) of the difference between some fixed age (a*) and the average age (a) of persons being treated

$$\ell = p\,(a^* - a) . \qquad (12)$$

Thus if a* is age 40 and p = .6 , addicts age 20 would be expected to remain addicted for another 12 years on the average. Those who were still addicted at age 30 would be expected to remain addicted for another 6 years. We further assume that the rate of success in withdrawal from addiction can be represented by a linear function of age

$$\sigma = s + t\,a , \qquad (13)$$

where $t > 0$.

Combining these equations we have

$$M = A\,s\,p\,(a^*-a) + A\,t\,a\,p\,(a^*-a). \qquad (14)$$

Multiplying through and differentiating with respect to age

$$\frac{dM}{da} = A\,t\,a^*\,p - A\,s\,p - 2\,A\,t\,p\,a\,. \qquad (15)$$

The second derivative is

$$\frac{d^2M}{d\,a^2} = -\,2\,A\,t\,p \qquad (16)$$

which, since A, t and p are positive, is negative indicating a maximum. Setting the first derivative equal to zero and solving, we derive the optimal level of a

$$a = \frac{a^*}{2} - \frac{s}{2t}\,. \qquad (17)$$

One inference that can be drawn from this result is that the effect of age differences in the success rate on the optimal age is important. If, for example, s were -.3 and t were .03, indicating a success rate of .4 at age 20, and .7 at age 30, the $-\frac{s}{2t}$ term would add five years to the optimal age. If s is positive, however, the optimal age is reduced. For example, if s = .2 and t = .02, indicating success of .6 at age 20 and .8 at age 30, the optimal age would be reduced by five years.

With the functional form we have chosen for ℓ, p does not enter into the result since it does not change the terms of trade between age groups. Unfortunately, we have little information with which to determine a*. Our curves of years of addiction remaining suggest that over the range of ages 22-32 the curves could be approximated by an a* at about 40, with different p's for each ethnic group.[11] In that age range, a figure of age 20 for $\frac{a^*}{2}$ would be valid. This is very approximate of course.

If $\frac{a^*}{2}$ were as high as 25 and a function such as $\sigma = -.3 + .03\,a$ were approximately correct, we would come out with an optimal age very close to the average age now being treated in the Methadone program. If, instead, as the program claims, success varies little with age so that a function such as $\sigma = .2 + .02\,t$ is appropriate, then to corroborate the ages now treated we would need an a* of around 70 which is completely impossible. Even if a* is as high as 50, $s \simeq .2$ and $t \simeq .02$, then the Methadone program should be concentrating on persons as much as 10 years younger than they are now, even though the proportion successfully withdrawing would be lower if it did. This would take into account the likelihood that addiction will be prevented for a larger number of years when the young withdraw, rather than relying on the more visible success rate alone.

Final Comments

The results of this analysis are illustrative and suggestive rather than definitive. They tend to support the efficacy of the Methadone program relative to the State program, if the goal is to minimize the number of addicts. However, they raise serious questions about the merits of the Methadone program's practice of treating a disproportionate number of older addicts. Combining information on the number of years of addiction remaining from successful withdrawal at each age with assumptions about success in achieving withdrawal at each age, suggests that it may be possible for the Methadone program to increase its impact on the number of addicts by the order of magnitude of at least one-fourth, with no additional expenditures, by shifting emphasis in treatment to younger groups.

Footnotes

1. Director of Research and Director of Health Planning, Office of Comprehensive Planning. The views expressed are those of the author and need not represent those of the City Planning Commission, the Community Renewal Program, or any of their affiliates or consultants. Beginnings on this study were made while at the Rand Corporation. The comments of Zili Amsel, Sidney Leveson and Clarence Teng were most helpful.

2. The term addict is used to imply users of hard drugs whether or not physiologically or psychologically addicted.

3. One study of addicts with a mean age 30.7 showed a mean age of most arrests of 22.8. See Bernard Greenfield, "The Riverside Study - Ten Years Later." Paper prepared for the New York City Department of Health, Health Research Training Program, Summer 1967.

4. If we make the assumption that the ratio of the benefits per man-year of addiction averted to the discount function is an inverse function of age

$$\frac{b_j}{(1+r)^j} = \frac{k}{a},$$

then

$$B = \sum_{j=1}^{\ell} \frac{k}{a}\,\ell_j\,N_j$$

or

$$B = \frac{k}{a}\,M\,.$$

The benefits of reducing addiction are proportional to the number of man-years of addiction averted and inversely related to the age of those assisted.

5. The model could easily be extended to allow for some positive benefit for nonsuccesses due to a period of withdrawal prior to recidivism.

6. See Irving Leveson, "Drug Addiction: Some Evidence on Prevention and Deterrence," paper presented at the meetings of the Econometric Society, Detroit, Michigan, December 1970 and John Holahan, The Economics of Drug Addiction in Washington, D.C.: A Model for Estimation of Costs and Benefits of Treatment and Rehabilitation, Report No. 33, District of Columbia Department of Corrections, October 1970.

7. Based on Bronx District Attorney Burton Robert's estimate that about four-tenths of the persons in detention are addicts.

8. New York State "Report of the Narcotic Addiction Control Commission for the First Twenty-One Month Period." The Commission did not release comparable figures in its second annual report. In the week of January 25-31, 1970, only 48 percent were known to be gainfully occupied.

9. John C. Kramer and Richard A. Bass, "Civil Commitment for Addicts: The California Program," Report of the Committee on Problems of Drug Dependency of the National Academy of Sciences and the National Research Council, 1968.

10. The program avoids constraints on the availability of professional manpower by using ex-addicts. However, retaining ex-addicts in the program delays the time they will have to re-enter the community. The adjustment process may be made more difficult in the future if large numbers of persons must be absorbed into the community in a relatively short period of time. If the ratio of the number of addicts to ex-addict employees is R, the rate of growth per two-year treatment period needed to retain S percent of persons treated as employees is S R. If the growth of the program were suddenly halted, $\frac{R}{1 + R}$ of those in the program would be ending treatment and have to find jobs.

11. At older ages the slopes are smaller and a*'s higher.

MEASURING THE QUALITY OF UNIVERSITIES

M. Melnyk and Michael Ferrari, Kent State University

Introduction. Attempts at evaluating universities (particularly graduate schools) according to various subjective and objective measures of quality certainly are not new. Beginning with the ratings assembled by Raymond Hughes in 1925 /6/ there have been repeated attempts at appraising the quality of graduate institutions. Studies by The American Council on Education in 1934 /10/, Keniston in 1958 /7/, Eels in 1960 /5/, Berelson in 1960 /2/ and Cartter in 1966 /3/ have provided a continual stream of evaluation methods and specific ratings. The Cartter study probably has become the most widely used evaluation of quality, and it has become a stimulus for further study, criticism, and refinement on matters of quality assessment. For example, recent articles by Lewis /9/, Knudsen and Vaughn /8/, and Shamblin /11/ concerning evaluation of quality in departments of sociology are indicative of the continuous interest in methods of measuring quality.

The present paper deals mainly with the following problems: dispersion in quality between the universities and the advisability of ranking universities according to quality; dispersion in quality between academic departments within schools; and the relation between quality and compensation bearing on the predictability of changes in quality. The analysis is based on data that are relevant to the problem. They are: quality indices, compensation of faculty, number of professors in Who's Who in America, and teacher-student ratios.

Quality indices were constructed from information contained in the Allan M. Cartter-American Council on Education /3/ study. Using this book's scores and quality classification of academic departments, two indexes were derived. One of them, named Average Rank Scores, makes possible, among other things, ranking of whole universities. Another, called Weighted Average Scores (not shown here), is used mainly to study the dispersion in quality among departments within a university.

The quality classification of schools can be accomplished by constructing quality indexes and by ranking. Whether it is accomplished by one or the other method, such classification is useful only when the quality of academic departments within a university is fairly uniform. If a considerable number of schools have both excellent and very poor departments, the classification of whole schools is misleading. The only guide to quality, then, would be the method adopted by Cartter, giving the classification of academic departments alone.

The present analysis suggests that there is enough uniformity among departments in a great many schools, and that, basically, classification of whole universities according to quality is justified.

However, there is a weighty problem of how to classify schools. A great many people have considerable predilection for ranking. Yet, in too many cases the difference in quality among bunches of schools is so small that ranking may be considered inappropriate.

Another important problem concerning the evaluation of schools is the method of evaluation. Cartter's grading is based on opinions of scholars. The analysis of data based on his evaluation of academic departments, strongly suggests the conclusion that such a method of evaluation exaggerates the quality of top schools and has the opposite effect on schools at the other end of the spectrum.

Finally, the relation between faculty compensation and quality on one hand and the dispersion in quality on the other has been studied. It makes possible approximate forecasting of quality and of changes in quality. It also suggests that it is almost an impossible task to build up a small number of excellent departments in an otherwise low quality school when the average compensation of this school is considerably below that of schools with predominantly excellent departments.

The Statistical Data. Let us show first how our quality indexes were derived from Cartter's study.

Cartter sent out his questionnaire to department chairmen and senior and junior scholars of universities granting doctor degrees. The questionnaire dealt not only with the quality of academic departments, but also with the quality of graduate programs. The present study is concerned with the quality of academic departments alone.

The basic question on this matter presented in the questionnaire was: "Which of the terms below best describes your judgment of the quality of graduate faculty in your field at each of the institutions listed? Consider only the scholarly competence and achievements of the present faculty. 1) Distinguished; 2) Strong; 3) Good; 4) Adequate; 5) Marginal; 6) Not sufficient to provide acceptable doctoral training; 7) Insufficient information."

The questionnaire was answered by over 4,000 scholars from about one hundred universities.

The tabulation of results in Cartter's book is somewhat different from the classification used in the questionnaire. It is restricted to "Distinguished," "Strong," "Good," and "Adequate plus" categories.

Departments in the first two categories were assigned both rank and quality coefficients. Departments in "Good" and "Adequate plus" categories have neither rank nor score. They are merely listed alphabetically. One has to assume that departments that were not listed fall into a "Less than adequate plus" class.

Let us now describe how our quality index for whole schools was obtained from this information.

This index is basically an "average rank" index. But before an Average Rank Score for, e.g., Chicago could be computed, Cartter's ranking had to be adjusted and rank numbers had to be assigned to departments below the "Strong" category.

The adjustment of Cartter's ranking was necessary because in his study the number of schools in various academic disciplines under "Distinguished" and "Strong" categories is not the same. Thus, in economics the number of such schools is 16; in classics, 12; in geology, 19; and so on.

The ranking was adjusted so that in the "Distinguished" category of every academic discipline the first school received No. 1 and the last No. 9 because the greatest number of schools in this category was 9.

In the "Strong" category the first department always received No. 10, and the last, No. 34. The ranks for intermediate departments in both categories were assigned by the formula for common difference of arithmetic progression $d = \frac{z_n - 1}{n-1}$ where z_n for, e.g., "Distinguished" category is $z_n=9$ and for, e.g., anthropology the value of n is n=4, which is the number of distinguished departments in this academic discipline. As a result, these were the rank numbers assigned to the 4 schools: Chicago 1, Harvard $3^2/_3$, Berkeley $6^1/_3$, and Michigan 9.

Departments which were not individually ranked by Cartter, that is departments falling into "Good," "Adequate plus," and "Less than adequate plus" categories, received, respectively, ranks of 35, 36, and 37.

Finally, the ranks of all departments of a university were averaged. The Average Rank Scores are not shown here, but the ranking based on these scores is given in column (1) of the Appendix.

To see how this quality index compares with an index based on Cartter's scores, average scores for all departments of the top 10 schools were obtained. The results of the comparison are as follows: There is 0 rank point difference in 3 cases, 1 rank point difference in 5 cases, 2 point difference in 1 case, and 3 point difference in 1 case. This difference between the two indexes should be considered insignificant.

The influence of Cartter's scores upon the Average Rank Scores index diminishes as we descend the ranking ladder toward schools with more and more departments below "Strong" category. Thus, the University of Washington ranks 22nd according to Average Rank Scores and has 50 per cent of departments in "Strong" and no department in "Distinguished" category. Rensselaer Polytechnic Institute ranks 53rd and is the highest ranked school that has all departments below "Strong" category. This means that approximately after the 22nd school the Average Rank Scores are determined more and more by the weights 35, 36, and 37 assigned to other than the top two categories. The Average Rank Scores of schools below the rank 53 are determined exclusively by the above weights.

For this reason the Average Rank Scores were used only for correlation and ranking purposes. They were not used to compute measures of dispersion or for any other comparisons based on measures of dispersion.

In order to study the dispersion in quality, Weighted Average Scores were computed. Weights 4, 3, 2, 1, 0 were assigned respectively to "Distinguished," "Strong," "Good," "Adequate plus," and "Less than adequate plus" categories. All the departments of a school were assigned these weights and the weighted averages computed. Ranking by these scores differs somewhat from ranking based on Average Rank Scores, the median difference being 2 rank points. In spite of these unavoidable differences, Weighted Average Scores were very valuable for comparisons based on dispersion of data because their computation is based on criteria that are numerically uniform for all the schools. The Weighted Average Scores are not shown here.

Figures on the number of professors listed in Who's Who in America /12/ were obtained for the departments of chemistry, economics, mechanical engineering, English, history, mathematics, and sociology. Only full and associate professors were considered and the numbers of those listed were expressed in percentage form.

The data on average compensation (and not on average salary) come, of course, from the AAUP Bulletins.

The student-faculty ratios were extracted from two publications. One of them is American Universities and Colleges /1/ and the other, The College Blue Book /4/.

As the reader may note, data are not always perfectly comparable and figures on compensation do not reflect regional differences in standards of living.

Ranking and Dispersion Between Schools. Most every professor ranks schools with regard to quality. Schools also are very often ranked by parents, students, and employers. It is, therefore, reasonable to assume that this ranking "business" is important and cannot be avoided. If so, then criteria and methods for less haphazard classification of schools should be developed.

When dealing with this problem, one should keep two things in mind. One is the method of measuring quality, and another is the choice of a most appropriate index of quality.

Ranking as an index of quality generally is used only if no other method is available or in cases where the quality index numbers increase or decrease uniformly. These principles apply very clearly to the classification of universities. If there is a sizeable number of schools of almost the same quality and another bunch in which the schools differ significantly among themselves then ranking is misleading. Ranking usually implies equal numerical or qualitative difference between the objects ranked.

To measure the amount of this kind of bunching, mean deviations of Weighted Average Scores were computed for the schools ranked 1-10, 11-20, ..., 51-60. Since these mean deviations are not comparable, coefficients of dispersion were obtained by dividing the mean deviations by their respective means. These coefficients are: .31, .095, .027, .022, .0057, .0047. They are significantly different. The dispersion of the ten top schools seems to be quite exaggerated. Is it due to differences in quality or reputation? Also, the bunching of schools with regard to quality does exist. The ranking, therefore, is not appropriate for the purpose of classifying schools, and, if possible, some kind of scores or index numbers should be used instead.

The ranking of schools, based on Average Rank Scores (and not on Weighted Average Scores) is shown in column (1) of the Appendix. The reason why ranking was used instead of Average Rank

Scores is that, due to the above mentioned weighting process, the first is superior to the second. These rank numbers should be used with caution. It is reasonable to assume that, on the average, our ranking may be in error by 1 to 5 points, and in some cases by more than that. Most of the time these are the sampling and statistical errors.

Additional information on the quality differences between schools is provided by columns (4) to (8) of the Appendix. The University of California at Berkeley has 100 per cent of its departments in two categories. West Virginia University has 93.3 per cent of departments in "Less than adequate plus" category. What a tremendous difference. Furthermore, 20 of 98 schools have 75 per cent or more of their departments that, according to Cartter's study, are not or are not quite adequate, and half the schools have 50 per cent or more departments in that category. This result does not make sense, no matter whether one compares schools or departments.

This conclusion is supported not only by common sense, but also by an analysis of the percentages of professors listed in Who's Who. The Who's Who percentages are listed in column (3) of the Appendix. As expected, there is only a moderate of correlation (r=.55) between these percentages and Average Rank Scores. It is, however, more important to note that schools with low ranks have relatively sizeable proportions of professors listed in Who's Who. For example, the 20 schools which have 75 per cent or more of departments in the "Less than adequate plus" category, have, on the average, 11.2 per cent of their professors listed in Who's Who, compared with 23.6 per cent for the top 20 schools. This difference is sizeable, but by no means is it as great as the difference between the best and worst schools based on the number of departments in various quality categories.

One may object to the present argument contending that listing in Who's Who is not always an indicator of scholarly standing. The response to this objection is that people listed by this publication did distinguish themselves in their fields of endeavor. These fields are most of the time compatible with the subjects that the listed professors teach.

The ranking of all schools, presented here, may not differ much from ranking that might be obtained from a more detailed and objective investigation. But the differences in quality (or the amount of dispersion) among schools, as indicated by the data in the Appendix, are without any doubt grossly exaggerated. Cartter's questionnaire demands a subjective evaluation of academic departments that can be achieved only from the knowledge of people who teach there. Since every participant of the survey had to evaluate about 100 departments, it is obvious that he had to consider only the nationally known scholars. However, there are many professors who are not nationally known, but who publish respectable pieces of work. They will certainly be overlooked by a participant of such a survey, although they are quite capable of training graduate students.

Ranking and Dispersion Within Schools. Ranking schools is meaningless if a sizeable portion of all the schools under consideration have departments of uniform quality and the quality of departments in other schools is widely dispersed. Some information on this problem is provided in column (9) of the Appendix. This column gives sums of percentages of two neighboring quality categories in which the number of departments is greatest. A frequency distribution has been made of these sums and it is shown in Table 1. Only the top 64 schools were enumerated since too many departments in the schools below that rank fall into "Less than adequate plus" category of column (8).

Table 1
Percent of Departments in Two Neighboring Quality Categories in which the Number of Departments is Greatest
(The distribution is made up of percentages for top 64 schools.)

Per cent	Number of schools
95-100	5
90-94.9	6
85-89.9	13
80-84.9	9
75-79.9	9
70-74.9	3
65-69.9	7
Less than 65	12
	64

Source: Column (9) of the Appendix.

The frequency distribution shows that 52 of 64, or 81 per cent of the schools under consideration have between 65 and 100 per cent of their departments in two neighboring quality categories. This result, together with a more detailed examination of the frequency distribution, leads us to an important conclusion: there is a great degree of uniformity between departments in the universities offering graduate degrees. One result of this analysis is that the interdepartmental dispersion in quality does not diminish the usefulness of classifying whole universities as to quality.

In addition to ranking, the information on interdepartmental dispersion of quality throws some light on the problem of change in quality. One has to suspect that most often change in quality does not occur in all departments simultaneously. Rather, a small number of departments lead the change. It means that schools in which a significant change in quality is taking place will have greater dispersion in quality between departments.

This point will be taken up in the following section. However, one already can make another interesting observation. Further analysis of columns (4) to (8) shows that only one school (University of Delaware) has one department two quality categories above the bulk of departments. Another school (M.I.T.) has four departments two categories below other departments. It is, therefore, most unlikely to find a very poor department in a high quality school and an excellent department in a low ranking university. It appears that an effort to upgrade a university, starting with one or two departments either was not present, or if it was, did not meet with success.

Forecasting Changes in Quality. The relationships between Average Rank Scores, average compensation, Who's Who percentages, and student-faculty ratios were analyzed with the help of regression analysis and some other statistical techniques.

The student-faculty ratios shown in column (2) of the Appendix are of interest to those teachers and students who like small classes. Since small classes are expensive, they are found predominantly in rich private schools. There is no correlation between quality of schools in general and the student-faculty ratios. The figures in column (2) are given merely to satisfy the reader's curiosity.

The regression analysis of the other three variables has been done with the help of parabolas. In regression analysis the quality of estimates was measured not by absolute, but by relative values of standard error of estimate. To make these comparisons, the standard errors were divided by the means of dependent variables. The coefficients of correlation and coefficients of dispersion of dependent variables around the regression curves are shown in Table 2.

Table 2
Coefficients of Correlation and
Coefficients of Deviation from the
Regression Line

The variables: 1 Average Rank Scores
2 Average Compensation Data
3 Per cent of Associate and Full Professors Listed in Who's Who in America

	1	2	3	
1	----	.777	.529	r
		.156	.184	$\sigma_{yx}/\bar{Y}$
2	.769	----	.634	r
	.100		.190	$\sigma_{yx}/\bar{Y}$
3	.556	.610	----	r
	.514	.475		$\sigma_{yx}/\bar{Y}$

Further information on the relationship between the three variables is provided by Figures 1 and 2. These are the polygons drawn from standardized ($z_i = x_i/\sigma_x$) figures representing Average Rank Scores, average compensations, and Who's Who percentages.

We note in Figure 1 that both polygons are quite similar and skewed to the right. Thus, not only more than 50 per cent of the schools are below the mean in quality and compensation, but the differences between the schools below the mean are much less than above the mean. The dispersion in quality below the mean is less than the dispersion in compensation, but not by much. As has been noted, this difference is due partly to weights used in the computation of Average Rank Scores. Similar results are obtained from Figure 2.

FIGURE 1

STANDARDIZED AVERAGE RANK SCORES AND AVERAGE COMPENSATION FIGURES

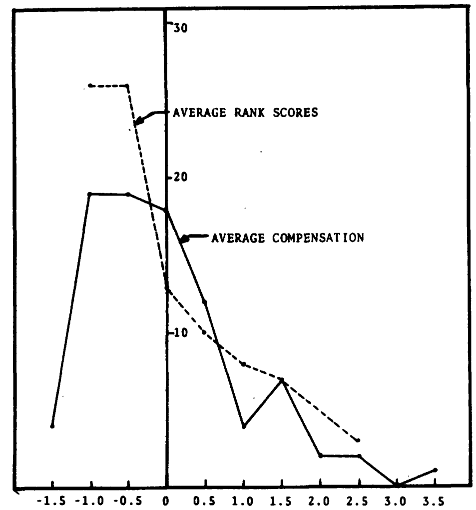
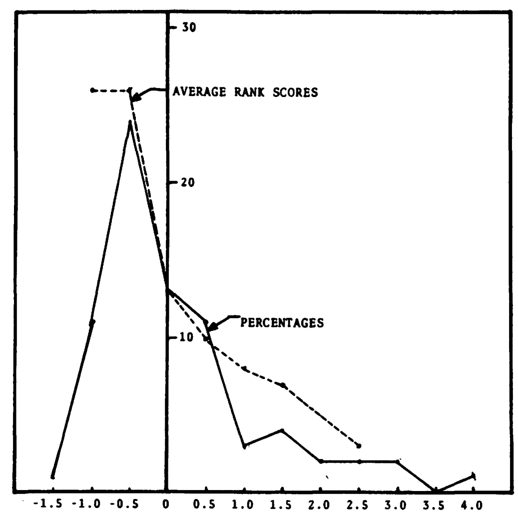


FIGURE 2

STANDARDIZED AVERAGE RANK SCORES AND PERCENTAGES OF ASSOCIATE AND FULL PROFESSORS LISTED IN WHO'S WHO IN AMERICA



To the extent that average compensation and Who's Who percentages are predictors of quality of schools, the two graphs tend to confirm our assumption that the classification of schools based on Average Rank Scores is not too bad. They also confirm our conclusion that the ranking of schools is not justified. If it were, the polygons would resemble a rectangular distribution.

Due to results of Table 2 average compensation can be used as a predictor of the quality of schools. This is important because average compensation figures are available every year, while any index of quality probably will be computed at considerable time intervals. This time interval is now and may continue so far into the future that it may be of interest to estimate significant changes in the quality of schools before the next quality index is computed. Further comparisons of quality and average compensation may provide additional clues in this respect.

Column (13) of the Appendix shows differences between quality and compensation ranks. The frequency distribution of these differences (without regard to signs) is shown in Table 3. The median difference is 10.7 rank points and the greatest difference is 55 points.

Table 3
Differences between Rank Numbers Expressing Classification as to Quality and Average Compensation of 89 Universities

Number of rank points	Number of universities
0-3	20
4-7	17
8-11	11
12-15	11
16-19	6
20-23	14
24-27	3
28-31	2
32-35	3
Less than 35	2
	89

Source: Column (13) of the Appendix.

