I-A

THE 1970 CENSUSES OF POPULATION AND HOUSING: NEEDS FOR THE 1970 CENSUSES

Chairman, ROBERT C. KLOVE, U. S. Bureau of the Census

Page

THE 1970 CENSUS: NATIONAL USES--CHALLENGE AND OPPORTUNITY

Otis Dudley Duncan, University of Michigan

Our census evolves in response to a complex of internal and external forces whose changes guarantee that no decennial enumeration will merely replicate its predecessors. By "internal" forces I mean the technical and technological developments that permit us to do old jobs better or more efficiently and to take on new jobs which were hitherto inconceivable. Much of today's discussion will doubtlessly include reference to innovations in geographic identification, sampling, field procedures, and automated data processing, and I shall not elaborate on these matters here. My concern is, rather, with the "external" forces which determine, not what kind of a census is economically and technically feasible, but what kind of a census we want--not to overlook the important interaction between our desires and our images of the possible.

Among these external forces, I suppose, are the very constitutional requirement for the head count itself; legislative and administrative determinations that certain information shall be collected; demands generated by various governmental, commercial, and academic interests; and the tradition of scientific census taking, as it is elaborated by its practitioners around the world and in our own statistical offices. The Bureau of the Census makes every reasonable effort to ascertain responsible opinion on what the enumeration should include, and reaches its final decision in the context of a variety of constraints, the nature of which is not the subject at hand. The point which is relevant here is that changes in all these forces--perhaps not even excluding the constitutional provision, if we think in terms of its interpretation and implementation--issue periodically in a new package of content and procedures. We confront not only a revolution in methods of taking the census, owing to technical developments, but a continuous evolution of our concepts of what its purposes are and should be.

I should like to emphasize three trends which seem to be shaping our orientation to the task of planning the 1970 Census in regard to distinctively national needs. Presumably there is something to be gained by facing as explicitly as possible the implications of such trends. I shall try to be specific about some of the implications of the following trends:

First, we are rapidly developing a much more crystallized commitment to the goal of equal opportunity for all citizens.

Second, we discern more and more clearly the necessity for prompt and efficient adjustments to rapid social change, and appreciate more profoundly the role of statistical intelligence in effecting such adjustments.

Third, the nascent concept of a statistical system forces us toward much bolder ideas for the operational integration of census statistics with other bodies or sources of data.

My task, then, is to suggest some issues that arise from a recognition of the bearing of these three trends on planning for the Nineteenth Decennial Census.

Statistics of Opportunity

By and large, there is little hope of securing in a census direct measures of effective opportunity as such or of inequality of opportunity. Moreover, the statistics themselves do not record the overt or covert denials of opportunity which it is the object of national policy to eliminate. What we must do is infer lack of opportunity or discriminatory denial of opportunity from variations in magnitudes that presumably would be or become equal if opportunities were in fact available and equal. It appears that even the diagnosis of inequality--not to mention the design of remedies--depends on inference from the observations, and not on the mere summarization of the observations themselves. The inference characteristically involves the methods of multivariate analysis. To show that there is discrimination in the housing market, you have to demonstrate (for example) that areal distribution of residences differs from what it would be if it were solely a function of ability to pay; or that the cost of housing varies systematically among social groups apart from variations in the quality of housing obtained. While instances of gross discrimination can be detected (or at least strongly suspected) on the basis of crude analysis, the specifics of the incidence and magnitude of discrimination may be estimated only after painstaking manipulations on whole sets of variables.

Census statistics, therefore, are or may become relevant to the problem of equal opportunity under the three conditions, (a) that they provide adequate measures of outcomes presumed to reflect opportunity or its absence, (b) that they provide a sufficient range of variables, the analytical control of which is indispensable in making an inference of unequal opportunities or discriminatory variation in access to opportunity, and (c) that these requisite data are available for the relevant population groupings.

On the last point, that of the relevant groupings, the criteria are reasonably clear in general terms and have in fact been written into the legislation expressing our determination to remedy the unequal availability of opportunities for individuals "by reason of race, color, religion, or national origin." To know whether individuals distinguished by race, by color, by religion, or by national origin, enjoy more or less of the fruits of opportunity, you have to classify individuals by race, color, religion, and national origin.

Although the movement to delete indications of race and color from statistical records seems a little less threatening now than it was a few years ago, no opportunity should be lost to point out to the partisans of racial equality that their cause and that of a nation committed to equality of opportunity is best served by having full information on present social and economic differences by race and changes therein.

In regard to color, as you know, census statistics are very frequently tabulated and presented for the two mutually exclusive and exhaustive categories, white and nonwhite. I, for one, hope that the 1970 Census will mark the demise of this practice. The supposition that the residual category of "nonwhite" is a homogeneous one is not seriously entertained by anyone; it is well understood to be a "heterogeneous classification used by the census to simplify its tabulations at the cost of providing confused information."[1] For the nation as a whole and for the North and South, "nonwhite" is but a