The reasons for these differences are, of course, the same as the reasons for a relatively high value of $\sigma_{yx}/\overline{Y}$=15.6 per cent for regression of Average Rank Scores on compensation in Table 2. One reason is considerable change in compensation at the time of comparison so that there was no time for readjustment between quality and compensation rank. Other reasons are differences in the dispersion of salaries within a school (permitting employment of high and low quality instructors), differences in the administrative efficiency, statistical errors, and others.

The comparison of compensation and quality rank numbers for the purpose of estimation is, to a certain extent, justified because the shapes of distributions of standardized Average Rank Scores and average compensation figures are similar. On the other hand, one has to bear in mind that, due to the shapes of these two polygons, a ten point difference for the 20 top schools is more significant than, say, for the 40 middle quality schools. The following percentile ratios, computed from 1963/64 compensation figures, are indicative of the problem: P_0/P_{20}=1.380, P_{20}/P_{40}= 1.128, P_{40}/P_{60}=1.080, P_{60}/P_{80}=1.081, P_{80}/P_{100}= 1.179, and P_0/P_{100}=2.148. The differences between these percentiles in dollars are respectively: 4,572, 1,358, 790, 738, and 1,388. The difference between P_0 and P_{100} is $8,842. However, the estimates made with the help of regression analysis would suffer from the same weaknesses.

It is also interesting to note that the correlation between quality and compensation for schools at the top and bottom is greater than for the schools in the middle. This statement is supported by medians of rank differences of column (13) for schools ranked 1-22, 23-44, 45-66, and 67-89. The medians are: 6.0, 15.0, 18.0, and 7.0, to be compared with the already quoted median of 10.7 points for all schools.

Once again, the magnitude of differences in column (13) may be due to the amount of dispersion in salaries within a school, or to administrative efficiency, or to both so that only a significant difference in this column indicates a change in quality due to the change in average compensation. How substantial should this difference be?

Considering the amount of standard error of estimate of "compensation on Average Rank Scores" in Table 2, the nature of the frequency distribution in Table 3, and the above analysis, there should be ten or more rank points difference for about 20 top schools and at least 15 points difference for other schools. This, coupled with the relatively small value in column (2) of the Appendix, would be an indicator of change. If the number in column (9) is large (say, 65 per cent or more), greater rank difference would be required to indicate the quality change of a school.

Column (14) is crucial in making estimates of change in quality for an individual school. A large value in column (2) and a large value in column (13) may indicate that the quality of a school is <u>about</u> to change. If this large value in column (13) is confirmed or reinforced by the corresponding value of column (14), then there is great probability that the quality of that school actually is undergoing a change. Of course, a value of equivalent magnitude but with an opposite sign in column (14) would mean a reversal of the trend.

If a school has a low value in column (9) and a large value in column (13), then it may be suspected that the change in quality actually is occurring. If the value in column (14) is small, the change will be consummated. A large value in column (14) means - depending on the sign - either reinforcement or arrest of change.

The differences in column (14) also indicate that changes in compensation are less drastic in private than in public schools. But they often are quite drastic. For example, within a five year span 11 schools changed compensation rank by 25 or more points.

The shape of the polygon of compensation data for 1968/69 (not shown here) is similar to 1963/64 polygon, but the dispersion is less. The percentile ratios for 1968/69 are: P_0/P_{20}=1.218, P_{20}/P_{40}=1.096, P_{40}/P_{60}=1.062, P_{60}/P_{80}=1.070, P_{80}/P_{100}=1.263, and P_0/P_{100}=1.916. These ratios can be compared with already quoted ratios for 1963/64: 1.380, 1.127, 1.080, 1.081, 1.179, and 2.148. The difference is significant and it may indicate a process toward diminution of the dispersion in quality between schools.

REFERENCES

1. American Universities and Colleges. 10th edition. Washington, D. C.: The American Council on Education, 1968.

2. Berelson, Bernard. Graduate Education in the United States. New York: McGraw-Hill Book Co., 1960.

3. Cartter, Allan M. An Assessment of Quality in Graduate Education. Washington, D. C.: American Council on Education, 1966.

4. The College Blue Book. New York: C. C. M. Information Corporation, 1965.

5. Eells, Walter C. "Leading American Graduate Schools, 1948-1958." Liberal Education, XLVI (March, 1960), 112-16.

6. Hughes, Raymond M. A Study of the Graduate Schools of America. Oxford, Ohio: Miami University, 1925.

7. Kensiton, Hayward. Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania. The Educational Survey of the University of Pennsylvania, June 1958.

8. Knudsen, D. D. and Vaughn, T. R. "Quality in Graduate Education: A Re-evaluation of the Rankings of Sociology Departments in the Cartter Report." American Sociologist, IV (February, 1969), 12-19.

9. Lewis, L. S. "On Subjective and Objective Rankings of Sociology Departments." American Sociologist, IV (May, 1968), 129-31.

10. Report of Committee on Graduate Instruction. Washington, D. C.: American Council on Education, April 1934.

11. Shamblin, Don H. "Prestige and the Sociology Establishment." American Sociologist (May, 1970), 154-6.

12. Who's Who in America. Chicago, Ill.: A. N. Marquis, Inc., 1968/69.

APPENDIX

SUPPORTING DATA

DESCRIPTION OF COLUMNS: (1) QUALITY RANKING OF SCHOOLS. RANK NUMBERS BASED ON "AVERAGE RANK SCORES" DESCRIBED IN THE TEXT. (2) STUDENT-FACULTY RATIOS. (3) PERCENT OF ASSOCIATE AND FULL PROFESSORS FROM SELECTED DEPARTMENTS IN WHO'S WHO IN AMERICA. PERCENT OF DEPARTMENTS IN THE FOLLOWING QUALITY CATEGORIES: (4) "DISTINGUISHED"; (5) "STRONG"; (6) "GOOD"; (7) "ADEQUATE PLUS"; (8) "LESS THAN ADEQUATE PLUS". (9) SUM OF PERCENTAGES OF TWO NEIGHBORING QUALITY CATEGORIES IN WHICH THE NUMBER OF DEPARTMENTS IS GREATEST. (10) QUALITY RANKING OF SCHOOLS FOR WHICH 1963/4 AND 1968/9 COMPENSATION FIGURES WERE AVAILABLE. (11) SCHOOLS RANKED ACCORDING TO AVERAGE COMPENSATION FOR 1963/4 IN JUNE 1964 AAUP BULLETIN. (12) SCHOOLS RANKED ACCORDING TO AVERAGE COMPENSATION FOR 1968/9 IN JUNE 1969 AAUP BULLETIN. (13) COLUMN (10) MINUS COLUMN (11). (14) COLUMN (11) MINUS COLUMN (12).

Name of School	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
U. OF CAL. BERKELEY	1	18.2	26.3	79.3	20.7	---	---	---	100.0	1	16	23	-15	- 7
HARVARD U.	2	3	68.9	92.3	3.8	3.8	---	---	96.1	2	1	1	+ 1	0
CAL. INST. OF TECH.	3	3	21.6	61.5	38.5	---	---	---	100.0	3	4	2	- 1	+ 2
M. I. T.	4	5	22.5	60.0	13.3	---	---	26.7	73.3	4	5	7	- 1	- 2
U. OF MICHIGAN	5	8.7	17.5	28.6	71.4	---	---	---	100.0	5	21	14	-16	+ 7
STANFORD U.	6	11	34.8	38.5	53.8	3.8	---	3.8	92.3	6	2	5	+ 4	- 3
PRINCETON U.	7	5.5	56.6	37.5	54.2	4.2	---	4.2	91.7	7	7	12	0	- 5
U. OF CHICAGO	8	10	47.2	33.3	66.7	---	---	---	100.0	8	3	3	+ 5	0
U. OF WISCONSIN	9	15.6	21.0	24.1	65.5	10.3	---	---	89.6	9	44	51	-35	- 7
YALE U.	10	5.5	49.7	35.7	46.4	7.1	7.1	3.6	82.1	10	13	9	- 3	+ 4
COLUMBIA U.	11	6	35.9	28.6	53.6	10.7	7.1	---	82.2	11	6	15	+ 5	- 9
U. OF ILLINOIS	12	15.7	10.1	20.7	55.2	17.2	---	6.9	75.9	12	53	37	-41	+16
CORNELL U.	13	4	21.2	3.6	78.6	7.1	7.1	3.6	85.7	13	8	10	+ 5	- 2
U. OF MINNESOTA	14	13	10.6	3.6	75.0	14.3	7.1	---	89.2	14	34	41	-20	- 7
JOHN HOPKINS U.	15	5.5	33.3	3.8	76.9	15.4	---	3.8	92.3	15	9	8	+ 6	+ 1
U. OF CAL. LOS ANG.	16	20.9	23.4	---	75.0	10.7	---	14.3	85.7	16	16	23	0	- 7
U. OF PENNSYLVANIA	17	7	23.1	3.6	60.7	17.8	7.1	10.7	78.6	17	19	17	- 2	+ 2
ROCKEFELLER INST.	18	0.3	44.4	33.3	---	8.3	---	58.3	---	--	--	--	---	---
PURDUE U.	19	11	6.5	---	47.1	35.3	11.8	5.9	82.4	18	27	32	- 9	- 5
NORTHWESTERN U.	20	7	32.1	---	59.2	22.2	14.8	3.7	81.5	19	10	6	+ 9	+ 4
INDIANA U.	20	17	12.0	---	54.2	37.5	---	8.3	91.7	20	23	33	- 3	-10
YESHIVA U.	22	3	15.0	---	42.8	---	14.3	42.8	---	21	41	34	-20	+ 7
U. OF WASHINGTON	23	15.7	12.0	---	50.0	32.1	3.6	14.3	82.1	22	29	36	- 7	- 7
U. OF TEXAS	24	19	15.5	3.6	32.1	32.1	25.0	7.1	64.2	23	57	47	-34	+10
U. OF DELAWARE	25	20.5	9.7	20.0	---	---	20.0	60.0	80.0	--	--	--	---	---
DUKE U.	26	7	14.3	---	35.0	45.0	5.0	15.0	80.0	24	12	16	+12	- 4
BROWN U.	27	6	15.7	---	45.0	30.0	20.0	5.0	75.0	25	25	13	0	+12
CARNEGIE TECH.	28	9	22.2	---	44.4	44.4	11.1	---	88.8	--	--	--	---	---
BRANDEIS U.	29	7	47.1	---	27.3	36.4	9.1	27.3	63.7	26	26	20	0	+ 6
BROOKLYN POLYTECH.	30	14	13.3	---	28.6	28.6	14.3	28.6	57.2	27	48	81	-21	-33
U. OF N. CAROLINA	31	11	13.9	---	37.5	25.0	20.8	16.7	62.5	28	83	38	-55	+45
N. Y. U.	32	8.4	23.3	4.3	21.7	47.8	17.4	8.7	69.5	29	22	30	+ 7	- 8
U. OF CAL. DAVIS	33	15.5	14.9	---	28.6	21.4	---	50.0	50.0	30	16	23	+14	- 7
WESTERN RESERVE U.	34	6	35.2	---	20.0	5.0	25.0	50.0	75.0	--	--	--	---	---
ROCHESTER U.	35	7	10.6	---	15.0	45.0	10.0	30.0	60.0	31	11	11	+20	0
OHIO STATE U.	36	14.8	14.9	---	17.2	44.8	24.1	10.3	68.9	32	24	31	+ 8	- 7
U. OF IOWA	37	15.4	12.7	---	8.0	40.0	28.0	24.0	68.0	33	35	26	- 2	+ 9
WASHINGTON U.	38	6	9.2	---	16.0	40.0	24.0	20.0	64.0	34	28	27	+ 6	+ 1
IOWA STATE U.	39	15.9	16.1	---	11.8	52.9	11.8	23.5	64.7	35	47	66	-12	-13
EMORY U.	40	10	27.3	---	---	25.0	16.7	58.3	75.0	36	61	61	-25	0
U. OF PITTSBURGH	41	7.9	11.9	---	12.5	16.7	54.2	16.7	70.9	--	--	--	---	---
BRYN MAWR	42	7	19.0	---	6.7	26.7	13.3	53.3	66.6	37	15	46	+22	-31
PENN. STATE U.	43	13	8.3	---	8.0	28.0	36.0	28.0	64.0	--	--	--	---	---
SYRACUSE U.	44	12	8.3	---	8.0	12.0	52.0	28.0	80.0	38	32	28	+ 6	+ 4
MICHIGAN STATE U.	45	18	10.8	---	11.5	42.3	23.1	23.1	65.4	39	43	43	- 4	0
CASE INST. OF TECH.	46	9	23.0	---	25.0	12.5	37.5	25.0	62.5	--	--	--	---	---
RICE U.	47	10	44.3	---	18.8	37.5	12.5	31.2	56.3	40	20	19	+20	+ 1
CLAREMONT GRAD.	48	7	15.5	---	14.3	28.6	28.6	28.6	57.2	41	14	4	+27	+10
U. OF UTAH	49	15	10.7	---	4.5	9.1	18.2	68.2	86.4	42	55	55	-13	0
U. OF KANSAS	50	8	13.5	---	7.4	14.8	33.3	44.4	77.8	43	65	60	-22	+ 5
U. OF CINCINNATI	51	8	12.0	---	4.8	4.8	33.3	52.4	85.7	44	67	59	-23	+ 8
RUTGERS STATE U.	52	12	9.9	---	3.8	26.9	15.4	53.8	69.2	45	56	29	-11	+27
U. OF VIRGINIA	53	8	30.7	---	5.3	15.8	26.3	52.6	78.9	46	30	22	+16	+ 8
RENSSELAER POLYTECH.	54	12	25.7	---	---	50.0	25.0	25.0	75.0	47	36	45	+11	- 9
U. OF MARYLAND	55	12	14.2	---	3.8	7.7	42.3	46.2	88.5	48	74	84	-26	-10
LEHIGH U.	56	10	10.3	---	9.1	---	27.3	63.6	90.9	49	45	40	+ 4	+ 5
U. OF OREGON	57	11	16.2	---	4.3	26.1	30.4	39.1	69.5	50	50	50	0	0
SUNY BUFFALO	58	7.9	12.1	---	6.7	6.7	13.3	73.3	86.7	51	69	18	-18	+51
OREGON STATE U.	59	16	4.3	---	---	28.6	28.6	42.8	71.5	52	62	71	-10	- 9
FLORIDA STATE U.	60	14	12.5	---	7.1	---	42.8	50.0	92.8	53	72	62	-19	+10
U. OF COLORADO	61	12	10.6	---	---	15.4	46.2	38.5	84.7	54	31	56	+23	-25
U. OF S. CALIFORNIA	62	10	14.4	---	---	12.5	50.0	37.5	87.5	55	66	39	-11	+27
TUFTS U.	63	9	19.1	---	---	10.5	10.5	78.9	89.4	56	33	54	+23	-21
GEORGIA TECH.	64	11	8.8	---	---	14.3	42.8	42.8	85.8	57	60	53	- 3	+ 7
VANDERBILT U.	65	5	18.8	---	4.3	17.4	21.7	56.5	78.2	58	39	21	+ 9	+18
CLARK U.	66	12	8.8	---	---	33.3	---	66.7	---	59	38	44	-21	+ 6
U. OF FLORIDA	67	12.1	10.3	---	---	13.6	36.4	50.0	86.4	60	77	72	-17	+ 5
TULANE U.	67	8	11.7	---	---	16.0	16.0	68.0	84.0	61	51	48	+10	+ 3
ILL. INST. OF TECH.	69	14	14.7	---	---	21.4	14.3	64.3	78.6	62	42	65	+20	-23
N. CAROLINA STATE	70	13.6	12.0	---	---	23.1	7.7	76.9	84.6	63	40	78	+23	-38
WASHINGTON STATE U.	71	17	16.1	---	---	15.8	21.0	63.2	84.2	64	36	45	+28	- 9
NOTRE DAME U.	72	13	23.9	---	---	7.1	35.7	57.1	92.8	65	46	57	+19	-11
KANSAS STATE U.	73	17	10.6	---	7.7	---	23.1	69.2	92.3	66	73	79	- 7	- 6
WAYNE STATE U.	74	24	6.6	---	---	5.3	26.3	68.4	94.7	67	52	52	+15	0
NEW SCHOOL FOR S. R.	75	6.1	8.7	---	---	---	40.0	60.0	100.0	--	--	--	---	---
FORDHAM U.	76	7	10.9	---	---	---	36.4	63.6	100.0	68	82	75	-14	+ 7
ST. LOUIS U.	77	8	10.5	---	---	5.9	17.6	76.5	94.1	69	76	88	- 7	-12
U. OF NEW MEXICO	78	23	11.1	---	---	8.3	16.7	75.0	91.7	70	58	73	+12	-15
LOUISIANA STATE U.	79	18	16.1	---	4.5	---	18.2	77.3	95.5	71	68	89	+ 3	-31
U. OF CONNECTICUT	80	17	9.8	---	---	---	31.2	68.8	100.0	72	64	35	+ 8	+29
U. OF ARIZONA	81	22	8.6	---	4.3	---	17.4	78.3	95.7	--	--	--	---	---
U. OF NEBRASKA	82	18	14.8	---	---	---	29.2	70.8	100.0	73	59	67	+14	- 8
CATHOLIC U.	82	10	7.8	---	---	7.1	14.3	78.6	92.9	74	70	63	+ 4	+ 7
U. OF OKLAHOMA	84	18	24.0	---	---	---	26.1	73.9	100.0	75	79	80	- 4	- 1
GEORGE PEABODY C.	85	11	10.5	---	---	---	25.0	75.0	100.0	76	89	87	-13	+ 2
U. OF MISSOURI	86	10	15.5	---	---	---	24.0	76.0	100.0	77	78	76	- 1	+ 2
TEXAS A & M	87	13	3.0	---	---	---	23.1	76.9	100.0	78	88	70	-10	+18
G. WASHINGTON U.	88	13.3	15.3	---	---	5.6	11.1	83.3	94.4	79	49	42	+30	+ 7
OKLAHOMA STATE U.	89	20	6.7	---	---	---	26.1	73.9	100.0	80	81	83	- 1	- 2
U. OF TENNESSEE	90	10	14.4	---	---	---	20.0	80.0	100.0	81	68	85	+13	-17
U. OF MASSACHUSETTS	90	15	8.8	---	---	---	20.0	80.0	100.0	82	75	69	+ 7	+ 6
VIRGINIA POLYTECH.	92	11	6.9	---	---	---	15.4	84.6	100.0	83	80	77	+ 3	+ 3
U. OF KENTUCKY	93	11	12.4	---	---	---	11.8	88.2	100.0	84	87	49	- 3	+38
BOSTON U.	93	10	12.1	---	---	---	12.5	87.5	100.0	85	63	64	+22	- 1
TEMPLE U.	95	16	10.5	---	---	---	9.1	90.9	100.0	86	54	68	+32	-14
GEORGETOWN U.	95	5	8.0	---	---	---	9.1	90.9	100.0	87	84	74	+ 3	+10
U. OF WYOMING	97	15	8.3	---	---	---	7.7	92.3	100.0	88	86	86	+ 2	0
WEST VIRGINIA U.	98	12	12.3	---	---	---	6.7	93.3	100.0	89	85	82	+ 4	+ 3

CROSS-SECTIONAL VS. COHORT ESTIMATES OF LIFETIME INCOME

Herman P. Miller and Richard A. Hornseth
U.S. Department of Commerce
Bureau of the Census

The purpose of this paper is to compare estimates of lifetime income based on cross-sectional data with those based on life-cycle data for selected cohorts of men. Estimates of lifetime income now available are based on the variation of income with age reported during a single survey year. The 1970 census will show the amount of income received in 1969 by men who, at the time of the census, were 18-24 years old, 25-34, etc., or other age groups. Estimates of lifetime income from, say, starting age 18, can be prepared from these figures if it is assumed that the variation of income with age for men 18 to 64 years old in 1969 approximates the pattern of income that will be received by 18 year olds as they age during their working lifetime provided some suitable adjustment can be made each year for the change in income due to inflation and economic growth. Until recently this has been an untested assumption because there are no data which trace a man's earnings from the time he starts working until he retires. It is now possible to simulate such data for selected cohorts of men on the basis of information on income and age collected by the Bureau of the Census in the 1950, 1960 and 1970 Censuses and in the Current Population Survey in each year since 1947. Men who were 25 to 34 years old in the 1950 census were 35 to 44 years old in 1960 and 45 to 54 years old in 1970. By comparing the incomes reported for each of these age groups in the past three censuses we can observe the actual variation of income by age for the cohort of men born in 1915-24 and we can compare those incomes with the data reported for men in the same age groups in the 1950 census. A similar comparison for the cohort born in 1925-34 would provide a test for the 35 to 44 year age group. These comparisons could be made for all men as well as for whites and for nonwhites; they could also be made for education groups if it is assumed that men do not complete additional years of schooling beyond age 25. The analysis could also be extended to occupation groups, but this involves the doubtful assumption that men remain in the same occupation during their working lifetime.

Since the 1970 census data have not been tabulated at present, analysis must be made on the basis of Current Population Survey data. Figures are available for 6 cohorts of men: those born in 1913-22, 1914-23, 1915-24, 1903-12, 1904-13, and 1905-14; i.e., those 25-34 in 1948, 1949, and 1950 and those 35-44 in those same years. On the basis of these data we can trace the average incomes of men 25 to 34 years old in the March 1948 CPS, 35 to 44 in the March 1958 CPS and 45 to 54 years old in the March 1968 CPS. The same information is available for men 35 to 44 years old, 45 to 54 years old and 55 to 64 years old in the same surveys. Similar data are also available from surveys conducted in March 1949, March 1959 and March 1969 and for March 1950, March 1960 and March 1970.

CROSS-SECTIONAL DATA ON INCOME BY AGE

Table 1 below presents the basic information now used to compute estimates of lifetime income. It shows for each year the average income reported for men in various age groups. These figures are taken directly from the published CPS reports. In the typical method of computing lifetime income, the averages for these age groups are converted to estimates of average income for single years of age by fitting a parabolic function to the age group data. The value of expected income from age 30 to 54 and from age 40 to 64 is then computed according to the procedure described below.[1] The results of this calculation are shown in Table 2, based on the cross-sectional income surveys conducted in March 1948, March 1949 and March 1950. The 25-year earnings figures on the lines for 1947, 1948 and 1949 in this table are based on the average incomes obtained in cross-section surveys with separate figures for 0, 4, 4 1/2 and 5 percent allowance for the increase in income due to inflation and economic growth. The figures on the lines for 1967, 1968, and 1969 are based on the average incomes for cohorts. Although it is customary to discount future incomes to present values, a zero discount rate is used in order to focus more directly on the issue under consideration. Also, it is customary to make allowance for the probability of death, but this too was deleted for the sake of simplicity of presentation.

COHORT DATA ON INCOME BY AGE

With respect to the data in table 2 along the diagonals showing income by cohorts, it should be noted that unlike the cross-section data which show a leveling off of income between the ages of 35 and 54 and a decline thereafter, the cohort data show a continuous rise in average income extending well into the mid-sixties. The reason for the difference is that the cross-section data show the variations in the payments made by the economy to men in different age groups at a given point in time. They reflect the employment conditions, demand for and supply of men with different amounts of experience and training, occupational wage differentials and other relevant factors as they exist at a given point in time. In contrast, the cohort data approximate the results that would be obtained by tracing a man's actual earnings over a period of time. They therefore reflect changes in income due to inflation and general increases in wage levels.

The figures in table 2 on expected earnings from age 30 to 54 and from age 40 to 64 based on cohort data are based on a parabola that was fit to the average incomes for the appropriate age groups. An interesting feature of table 2 is the comparison of the percent change in income for an age group over one and two decades and the allowances for growth and inflation in the cross-sectional estimates of 25 year earnings. In the first block we see that income rose at 4.7% per year not only in the ten year period from 1947 to 1957 for the age groups 35-44, but also in the twenty year period from 1947 to 1967 for the 45-54 year age group. It is remarkable that the estimate of $149,000 for 25 year earnings based on fitting a parabola to the cohort data is very close to the estimate of $154,000 based on fitting a parabola to the cross-sectional data with an allowance of 4.5% for economic growth and inflation. This relation holds well throughout the whole table. It appears then that for at least over the past 20 years or so, the relation of income to age and the rates of inflation and economic growth have been so surprisingly stable that estimates of lifetime earnings from cross-sectional data are quite reliable.

One indication of the stability of the relationship between income and age is the similarity of the growth rates for each age group. A curve, $Yn = A\ (1 + X)^T$ was fitted by least squares (logarithms) to the 23 years of income data shown in Table 1 for each of the age groups 25-34, 35-44, 45-54, and 55-64. The results show that the least squares estimate of X, the rate of inflation and growth over the 23 year period was 4.93% for the 25-34 year group, 5.03% for the 35-44 year group, 4.91% for 45-54, and 4.96% for the 55-64 group.

The figures below are the coefficients of a least square parabola ($Yn = A + BN + CN^2$) fitted to the mean incomes deflated by 5% per year, for the 4 age groups 25-34, 35-44, 45-54, and 55-64, for each year 1947 to 1969. In this equation, N is age, scaled so that $N = \frac{\text{actual age}-1}{10}$.

For example, for 1947:

$$Yn = 1511 + 1493\ N - 293.5\ N^2.$$

The change coefficient B (Col. 2) and the rate of change C (Col. 3) are both quite stable over time. This standardization process reveals clearly that the parabolic coefficients are remarkably similar over the 23 years; that is, the age curve for income is very stable.

Thus, it appears that estimates of lifetime income based on cross-sectional data can be quite adequate if the distributions of income by age and the rates of inflation and economic growth are as stable as they have been over the past twenty years.

	A	B	C
1947	1,511	1,493	-293
1948	1,787	1,241	-248
1949	1,680	1,112	-225
1950	1,505	1,439	-285
1951	2,038	1,104	-232
1952	2,057	1,031	-205
1953	2,118	1,007	-207
1954	1,716	1,210	-243
1955	1,808	1,087	-205
1956	1,897	1,202	-248
1957	1,813	1,237	-257
1958	1,849	1,121	-231
1959	2,003	984	-193
1960	1,713	1,310	-267
1961	1,843	1,113	-219
1962	1,526	1,371	-272
1963	1,695	1,157	-226
1964	1,756	1,144	-236
1965	1,649	1,323	-269
1966	1,744	1,264	-256
1967	1,689	1,244	-251
1968	1,781	1,200	-236
1969	1,805	1,335	-267

[1]For example, we take the incomes as centered on the mid ages of the groups (see below) and

Age	1947 Income
30	2704
40	3344
50	3329

fit a parabola, $Yn = A+BN+CN^2$, where N represents age; Yn represents the parabolic estimated value of income for any age N; and A, B, and C are coefficients obtained by fitting to the observed (above) 3 points. For the above values the equation becomes

$$Yn = 3146 + 293.25N - 3.275\ N^2.$$

From this we can estimate the incomes at every age N from say 30 to 54 and add them together to estimate 25-year earnings for a man age 30, without of course, allowance for death and for the effect of inflation or economic growth on income during the 25 years. Incidentally the result here is $80,577. See the first entry under "expected income" in table 2.

We make allowance for inflation and economic growth by multiplying the estimated income (Yn) at each age (N) by $(1 + X)^{N-30}$ where X is an assumed constant rate of inflation and growth per year and N-30 is the number of years that intervene before a man now age 30 reaches age N. Applying an allowance of 4.0% per year to the above data we obtain $142,000, the second figure in table 2.

Table 1.—Mean Income in 1947 to 1969 of Males 14 Years Old and Over, by Age, for United States

(In current dollars)

Year	All ages	14-19	20-24	25-34	35-44	45-54	55-64	65 and over
1947.............	2,636	746	1,641	2,704	3,344	3,329	2,795	1,906
1948.............	2,755	692	1,911	2,898	3,508	3,378	2,946	1,778
1949.............	2,677	638	1,808	2,842	3,281	3,331	2,777	1,827
1950.............	2,994	634	1,969	3,068	3,782	3,733	3,123	1,894
1951.............	3,275	744	2,326	3,511	4,112	3,884	3,356	1,891
1952.............	3,406	745	2,311	3,707	4,130	4,292	3,664	2,211
1953.............	3,549	803	2,285	3,909	4,435	4,382	3,796	2,217
1954.............	3,565	737	2,310	3,774	4,461	4,437	3,755	2,151
1955.............	3,748	668	2,439	3,984	4,644	4,786	4,231	2,295
1956.............	4,005	708	2,670	4,403	5,200	5,012	4,270	2,351
1957.............	4,068	678	2,614	4,552	5,300	5,227	4,295	2,233
1958.............	4,162	671	2,643	4,681	5,433	5,345	4,515	2,283
1959.............	4,487	708	2,783	4,956	5,926	5,587	5,166	2,559
1960.............	4,659	693	2,808	5,160	6,271	5,994	5,089	2,775
1961.............	4,893	680	2,960	5,389	6,424	6,263	5,563	3,229
1962.............	4,970	708	2,959	5,379	6,847	6,387	5,587	2,943
1963.............	5,125	708	2,928	5,705	6,872	6,747	5,938	3,097
1964.............	5,329	773	3,172	6,073	7,211	6,914	5,884	3,534
1965.............	5,692	901	3,532	6,458	7,878	7,538	6,360	3,241
1966.............	6,019	942	3,728	6,935	8,257	8,098	6,825	3,335
1967.............	6,159	1,015	3,669	7,107	8,443	8,342	7,004	3,667
1968.............	6,626	1,042	3,896	7,617	9,115	8,967	7,828	3,988
1969.............	7,202	1,099	4,149	8,378	10,042	9,873	8,405	4,306

SOURCE: U.S. Bureau of the Census, Current Population Reports, Series P-60, annual issues.

Table 2.—Expected Income From Age 30 to 54 and From Age 40-64 Based on Cross-Sectional and Cohort Data

(Expected income in thousands of dollars)

Year	Mean Income by age			Expected income from age 30-54 with assumed annual growth rate of:			
	25-34	35-44	45-54	0%	4.0%	4.5%	5.0%
1947................	2,704	3,344	3,329	80	142	154	166
1957................		5,300					
1967................			8,342	149			
Growth rate.........		4.7%	4.7%				
1948................	2,898	3,508	3,378	83	147	158	171
1958................		5,433					
1968................			8,967	157			
Growth rate.........		4.5%	5.0%				
1949................	2,842	3,281	3,331	80	142	154	166
1959................		5,926					
1969................			9,873	171			
Growth rate.........		6.1%	5.6%				
Year	35-44	45-54	55-64	Expected income from age 40-64			
1947................	3,344	3,329	2,795	78	132	142	153
1957................		5,227					
1967................			7,004	139			
Growth rate.........		4.6%	4.7%				
1948................	3,508	3,378	2,946	80	137	148	159
1958................		5,345					
1968................			7,828	148			
Growth rate.........		4.7%	5.0%				
1949................	3,281	3,331	2,777	77	131	141	152
1959................		5,587					
1969................			8,405	156			
Growth rate.........		5.3%	5.7%				

A TEST OF LOCAL INDEPENDENCE IN LATENT STRUCTURE ANALYSIS

H. I. Patel, University of Wisconsin-Milwaukee

1. Introduction

In many areas, especially in the social and behavioral sciences, we are interested in assessing some trait or characteristic of individuals which cannot be observed directly. Suppose each individual belongs, unknown to us, to one of several classes. Then since these classes are unobservable, they are called latent classes. In this situation we usually observe the response items which provide the information pertinent to the latent classes. The latent classes are characterized by a set of latent parameters which can be estimated by solving a series of accounting equations linking the latent parameters and the empirically observed response frequencies. The main assumption to be made for solving this system of accounting equations is the axiom of local independence which establishes the independence of the responses of different items within a given latent class. In other words the dependence among the response-items is explained by the latent variable.

The original work of latent structure analysis is due to Lazarsfeld [1950, 1954]. Various techniques for estimating the latent parameters have been proposed by Lazarsfeld and Dudman [1951], Green [1951], Gibson [1955, 1959], McHugh [1954, 1956], Anderson [1954, 1959], Madansky [1958, 1960] and others. The relation of latent structure analysis to factor analysis has been studied by Bargmann [1957], Gibson [1951] and Green [1952]. An excellent treatment of latent structure analysis can be found in Lazarsfeld and Henry [1968].

Under the assumption of local independence the maximum likelihood estimates of the latent parameters are obtained using an iterative procedure and a test of goodness fit is proposed.

2. Formulation of the Model

We define a set of dichotomous variables $X_1, X_2, \ldots, X_p$ to be observed on a randomly selected individual by $P(X_i=1) = p_i$ and $P(X_i=0) = 1-p_i$, $i = 1,2,\ldots,p$. Then the vector $\underset{\sim}{X}$ has a point multivariate binomial distribution whose f.m.g.f. is given by (Krishnamoorthy [1947])

$$(1)\ \varphi(t_1,\ldots,t_n) = (1 + \Sigma\, p_i t_i + \sum_{ij} p_{ij} t_i t_j + \ldots + p_{12\ldots p} t_1 t_2 \ldots t_p),$$

where

$$p_{ij} = P(X_i=1, X_j=1)$$

$$\ldots$$

$$p_{12\ldots p} = P(X_1=1, X_2=1, \ldots, X_p=1).$$

We now consider that each individual in this population can be classified into one of the m+1 latent classes B^α ($\alpha = 0,1,\ldots,m$), so that $P(B^\alpha) = \pi_\alpha$ and $\pi_0+\pi_1+\ldots+\pi_m=1$. Let H be the hypothesis of local independence which establishes independence of $X_1, X_2, \ldots, X_p$ within a latent class B^α (see Lazarsfeld and Henry [1968]). The hypothesis H can be stated as

$$(2)\ P(X_{i_1}=1, X_{i_2}=1, \ldots, X_{i_k}=1 \mid B^\alpha) = p^\alpha_{i_1} p^\alpha_{i_2} \ldots p^\alpha_{i_k},\ 1 \le i_1 < i_2 < \ldots < i_k \le p,\ k=2,\ldots,p,$$

where $P(A \mid B^\alpha)$ denotes the conditional probability of the event A given that an individual belongs to class B^α and $p^\alpha_j = P(X_j=1 \mid B^\alpha)$. Hence the characteristic function of $\underset{\sim}{X} = (X_1, \ldots, X_p)'$, under H, is

$$(3)\ \phi(t_1,\ldots,t_p) = \sum_{\alpha=0}^{m} \pi_\alpha \prod_{j=1}^{p} (1-p^\alpha_j + p^\alpha_j e^{it_j}),$$

and by the inversion theorem it can be shown that the probability function (p.f.) of $\underset{\sim}{X}$ is

$$(4)\ f(x_1,\ldots,x_p) = \sum_{\alpha=0}^{m} \pi_\alpha \prod_{i=1}^{p} (p^\alpha_i)^{x_i} \times (1-p^\alpha_i)^{1-x_i},\ x_i=0,1\ (i=1,2,\ldots,p).$$

The p.f. in (4) may be considered as a mixture of $f^\alpha(x_1,\ldots,x_p)$ for $\alpha=0,1,\ldots,m$. where

$$f^\alpha(x_1,\ldots,x_p) = \prod_{i=1}^{p} (p^\alpha_i)^{x_i}(1-p^\alpha_i)^{1-x_i}.$$

The alternative hypothesis H_a is that the dependence among $X_1, X_2, \ldots, X_p$ cannot be explained by the latent variable, i.e., the p.f. of $\underset{\sim}{X}$ cannot be written as in (4).

3. Maximum Likelihood Estimates under H

If $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_N$ is a random sample from the population with p.f. (4), the log of the likelihood function, under H, is given by

$$(5)\ \log L = \sum_{t=1}^{N} \log \sum_{\alpha=0}^{m} \pi_\alpha W_{\alpha t},$$

where

$$W_{\alpha t} = \prod_{i=1}^{p} (p^\alpha_i)^{x_{it}} (1-p^\alpha_i)^{1-x_{it}},$$

$$\alpha = 0,1,\ldots,m;\ t=1,2,\ldots,N.$$

Setting the first partial derivatives of log L with respect to the parameters equal to zero, we get

(6) $\sum_{t=1}^{N} W_{\alpha t}/g_t = \sum_{t=1}^{N} W_{ot}/g_t$, $\alpha=1,2,\ldots,m$,

(7) $p_i^\alpha = (\sum_{t=1}^{N} W_{\alpha t}x_{it}/g_t)/\sum_{t=1}^{N}(W_{\alpha t}/g_t)$,

where $g(t)=\pi_o W_{ot}+\pi_1 W_{1t}+\ldots+\pi_m W_{mt}$. Multiplying both sides of equation (6) by π_α, and summing over α, we get

$$\sum_{\alpha=0}^{m}\sum_{t=1}^{N}\pi_\alpha W_{\alpha t}/g_t = \sum_{\alpha=0}^{m}\sum_{t=1}^{N}\pi_\alpha W_{0t}/g_t$$

or

(8) $\pi_\alpha = (\pi_\alpha/N)\sum_{t=1}^{N} W_{\alpha t}/g_t$, $\alpha=1,2,\ldots,m$.

If we have a set of initial estimates of $\pi_\alpha(\alpha=1,2,\ldots,m)$ and p_i^α $(\alpha=0,1,\ldots,m;i=1,2,\ldots,p)$, the equations (7) and (8) can be used as the basis for iteration scheme. A method for obtaining initial estimates is given in the next section.

It should be observed that the distribution given by the p.f. W_α is a member of the exponential family and hence the population in (4) is a mixture of m+1 sub-populations each being a member of the exponential family. Hasselblad [1969] has given an algorithm for obtaining the maximum likelihood estimates in finite mixtures of distributions. Based on over three hundred trials with several mixtures of distributions, he concludes that (i) the likelihood function increases monotonically, ignoring the oscillation, (ii) the procedure attains a relative maximum of the likelihood function, and (iii) in general, different sets of initial estimates give the same final estimates to a certain accuracy. The same convergence properties will be possessed by our algorithm.

4. Initial Estimates

Various methods are available for estimating the latent parameters. Such estimates can be used as initial estimates in ite-·tive procedure of maximum likelihood ·.nod. In this section we propose a method based on a factor analysis solution. The estimates obtained by this method are not unique and perhaps are less efficient than the present methods available. However, they can be used as initial estimates in the iterative procedure given in the previous section.

We have

$$\sigma_i^2 = \mathrm{Var}(X_i) = p_i(1-p_i), i=1,2,\ldots,p$$

(9)

$$\sigma_{ij} = \mathrm{Cov}(X_i,X_j) = p_{ij} - p_ip_j,$$

$$i \neq j=1,2,\ldots,p.$$

Further, $p_i=\pi_o p_i^o+\pi_1 p_i^1+\ldots+\pi_m p_i^m$.

The axiom of local independence implies the pairwise independence. Writing

$$p_{ij}^\alpha = P(X_i=1,X_j=1 \mid B^\alpha),$$

we get

(10) $p_{ij}^\alpha = p_i^\alpha p_j^\alpha$, $\alpha=0,1,\ldots,m$,

$i \neq j=1,2,\ldots,p$.

Hence $p_{ij} = \sum_{\alpha=0}^{m}\pi_\alpha p_i^\alpha p_j^\alpha$.

With these notations, under the assumption of local independence, we get the covariance matrix of $\underset{\sim}{X}$ as

(11) $\Sigma = D_u+YVY'$,

where D_u is a diagonal matrix with the ith diagonal element equal to

$$\sum_{\alpha=0}^{m}\pi_\alpha p_i^\alpha(1-p_i^\alpha),$$

$$Y = \begin{bmatrix} p_1^o-p_1^1 & p_1^o-p_1^2 & \cdots & p_1^o-p_1^m \\ p_2^o-p_2^1 & p_2^o-p_2^2 & \cdots & p_2^o-p_2^m \\ \cdot & \cdot & \cdots & \cdot \\ p_p^o-p_p^1 & p_p^o-p_p^2 & \cdots & p_p^o-p_p^m \end{bmatrix}$$

and the elements of matrix V are

$$v_{ii} = \pi_i(1-\pi_i),\ i=1,2,\ldots,m$$

$$v_{ij} = -\pi_i\pi_j,\quad i \neq j=1,2,\ldots,m.$$

Since V is a symmetric and positive definite matrix, it can be written as TT', where T is a triangular matrix with real elements. Thus the correlation matrix, P, of $\underset{\sim}{X}$ can be written as

(12) $P = D_w + \Phi\Phi'$,

where $D_w = D_\sigma^{-1}D_uD_\sigma^{-1}$, $\Phi = D_\sigma^{-1}YT$ and D_σ is a diagonal matrix with elements $\sigma_1,\sigma_2,\ldots,\sigma_p$.

Thus the correlation matrix of $\underset{\sim}{X}$ has the structure of factor analysis which we arrived at through the assumption of local independence. The matrix Σ should be estimated by the sample covariance matrix of $\underset{\sim}{X}$. Let F be any admissible factor analysis solution for Φ, such that $-1 \leq f_{i\alpha} < 1$ for all elements of F and $0 \leq \sum_{\alpha=1}^{m} f_{i\alpha}^2 < 1$ for all i.

Then $F^* = FA$ satisfying the above admissible constraints is also a solution, where A is an orthogonal matrix. If some knowledge about the parameters π_α's is available from the past experience, we can estimate Y as $\hat{Y}=D_\sigma F^*T^{-1}$, such that each element of $\hat{Y}$ is less than 1 in absolute value. Thus $p_i^0 - p_i^\alpha$ can be estimated for all i and $\alpha = 1,2,\ldots,m$. Using the relation $\sum_{\alpha=0}^{m} \pi_\alpha p_i^\alpha = p_i$, we can estimate all p_i^α. p_i are to be estimated from the sample. It should be noted that all estimates p_i^α must lie on the open interval (0,1).

It is clear that when m=1 (i.e., dichotomous latent variable), the matrix Φ can be estimated uniquely, if π_0 is known and hence all p_i^α can be estimated uniquely. When $\pi_0,\pi_1,\ldots,\pi_m$ are not known, it is recommended to take equal values for all π_α's. A drawback of this method of obtaining initial estimates is that they are based on a derived solution of the factor analysis. Further, it may be noted that any factor analysis method can be used to obtain an estimate of Φ.

5. Test of Goodness of Fit

Let N be the number of individuals interviewed and $N_{j_1 j_2 \ldots j_p}$ the number of individuals with response vector $(j_1,j_2,\ldots,j_p)$, where $j_i = 0$ or 1. If $p_{j_1 j_2 \ldots j_p} = P(X_1=j_1, X_2=j_2,\ldots,X_p=j_p)$, the observed probability of the response vector $(j_1,j_2,\ldots,j_p)$ is equal to $N_{j_1 j_2 \ldots j_p}/N$. The maximum likelihood estimate of $p_{j_1 j_2 \ldots j_p}$ is given by

$$\hat{p}_{j_1 j_2 \ldots j_p} = \sum_{\alpha=0}^{m} \hat{\pi}_\alpha \prod_{i=1}^{p} (\hat{p}_i^\alpha)^{j_i} (1-\hat{p}_i^\alpha)^{1-j_i},$$

where $\hat{\pi}_\alpha$ and $\hat{p}_i^\alpha$ are the maximum likelihood estimates of π_α and p_i^α. In sample data one may not observe all of the 2^p distinct vectors $(j_1,j_2,\ldots,j_p)$. Suppose k is the number of distinct observed vectors. If $k < 2^p$, the probabilities of the observed vectors should be adjusted so as to make the total expected frequency N. Finally

$$\chi^2 = \sum_{j_1,j_2,\ldots,j_p} \left(N_{j_1 j_2 \ldots j_p} - N\hat{p}_{j_1 j_2 \ldots j_p}\right)^2 / N\hat{p}_{j_1 j_2 \ldots j_p}$$

with k-1-[(m+1)p+m] d.f. can be used as a test statistic.

Acknowledgement

This research was supported, in part, by the office of Naval Research, Contract Number N00014-69-A-0423, Department of Statistics, University of Georgia.

REFERENCES

[1] Anderson, T.W., "On estimation of parameters in latent structure analysis", *Psychometrika*, 19, (1954) 1-10.

[2] ____________, "Some scaling methods and estimation procedures in the latent class model", Probability and Statistics, Ulf Grenander, ed., New York: Wiley, 1959, 9-38.

[3] Bargmann, R. E., "A study of independence and dependence in multivariate normal analysis", Inst. of Stat., University of North Carolina, 1957, No. 186.

[4] Gibson, W. A., "Applications of Mathematics of multiple factor analysis to problems of latent structure analysis", Unpublished Ph.D . dissertation, University of Chicago, 1951.

[5] ____________, "An extension of Anderson's solution for the latent structure equations", Psychometrika, 20, (1955), 69-73.

[6] ____________, "Three multivariate methods: factor analysis, latent structure analysis and latent profile analysis", Psychometrika, 24, (1959), 229-52.

[7] Green, Bert F., "A general solution for the latent class model of latent structure analysis", Psychometrika, 16, (1951), 151-66.

[8] ____________, "Latent structure analysis and its relation to factor analysis", J. A. S. A., 47, (1952) 71-76.

[9] Hasselblad, Victor, "Estimation of finite mixture of distributions from the exponential family", J.A.S.A., 64, (1969), 1459-71.

[10] Krishnamoorthy, A. S., "Multivariate binomial and Poisson distributions", Sankhye, 11, (1947), 117-24.

[11] Lazarsfeld, Paul F., "The logical and mathematical foundation of latent structure analysis", Measurement and Prediction, S.A. Stouffer, et al., Princeton; Princeton University Press, 1950.

[12] ________________, "A conceptual introduction to latent structure analysis", Mathematical Thinking in Social Sciences, Paul F. Lazarsfeld ed., New York: The Free Press, 1954.

[13] ________ and Dudman, Jack, "The general solution of the latent class case", The Use of Mathematical Models in the Measurement of Attributes, Paul F. Lazarsfeld, ed. Rand Research Memorandum, 1951, No. 455.

[14] ________ and Henry N. W., Latent Structure Analysis, Houghton Mifflin Company, 1968.

[15] Madansky, A., "Identification and estimation in latent class analysis", Unpublished Ph.D. dissertation, University of Chicago, 1958.

[16] ______, "Determinantal Methods in latent class analysis" Psychometrika, 25, (1960), 183-98.

[17] McHugh, R. B., "On the scaling of psychological data by latent structure analysis", Unpublished doctoral dissertation, University of Minnesota, 1954.

[18] McHugh, R. B., "Efficient estimation and local identification in latent class analysis", Psychometrika, 21, (1956), 331-47; correction, 23, (1958), 273-74.

AN INTERVIEW SURVEY OF STATE FAIR GOERS

C.H. Proctor, North Carolina State University

1. BACKGROUND

As a class exercise in sample survey design and execution, a group of students counted and interviewed a portion of the exiting fair goers during North Carolina's 1969 State Fair. Aside from this intrustion of the classroom into the society, the survey can be counted as a realistic attempt to sample and measure people's reactions to recreation experiences, so that describing it might be of some interest.

The fair management had been looking at reports of other fair surveys and thinking about their own needs so that, when requested, they drafted a questionnaire form that incorporated some of their special interests as well as the more conventional material. This is a most painless, as well as concise, way to have the client express the purposes of the survey. Some marginal adaptations of the form were made so that it could both be self-administered as well as presented by the interviewer, depending on the circumstances.

Another early decision was similarily disposed of with ease, and that was the sample size or, in this case, the size of the budget for the survey. There were 18 people in the class who could each be asked to do one hour of interviewing at two different times during the 9 days of the fair. The instructor (the writer, that is) felt this to be a reasonably just assignment in that it provided actural enumerating experience, without undue exploitation.

2. FRAME CONSTRUCTION

The issue of most concern in early planning of the sample design was the choice of the sampling unit and enumeration procedure. Interviewing on the fairgrounds was ruled out because of the possible biasing influence of traffic patterns, and thus it was decided to interview people as they left the fair. Just outside the fairground's gates one faces an outgoing stream of persons grouped into parties. The stream's density rises and falls from time to time and varies from gate to gate. The fair's management gave their judgements of times of the day when the stream increases. These were the afternoon from 4 to 6 and the evening from 9 to 11, while the times 1 to 4, 6 to 9 and 11 to 12 were guessed to have less traffic. Also gates numbered 1, 9, 10 and 11 were judged to have heavier traffic than gates numbered 4, 8, and 12.

Two of us visited each of the gates, as workmen were busy putting the grounds in order, a week before the opening. Being caught up in the enthusiasm and optimism of the occasion, we decided that the heavily used gates could be split and the busy hours chopped and interviewers would be able, nay, they would be required, to interview <u>everybody</u> exiting.

Under this policy the sampling unit became a day-place-time unit, a DPT. At each of the hours 1 to 2, 2 to 3, 3 to 4, 6 to 7, etc. there was listed one DPT for each of the gates 4, 8, and 12 and two DPT's for each of the split-halves of gates 1, 9, 10 and 11. The A part of gate 1 was to be the lefthand two exits and the B-part was defined as the right-hand ones, and similarly for the other gates to be split. During the busy times, each hour was chopped into an early and a late portion. The early portion contained the first, the third and the fifth ten-minute period and the late portion was the remainder. To enumerate a selected DPT required interviewing <u>all</u> persons in parties whose first-exiting member crosses the exit line within the time assigned and leaving from the portion of the gate assigned.

3. SAMPLE DESIGN

The frame in its final form consisted of 1,452 DPT's divided into an afternoon (before 9) stratum of 957 DPT's and an evening stratum of 495 DPT's. The class was divided into 3 teams, each of 6 students [1]. Each team selected their DPT's using their own random number tables until the sample contained 4 afternoon and 2 evening DPT's. These selections were controlled in that all were on different days. Thus the original design was proportional stratified with controlled selection in 3 replicated subsamples. A very neat and promising design, or so we thought.

The fair began on Friday and field work done on that first weekend was a study in confusion. Too many people were exiting ever to begin to interview them all with the limited manpower of 2 persons at a gate. In addition, the questionnaire, even when self-administered, required that the interviewer work with only one party at a time. Consequently, when the class met on Monday, we had to redefine the measurement operation at a DPT. Then the data already collected had to be converted to the new basis. This adjustment was made rather than just throwing away the data on the first 3 days and redefining the population to consist of the last 6 days.

Assignment was made to a whole hour and to the whole gate (the former A and B divisions of the gate did not correspond to anywhere near equal volumes of traffic). The frame was left unchanged and thus the sample selection was not affected. Only the measurement operation was changed. Two enumerators were assigned to each DPT. They randomly selected a number from 1 to 10 to use as a start minute. After the start time they both counted persons exiting for 2 minutes (their watches were synchronized); then they interviewed the next two parties (one party for each enumerator) exiting after the count period. If a party or a person refused to be interviewed the enumerator marked questionnaires for such persons with an "R" and interviewed the next party he could contact. The enumerators resumed counting after 8 minutes of interviewing and repeated this pattern of count-2-minutes-interview-8-minutes 6 times. Occasionally one enumerator would not finish his interviewing in time to resume counting but his partner was always there. The scheme was workable under the heaviest of traffic conditions and became worrisome only with sparse traffic. Then one could revert to interviewing every party [2].

There may have been some "edge effects" introduced because the counting always preceeded the interviewing of a DPT but they were minor indeed. The arbitrariness in phase of second hands of the enumerator's watches would seem to make the probability of a given party being interviewed by either the first or second enumerator roughly proportional to the time gap from the leader of two parties ahead to its leading member. It is reasonable to judge that the results were not biased in any practical degree by this introduction of slightly unequal selection probabilities.

4. IMPROVING SAMPLE DESIGN

The form of an estimated proportion was essentially:

$$(1) \qquad \hat{P} = \sum_{t=1}^{n} w_t \bar{y}_t$$

where w_t is a measure of traffic through the DPT based on the counting and $\bar{y}_t$ is the average proportion of the variable of interest based on interviewing. After accounting for the replication into 3 surveys, the stratification, and the controlled selection of the actual design, the expression for the variance of $\hat{P}$ was reduced to:

$$(2) \qquad V(\hat{P}) \doteq \sum_{t=1}^{n}[(\frac{1}{n})^2 V(\bar{y}_t) + P^2 V(w_t)].$$

The quantity $V(\bar{y}_t)$ becomes $\sigma_p^2 + \sigma_a^2/E + \sigma_g^2/EP + \sigma_e^2/EPI$ where E is the number of enumerators at a DPT, P is the number of parties interviewed by each enumerator at a DPT and I is the number of individuals interviewed in each party, while σ_p^2 is the DPT-to-DPT variance in the proportion and the other variance components correspond to their divisors. Because we did not re-interview people it was not possible to separate measurement error variance from true score variance and the definitions of the σ^2 are for infinite populations. Even more, we treated the 0,1 data as numerical for ANOVA computations, and just how unrealistic this is is not known. We formulated a cost function for interviewing as:

Cost in man-minutes = 10n + 100nE + 6nEP + 2nEPI,

and were thus able to find an optimal allocation of effort. Only one person per party should be interviewed, more interviews should be made (lengthen the DPT in time?) but only one enumerator should be sent (we had sent 2).

Another derivation led to the expression for the variance of the counts as:

$$V(w_t) = (236n)^{-2}(30/r)^2 \times$$
$$\times [S_x^2(sr/30)^2 + rS_e^2(1 - r/30) + r\sigma_d^2/s],$$

where r is the number of 2-minute counting periods in a DPT and s is the number of enumerators making the counts. Here we did duplicate measurements so the measurement error can be separated. The component of measurement error variance σ_d^2 carries no fpc, while S_C^2, the sampling component does. One enumerator is found to be the optimal design in this case also.

When these recommendations are incorporated into an improved design it would appear that we could have reduced the variance by 28% for the same cost. Because of the very small sample sizes (3 teams) and the consequent uncertainty in estimating the variance components, one cannot take these results too seriously.

5. CONCLUSIONS

The usefulness of the study would seem to lie in:

1) Giving students experience in sample surveys.
2) Developing a workable sampling unit and measuring operation.
3) Verifying the usefulness of replication for detecting the importance of response error components of variance and to allow standard error calculations.
4) Illustrating the kinds of arguments needed for developing approximating formulas for complex designs.
5) Illustrating the kinds of computations that could be made with more data (and realistic data) to estimate the variance components.
6) Showing how a more elaborate design would be needed to estimate all variance components of interest [3].

[1] In each team one person served as chairman or communications hub, another selected the sample, one made field work assignments, another tabulated questionnaire responses, one calculated estimates and the last wrote the reports. This division of labor, as well as that into teams, was randomly made, of course.

[2] A strikingly similar survey situation and measurement operation is that described in P.V. Sukhatme, V.G. Panse and K.V.R. Shastry (1958), "Sampling techniques for estimating the catch of sea fish in India," Biometrics, 14, 78-96.

[3] These last four points were considered in detail in the original version of this paper, available from Dept. of Statistics, NCSU, Raleigh, N.C.

THE RELATION OF CHILD MORTALITY TO FERTILITY IN TAIWAN*

Shea Rutstein, The University of Michigan

I. Introduction

It has been postulated by several social scientists that high levels of infant and child mortality and the fear of infant and child mortality induce couples in less developed countries to have more children than they might otherwise prefer.[1, 2] In fact it has been implied that child mortality not only induces replacement fertility but also additional fertility as a form of insurance.[3] There are two ways in which child mortality may influence fertility: a) that resulting from community experience of child mortality and b) that resulting from personal experience with child mortality. Community experience may have a bearing on fertility behavior in two ways: first, through the effect on the couple's fear of child mortality, which in turn may influence its decision as to the number of births necessary to obtain an ideal number of children; and second, through its effect on community norms about the ideal number of children.[4] Personal experience with child mortality should raise fertility directly through the desire to replace the lost child and indirectly through increasing the fear of child mortality.

The data used to investigate the relation between child mortality and fertility consists of two surveys conducted by the Taiwan Provincial Institute of Family Planning. Taiwan is particularly suited to the study of demographic problems as it is in the midst of an economic and demographic transition. During the last two decades national income per capita has risen at an average annual rate of 4.6 percent in constant Taiwanese dollars. The growth of real per capita national income accelerated during the last eight years, when it rose at a rate of 6.26 percent per year to a level of US$258.[5] The general fertility rate has fallen from 211 per thousand in 1951[6] to 124 per thousand in 1969[7] and the infant mortality rate from 125.2 per thousand in 1942[8] to 38.7 per thousand in 1968.[9]

The first survey interviewed a cross-section of currently married women aged 20 to 44 during the period October 1967 to February 1968. They were asked about their demographic status, attitudes, and social status. Approximately twenty months later, the husbands of these women were interviewed, and information was obtained as to the current economic status, the husband's attitudes, and changes in demographic status. Overall 18.9 percent of the 2,277 couples included had experienced at least one child death.

The data are analyzed by the method of multiple classification analysis (MCA), which is a form of dummy variable multiple regression. The advantage of MCA is that the independent variables need not be continuous but may consist of sets of subclasses which represent only nominally scored variables (e.g., occupation). MCA makes no assumption as to the form of the relationship between the dependent and the independent variables and hence allows it to be determined by the data. Two sets of coefficients are produced: (1) unadjusted means for each subclass of the independent variables and (2) adjusted means controlled for the effects of the other independent variables.[10]

II. Characteristics of Couples Experiencing Child Mortality

Some correlates of child mortality are shown in Tables 1 and 2. In general Table 1 shows that couples who had experienced child mortality were older and had been married much longer than couples who had not experienced child mortality. There are two explanations for the differences we find: couples who have been married longer have more and older children and on that account have been more exposed to the possibility of experiencing child mortality; they were also married at a time when child mortality rates were substantially higher than they were at the time of the interviews.

Table 2 reveals a strong relation between child mortality and socio-economic status. Child mortality declines with increasing education, while husband's occupation shows similar strong differences in child mortality experience. Couples classified by husband's ancestry and farm background also show substantial differences in child mortality experience. Some other characteristics are presented in column two of Table 3. Total family income per adult shows a strong inverse relationship to experience with child mortality. Quality of housing may have an independent effect on child mortality since poor construction (e.g., mud brick walls, dirt floors, absence of indoor plumbing, etc.) could lead to a higher incidence of disease. An index of quality of housing was therefore constructed which takes material used for walls and the floor as symptomatic of housing quality. This index is highly correlated with income and in fact may be a better indication of permanent income available to the couple than the current income measure. The index shows that couples with housing in the lowest category had two and a half times the incidence of child mortality of couples in the highest category.

While Tables 1 and 2 and column two of Table 3 are of some descriptive interest, they do not isolate the effect that economic status has on child mortality. Since child mortality is in part due to the joint effects of longer exposure and low socio-economic status, MCA was used to determine the separate effects of exposure and status on a family's experience with child mortality. The dependent variable in this analysis was the deviation from the expected probability of having had at least one child death, based on the number of children a couple has had and their birthdates. The formula used to calculate the deviation for each couple is

$$\delta = \nabla - (1 - \prod_{i=1}^{i=N} \sigma_i)$$

where δ is the deviation from the expected probability of having had at least one child death; ∇ is equal to one if one or more child deaths occurred and zero if none occurred; N is the number of births the couple has had; and σ_i is the probability of survival to the date of interview for

Table 1--Means of Characteristics Related to Exposure to Child Mortality in Taiwan

Item	At least one child died	No child died
Husband's age, yrs.	39	36
Wife's age	35	32
Wife's age at marriage	19.4	20.3
Marriage duration, mos.	159	114
Number of births	5.2	3.6
Number of cases	430	1847

Source: The source for all tables are two sample surveys conducted in Taiwan in 1967-68 and 1969.

child i. The probabilities of survival were derived internally from the sample data. First the probability of a child dying during a given age interval[11] was determined by dividing the number of deaths occurring during that age interval by the number of people at risk at the beginning of the interval, and then subtracting from one to get the probability of survival to the beginning of the next interval. The product of these probabilities, period by period up to a child's potential current age (interview date), yields σ_i.[12]

Table 3 presents the results of an MCA, using these deviations as the dependent variable. The third column shows the results after removing the effects of exposure by use of the deviations calculated as described in the previous paragraph. This procedure was performed directly on the dependent variable. To remove the effects of other variables, MCA was used, and the combined results are shown in column 4. Exposure accounts for about 19 percent of the variance, and the other variables account for an additional two percent.

Housing quality, adjusted for exposure to child mortality, shows a pronounced relation to experience with child mortality. Adjusting in addition for the date of the first birth, income per adult, and wife's education somewhat weakens this relation (the progression becomes less regular). After exposure to child mortality, the housing index provides the largest increment to the proportion of variance explained.

The date of the first birth, which represents the point in time when the couple's exposure to child mortality began, also shows a strong relationship to child mortality. The means in column 2 show the effects of two sets of factors: 1) a greater amount of exposure through higher numbers of children and the longer time exposure and 2) the general decline in child mortality over the thirty-year period that is represented. Removing the effects of exposure directly eliminates the first factor and thus reduces the difference between the earliest and latest cohort to a ratio of only 1.5 to 1 (column 3). Holding constant socio-economic factors does not produce much further change in the relationship. One reason that the percent of couples with child mortality rises for the latest cohort after adjusting for exposure may be that they do not suffer from as much recall error as other couples in the sample.

Controlling for exposure and holding constant, in addition, the other socio-economic variables, both total income/adult and wife's education show somewhat weakened relationships to

Table 2--Socio-Economic Status of Couples with Child Mortality Experience

Item	Relative frequency of experience	No. of cases
ALL	18.9	2277
HIGHEST EDUCATIONAL LEVEL ATTAINED		
HUSBAND:		
None	27.1	240
Primary	20.3	1376
Junior h.s.	16.0	269
Senior h.s. grad.	10.6	254
College or univ. grad.	7.2	83
WIFE		
None	25.1	817
Primary attended	19.5	210
Primary completed	16.6	940
Junior h.s.	6.7	164
Senior h.s. grad.	6.4	109
HUSBAND'S OCCUPATION		
Prof., tech., manager	7.2	181
Clerk, office worker	14.2	155
Protective services	12.1	91
Small prop., sales w.	16.1	360
Skilled worker	16.3	276
Personal services	20.0	145
Unskilled worker	23.5	255
Farming, fishing, etc.	23.3	772
EVER LIVED ON FARM		
Never	12.0	357
Wife only	12.6	183
Husband only	14.8	81
Both	21.1	1627
HUSBAND'S ANCESTRY		
Hakka	22.6	297
Fukienese	18.8	1668
Mainlander	12.2	237
Other	25.0	48

differences in child mortality experience. For both adjusted variables, the lowest category has 1.5 times the experience with child mortality of the highest.

Next we may look at the antecedants of fear of child mortality. We would expect to find that people concerned about child mortality have low income, are poorly educated and badly informed, live in areas with high child mortality rates and in poor housing, and perhaps have experienced the death of one of their children.

As an indication of attitudes toward child mortality, the Child Mortality Fear Index (CMFI) was constructed from the following three survey questions asked of the husband: a) "Generally speaking, in the past children often died, and therefore it was a great advantage to have at least three or four sons. Do you think this is equally true today?", b) "Most people feel that a couple with 5 or more children have a large family. In your view, what are the main advantages of having such a large family?", and c) "Are there any important disadvantages to having only 2 children?". If the husband answered "equally true" to a), he scored a one; if he answered "less true," he scored a zero. He also scored one on b) and c) if he mentioned child mortality spontaneously. The scores from the three questions were summed to

Table 3--Results of MCA on Deviations from Probability of Child Mortality Experience Expressed as Percent of Couples with Experience, by Date of First Birth(B), Total Family Income/Adult(I), Wife's Education(E), and Housing Quality Index(H)

Mean Percent with Experience: 18.9					
	No. cases	Un-adj. mean	Mean adj. for exposure to child mortality[a] only	and--	
BIRTH DATE				I,E	I,E,H
1940-50	215	37.7	25.6	24.7	24.9
1951-55	501	28.9	22.7	22.4	22.4
1956-60	626	20.4	18.8	18.5	18.3
1961-65	635	9.4	14.5	14.9	14.9
1966-69	277	5.8	17.1	18.2	18.4
INCOME				B,E	
<NT$6,000	798	22.6	21.1	19.6	
6-9,000	425	22.9	20.0	19.2	
9-12,000	317	20.3	21.0	21.4	
12-16,000	286	15.1	16.7	18.4	
16-20,000	170	13.5	15.6	16.9	
>20,000	207	6.8	10.5	13.5	
WIFE'S EDUC.				B,I	
None	820	25.5	23.1	22.1	
<Primary	212	20.2	19.7	19.7	
Primary	940	17.0	16.8	17.0	
>Primary	274	7.0	12.5	15.0	
HOUSING INDEX					B,I,E
Lowest	328	26.7	25.3		24.4
	348	25.3	24.9		24.1
	500	17.8	17.1		17.7
	676	17.1	16.7		17.8
	234	13.2	15.5		17.8
Highest	168	10.7	12.8		15.4

a-Adj. for exposure to child mortality directly upon dpendent variable, adj. for other variables by MCA.

arrive at the husband's score on the index. In all, 66.3 percent of the respondents scored zero on the index (i.e., they expressed no fear of child mortality at all), 26.2 scored one, 2.0 scored two, and 5.5 percent were excluded for uncertainity. The mean score was .32.

For the most part, the unadjusted means in Table 4 bear out our expectations. Column 2 shows us that the fear of child mortality rises substantially with the number of child deaths and declines with increases in income per adult, wife's education, exposure to media, and housing quality. The local area post-neo-natal mortality rate, averaged over the years 1965, 1966, and 1967, serves as the measure of the prevalence of child mortality in that local area.[13]

Exposure to mass media displays a strong relationship to fear of child mortality even when number of child deaths, date of first birth, income per adult, and wife's education are controlled. It had the strongest relationship to fear of child mortality of all the variables. This finding indicates that information is able to dispel fear of child mortality by either changing perceptions about the general child mortality conditions in Taiwan or indicating greater accessibility to health services that allay such fear. Total family income per adult and wife's education, adjusted for number of child deaths and date of first birth, still show inverse relations to CMFI. Additionally controlling for mass media exposure, and wife's education for income per

Table 4--Results of MCA on Child Mortality Fear Index(CMFI), by No. of Child Deaths(D), Date of First Birth(B), Total Family Income/Adult(I), Wife's Education(E), Mass Media Exposure Index(M), Housing Quality Index(H), and Area Post-Neo-Natal Mortality Rate(P)

Mean Index Score for CMFI: .32					
	No. cases	Unadj. score	Mean score adjusted for		R^2 at entry
NO. OF DEATHS			B	B,I,E,M	.006*
0	1750	.30	.31	.31	
1	320	.36	.34	.33	
2	59	.44	.41	.36	
3+	26	.54	.50	.47	
BIRTH DATE			D	D,I,E,M	.008
1940-50	206	.41	.39	.39	
1951-55	481	.37	.36	.35	
1956-60	583	.32	.31	.31	
1961-65	601	.27	.27	.28	
1966-69	284	.27	.28	.30	
INCOME			B,D	B,D,E,M	.016
<NT$6,000	754	.35	.35	.30	
6-9,000	412	.36	.35	.34	
9-12,000	304	.31	.31	.33	
12-16,000	270	.32	.33	.37	
16-20,000	167	.29	.29	.35	
>20,000	203	.15	.17	.24	
WIFE'S EDUC.			B,D,I	B,D,I,M	.023
None	768	.40	.38	.35	
<Primary	206	.30	.29	.29	
Primary	910	.29	.29	.30	
>Primary	264	.19	.24	.28	
MEDIA INDEX			B,D,I,E		.032
Lowest	382	.45	.43		
	313	.35	.34		
	312	.38	.36		
	261	.30	.29		
	500	.25	.25		
Highest	284	.21	.24		
HOUSING INDEX			B,D,I,E,M		.031
Lowest	306	.38	.34		
	338	.33	.30		
	465	.35	.34		
	660	.30	.30		
	222	.26	.31		
Highest	164	.23	.31		
AREA PNN MR			B,D,I,E		.039
4.0-7.9(/000)	197	.34	.35		
8.0-10.9	486	.30	.32		
11.0-13.9	936	.29	.29		
14.0-16.9	135	.47	.43		
17.0-19.9	277	.27	.25		
20.0+	124	.45	.44		

*ETA-squared

adult, weakens the relationships so that only the highest categories differ from the rest. After considering the other variables, housing quality appears to have even less impact. A priori, the post-neo-natal mortality rate might be expected to display a strong association with the CMFI, but no clear relationship is shown for either unadjusted or adjusted mean scores. It is possible that in as small a country as Taiwan information about national conditions, as communicated by the mass media, is more relevant for attitude formation than perceptions about local conditions. Or it is possible that we have not looked at the most appropriate local mortality rate since the recency of the rate may not sufficiently allow for the

Table 5--Unadjusted Parity Progression Ratios by Experience with Child Mortality

Mortality experience	Parity level reached 0	1	2	3	4	5	6	7
NO EXP.								
Number:								
Eligible	2277	2105	1742	1329	884	446	210	79
Prog.	2258	2036	1574	1064	573	241	110	33
Ratio	99.2	96.7	90.4	80.1	64.8	54.0	52.4	41.8
SOME EXP.								
Number:								
Eligible	0	91	200	264	279	240	167	88
Prog.	0	90	198	249	224	175	89	36
Ratio	--	98.9	99.0	94.3	80.3	72.9	53.3	40.9
ALL								
Number:								
Eligible	2277	2196	1942	1593	1163	686	377	167
Prog.	2258	2126	1772	1313	797	416	199	69
Ratio	99.2	96.8	91.2	82.4	68.5	60.6	52.8	41.3

gestational period necessary in formation of attitudes toward child mortality. These questions will be examined in later work. The proportion of variance explained by all the variables jointly is quite low although some of the relationships reveal a regular and meaningful progression of the means. We must keep in mind that our CMFI may be a weak measure of the actual fear of the death of a child. This is a first attempt at measuring this attitude, and the survey questions used may not be optimal in fully eliciting the fear that may exist. Conceivably, couples who act on a fear of child mortality may not want to acknowledge this fear even to themselves.

III. Fertility and Child Mortality

We are now ready to examine the major question with which this paper is concerned: What are the effects of child mortality and fear of child mortality on fertility? To take a first look at this, we use parity progression ratios to control for the high correlation between number of children and exposure to child mortality. For our measure of mortality we take the number of child deaths occurring more than nine months prior to the date of the next birth or the total number of deaths that had occurred by the date of the first interview if the next birth did not occur. Table 5 shows the rates of progression to the next parity by the couple's experience with child mortality as just described. All couples who had arrived at a given parity level prior to the first interview were eligible to progress to the next level. They were considered to have progressed if they reached or exceeded the next parity level by the second interview. Table 5 shows that couples without mortality experience have lower progression ratios than couples with a child death at each parity level until the sixth level is reached. However, we learned earlier that couples without child mortality experience are of higher economic status than others. Conceivably, this difference could account for their lower progression ratios. Therefore we need to use MCA to see what happens to the parity progression ratios in Table 5 when they are adjusted for the normal likelihood of couples having another birth for reasons other than experiencing a child death.[14]

The first factor to be adjusted for in

Table 6--Results of MCA on Parity Progression Ratios by Number of Child Deaths Experienced, Adjusted for Date of Birth from which Couple is Eligible to Progress(T), CMFI Score(F), Total Family Income/Adult(I), Wife's Education(E), and Husband's Occupation(O).

Parity Progression	No. cases	Unadj. mean	Mean adjusted for T	T,F	T,F,I,E,O
2ND TO 3RD					
No. Deaths					
0	1741	90.4	90.9	90.9	91.0
1	183	98.9	94.2	94.4	93.8
2	17	*	*	*	*
Adj. R^2		.009[a]	.135	.136	.160
3RD TO 4TH					
No. Deaths					
0	1329	80.1	80.9	80.9	81.3
1	236	94.1	89.5	89.3	87.8
2	28	96.4	94.2	94.7	90.7
Adj. R^2		.019[a]	.125	.130	.195
4TH TO 5TH					
No. Deaths					
0	884	64.8	66.1	66.0	66.4
1	232	78.0	73.7	74.0	73.1
2	44	93.2	89.7	89.8	87.0
Adj. R^2		.024[a]	.119	.124	.154
5TH TO 6TH					
No. Deaths					
0	446	54.0	55.7	55.7	56.0
1	181	72.9	69.3	69.5	69.6
2	47	70.2	68.5	68.2	66.4
Adj. R^2		.035[a]	.174	.176	.216

*Base less than 25. a-ETA-squared.

Table 6 is the length of time following a given birth in which a couple could have had another birth. We use the date of the birth from which they are progressing to control for this exposure to subsequent fertility. Further, we may control for their fear of child mortality by holding constant the CMFI. The effects of socio-economic status are adjusted for by using income per adult, wife's education, and husband's occupation. Table 6 indicates that after adjusting for these characteristics, parity progression ratios for couples with mortality experience are still higher than those for other couples at each parity level but that the differences have been substantially reduced. We can see that the largest reduction in the progression differentials between couples with and without mortality occurs when exposure to subsequent fertility is taken into account. Controlling for the socio-economic variables makes for some small further reduction, white controlling for fear of child mortality affects the differentials very little.

Table 7 enables us to examine the impact of fear of child mortality on fertility. The unadjusted ratios show substantial differences in parity progression for different levels of the fear index except the last. Controlling for the number of child deaths experienced and the length of time exposed to subsequent fertility mainly reduces the parity progression ratios for couples with a score of two on the CMFI. Taking account of socio-economic status reduces the differences between the ratios by a small additional amount.

We can see in both Tables 6 and 7 that the differentials in the ratios rise with parity level.

Table 7--Results of MCA on Parity Progression Ratios by Fear of Child Mortality, Adjusted for Number of Chil Deaths(D) Date of Birth from which Couple is Eligible to Progress(T), Total Family Income/Adult(I), Wife's Education(E), and Husband's Occupation(O)

Parity progression	No. cases	Unadj. mean	Mean adjusted for D,T	D,T,I,E,O
2ND TO 3RD				
CMFI				
0	1281	90.1	90.7	90.9
1	511	93.5	92.6	92.0
2	42	97.6	93.7	93.1
Adj. R^2		.004[a]	.135	.160
3RD TO 4TH				
CMFI				
0	1028	80.4	80.9	81.9
1	435	87.4	86.3	84.3
2	39	89.7	86.7	84.3
Adj. R^2		.008[a]	.125	.195
4TH TO 5TH				
CMFI				
0	726	66.5	66.7	67.2
1	345	72.8	72.5	71.5
2	31	80.6	76.4	75.4
Adj. R^2		.006[a]	.119	.154
5TH TO 6TH				
CMFI				
0	409	60.1	60.3	61.3
1	221	60.2	60.2	58.9
2	23	*	*	*
Adj. R^2		.005[a]	.174	.216

*Base less than 25 a-ETA-squared

This may be due to the fact that at low parity levels most couples continue to have children irrespective of child mortality or their fear of child mortality, while at higher levels the number of living children has an important bearing on the decision to continue or not to continue. The rise in the differentials is caused by large drops in the ratios for couples with no child deaths and for those scoring zero on the CMFI as parity level increases, while couples with mortality experience and couples with more fear of child mortality remain at high ratios until relatively higher parity levels are reached.

Table 8 enables us to compare couples with equal numbers of living children. According to the adjusted ratios in Tables 6 and 8, while couples with child mortality experience are more likely to have another birth than couples without such experience, as a whole they do not make up for the lost child. The exception is progression from four living children. The data seem to indicate that couples with child mortality experience may end up with smaller numbers of living children than couples without such experience.

To check this indication, we used the parity progression ratios to predict the mean number of additional births occurring to couples with given numbers of living and dead children. This is done in Table 9 for those with zero and one child death. From the unadjusted ratios we might infer that the couples experiencing a child death would slightly more than replace that dead child, up to four living children. However, when we use the adjusted ratios, we see that couples with a child death end up with substantially

Table 8--Unadjusted and Adjusted[a] Parity Progression Ratios by Number of Child Deaths

Number of living children	Number of child deaths 0	1	2
ONE			
Unadjusted	96.7	98.9	96.4
Adjusted	96.8	93.8	90.7
TWO			
Unadjusted	90.4	94.1	93.2
Adjusted	91.0	87.8	87.0
THREE			
Unadjusted	80.1	78.0	70.2
Adjusted	81.3	73.1	66.4
FOUR			
Unadjusted	64.8	72.9	50.0
Adjusted	66.4	69.6	--
FIVE			
Unadjusted	56.0	58.9	*
Adjusted	--	--	--

[a]Adjusted for period of time exposed to subsequent fertility, fear of child mortality, total family income per adult, wife's education, and husband's occupation.

--Not available *Base less than 25

smaller numbers of additional births at every level of achieved family size. This result is consistent with earlier findings that a significant proportion of couples in Taiwan[15] have more children than they consider ideal.

Table 10 shows projections of mean number of additional births at each achieved parity for couples with differing scores on the Child Mortality Fear Index. Projections based on unadjusted and adjusted parity progression ratios show similar patterns. The higher the score on the CMFI, the greater is the number of additional births. Adjustment of the progression ratios for number of child deaths, length of time exposed to subsequent fertility, and socio-economic status substantially decreases the amount of excess fertility attributable to fear of child mortality but does not eliminate it altogether.

IV. Conclusion

To summarize our findings, we have shown that in Taiwan the poorer and less well educated and informed couples are most likely to experience child mortality and have fear of child mortality. Differences in child mortality by socio-economic status persist after differences in demographic exposure have been taken into account.

The hypotheses that high levels of child mortality and fear of child mortality lead to additional births is supported by the foregoing analysis. However, the idea that couples who experience a child death may have more children overall as a form of insurance against additional deaths is not substantiated by the data. In fact they do not even replace the lost child. Fear of child mortality seems to affect only about 30 percent of the couples in present day Taiwan, and these couples have only minimally higher fertility (after adjustments) than couples without such fear. Thus, we conclude that while child mortality has a small negative impact on population growth, fear of child mortality has a small positive impact. Together these two factors do not seem to have more than a negligible effect on the population growth rate in Taiwan today.

Table 9--Number of Additional Births by Number of Living Children for Couples with No and One Child Death, Predicted by Unadjusted and Adjusted[a] Parity Progression Ratios

Number of living children	Unadj. ratios by no. of deaths		Adj. ratio by no. of deaths	
	0	1	0	1
0	4.38	4.57	4.47	4.08
1	3.42	3.62	3.51	3.13
2	2.54	2.66	2.62	2.32
3	1.81	1.83	1.89	1.65
4	1.32	1.34	1.32	1.28
5	.94	.84	.98	.84
6	.74	.58	.74	.58

[a]Ratios were adj. ratios from Table 7 where available. Unadj. births shown for five and six living children since MCA could not be used due to small no. of cases and similarity of ratios.

NOTES

*This research is part of a larger project supported in part by the Population Council and through a fellowship from The University of Michigan Population Studies Center. The author is grateful to Eva Mueller, Ronald and Deborah Freedman, and Jeremiah Sullivan for valuable suggestions and comments. The author alone is responsible for any remaining shortcomings.

1. For example see R. Freedman and J. Y. Takeshita, Family Planning in Taiwan, Princeton: Princeton University Press, 1969; H. Frederiksen, "Feedbacks in Economic and Demographic Transition," Science, 166 (1969), 837-847; Taiwan: "An Area Analysis of the Effect of Acceptances on Fertility," Studies in Family Planning, 33 (1968), 7-11; and D. Heer, "Economic Development and Fertility," Demography, 3 (1966), 423-444.

2. Henceforth infant mortality and childhood mortality will be referred to as child mortality.

3. For example see T. Paul Schultz, "A Family Planning Hypothesis: Some Empirical Evidence from Puerto Rico," RAND Corp. Memorandum RM-5405-RC/AID, (1967); ----- and J. DaVanzo, "Analysis of Demographic Change in East Pakistan: A study of Retrospective Survey Data," RAND Corp., R-464-AID, (1970); and A. Harman, "Fertility and Economic Behavior of Families in the Philippines," RAND Corp., RM-6385-AID, (1970).

4. For example see R. Freedman, "Norms for Family Size in Underdeveloped Areas," Proceedings of the Royal Society, B, 159 (1963), 220-245.

5. Industry of Free China, 34 (July 1970), Taipei.

6. Demographic Reference: Taiwan, Republic of China, 2 (1965), Taipei.

7. Taiwan Demographic Fact Book, 1969, Taipei.

8. Statistical Bureau of Taiwan, Administrative Office, (Past) Fifty-One Year Statistical Abstract of Taiwan, (1946), Taipei.

9. J. M. Sullivan, "A Review of Taiwanese Infant and Child Mortality Statistics, 1961-1968," Taiwan Population Studies Working Paper No. 10, Ann Arbor: Population Studies Center, University of Michigan.

Table 10--Number of Additional Births by Achieved Parity Level for Couples with Various Fears of Child Mortality (CMFI), Predicted by Unadjusted and Adjusted[a] Parity Progression Ratios

Parity level achieved	Unadjusted ratios by CMFI score			Adjusted ratios by CMFI score		
	0	1	2	0	1	2
0	4.46	4.80	5.32	4.53	4.66	4.79
1	3.49	3.84	4.36	3.57	3.70	3.88
2	2.61	2.97	3.51	2.69	2.82	3.02
3	1.90	2.18	2.60	1.96	2.07	2.24
4	1.36	1.49	1.91	1.39	1.45	1.65
5	1.05	1.05	1.37	1.07	1.03	1.19
6	.75	.75	.75	.75	.75	.75

[a]Ratios were adjusted ratios from Table 7.

10. For a more detailed explanation of MCA, see F. Andrews, J. Morgan, and J. Sonquist, Multiple Classification Analysis, Ann Arbor: Institute for Social Research, University of Michigan, 1969.

11. For ages less than 2 years, one month was used as the age interval to accurately reflect the sharply decreasing mortality rate with age. For ages 2 to 15 years, a one-year interval was used.

12. The use of a single schedule of age-specific survival probabilities biases the deviations for couples with early dates of first birth by underestimating their expectation of mortality experience. The opposite is true for couples with recent dates of first birth. Since the date of the first birth is included in the MCA, it should remove the bias from the other explanatory variables.

13. The surveys covered 56 local administrative units chosen to be representative of the entire Chinese population: 23 rural townships (Hsiang), 16 urbanized townships (Chen), 4 small cities (Shih), and 13 precincts (Chu) of the five major cities. The additional data for these areas come from the Taiwan Demographic Fact Book for the relevant year.

14. MCA is not shown for the first-to-second progression since nearly everyone progressed, and it was not carried out on the sixty-to-seventh and seventh-to-eighth progressions since the numbers of cases involved were small and the unadjusted ratios very similar.

15. R. Freedman and J. Takeshita, op. cit.

"LONG- VERSUS SHORT-TERM IMPACTS OF MOBILITY ON INCOME AND EARNINGS"

Larry Schroeder and Martin David
Georgia State University and University of Wisconsin

The data source--An alphabetic cluster sample of surnames was used to extract tax records on an identical sample of individual taxpayers from the archives of the State of Wisconsin (1). The resulting sample is representative of taxpayers in any given year from 1947-1959 and includes observations on some persons for 1946-1960.[1] A time series of data on individuals is embedded in the sample so that as many as fifteen consecutive tax returns may be available for some individuals. Individuals may disappear from (or reappear in) the sample because of years of low income, residence outside the State of Wisconsin, marriage (and change of surname), or death. Some information on the residence prior to a move into the State is available, as is some information on the reasons why tax returns were not filed in the prior year. Lack of symmetric data on persons who leave the sample limits the utility of that information for studies of residential mobility.

In this study we restrict our attention to males who filed at least two tax returns in the period 1946-1960. A comparison of residence codes and occupation codes for any two years provides measures of mobility for persons who remain in the State of Wisconsin. None of these findings can be generalized to inter-state mobility; those moves whose origin or terminus lies outside Wisconsin are not observed. The observed mobility measures also suffer from the random variation in taxpayers' descriptions of their occupations; occupation could not be reliably coded in as much detail as is provided by the 2-digit Census occupational titles. Detailed 2-digit occupations are differentiated for professional workers, but semi-skilled and unskilled occupations could not be differentiated from information given by the taxpayer (9). A description of the occupational classification used to define mobility status and the frequency of those observed occupations is given in Table 1. The distribution of major (1-digit) occupations reported for the tax record sample conforms closely to Census distributions. Nonetheless, taxpayer idiosyncracies in reporting the same occupation may appear in these findings as spurious occupational changes.

Residential location was both more reliably reported and more precisely defined. The State uses information on the residence of taxpayers as the basis for income tax distribution to municipalities; and therefore took great care to validate the taxpayer's correct city and county of residence in the current and prior years (the year for which income data are reported). Inter-county moves can be accurately measured from observed data, while intra-county moves are measured by moves among the nine largest towns and cities within the county and moves between those towns and cities and the remainder of the county.

The biases of the measures of residential mobility can be seen in Table 2. Intra-county moves bear no resemblance to the CPS data for the U.S. Wisconsin intra-county moves are probably understated by a factor of four. This is of interest for itself, since it indicates that "very small" moves (within defined units such as major towns) constitute about three-fourths of all intra-county changes in residence. We conjecture that such moves are less likely to be associated with a change in employment and are therefore of less interest to a study of the labor market. Inter-county, intra-state mobility reported to the Current Population Survey for the entire U.S. (11) correlates roughly with the measure derived from the tax data.[2] (Given the somewhat atypical economic structure of Wisconsin it is not clear that a higher degree of correlation could be expected.)

Four different samples of tax records for male taxpayers were drawn from the basic sample for studies of occupational and residential mobility:

Sample A--at least two tax records in the years 1946-60
Sample B--at least two tax records in the years 1947-59
Sample C--at least three tax records in the years 1947-59
Sample D--at least four pairs of consecutive tax records in the years 1947-59, sampled at a rate of 50 percent of the basic sample.

As interstate mobility will cause individuals to be excluded from the subsamples, a bias exists in our observations to the extent that interstate and intrastate mobility are different facets of the same underlying economic and social processes. The data must be regarded as a truncated sample of movers. However, not all differences between the four samples are associated with interstate mobility. Entry into the labor force, exit from the labor force, and death are also reasons why we will fail to observe tax records for an individual over the entire period under observation. Table 3 gives a picture of the attrition in the sample that occurs when attention is focussed on individuals with long time series of tax records.

Principal hypotheses--We hypothesize that change in occupational and residential location are motivated primarily by economic incentives: both types of mobility cause a subsequent improvement in earnings or a reduction in unemployment (4, 5, 12). The null hypothesis is that mobility is motivated by personal idiosyncracies and tastes and does not cause significant economic improvement.

Evidence to support the hypothesis is crude, but consistent. Four types of findings are reported below:

1) immediate impacts of mobility on earnings. These findings correspond to recall data that can be extracted from sources such as the Current Population Survey.

2) two-year results of mobility on earnings. Our hypothesis is that the improvement in earnings expected from mobility is not likely to appear until the year following the reported change.

3) cumulative effects of occupational mobility on trends in relative income. Our hypothesis here was that mobility will lead to relatively greater trends in earnings than those observed for non-mobile per-

sons. This result will only hold if voluntary upward mobility outweighs involuntary downward movements.

4) indirect evidence that mobility is enhanced by favorable economic conditions.

Annual mobility and earnings--Table 4 shows the earnings reported by male taxpayers for the tax year that most closely corresponds to the interval between dates of reporting a change in occupation or residence. It is evident that all but one of the reported changes are associated with lower levels of earnings than non-mobile persons. Only movers within the county who did not change occupation report higher mean earnings in the year of the move than the non-movers. This finding suggests that a large component of intra-county moves represents upgrading of residence by persons who move to suburban areas. The contrast between the high earnings of immobile individuals and the relatively lower earnings of movers is deceptive. Any period of unemployment or hiatus in employment associated with a job change would lead us to suspect lower earnings for the mover. Conversely, earnings reported by persons who did not move last year include beneficial (or detrimental) effects of moves made by some individuals in prior years. Hence the interpersonal comparison does not offer evidence on the impact of mobility on earnings, *per se*.

Table 5 indicates the mobility rates underlying Table 4 and demonstrates that more occupation changers also change residence than would be expected on the basis of chance. This finding is at variance with the time series analysis of mobility rates that indicates a negative correlation. (See footnote 2.) Disaggregating the observed data into birth cohorts, displays both significantly higher rates of residential mobility for those who change occupations and an expected decline in the probability of mobility with the increasing age of the cohort. (See Table 6.)

Two-year impact of mobility on earnings--As we hypothesize that mobility will have a positive effect on earnings in years subsequent to the move, we have computed earnings changes for individuals. The difference calculated from the year t-1 to t reflects the change in earnings from the year prior to the reported move to the year of the move. The change in earnings from t to t+1 reflects experience in the year subsequent to the move. Table 7 contrasts earnings changes for persons who did not move between years t-1 and t with the changes that occurred for those who shifted occupation or changed their labor force status. In each birth cohort mean income improvement in the year of the occupational change was less than improvement for those who made no change. (The unknown group is an exception to that finding.) At the same time occupational mobility results in substantially larger average increases in earnings for all cohorts in the year subsequent to their move. The cumulative impact of the occupational move over two years is most favorable for the 1905-24 cohorts, who earned a substantial excess over their immobile cohorts. This positive impact on earnings registers on a group of individuals whose initial income position was worse than the position of immobile workers. (Cf. Table 4.)

Changes in county of residence show similar but slightly less systematic results. Earnings improvement for persons whose residence shifted across county lines was larger in the year following the move than for persons who did not cross county boundaries. The 1905-14 cohort experienced a net loss in earnings, but the cohort pattern of net gain over the two-year period otherwise parallels the pattern shown for occupational cohorts. The large negative effect of locational mobility for the oldest cohort can readily be explained by the association between retirement and locational change that is shown in Table 5.

Cumulative impact of mobility on earnings--The obvious extension of Table 7 is to tabulate changes in earnings for k years following the change in occupation or location. Such tabulations are not yet available and we adopt two alternative approaches. In this section inter-personal differences in mean earnings are presented according to the cumulative mobility experienced between the first and the last tax record filed. Thus Table 8 corresponds to Table 4, except that change in occupation is measured over a period of at least 5 years, and possibly as many as 12 years. (Sample D was used for the tabulation.) Those who did not change occupation during the period under observation are shown on the first line, while those who changed major occupation grouping over the period are shown on the second line. An individual's earnings in any given year might be included in any of the mobility categories of Table 4; in Table 8 all the earnings for an individual must be included in one of the four rows. There is another difference between the two tables. Table 8 shows the average of mean earnings for an individual; Table 4 shows the mean earnings for all tax records. Individuals with relatively short time series of tax records thus receive greater weight in Table 8 than in Table 4. On the other hand individuals who do not meet the criterion for Sample D, at least four pairs of tax records, are excluded altogether and receive no weight in the average at all.

Persons who change major occupation have lower average earnings than those with a stable affiliation. The difference is only on the order of four hundred dollars or nine percent of the average earnings of immobile workers. The underlying mobility rates on which these averages of individual earnings are based are shown in Table 9. (The cumulative long term probability of remaining in a given occupation is significantly higher than the probability forecast from the annual average rate of occupational change for the cohort. This tangential finding clearly implies a need for a stochastic model of movement that includes some inhibiting effect for prior mobility.) Just as with the earlier interpersonal comparisons in Table 4 we can not easily assess the meaning of this mean difference in earnings.

To gain further insight into the impact of mobility we must remedy several problems in the data presented: 1) the bias in choosing longer time series for an individual is not controlled in the estimates, 2) only changes in major occupation group are reflected in the mobility measure, and 3) the averages presented combine data from moves in the period 1947-59 without explicitly introducing the date of the move.

Relative income position and occupational mobility--To meet the problems inherent in the two- and three-year time series, we computed a simple trend on the relative income position of male taxpayers (3). Relative income position was determined as the ratio to mean income estimated for the cohort from CPS and Census data. The ratio is thus independent of the mean income observed for the cohort in the tax record sample. A fortiori relative income position will reflect the bias due to omission of individuals with few observed tax records. By estimating the income of the individual relative to his cohort, systematic influences of aging, price inflation, and general increases in productivity and interest rates are removed from the data. What remains is a variable that may be thought of as the heterogeneity of income experiences within birth cohorts.[3]

If change in occupation is associated with a once and for all, or a continuing increase in earnings greater than the average for the birth cohort, the trend estimated for those who change occupation will be higher than for those who do not. This finding is confirmed in Table 10, except for those individuals who reported three or more major occupation groups during the period for which their tax records are available.

Table 10 indicates that the most rapid growth of income occurred for individuals who made minor changes in occupation that did not cross major occupational groups as defined in Table 1. Individuals who changed their major occupation also reported more rapid improvements in relative income position than those with a unique occupation, but that growth was not sufficient to place them in a higher relative income position in 1959 (as indicated by comparing the intercepts for the two groups).

Some insight into the meaning of these associations between relative income trends and occupation change can be gleaned from further classification of the group in Table 10 by birth year cohorts (Table 11). Our a priori hypothesis would be that occupational mobility would be likely to produce the most favorable change in income early in a career. (Mobility between ages 55-64 is also likely to produce favorable impacts on income, given the manner in which the table is generated. Persons in this age group who attempt a change in occupation and do not succeed will retire from the labor force altogether and will no longer file tax returns. Hence such individuals are excluded from the tabulation.) Relative to those with no change in occupation, workers who made a limited move (change in detail or one change in major occupation) achieved a higher relative income position within the 1915-34 birth year cohorts. For the cohorts in the peak of their careers, 1895-1914, a change in detailed occupation was advantageous relative to those who made no move; a change in major occupation group did not enhance income by a sufficient amount to provide a relative income position higher than those with a stable occupation.

The evidence in Tables 10 and 11, in conjunction with Table 7, unequivocally shows that occupational mobility enhances income. The larger trend observed for those who change occupation is consistent with two interpretations: a) occupation changers have extremely high rates of income increase prior to and after the change in occupation combined with a drop in income position as the result of change; or b) occupation changers have a trend in earnings following occupation change that exceeds that of immobile persons either because of a once and for all jump or because of subsequent improvements. Table 7 documents the increase in income position associated with occupation change so that the higher trend rates of growth for occupation changers must be taken as evidence of permanent income increases, interpretation (6).

Evidence on macroeconomic conditions and mobility--If improvement in earnings motivates occupational and inter-county mobility one would expect systematic relationships to vacancy rates and the duration of unemployment. Given that a worker is employed at time t-1, occupational mobility requires him to search, and possibly to quit his current position (7, 13). "Daring" search involving a quit without a definite job offer will be inhibited by high unemployment rates. Continued high unemployment rates in turn will be reflected in an increase in the average duration of unemployment. In fact that relationship is not strong:[4]

$$Z_1 = \underset{(-17.6)}{-2.49} - \underset{(-1.55)}{.0196}\, D \qquad \bar{R}^2 = .19 \qquad \sigma_\epsilon = .09$$

where D is the average duration of unemployment for the year (U.S.). t-ratios are shown in parentheses.

Conflicting hypotheses can be raised for the dependence of residential mobility on unemployment. If the local labor market suffers severe unemployment, migration elsewhere may be the only possible alternative for a worker who seeks to improve skill level, hours worked or both. The same slack in the labor market can inhibit moves by raising the cost of selling a home and coloring the workers' views about opportunities available elsewhere. For the inter-county moves the former effect predominates

$$Z_3 = \underset{(-33.5)}{-4.20} + \underset{(3.84)}{.0431}\, D \qquad R^2 = .60 \qquad \sigma_\epsilon = .08$$

A second approach to this problem is to disaggregate mobility flow and study the dependence of specific types of moves on macroeconomic variables. We study the relationship

$$P_t = (\underset{(10X10)}{a_1 \Delta E_t} + \underset{(10X1)(1X10)}{a_2 O_t \iota'} + \underset{(10X1)(1X10)}{a_3 U_t \iota'} + \underset{(10X10)}{a_4 \Delta T} + \underset{(10X1)(1X1)}{a_5 \iota \iota'}) + \underset{(10X10)}{\epsilon_t}$$

The variables are defined as follows:

$P_t = \| P_{ijt} \|$ the matrix of probabilities that workers will move from the j^{th} to the i^{th} market from t-1 to t; i, j defined by

	Milwaukee SMSA	Other
Professional, managerial, sales, clerical	1	6
Service, skilled	2	7
Semi- and unskilled	3	8
Self-employed	4	9
Farmers	5	10

$\Delta E_t = \| \bar{E}_{i,t} - \bar{E}_{j,t} \|$ where $\bar{E}_{i,t}$ is the mean earnings in the i^{th} market during period t.

$O'_t = (O_{1t}, O_{2t}, \ldots, O_{10t})$ where O_{it} is the number of job openings listed with the employment service at time t in the i^{th} market.

$U_t = (U_{1t}, \ldots, U_{10t})$ where U_{it} is the unemployment rate in the geographic market of which i is a part.

$\Delta T = \| T_i - T_j \|$ where T_i are the estimated training requirements for the i^{th} job (10).

ι is a 10X1 vector of "ones"

Because of the definitional constraints on the Markov matrix P_t, we can restrict our attention to the off-diagonal elements. Furthermore as the a_k are constants, the information for all off-diagonal estimates can be pooled for estimation.

The results of six pooled regressions for the period 1951-59 are shown in Table 12.[5] As indicated by the R^2 statistic the independent variables listed above explain very little of the variation in P_t. However, the regressions do indicate findings of interest. Only in Model 2 are the coefficients quite insignificant and a sign (on $\Delta\bar{E}_{t-1}$) not as expected. The imprecision there is undoubtedly due to the collinearity between the explanatory variables. (A correlation of .98).

Model 3 shows unemployment retarding mobility. The crude proxy for vacancies (O_{it}) indicates the attractive force of this variable on mobility. Models 5 and 6 suggest that differentials in training requirements retard mobility with Model 6 indicating that the "higher on the training scale" one is, the less likely he is to change occupations.

Of most interest to this study is the attractive force of earnings differentials on mobility. The coefficients are admittedly small but positive and significantly different from zero. This is surprising since mobility between markets is most often observed in both directions,[6] yet the independent variable, ΔE_t, is a summary measure of simple differences in mean earnings between the markets of destination and origin. Thus gains in earnings are seen as a factor in the mobility process in a macro sense as well as on the micro level.

FOOTNOTES

1. We are in the process of bringing the sample up to date to include the years 1960-64 on a representative basis. A detailed discussion of the data appears in (8) and (9).

2. The mobility rates were transformed to $Z_i \triangle \log_{10}(x_i/(1-x_i))$ where

Tax Record Data	CPS, U.S.	
Z_1	--	logit of the occupational mobility rate
Z_2	Z_4	logit of the intra-county mobility rate
Z_3	Z_5	logit of the inter-county, intra-state mobility rate
	Z_6	logit of the interstate mobility rate

The resulting correlation matrix is

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
1.00	.65	-.18	.24	-.15	-.15
	1.00	-.26	<u>-.19</u>	-.45	-.24
		1.00	<u>-.17</u>	<u>.42</u>	-.04

The underlinings indicate correlations between similarly defined mobility rates.

3. Note that income, not earnings, is used to obtain the relative position within the cohort.

4. Z_i are defined in footnote 2. The entries in parentheses are t values.

5. The years 1947-50 were excluded since no data on openings were available for that period.

6. See, for example (12).

REFERENCES

1. Bauman, R., M. H. David and R. F. Miller, "The Wisconsin Assets and Incomes Studies Archive" Social Science Information (Dec. 1967) 49-70.*

2. David, M., "Time-Series Versus Cross-Section Lifetime Earnings Patterns in Different Occupational Groups", Proceedings of the American Statistical Association (1969).*

3. David, M. and R. F. Miller, "A Naive History of Individual Incomes in Wisconsin, 1947-1959", J. of Income and Wealth, March 1970.*

4. Gallaway, L., Interindustry Labor Mobility in the United States, 1957-1960, Social Security Administration. Washington: GPO, 1967.

5. _____, Geographic Labor Mobility in the United States, 1957-1960, Social Security Administration. Washington: GPO, 1969.

6. Lansing, J. B. and E. Mueller, "The Geographic Mobility of Labor", Ann Arbor: Institute for Social Research, 1967.

7. Matilla, J., "Quit Behavior in the Labor Market", Proceedings of the American Statistical Association, 1969.

8. Moyer, M. E., The Validity of Income Distributions from a Multi-Year Sample of Wisconsin Income Tax Returns, 1965.*

9. _____, The Processing of Data from a Multi-Year Sample of Wisconsin Tax Returns, 1966.*

10. Scoville, J. G., "Education and Training Requirements for Occupations", R. Econ. Stat., 48 (Nov., 1966).

11. Shryock, H. S., Jr., Population Mobility Within the United States, Chicago: Community and Family Study Center, 1964.

12. Sjaastad, L., "Costs and Returns of Human Migration", JPE, 70 (Oct. 1962).

13. Toikka, R., "Supply Responses of the Unemployed". (Unpublished Ph.D. Dissertation, Univ. of Wisconsin 1971).

14. U.S. Census, Current Population Survey Reports, Series P-20, #104.

*Reprints available from the Social Systems Research Institute, Madison, Wisconsin 53706.

(This research was undertaken under grants from NSF and WARF. We appreciate critical comments by R. F. Miller.)

TABLE 1--Distribution of Major Occupations in Wisconsin, 1950 and 1960 (Male workers only)

Occupation	1950 Census[a]	1950 Wis[b]	1960 Census[a]	1960 Wis[b]
Professional, technical and kindred	7%	(7)%	9%	(9)%
Professional *		5		7
Semi-professional *		2		2
Managers, Proprietors	11	(15)	9	(15)
Managerial *		6		8
Self-employed businessmen *		8		8
Farmers, farm managers *	11	12	11	12
Clerical & kindred *	7	5	6	4
Sales *	6	6	6	5
Service workers, excluding private household	6		5	
Private household	c		c	
All service *	(6)	4	(5)	4
Craftsmen, foremen, and kindred *	19	17	20	16
Operatives and kindred	20		24	
Laborers, excluding farm and mine	8		6	
Farm laborers and foremen	5		4	
All laborers and operatives *	(33)	34	(33)	34
Total employed, Occupation not ascertained excluded	100%	100%	100%	100%
% of employed with occupation not ascertained	1%	3%	4%	3%

[a]Source: 1960 Census of Population PC(1), 51C-Wisconsin. [b]Tax record data are for the tax years 1949, 1959. [c]less than 0.5%.

*A change in major occupation group is defined as any movement that causes an individual to be classified in two distinct groups on successive tax returns. (For this purpose the aggregates in cols. 2 and 4 that are enclosed in parentheses are ignored.) For a limited number of persons successive tax returns may not be consecutive years. An investigation of those cases indicated that no bias was involved in ignoring "gaps" in the filing of tax returns.

TABLE 4--Mean Earnings within Annual Occupational and Residential Mobility Status (Males) 1948-1959[a]

Occupational Mobility, year t	Geographic Mobility, Year t: No Geographic Move	Intra-County Move	Inter-County Move	All
None	$4,203	$4,288	$3,923	$4,200
Change occupation	3,424	3,111	3,178	3,387
Enter labor force	1,801	2,406	2,250	1,886
Exit labor force	1,414	1,215	1,604	1,417
All, including N.A.	4,116	4,054	3,665	4,104

[a]Sample B

TABLE 2--Annual Observed Geographic and Occupational Mobility Rates (Males)

Year	Occupational Change Wis[b]	Intra-County Move Wis[b]	Intra-County Move CPS-U.S.[a]	Inter-County Move Wis[b]	Inter-County Move CPS-U.S.[a]	Inter-State Move CPS-U.S.[a]
1947-48	7.0%	3.4%	13.6%	2.1%	3.3%	3.2%
1948-49	6.2	2.8	12.9	2.0	2.8	3.2
1949-50	7.0	3.6	13.1	2.3	3.1	2.7
1950-51	6.6	2.8	14.1	2.3	3.6	3.6
1951-52	7.0	3.1	13.2	2.1	3.3	3.5
1952-53	6.5	3.2	13.4	2.3	3.1	3.7
1953-54	5.6	3.4	12.3	2.3	3.1	3.3
1954-55	6.4	2.8	13.4	3.0	3.5	3.2
1955-56	5.9	2.6	13.6	2.4	3.8	3.2
1956-57	5.7	2.4	13.3	2.2	3.3	3.1
1957-58	5.4	2.3	13.3	2.5	3.5	3.5
1958-59	6.4	2.8	---	2.8	---	---

[a]Source: (Shryock) Table 11.7. [b]Wisconsin tax record sample (8).

TABLE 3--Data Used in the Mobility Study

No. of Records per Taxpayer	No. of Tax Records 1946-60	No. of Tax Records 1947-59	No. of Moves Observed 1946-60	No. of Moves Observed 1947-59	Taxpayers 1947-59
1 or more	93894	87814			10934
2 or more		87765	83853[a]	76880[b]	9885
3 or more		85169		76082[c]	9087
4 pairs of consecutive years[d] and					
5-6 records					1088
7-8 records					1254
9-10 records					986
11-12 records					1116
13 records					3036

[a]Sample A; [b]Sample B; [c]Sample C; [d]Sample D; totals are estimated for the population from a 50% random sample.

TABLE 5--Rate of Annual Occupational and Residential Mobility (Males) 1948-1959[a]

Occupational Mobility, Year t	Geographic Mobility, Year t: No Geographic Move	Intra-County Move	Inter-County Move	All
None	87.8%	2.4%	1.7%	91.9%
Change occupation	5.6	.4	.4	6.4
Enter labor force	.5	---	.1	.6
Exit labor force	.9	---	.1	1.0
All[b]	94.8%	2.8%	2.3%	100.0%

[a]Sample B; [b]Totals may not be exact due to rounding. --- is less than .1%.

TABLE 6--Rate of Annual Occupational and Residential Mobility Within Birth Year Cohorts, 1947-1960[a]

Type of Mobility, Year t	Birth Year Cohort 1860-1894	1895-1904	1905-1914	1915-1924	1925-1959
No occupation change					
No location change	94.5%	93.2%	91.5%	88.0%	81.5%
Intra-county move	1.4	1.8	2.0	2.6	3.8
Inter-county move	.8	.8	1.1	2.1	3.3
Occupation change*					
No location change	2.9	3.6	4.7	6.4	9.7
Intra-county move	.2	.2	.3	.4	.8
Inter-county move	.2	.3	.4	.4	.9
Total	100.0	100.0	100.0	100.0	100.0
Number Records	10854	14809	17579	17326	14631

*Excludes movement into and from labor force.

[a]Sample A.

TABLE 7--Two-Year Changes in Earnings Associated with Annual Mobility Within Birth Year Cohorts, 1948-1959[a]

Birth Year Cohort	Earnings Change From	Occupational Mobility, Year t: None	Change	Gain	Residential Mobility, Year t: Within County	Inter-County	Gain
1860-	t-1,t	$-27	$-132	$	$-98	$-362	$
1894	t,t+1	-122	-109	+13	-135	-107	28
	2-years	-149	-241	-92	-233	-469	-236
1895-	t-1,t	130	99		125	28	
1904	t,t+1	99	121	22	97	213	116
	2-years	229	220	-9	222	241	19
1905-	t-1,t	218	217		225	-247	
1914	t,t+1	182	243	61	184	506	322
	2-years	400	450	50	409	259	-150
1915-	t-1,t	302	259		303	169	
1924	t,t+1	245	363	118	244	657	413
	2-years	547	622	75	547	826	279
1925-	t-1,t	468	356		449	431	
1959	t,t+1	350	478	128	379	315	-64
	2-years	818	834	16	828	746	-82

[a]Sample C (males only)

TABLE 8--Means of Individual Mean Earnings within Long Term Occupational Mobility and Birth Year Cohorts 1947-1959, Sample D

Long-term Mobility[a]	Birth Year Cohort 1860-1894	1895-1914	1915-1924	1925-1959	Entire Sample
None	$3351	$4744	$4376	$3597	$4292
Change	3039	4139	4145	3404	3807
Entered L. F.	*	3905	4471	3052	3387
Exit L. F.	2861	2738	2863	1867	2701

* No Entries

[a]See Table 9 for definitions.

TABLE 9--Rate of Long-Term Occupational Mobility By Labor Force Participation Within Birth Year Cohorts 1947-1959, Sample D

Labor Force Participation	No. of Occupations	Birth Year Cohort 1860-1894	1895-1914	1915-1924	1925-1959	Total
Continuously Employed	*One	81.6%	69.5%	58.1%	50.4%	63.6%
	**Two	16.7	24.4	31.2	38.7	28.7
	** >2	1.8	6.0	10.7	10.9	7.8
	Total	100.0	100.0	100.0	100.0	100.0
	N	(342)	(1415)	(766)	(613)	(3341)
Entered and Remained in labor force	#One	--	75.0	50.0	65.8	59.4%
	#Two	--	25.0	35.7	28.9	31.2
	# >2	--	--	14.3	5.3	9.4
	Total	100.0	100.0	100.0	100.0	100.0
	N	(0)	(4)	(14)	(38)	(64)
Exited and Remained out of labor force	##One	84.8	77.8	100.0	88.9	84.3
	##Two	14.5	16.7	--	11.1	14.8
	## >2	.6	5.6	--	--	1.0
	Total	100.0	100.0	100.0	100.0	100.0
	N	(165)	(18)	(1)	(9)	(210)
Other	*One	50.0	15.0	25.0	22.2	29.3
	**Two	37.5	50.0	50.0	58.3	49.5
	** >2	4.2	20.0	25.0	13.9	13.1
	**Other	8.3	15.0	--	5.6	8.1
	Total	100.0	100.0	100.0	100.0	100.0
	N	(24)	(20)	(4)	(36)	(99)
Continuously Non-Labor Force		20	1	0	2	26
Total		551	1458	785	698	3740

*row one of Table 8. **row two of Table 8. #row three of Table 8. ##row four of Table 8.

TABLE 10--A Comparison of Income Trends of Persons Who Change Occupation with Those Who Do Not Change Occupational Affiliation (Individuals who were continuously in the labor force, Sample D)

Cumulative Occupational Mobility	Number of Individuals	Relative Income Position Forecast for 1959[a]	Trend	Std. Error of Estimate
Unique detail occupation	1821	1.12	-0.0108	.36
Unique major occupation, change in detailed occupation	71	1.16	0.0014	.06
Two major occupation groups	801	1.05	-0.0058	.14
Three or more occupation groups	204	.92	0.0165	.15
All employed	2897	1.09	-0.0095	.28

[a]Intercept of $y_{it}=\alpha_i+\beta_i(t-1959)+u_{it}$ where α_i and β_i were subsequently pooled to the above groups; y_{it} is the income relative to his birth cohort. (cf. 3)

TABLE 11--Differential Impact of Occupational Mobility on Different Birth Cohorts
(Individuals who were continuously in the labor force, Sample D)

Cumulative Occupational Mobility	1860-1884	1885-1894	1895-1904	1905-1914	1915-1924	1925-1929	1930-1934
			Estimated Relative Income Forecast for 1959				
Unique detail occupation	1.01	1.24	1.27	1.09	0.99	0.89	0.89
Unique major occupation, change in detail	*	*	1.40	1.15	1.01	*	*
Two major occupation groups	1.94	1.33	1.09	1.01	0.98	0.95	1.08
Three or more major occupation groups	*	*	0.86	0.83	0.96	0.99	0.81
All	1.09	1.27	1.22	1.05	0.99	0.92	0.96
Trend in Relative Income Position for Different Birth Cohorts by Occupational Mobility							
Unique detail occupations	-.072	.010	.011	.004	-.011	-.031	-.107
Unique major occupation, change in detail	*	*	.035	-.007	-.005	*	*
Two major occupation groups	.010	.030	-.002	.013	-.002	-.027	-.048
Three or more major occupation groups	*	*	.012	.001	.001	.001	-.143
All	-.063	.014	.009	.006	-.007	-.026	-.085

*Less than 10 observations

TABLE 12--Pooled Regressions on p_{ijt} for 1951-1959 Coefficients on Independent Variables
(t values in parentheses)

Model	Constant	$\overline{\Delta E}_t$	$\overline{\Delta E}_{t-1}$	O_{it}	U_{it}	T_i	ΔT	R^2	F (d.F.)	Std. Error Est.
1	$.710 \times 10^{-2}$	$.590 \times 10^{-8}$						.010	8.34*	.012
	(17.35)*	(2.89)*							(1808)	
2	$.710 \times 10^{-2}$	$.120 \times 10^{-7}$	$-.674 \times 10^{-8}$					.011	4.37	.012
	(17.34)*	(1.21)	(-.63)						(2807)	
3	$.899 \times 10^{-2}$	$.591 \times 10^{-6}$			$-.260 \times 10^{-3}$			.014	5.82*	.012
	(8.36)*	(2.90)*			(-1.90)				(2807)	
4	$.623 \times 10^{-2}$	$.484 \times 10^{-8}$		$.775 \times 10^{-5}$				.015	5.97*	.012
	(10.11)*	(2.28)*		(1.89)					(2807)	
5	$.610 \times 10^{-2}$	$.105 \times 10^{-7}$		$.887 \times 10^{-5}$			$-.510 \times 10^{-4}$	.032	8.80*	.011
	(9.97)*	(4.07)*		(2.17)*			(-3.78)*		(3806)	
6	$.220 \times 10^{-1}$	$.992 \times 10^{-8}$		$.140 \times 10^{-4}$		$-.151 \times 10^{-3}$	$-.128 \times 10^{-3}$	.077	16.84	.011
	(8.48)*	(3.94)*		(3.45)*		(-6.30)*	(07.11)*		(4805)	

*Denotes significance at less than .01 level.

REASONS DEATHS ARE MISSED IN HOUSEHOLD SURVEYS OF POPULATION CHANGE

Monroe G. Sirken and Patricia Nellans Royston
National Center for Health Statistics

1. Introduction

The primary objective of this report is to investigate reasons for incomplete reporting of noninstitutional deaths in household surveys when the household reports:[1]

(1) deaths that occurred in the dwelling unit the household occupies,

(2) deaths that occurred in the unit adjacent to the dwelling unit the household occupies,

(3) deaths of parents that occurred in a different dwelling unit than the one the household occupies,

(4) deaths of siblings that occurred in a different dwelling unit than the one the household occupies.

Why investigate the incomplete reporting of deaths in household surveys under the alternative conditions listed above? This question needs to be discussed first. Then, we will describe the design of a survey experiment to investigate incomplete reporting of deaths. Finally, we will analyze some preliminary statistics derived from the survey experiment.

2. Counting rules

In mortality surveys, households report retrospectively about deaths that occurred during a prior calendar period. Since decedents are not residents of households when the survey is conducted, it is essential to adopt counting rules for linking deaths to households or dwelling units that will report them in the survey. The counting rule conventionally adopted in the mortality survey links each death to the decedent's former dwelling unit, that is, the dwelling unit that was the decedent's place of residence at the time of death. Thus, in the conventional survey, a household reports about deaths that occurred in the dwelling unit it is occupying.

In addition to the counting rule of the conventional survey, there are other possible rules for linking deaths to dwelling units. For example, deaths may be linked to the dwelling units of neighbors. According to this rule, a household reports deaths of persons who formerly lived in a neighboring dwelling unit. There are also a large number of possible rules for linking deaths to households. For example, consanguine rules link deaths to households of specified surviving relatives, such as siblings or children, of the decedent. According to consanguine rules, a household reports deaths of surviving relatives who live in the household.

The survey estimator is unbiased, however, if and only if the counting rule assures that every death is linked to an existing household or dwelling unit when the survey is conducted. The conventional rule would appear to satisfy this condition for noninstitutional deaths barring the possibility that decedents were homeless or that their former dwelling units were demolished. The rule linking deaths to dwelling units of neighbors also comes close to satisfying the condition for an unbiased estimator since virtually every dwelling unit has a neighboring dwelling unit assuming that it is operationally feasible to uniquely define a "neighboring" dwelling unit. Neither of the consanguine rules previously mentioned would satisfy the condition because many decedents do not have surviving children or siblings. However, if a consanguine rule and the conventional counting rule were jointly adopted in a survey the condition for an unbiased estimator would be satisfied. Based on a multiplicity rule of this type, a household would report deaths that occurred in the dwelling unit it occupies as well as deaths of relatives that occurred in other dwelling units. A noteworthy difference between conventional and multiplicity counting rules is that the conventional rule links each death to one and only one household while the multiplicity rule links each death to at least one household.[1/]

Incomplete reporting of deaths in household surveys based on conventional counting rules has been substantial. In household surveys that were conducted in several Asian countries the incompleteness ranged between 10 and 80 percent.[2/] In a North Carolina survey experiment conducted by Horvitz,[3/] 15-20 percent of the deaths were not reported. The underreporting would have been greater except that neighbors were used as proxy respondents whenever the information could not be obtained from households that occupied the decedents' former dwelling units.

We believe that bias due to incomplete reporting of deaths is greater in surveys based on a conventional counting rule than it would be in surveys based on alternative counting rules involving households of neighbors and/or relatives. However, this report has a more limited objective. We hypothesize that the level of incomplete reporting and the factors contributing to incomplete reporting of deaths are related to the kinds of deaths that households are eligible to report in the survey.

3. Survey experiment

Estimates presented in this report are based on a survey experiment involving a sample of 142 white noninstitutionalized adult deaths that occurred in Los Angeles County during the four month period July-October 1969. The sample was selected from the records of registered deaths filed in the County Health Department.

The field work, completed during the first three months of 1970 involved interviews with three kinds of households:

(1) For each of the 142 sample deaths the interviewer visited the household occupying the decedent's former dwelling unit. Addresses for these dwelling units were obtained from the death records.

(2) For a random subsample of 26 deaths the interviewer visited a household located adjacent to the decedent's former dwelling unit.

(3) For 46 deaths the interviewer visited the household of a surviving sibling or child that had been reported by the household occupying the decedent's former dwelling unit.

In the survey experiment a household occupying the decedent's former dwelling unit and the household adjacent to it reported deaths that occurred in the decedent's former dwelling unit. The household of relatives not occupying the decedent's former dwelling unit reported deaths of siblings and parents. Any adult in a selected household was an acceptable respondent to report the eligible deaths that occurred during calendar year 1969. In this manner, the survey experiment simulated household surveys based on the conventional counting rule and on rules linking deaths to dwelling units of neighbors and to households of siblings and children.

Some limitations of the survey experiment design are noteworthy. (1) The sample is not representative of a well defined population because the selection was purposively limited to 10 sets of Census tracts in order to reduce field traveling costs. (2) The sample sizes are very small particularly with regard to the number of interviews that were attempted with neighbors and with relatives that did not reside at the decedents' former dwelling units. (3) Interviews were possible with relatives if and only if the decedent's former dwelling unit reported the death and reported the name and address of the decedent's relative who lived in Los Angeles. (4) Finally, the survey experiment did not measure erroneously reported deaths. It was limited to an investigation of missed deaths.

4. Findings

Estimates of incomplete reporting of white deaths by resident and nonresident respondents were derived from the survey experiment. These estimates as well as statistics on the reasons that deaths were missed are presented in Table 1.

Nearly 40 percent of the deaths were missed in the survey of households occupying the decedents' former dwelling units. A smaller percentage of deaths was missed by households of neighbors. Households of children and of siblings respectively that were not occupying the decedents' former dwelling units missed about 20 percent of their parents' deaths and about 40 percent of their siblings' deaths.

Basically one of two conditions accounted for deaths that were missed in the survey. Either the interview was completed and the death was not reported or the death was not reported

Table 1. Reasons white deaths were missed by type of deaths that households reported in the survey.

Reasons deaths were missed	Deaths reported by the household				
	Deaths in the dwelling unit occupied by the household	Deaths in the dwelling unit adjacent to the household	Deaths of relatives in a dwelling unit not occupied by the household		
			Total	Sibling	Parent
Number of deaths	142	26	46	16	30
Total percent	100	100	100	100	100
Deaths reported	63	69	74	62	80
Deaths missed	37	31	26	38	20
Interview completed	19	27	13	25	7
Interview not completed	18	4	13	13	13

because the interview was not completed. Noninterviews account for about half the deaths missed by households occupying the decedents' former dwelling units, for more than half the deaths missed by neighbors and by households of siblings, and for less than half the deaths missed by households of children.

Principal reasons for noninterview were refusals and no one was found at home. The refusals and not-at-home rates were about six and ten percent respectively for households occupying decedents' former dwelling units. The comparable rates were smaller for households of neighbors and relatives.

The proportion of deaths that were missed in interviews completed with each type of household is presented in Table 2. Deaths of parents were reported in over 90 percent of the interviews completed with childrens' households that were not occupying the decedents' former dwelling units. Deaths were reported in 70 to 80 percent of the interviews completed with each of the other kinds of households. The reasons deaths were missed in completed interviews varied considerably by the kind of deaths the household reported.

Table 2. Percent of white deaths that were missed in completed interviews by type of deaths that households reported in the survey.

Reporting status of death	Deaths reported by the household				
	Deaths in the dwelling unit occupied by household	Deaths in the dwelling unit adjacent to the household	Deaths of relatives in a dwelling unit not occupied by the household		
			Total	Sibling	Parent
Number of completed interviews	117	25	40	14	26
Total percent	100	100	100	100	100
Deaths reported	77	72	85	71	92
Deaths not reported	23	28	15	29	8

Of the 117 completed interviews with households occupying the decedent's former dwelling units, 91 were former households of the decedent and 26 were households that had moved into the dwelling units after the deaths occurred. All but one of the former households reported the death but none of the latter households reported the death in the survey. Thus, changes in the household occupying the decedents' former dwelling units between the date of death and the date of the interview accounted for all but one of the 27 deaths that were not reported in completed interviews with these dwelling units. (In addition, three interviews were not completed because the decedents' former dwelling unit was vacant when the survey was conducted.)

If neighbors did not know of the death next door, it was not because they had moved there after the death occurred. Twenty-three of the twenty-five neighbors who did not report deaths in completed interviews had lived in the same dwelling unit prior to July 1969, the earliest month of any death in the survey experiment.

Why were deaths of siblings and parents missed in interviews completed with relatives' households not occupying the decedents' former dwelling units? Since these relatives were reported by the decedents' former household, it seems almost certain that the relatives knew of the death. It is noteworthy that proxy respondents such as in-laws were interviewed in about half the relatives' households. However, proxy respondents missed a smaller fraction of deaths than self-responding relatives although the difference is not statistically significant.

We can only speculate why deaths of parents were reported more completely than the deaths of siblings. The small number of households of siblings and parents in the survey experiment precludes reaching any statistical conclusions. A case study review indicated, however, that the age of the respondent may be a factor. For example, neither of the two respondents in households of surviving sibling who were over 70 years old reported the sibling's death. On the other hand, none of the respondents in households of surviving children were in the oldest age grouping.

5. Conclusions

We have shown that the completeness of death reporting and the reasons deaths are missed in single retrospective surveys vary according to the kinds of deaths that households are eligible to report. These findings imply that the completeness of death reporting can be improved by adopting appropriate rules for linking dwelling units to the deaths they report in the survey. The appropriateness of the rules would probably vary from survey to survey depending on the population being studied.

Footnote

[1] An understanding of the difference between a household and a dwelling unit is critical in reading this report. A dwelling unit is a separate and independent enclosure such as a house or apartment occupied or intended for occupancy as living quarters by a group of persons living together or by a person living alone. The person or group of persons residing in the dwelling unit comprise the household. A dwelling unit may be occupied by the same or different households at two points in time.

References

1/ Sirken, Monroe G. "Household Surveys with Multiplicity" Journal of the American Statistical Association, Theory and Methods Section, March 1970, Volume 65, Number 329, pages 257-266.

2/ Seltzer, William "Some Results from Asian Population Growth Studies" Population Studies, Volume 23, November 1969, pages 395-406.

3/ Horvitz, D.G. "Problems in Designing Surveys to Measure Population Growth," 1966 Social Statistics Section Proceedings of the American Statistical Association, pages 245-249.

Acknowledgements

This study would not have been possible without the cooperation and assistance of Martin Donabedian and his staff of the Records and Statistics Division, Los Angeles County Health Department, who provided the sample of death records.

The field work of the survey experiment was conducted for the National Center for Health Statistics by the Survey Research Center, UCLA.

LOCAL MINIMUM SOLUTIONS IN NONMETRIC MULTIDIMENSIONAL SCALING

Ian Spence, University of Toronto

Although it has been widely recognized that nonmetric scaling algorithms are potentially susceptible to local minimum problems, there is little systematic data on the relative frequency with which these programs will become trapped in nonoptimal stationary positions. This study provides detailed information on this question.

An empirically obtained measure of the dissimilarity of any two objects in a set P of n objects gives rise to an order relation on P x P. Nonmetric multidimensional scaling seeks to represent this dissimilarity ordering as a geometrical model by mapping the objects of P into a metric space, where the ordering of the distances corresponds to the observed ordering of the dissimilarities. Thus, if s_{ij} is the dissimilarity measure obtained for objects i and j,

(1) $$d_{ij} \leqq d_{kl} \text{ iff } s_{ij} \leqq s_{kl} \quad ,$$

where d_{ij} is the metric distance between points i and j.

Practical computer algorithms (e.g. Kruskal [2], McGee[4], Guttman[1], Young[7]) for obtaining a configuration of n points in m-dimensional space have been based on an iterative procedure designed to minimize the residual sum of squares (called "stress")

(2) $$S = \sum_{i,j}^{n} (d_{ij} - \delta_{ij})^2 \quad ,$$

or some other quantity which is monotonically related to S (see Spence[6]). The $\{d_{ij}\}$ are <u>metric distances</u>, and the $\{\delta_{ij}\}$ are <u>pseudo distances</u>. The pseudo distances have the property of being order isomorphic to the dissimilarities $\{s_{ij}\}$, and have the same scale as the $\{d_{ij}\}$. Further, if the $\{\delta_{ij}\}$ are chosen to minimize S, given a particular set of $\{d_{ij}\}$, the resulting pseudo distances are known as the $\{\hat{d}_{ij}\}$ (Kruskal [2]). The minimization of S is not trivial since an optimal set of $\{\delta_{ij}\}$ is not known. Consequently, an alternating algorithm has been proposed (Kruskal[2], Guttman[1]): this procedure switches between satisfying the metric distance requirement and the order isomorphy requirement, with the hope that satisfying both in turn will eventually result in the algorithm arriving at a stationary position where both requirements will be optimally satisfied. In the Euclidean case, the following algorithm may be used:

1. Choose an initial $X_{n \times m}$: a configuration of points in m-space.
2. Compute $D_{n \times n}$ from $X_{n \times m}$, where
$$d_{ij} = [\sum_{a=1}^{m} (x_{ia} - x_{ja})^2]^{1/2} \quad .$$
3. Choose $\Delta_{n \times n}$ with general entry δ_{ij} .
4. Move $D_{n \times n}$ towards $\Delta_{n \times n}$ by use of a gradient algorithm to minimize S. Typically only one step is taken, viz.,

(3) $$x'_{ia} = x_{ia} - \frac{1}{n} \sum_{j=1}^{n} \left(1 - \frac{\delta_{ij}}{d_{ij}} \right) (x_{ia} - x_{ja}) \quad ;$$

however, if desired, more than one step may be taken--with a fixed set of $\{\delta_{ij}\}$--as in Guttman[1].
5. Test for termination: if further improvement is desired, go to 2 with the new X. Else,
6. End.

TABLE I

Summary of the Essential differences among the Algorithms

	Algorithms		
Options	M-D-SCAL	SSA-I	TORSCA
Choice of initial configuration.	does not make full use of information in the input data.	attempts to make "optimal" use of the input data.	attempts to make "optimal" use of the input data.
Choice of pseudo distances, $\{\delta_{ij}\}$.	to minimize S, hence these are the $\{\hat{d}_{ij}\}$.	ascending permutation of the $\{d_{ij}\}$ called $\{d^*_{ij}\}$. Does not minimize S.	to minimize S, hence these are the $\{\hat{d}_{ij}\}$.

There are two arbitrary points in the specification of the above algorithm: both the initial configuration and the pseudo distances may be chosen in a variety of ways. These choices may be expected to have some effect on the overall performance of the algorithm, especially with respect to the avoidance of nonoptimal stationary solutions.

In a study which is reported in greater detail elsewhere (Spence[6]), three widely used algorithms were compared using Monte Carlo techniques. The programs were Kruskal's M-D-SCAL (Kruskal[3]), Guttman-Lingoes's SSA-I (Guttman [1]), and Young-Torgerson's TORSCA-9 (Young[7]). The essential differences among these procedures are summarized in Table I. Known configurations were used as the basis for computing sets of dissimilarities which were then scaled by each of the algorithms. The recovered configurations were then compared with the known generated configurations. Specifically, the following method was used:

1. Choose n and m and generate the n x m coordinates of X by sampling from the rectangular interval [-1, +1] with the additional constraint that all points lie within a hypersphere of unit radius.

2. Compute $d^e_{ij} = [\sum_{a=1}^{m} (x^e_{ia} - x^e_{ja})^2]^{1/2}$,

where $x^e_{ia} = x_{ia} + N(0, \sigma_k^2)$, and

σ_k = 0.00; 0.15; 0.25; 0.35; is the set of error values used. This method of injecting error corresponds to a multidimensional analogue of Thurstone's Case V (see Ramsay[5]). The dissimilarity matrices were computed with entry

$$(4) \quad s_{ij} = \begin{cases} 1.8\,(d^e_{ij})^2 + 5.5 & \text{if } \sigma_k = 0.00, \\ d^e_{ij} & \text{otherwise.} \end{cases}$$

3. Obtain scaling solutions in m = 1, ... , 5 recovered dimensions using each of the algorithms. Compute $\hat{S}_1 = [\sum (d - \hat{d})^2 / \sum d^2]^{1/2}$ and $r(d,d_+)$ --the product moment correlation coefficient between the recovered and true distances--as measures of goodness of fit and metric recovery respectively.

In a single computer run n was varied from 6 to 36 and m from 1 to 4; in total, each of the three algorithms processed:

DISTINCT CONFIGURATIONS(18) x ERROR LEVELS(4)

x RECOVERED DIMENSIONS(5) = 360 SOLUTIONS .

Two replications were obtained; hence, 2160 solutions were computed.

Results and Discussion

The initial configurations generated by the TORSCA program were invariably much better, in terms of goodness of fit and metric recovery, than the approximations generated by the other two algorithms. SSA-I and M-D-SCAL did not differ significantly in their abilities to produce a starting configuration. The final solutions produced by the three procedures were, in the vast majority of cases, virtually identical; indeed, analysis of variance showed that the hypothesis of no differences among the programs could not be rejected (see Spence[6]). However, in some of the solutions, one or more of the algorithms obtained a solution which deviated considerably from the best attempt.

A simple method was used to investigate this local minimum problem: it was assumed that at least one of the three algorithms would, in all cases, get very close to the global minimum. This seems to be a plausible assumption, since these programs use rather different methods to produce a solution. Furthermore, this assumption is reinforced by the fact that inspection of the data showed that in almost all cases at least two of the algorithms finished in virtually identical positions. Using the criterion of goodness of fit defined above, $\hat{S}_1$, and considering each of the 720 solutions attempted per algorithm separately, the deviations of the values obtained by the other two programs from the lowest stress value were computed. If any deviation exceeded a preset threshold criterion, then the deviating algorithm was considered to be in a local minimum position.

The results of this analysis are shown in Table 2 for different values of the threshold (varying from 0.005 to 0.050). The 0.005 threshold is probably too stringent a criterion since a deviation of this magnitude may simply indicate that the offending algorithm was moving slowly in the region of the minimum, and had not quite converged. However, as the threshold size is increased to 0.010, and above, an interesting pattern emerges. It seems to be clear that TORSCA is least troubled by local minimum problems (in terms of the above operational definition), and SSA-I is only a little worse; although, it does appear from the data that SSA-I may not have fully converged, since the vast majority of SSA-I deviations are of the order of 10^{-3}. M-D-SCAL appears to be much more sensitive to local minimum problems than the other two algorithms, and, more importantly, 27 of its solutions deviated by more than 0.050 from the presumed minimum stress value. By contrast, only one of the 1440 SSA-I and TORSCA solutions deviated by more than this amount from the minimum value, and, in fact, only five of the SSA-I and TORSCA solutions deviated by more than 0.030. In one dimension the problem seems to be especially severe, and consequently, in practical situations, it would be prudent to regard any one dimensional solution with some suspicion, irrespective of the method used to derive the initial configuration.

Conclusions

It is reasonable to attribute the excellent performance of TORSCA to its very good initial configuration; as has already been noted, the TORSCA starting position was easily the best. Likewise, the poorest performance (by M-D-SCAL) may be partially attributed to its often unsatisfactory initial configuration. The fairly good performance of SSA-I is rather more difficult to explain on the basis of the quality of the initial configuration; the SSA-I initial configuration was not much better, in general, than the M-D-SCAL starting configuration. Indeed, it is

TABLE 2

Frequency of Deviation from the Presumed Minimum Stress Value

Threshold	Program	Dimensionality					Totals
		1	2	3	4	5	
0.005	TORSCA	30	11	8	8	21	78
	SSA-I	25	32	36	44	38	175
	M-D-SCAL	64	29	10	10	10	123
0.010	TORSCA	18	0	4	1	2	25
	SSA-I	14	12	12	16	8	62
	M-D-SCAL	55	22	5	3	5	90
0.020	TORSCA	5	0	0	0	0	5
	SSA-I	3	2	4	5	0	14
	M-D-SCAL	43	9	4	1	4	46
0.030	TORSCA	2	0	0	0	0	2
	SSA-I	2	0	0	1	0	3
	M-D-SCAL	32	5	4	1	4	46
0.050	TORSCA	0	0	0	0	0	0
	SSA-I	1	0	0	0	0	1
	M-D-SCAL	19	0	3	1	4	27

Note.--The total number of solutions per algorithm was 720, hence there were 144 solutions per algorithm per dimension.

probable that the use of the $\{d^*_{ij}\}$ produces a more "jumpy" algorithm which tends to step over small local depressions on its way to the minimum.

Acknowledgement

This research was partially supported by a predoctoral fellowship from the National Research Council of Canada, and also by an NRC grant (APA-176) to John C. Ogilvie.

REFERENCES

[1] Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 33 (1968), 469-506.

[2] Kruskal, J.B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 29 (1964), 115-129.

[3] Kruskal, J.B. How to use M-D-SCAL, a program to do multidimensional scaling and multidimensional unfolding. Unpublished MS (1968). Bell Laboratories, Murray Hill, New Jersey.

[4] McGee, V.E. The multidimensional analysis of "elastic" distances. British Journal of Mathematical and Statistical Psychology, 19 (1966), 181-196.

[5] Ramsay, J.O. Some statistical considerations in multidimensional scaling. Psychometrika, 34 (1969), 167-182.

[6] Spence, I. Multidimensional scaling: An empirical and theoretical investigation. Unpublished Ph. D. Thesis. University of Toronto (1970).

[7] Young, F.W. A FORTRAN IV program for nonmetric multidimensional scaling. L.L. Thurstone Psychometric Laboratory Report No. 56 (March, 1968). University of North Carolina, Chapel Hill, North Carolina.

ALLOCATION IN STRATIFIED SAMPLING BASED ON PRELIMINARY TEST OF SIGNIFICANCE

B. V. Sukhatme and Victor K. T. Tang
Iowa State University and Humboldt State College

1. Introduction

Consider a population of N units classified into k strata, the i-th stratum having N_i units so that $\sum_{i=1}^{k} N_i = N$. Let Y be the characteristic under study and consider the problem of estimating the mean $\bar{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i$ from a stratified random sample of size $\sum_{i=1}^{k} n_i$ where n_i units are drawn by simple random sampling without replacement from the i-th stratum i = 1, 2, ..., k. An unbiased estimate of the mean $\bar{y}_N$ is given by

$$\bar{y}_W = \sum_{i=1}^{k} W_i \bar{y}_{n_i} \qquad (1.1)$$

where W_i is the proportion of units in the i-th stratum and $\bar{y}_{n_i}$ is the simple mean estimate of $\bar{y}_{N_i}$, the mean for the i-th stratum. If N_i is so large that $\frac{N_i}{N_i-1} \simeq 1$, $V(\bar{y}_W)$ can be written as

$$V(\bar{y}_W) = \sum_{i=1}^{k} \frac{W_i^2\sigma_i^2}{n_i} - \frac{1}{N}\sum_{i=1}^{k} W_i\sigma_i^2 . \qquad (1.2)$$

If the total sample size n is fixed in advance, the classical problem of allocation of sample sizes in stratified sampling is to determine a vector $(n_1, n_2, \ldots, n_k)$ of k non-negative integers such that $\sum_{i=1}^{k} n_i = n$ and for which $V(\bar{y}_W)$ is minimum. The allocation so determined, commonly known as Neyman allocation (Neyman, 1934) is given by

$$n_i = nW_i\sigma_i / \sum_{i=1}^{k} W_i\sigma_i \quad . \qquad (1.3)$$

Neyman allocation however depends on strata variances σ_i^2 which are generally not known. One way out of this difficulty (Sukhatme, 1935) is to draw an initial sample of fixed size m from each stratum to estimate σ_i^2 which in turn are used to estimate n_i from (1.3). In this case, n_i is estimated by

$$n_i = nW_i s_i / \sum_{i=1}^{k} W_i s_i \qquad (1.4)$$

where s_i^2 is an unbiased estimate of σ_i^2. The allocation (1.4) will be called Modified Neyman allocation. Another allocation which is frequently used in practice and does not require knowledge of strata variances σ_i^2 is proportional allocation. If the strata variances σ_i^2 do not differ significantly among themselves, modified Neyman allocation may turn out to be less efficient than proportional allocation (Evans, 1951).

Before deciding on the method of allocation, it is therefore proposed to carry out a preliminary test of significance concerning the homogeneity of strata variances. If on the basis of the test of significance the strata variances are found to be homogeneous, the sample sizes to be drawn from the different strata will be determined according to proportional allocation. This allocation based on preliminary test of significance will be called 'sometimes proportional allocation'. This paper will consider in detail the sometimes proportional allocation and discuss its efficiency with respect to proportional allocation and modified Neyman allocation for the relatively simple case of two strata when $\sigma_1^2 \leq \sigma_2^2$. The results for three or more strata will be presented in a separate communication.

2. Variance of $\bar{y}_W$ under sometimes proportional allocation.

The sometimes proportional allocation may be defined as

$$\left.\begin{aligned} n_i &= nW_i && \text{if } \frac{s_2^2}{s_1^2} < \lambda \\ &= nW_i s_i / \sum_{i=1}^{2} W_i s_i && \text{otherwise} \end{aligned}\right\} \qquad (2.1)$$

where λ is a known constant. Clearly, the variance of $\bar{y}_W$ is given by

$$V(\bar{y}_W)_S = E\{V(\bar{y}_W \mid \frac{s_2^2}{s_1^2} < \lambda)\}\, P(\frac{s_2^2}{s_1^2} < \lambda)$$

$$+ E\{V(\bar{y}_W \mid \frac{s_2^2}{s_1^2} \geq \lambda)\}\, P(\frac{s_2^2}{s_1^2} \geq \lambda) \qquad (2.2)$$

where the expectation in each term is taken with reference to the corresponding set and S stands for sometimes proportional allocation. To evaluate the various terms, it will be assumed that $\frac{(m-1)s_i^2}{\sigma_i^2}$ is distributed as chi-square with f = m-1 degrees of freedom. It can then be seen that

$$V(\bar{y}_W)_S = \frac{\sigma_1^2}{n}(W_1^2 + W_2^2\,\theta_{21}) - \frac{\sigma_1^2}{n}(W_1 + W_2\,\theta_{21})$$
$$+ \frac{W_1W_2\sigma_1^2}{n}\,[(1+\theta_{21})\,I_{q_{21}}(\tfrac{f}{2}, \tfrac{f}{2})$$
$$+ G\,\theta_{21}^{1/2}\,I_{p_{21}}(\tfrac{f-1}{2}, \tfrac{f-1}{2})] \qquad (2.3)$$

where $\theta_{21} = \sigma_2^2/\sigma_1^2$, $p_{21} = \theta_{21}/(\lambda + \theta_{21})$, $q_{21} = 1 - p_{21}$, $G = 2\Gamma(\frac{f-1}{2})/[\Gamma(\frac{f}{2})]^2$ and $I_{\cdot}(\cdot,\cdot)$ is the incomplete beta distribution.

If we let $\lambda \longrightarrow \infty$, we obtain the variance under proportional allocation, namely,

$$V(\bar{y}_W)_P = (\frac{1}{n} - \frac{1}{N})\sigma_1^2(W_1 + W_2\,\theta_{21}) \qquad (2.4)$$

where P stands for proportional allocation.

If we put $\lambda = 0$, we get the variance under modified Neyman allocation, namely

$$V(\bar{y}_W)_{\underline{N}} = \frac{\sigma_1^2}{n}(W_1^2 + W_2^2\,\theta_{21}) - \frac{\sigma_1^2}{N}(W_1 + W_2\theta_{21})$$
$$+ \frac{W_1W_2}{n}\,\sigma_1^2\,G\,\theta_{21}^{1/2} \qquad (2.5)$$

where $\underline{N}$ stands for modified Neyman allocation.

3. Efficiency of sometimes proportional allocation.

We shall first discuss the relative efficiency of sometimes proportional allocation with respect to proportional allocation. If $e_1(\lambda, \theta_{21})$ denotes the relative efficiency of sometimes proportional allocation with respect to proportional allocation, it is easy to see that

$$e_1(\lambda, \theta_{21}) = \frac{V(\bar{y}_W)_P}{V(\bar{y}_W)_S}$$
$$= 1/\,[1 - \frac{W_1W_2}{W_1+W_2\theta_{21}}\,\{(1+\theta_{21})I_{p_{21}}(\tfrac{f}{2}, \tfrac{f}{2})$$
$$- G\,\theta_{21}^{1/2}\,I_{p_{21}}(\tfrac{f-1}{2}, \tfrac{f-1}{2})\}\,] \qquad (3.1)$$

Clearly, if $e_1(\lambda, \theta_{21}) \geq 1$, sometimes proportional allocation is at least as efficient as proportional allocation. We shall now obtain some results concerning the behavior of the efficiency function. We shall first consider the case when λ is an arbitrary but fixed number such that

$$1 - \frac{G}{2} + \frac{\lambda^{\frac{f-1}{2}}(\lambda^{\frac{1}{2}}-1)}{B(\frac{f}{2}, \frac{f}{2})} > 0\,.$$

Then it can be seen that

i) $\lim_{\theta_{21} \to 1} e_1(\lambda, \theta_{21}) \leq 1$

ii) $\exists\, \theta' \ni \frac{\partial}{\partial\theta_{21}} e_1(\lambda, \theta_{21}) > 0$ for every $\theta_{21} > \theta'$

and

iii) $\lim_{\theta_{21} \to \infty} e_1(\lambda, \theta_{21}) > 1\,.$

As a consequence of the above, we obtain the following result.

Theorem 3.1. Let λ be an arbitrary but fixed number in the set $[0, \infty)$ such that

$$1 - \frac{G}{2} + \frac{\lambda^{\frac{f-1}{2}}(\lambda^{\frac{1}{2}}-1)}{B(\frac{f}{2}, \frac{f}{2})} > 0\,.$$

Then $\exists\, \theta_0 \ni e_1(\lambda, \theta_0) = 1$ and $e_1(\lambda, \theta_{21}) > 1$ for every $\theta_{21} > \theta_0$.

Theorem 3.1 assures us that there exists a θ_0 such that for each $\theta_{21} > \theta_0$, sometimes proportional allocation is always more efficient than proportional allocation.

We shall now consider the case when θ_{21} is an arbitrary but fixed number greater than or equal to $\frac{1}{2}(G^2 - 2 + G\sqrt{G^2 - 4})$. Then it is easy to see that $e_1(0, \theta_{21}) > 1$. Further, it can be shown that $e_1(\lambda, \theta_{21})$ is increasing if $\lambda < 1$ or $\lambda > \theta_{21}^2$ and decreasing for $1 < \lambda < \theta_{21}^2$. It follows that $e_1(\lambda, \theta_{21})$ reaches its maximum at $\lambda = 1$ and its minimum at $\lambda = \theta_{21}^2$. Also, it is not difficult to see that $\lim_{\lambda \to \infty} e_1(\lambda, \theta_{21}) = 1^-$. It is now clear that there exists λ_0 such that $e_1(\lambda, \theta_{21}) > 1$ for every $\lambda < \lambda_0$. We have thus proved the following result.

Theorem 3.2. Let θ_{21} be an arbitrary but fixed number greater than or equal to $\frac{1}{2}(G^2 - 2 + G\sqrt{G^2 - 4})$. Then $\exists\, \lambda_0 \ni e_1(\lambda_0, \theta_{21}) = 1$ and $e_1(\lambda, \theta_{21}) > 1$ for every $\lambda < \lambda_0$.

Armed with the above results, it is now possible to prove the existence of a pair $(\lambda_1^*, \lambda_2^*)$ with $\lambda_1^* \leq \lambda_2^*$ such that for every λ outside the interval $(\lambda_1^*, \lambda_2^*)$, the relative efficiency of sometimes proportional allocation with respect to proportional allocation is never less than a preassigned value $e_0 < 1$. This result is stated in Theorem 3.3.

Theorem 3.3. Let e_0 be a real number such that $0 < e_0 < 1$. Then $\exists \lambda_1^* \leq \lambda_2^* \ni e_1(\lambda, \theta_{21}) \geq e_0$ for every λ outside the interval $(\lambda_1^*, \lambda_2^*)$.

Proof: To prove the theorem, let θ_{21} be a fixed number greater than or equal to 1.

First consider the case when $\inf_{\lambda} e_1(\lambda, \theta_{21}) \geq e_0$. Then $e_1(\lambda, \theta_{21}) \geq e_0$ for every λ. If we take $\lambda_1^* = \lambda_2^*$ to be any real number greater than 1, then the theorem is true.

Now consider the case when $\inf_{\lambda} e_1(\lambda, \theta_{21}) < e_0$. Then for some values of λ, $e_1(\lambda, \theta_{21}) < e_0$. But $e_1(\lambda, \theta_{21})$ is decreasing when $1 < \lambda < \theta_{21}^2$. Also $\lim_{\lambda \to 1} e_1(\lambda, \theta_{21}) > 1$. It follows that $\exists \underline{\lambda} \ni e_1(\lambda, \theta_{21}) \geq e_0$ for every $\lambda < \underline{\lambda}$.

Let $L_1 = \{\underline{\lambda} \mid e_1(\lambda, \theta_{21}) \geq e_0 \text{ for every } \lambda < \underline{\lambda} \text{ and } \theta_{21} \geq 1 \text{ and fixed}\}$.

Clearly $\inf_{\theta_{21} \geq 1} L_1$ is the required λ_1^*. Q.E.D.

On the other hand, since $e_1(\lambda, \theta_{21})$ is increasing when $\lambda > \theta_{21}^2$ and $\lim_{\lambda \to \infty} e_1(\lambda, \theta_{21}) = 1^-$, $\exists \bar{\lambda} \ni e_1(\lambda, \theta_{21}) \geq e_0$ for every $\lambda > \bar{\lambda}$. Let

$L_2 = \{\bar{\lambda} \mid e_1(\lambda, \theta_{21}) \geq e_0 \text{ for every } \lambda > \bar{\lambda} \text{ and } \theta_{21} \geq 1 \text{ and fixed}\}$

clearly, $\sup_{\theta_{21} \geq 1} L_2$ is the required λ_2^*.

We shall now discuss the relative efficiency of sometimes proportional allocation with respect to modified Neyman allocation given by

$$e_2(\lambda, \theta_{21}) = V(\bar{y}_W)_{\underline{N}}/V(\bar{y}_W)_S = \frac{1}{D}\left[W_1^2 + W_2^2\theta_{21} + W_1W_2\, G\, \theta_{21}^{1/2}\right] \tag{3.2}$$

where

$$D = (W_1^2 + W_2^2\theta_{21} + W_1W_2\, G\, \theta_{21}^{1/2}) - W_1W_2\, G\, \theta_{21}^{1/2}\, I_{q_{21}}\left(\frac{f-1}{2}, \frac{f-1}{2}\right) - W_1W_2\,(1 + \theta_{21}) I_{q_{21}}\left(\frac{f}{2}, \frac{f}{2}\right).$$

The results concerning the behavior of $e_2(\lambda, \theta_{21})$ can be obtained in a similar manner and are given below.

Theorem 3.4. Let λ be an arbitrary but fixed number in the set $[0, \infty)$. Then $\exists\, \theta_0 \geq 1 \ni e_2(\lambda, \theta_0) = 1$ and $e_2(\lambda, \theta_{21}) \geq 1$ for every $\theta_{21} \leq \theta_0$.

Theorem 3.5. Let θ_{21} be an arbitrary but fixed number greater than or equal to $\frac{1}{2}[G^2 - 2 + G\sqrt{G^2 - 4}\,]$. Then $\exists\, \lambda_0 \ni e_2(\lambda_0, \theta_{21}) = 1$ and $e_2(\lambda, \theta_{21}) \geq 1$ for every $\lambda \leq \lambda_0$.

Theorem 3.6. Let e_0 be a real number such that $0 < e_0 < 1$. Then $\exists\, \lambda_1^* \leq \lambda_2^* \ni e_2(\lambda, \theta_{21}) \geq e_0$ for every λ outside the interval $(\lambda_1^*, \lambda_2^*)$.

4. Numerical illustration

For the purpose of illustration, consider the problem of sampling households in a town in order to estimate the average amount of assets per household that are readily convertible into cash. The households are stratified into a high-rent and a low-rent stratum. The variance σ_2^2 in the high-rent stratum is considerably larger than the variance σ_1^2 in the low-rent stratum. On the basis of preliminary evidence, it is guessed that $\theta_{21} \leq 9$. It is known that

$N = 24{,}000$, $W_1 = 5/6$ and $W_2 = 1/6$

N_1 and N_2 are sufficiently large, so that finite correction factors can be ignored. Further, let $f = 7$ and $\lambda = 2$. The table below gives the relative efficiency of sometimes proportional allocation with respect to proportional allocation as also with respect to modified Neyman allocation for different values of θ_{21}.

Relative efficiency of sometimes proportional allocation

With respect to	$\theta_{21}=1$	$\theta_{21}=3$	$\theta_{21}=5$	$\theta_{21}=7$	$\theta_{21}=9$
Proportional Allocation	0.99	1.02	1.10	1.18	1.26
Modified Neyman Allocation	1.014	0.998	0.995	0.997	0.998

It is seen that for appropriate choice of the level of significance as determined by λ (in this case λ=2), sometimes proportional allocation is almost as efficient as modified Neyman allocation. It is also seen that sometimes proportional allocation is almost as efficient as proportional allocation for values of θ_{21} close to 1 while it is considerably more efficient than proportional allocation for values of θ_{21} closer to 9.

Acknowledgement

This work was supported under OEC Contract No. 0-70-3655 by the U.S. Office of Education, Department of Health, Education and Welfare.

References

[1] Evans, W. D. (1951). "On stratification and optimum allocation," Journal of the American Statistical Association, 46, 95-104.

[2] Neyman, J. (1934). "On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection," Journal of the Royal Statistical Society, 97, 558-606.

[3] Sukhatme, P. V. (1935). "Contribution to the theory of the representative method," Journal of the Royal Statistical Society, Supplement 2, 253-268.

BENEFIT-COST ANALYSIS AND THE MODELLING PROBLEM

Gordon A. Welty, University of Pittsburgh and Temple University

In the following essay we address a point that appears germaine to the further development of benefit-cost analysis. After establishing the nature of the technique, we discuss the concept of <u>scope</u>. This is a property of a predicate such as net social benefit of a project or cost-effectiveness ratio; it specifies the entities for which the predicate is relevant. Considerations of scope lead us to a discussion of the modelling problem, where the nature of the social production function becomes a concern of the benefit-cost analyst.

Benefit-cost analysis is an application of microeconomics when a comparison is made of program variants when a direct comparison of the variants in the marketplace is impossible. While it may be plausible directly to compare the efficiency of two techniques of production in the automobile industry in terms established in the marketplace, it is less plausible to make this comparison, say, for agencies in the Department of Health, Education and Welfare. In a market, the rates of exchange of commodities and services of all sorts are transformed by an adjustment process from subjectively held ratios which express personal preferences or judgments into social ratios or prices. In essence, benefit-cost analysis attempts to find surrogates or "shadows" for these prices, for the case of the bureau.

More precisely, the technique measures direct and indirect costs of each program variant. If the cost of the variant in the first year is c_1, in the second year c_2, etc. for m years, and i is the (constant) interest rate, then the present value of future costs of that variant is given by

$$c = \sum_{t=1}^{m} \frac{c}{(1+i)^t} \ .$$

A similar calculation is made which estimates the present value of net benefits, both direct and indirect, of each project variant as well. If the net benefit of a given variant in the first year is b_1, in the second year b_2, etc., again for m time periods, then the present value of future net benefits is given by

$$b = \sum_{t=1}^{m} \frac{b}{(1+i)^t} \ .$$

The net social benefit is the difference between the present values of all benefits and all costs, b-c. The decision rule prescribes that alternative where the net social benefit is greatest, while still non-negative. As we have addressed issues surrounding the estimation of costs and benefits elsewhere [13], we will merely note that benefit-cost analysis assumes that the analyst can specify the entity called a program variant and its properties.

Thus we turn from the problems of benefit-cost analysis to the prior problem of the analytical model of microeconomic theory. Clearly this model is presupposed: there is an entity which transforms inputs into outputs, the evaluation of which is given by costs and benefits. What bearing does the typical approach to benefit-cost analysis, to proceed on the basis of more or less plausible assumptions where evidence is lacking, have on this implicit theory?

Suppose we have n factors comprising a theory. For classical mechanics, n would be relatively small, including only mass, location, and time. On the other hand, the number of factors in a theory of human behavior are manifold. A general theory of price formation, for instance, would include a vast number of factors, as Professor Krelle has recently noted [3, pp. 148-149].

In making a prediction P based on the required social theory, we typically argue from hypotheses H and initial conditions C:

$$\prod_{i=1}^{n} H_i \cdot C \rightarrow P \ .$$

By <u>modus tollens</u>, we identify and rectify any inappropriate elements of the theory, vis.

$$\overline{P} \rightarrow \overline{C} \vee \sum_{i=1}^{n} \overline{H}_i \, .$$

The question before us is that of the identification of the erroneous element(s). If such a theory is <u>not</u> rectified, the analyst will continually make incorrect recommendations for action. If the backwaters of the Aswan High Dam become the breeding place for malarial parasites, this will diminish the benefit stream realized from the dam. This sort of error appears to occur in most benefit-cost analyses. Clearly, the rejection status of a hypothesis representing a particular factor (or Grünbaum's "component of the total theory"[1]) H_j is indeterminate in this case. In fact, hypotheses about all n factors may be true, if the initial conditions C are not fulfilled. Let us suppose, however, that the initial conditions are known to be fulfilled. Then,

$$\overline{P} \cdot C \rightarrow \sum_{i=1}^{n} \overline{H}_i$$

where at least one of the factors (or hypotheses) is false. Of the conditions of such a theory are the so-called scope conditions [2;9]. These conditions specify the kind of entities for which the hypotheses are to hold. For instance, a social action program might be based on a learning theory for which the scope conditions exclude children with specific learning disabilities. A factor critical to the functioning of the program is "learning ability": it is crucial to realize that it is <u>hypothesized</u> that this factor is present. We will argue that the failure of such a project in a comparative benefit-cost analysis, may be due to the absence of "learning ability," due to the unfulfilled scope condition, rather than a failure of the project, curriculum, etc.

In the case mentioned, if all n factors are mutually exclusive (independent), let the (real valued) probability of hypothesis X on the basis of datum Y be $0 \leq Pr(X/Y) \leq 1$. Then,

$$Pr(\prod_{i=1}^{n} H_i/P \cdot C) =$$

$$Pr(H_1/P \cdot C) \times Pr(H_2/P \cdot C) \times \ldots$$

$$\ldots \times Pr(H_n/P \cdot C),$$

which is to say the conditional probability of a conjunction of hypotheses on given data is equal to the product of the conditional probabilities of each hypothesis on the given data. As we consider the case where the prediction was false, we have

$$Pr(\prod_{i=1}^{n} H_i/P \cdot C) = 0$$

or, alternately,

$$Pr(\prod_{i=1}^{n} H_i/\bar{P} \cdot C) = 1.$$

Now we can prove Professor Grünbaum's <u>fundamental epistemological conclusion</u> [1, p. 1067]. Consider, for a factor of interest H_j (Grünbaum's "main component"), that

$$Pr(H_j/\bar{P} \cdot C) = \frac{Pr(\prod_{i=1}^{n} H_i/\bar{P} \cdot C)}{\prod_{k=1}^{n-1} Pr(H_k/\bar{P} \cdot C)} \ ;$$

$$Pr(H_k/\bar{P} \cdot C) > 0, (k=1,2. \ . \ .n-1) \ .$$

The lower the conjoint conditional probability of the n-1 other factors (or auxiliary hypotheses with respect to H_j), which is to say the smaller the denominator of the right-hand expression, the higher the conditional probability of H_j in the face of a false prediction.

In the limiting case, suppose all of the auxiliary hypotheses were known true on the basis of independent supporting evidence; if they were thus conclusively <u>verified</u>, then

$$\prod_{k=1}^{n-1} Pr(H_k/\bar{P} \cdot C) = 1$$

and H_j is categorically <u>refuted</u>, which is to say $Pr(H_j/\bar{P} \cdot C) = 0$. In general,

$$1 - \prod_{k=1}^{n-1} (Pr\ H_k/\bar{P} \cdot C) > \epsilon \longrightarrow$$

$$Pr(H_j/\bar{P} \cdot C) > \epsilon.$$

$\epsilon > 0$ indicates an "irremediable inconclusiveness of the verification of [the] auxiliary component[s]" and implies "a corresponding limitation on the deducibility of the categorical <u>falsity</u> of the main component.[1, p. 1067]. Notice that as the threshold ϵ increases in magnitude, the possibility of falsifying H_j diminishes; hence the empirical status of the total theory vanishes in a murk of mystery.

Perhaps most frequently it is the social scientist who denies (in principle or practice) the conditions for Grünbaum's conclusion. While Professor Zeuthen, for example, explicitly accepts that "direct or indirect measurement (or the possibility of other factual testing) is a necessary condition for the avoidance of mystery," [14] others as explicitly reject the conditions. Let us examine this other viewpoint.

Of those subscribing to another viewpoint, the Austrian marginalist, Carl Menger, comes immediately to mind. Menger argued for a methodological distinction between pure economics and empirico-realistic economics. There was to be no empirical testing of the former, and of course there is little predictable in the latter [8, pp. 58-59]. More recently, Professor Machlup, in a well-known essay, proposes that the "postulate of rational action," which can be understood as "the fundamental assumption" of microeconomic theory, can be considered "as an idealization with constructs so far removed from operational concepts that contradiction by [empirical] testimony is ruled out." [4, p. 11] Thus, Machlup maintains that an economic assumption H_e postulating rational action has the conditional probability

$$Pr(H_e/\bar{P} \cdot C) = 1 \ .$$

Thus the rationality factor can become the basis of a scope condition: rational behavior is necessary for the applicability of economic theory.

Indeed, the probability of H_e is, for Machlup, <u>unconditional</u>; he goes on that

> the fundamental assumption is a resolution to proceed in the interpretation of all data of

observation as if they were the result of the postulated type of behavior. [4, p. 11]

That is to say, the assumption has a tautological implicate, as we find $Pr(H_e/\bar{P} \vee P) = 1$, which is guaranteed by the rules of valid inference alone. We must note the rather obvious point that if H_e bears any substantive weight whatever, on Grünbaum's argument the empirical content of microeconomic theory vanishes. The assumption surely does bear weight in typical economic theory, and on Machlup's argument is subject to no independent empirical verification.

But Machlup presents an alternative, by resorting to Verstehende Sozialwissenschaft. States he

> the fundamental assumptions of economic theory are not subject to a requirement of independent empirical verification, but instead to a requirement of understandability in the sense in which man can understand the actions of fellowman [4, p. 17; 5, p. 487].

It is unclear what such a requirement as this implies for the social sciences, other than a methodological dualism. Indeed, Machlup later appears to qualify his dualism, citing approvingly both Max Weber and Alfred Schütz in the qualification [6, p. 291].

We find that both Weber and Schütz explicitly recognized that, quite independent of the value of Verstehen to the framing of hypotheses,

> verification of subjective interpretation by comparison with the concrete course of events is, as in the case of all hypotheses, indispensible. [12, p. 97]

Thus Weber realized that empirical verification was necessary. Even in the ideal-typical case, a necessary condition for objective meaning is that the ideal type be "causally adequate."

> An ideal-typical construct is said to be causally adequate when it turns out to predict what actually happens, in accord with all the rules of frequency. [11, p. 233]

We can understand "the rules of frequency" to consist for Schütz in the conventional procedures of the statistical testing of hypotheses.

Thus it appears that in the last analysis, even the foremost proponents of Verstehen as method carefully qualify themselves to methodological monism. Which returns us to Grünbaum's conclusion. To date, no one has been able to avoid the implication of vanishing empirical content of a theory, when they seek to avoid the necessity of empirical testability of the components of that theory.

Let us now return to our supposition that the initial conditions C are known to be fulfilled. If they are not known true, hence are contingencies, we must explicitly consider the probability of the initial conditions Pr(C). From Bayes' Theorem, we have for any X, Y, Z,

$$Pr(X/Y \cdot Z) = \frac{Pr(X/Z) \times Pr(Y/X \cdot Z)}{Pr(Y/Z)}.$$

By appropriate substitution,

$$Pr(H_j/\bar{P} \cdot C) = \frac{Pr(H_j/\bar{P}) \times Pr(C/H_j \cdot \bar{P})}{Pr(C/\bar{P})}.$$

Let us examine the right-hand expression. Since C may be false (unfulfilled), we must concede that $Pr(H_j/\bar{P}) > 0$. Similarly, since C may be true (fulfilled), we must concede both $Pr(C/H_j \cdot \bar{P}) > 0$ and $Pr(C/\bar{P}) > 0$. Thus,

$$Pr(C/P) \times Pr(H_j/\bar{P}) \times Pr(C/H_j \cdot \bar{P}) > 0,$$

and $Pr(H_j/\bar{P} \cdot C) > 0$ is guaranteed. In short, if we relax the supposition that initial conditions or scope conditions are known true, no hypothesis can be categorically refuted. This corresponds to our earlier suggestion that all n factors may be true. (It is readily apparent that the same conclusion follows if we relax the implicit supposition that the truth value of P is unequivocal.) As Sir Peter Medawar has emphasized,

> The act of falsification is not immune to human error. [7, pp. 53-54]

On the one hand, alternatives to empirical testing are as yet a program. On the other hand, it is naive to assume that falsification alone can be a comprehensive methodological basis for the social sciences. It would appear that a more subtle view of the nature of hypothesis testing, one recognizing the potential epistemic parity of several conjoined hypotheses, must be taken by the investigator. What does all this portend for benefit-cost analysis?

An even fleeting acquaintance with benefit-cost analyses shows that they are a conglomerate of assumptions. We are immediately faced with the circumstance that initial conditions are not known to be true. Then the epistemic parity of the constituents of the theory arises.

Under such circumstances, it is likely that program variant A will be chosen over variant B because rates of transformation in social production for A appear more favorable than for B, on the basis of false assumptions about A (or B). Instead of the false assumption being rejected, the better project is rejected. This becomes particularly acute when outputs are inferred from inputs, frequently the case in bureaus. The solution to this problem appears to lie in the careful specification of the social production function [10, pp. 56-64] for each project variant prior to benefit-cost analysis.

REFERENCES

[1] Grünbaum, Adolf, "Can We Ascertain the Falsity of a Scientific Hypothesis?" Studium Generale, Vol. 22 (1969).

[2] Hempel, Carl, Aspects of Scientific Explanation, Free Press: New York (1965).

[3] Krelle, Wilhelm, "Economic and Non-economic Factors of Price Formation," Universitas, Vol. 12 (1970).

[4] Machlup, Fritz, "The Problem of Verification in Economics," Southern Economic Journal, Vol. 22 (1955).

[5] Machlup, Fritz, "Rejoinder to a Reluctant Ultra-Empiricist," Southern Economic Journal, Vol. 22 (1956).

[6] Machlup, Fritz, "If Matter Could Talk," Morgenbesser et al (eds), Philosophy Science and Method, St. Martin's Press: New York (1969).

[7] Medawar, Peter, Induction and Intuition in Scientific Thought, Vol. 75, Memoirs of the American Philosophical Society: Philadelphia (1969).

[8] Menger, Carl, Untersuchungen über die Methode der Socialwissenschaften und der Politischen Oekonomie insbesondere, Dunker and Humblet: Leipzig (1883).

[9] Quine, Willard Van Orman, Methods of Logic, Holt, Rinehart and Winston: New York (1959).

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

[11] Schütz, Alfred, Phenomenology of the Social World, Northwestern University Press: Evanston (1967).

[12] Weber, Max, Theory of Social and Economic Organization, Oxford University Press: New York (1947).

[13] Welty, Gordon A., "Cost-Benefit Analysis as a Guide to Action in Water Resource Projects," Highlights 1970: Congress for Recreation and Parks, National Recreation and Park Association: Washington, D. C. (1971).

[14] Zeuthen, Frederik, Economic Theory and Method, Cambridge, Massachusetts: Harvard University Press (1955).

MINUTES OF THE ANNUAL MEETING OF THE SOCIAL STATISTICS SECTION

Detroit, Michigan, December 28, 1970

The meeting was opened by Otis Dudley Duncan, the Chairman-Elect, at 5:30 p.m. The Chairman, Daniel O. Price, was not able to attend. As a result of the recent elections, the list of officers for 1971 is:

Chairman	Otis Dudley Duncan
Chairman-Elect	Eva Mueller
Vice-Chairman	Philip C. Sagi
Vice-Chairman	Charles B. Nam
Secretary	Regina Loewenstein
Section Representative on the Board of Directors	John D. Durand
Section Representative on the Council	Hubert Blalock
Publications Liaison Officer	Karl E. Taeuber

Edwin D. Goldfield was reappointed Editor of the Proceedings.

Two changes in the charter were approved. The Publications Liaison Officer will be included in the list of officers and the term of office will be two years. The Editor of the Proceedings will be appointed by the Chairman for one year and can be reappointed.

Karl Taeuber will function as Program Chairman for the Annual Meeting in August 1971.

Based on her experiences as Program Chairman for the 1970 Meeting, Eva Mueller made a few suggestions. The American Statistician asks for suggestions for program topics too late to be useful. She suggested that the request be published earlier or be omitted.

She also suggested that contributed papers be screened in order to reduce the number of papers. Margaret Martin said that it would be difficult to screen contributed papers, and recommended that all papers be accepted and more sessions of contributed papers be held at the same time. Other suggestions with regard to this problem were presentation of papers by title only or printing the exact time for each paper in the program.

Suggested topics for the program for 1971 included The National Health System, Drug Abuse, National Assessment of Educational Progress, Environmental Data Banks, Legal Enforcement Statistics, Federal Statistics Commission, and Training Statisticians to Work with Social Scientists.

The Program Chairman should request speakers to prepare handouts that include graphs, tables, mathematical expressions and/or parts of the text. Blackboards and projection equipment should not be used.

A special event, such as a luncheon, can be arranged by the Chairman of the Section.

1970 Officers of the Social Statistics Section

Chairman	Daniel O. Price
Chairman-Elect	Otis Dudley Duncan
Vice-Chairman	Eva Mueller
Vice-Chairman	Philip C. Sagi
Secretary	Regina Loewenstein
Section Representative on the Board of Directors	John D. Durand
Section Representative on the Council	Hubert Blalock
Publications Liaison Officer	Denis F. Johnston
Editor of Proceedings	Edwin D. Goldfield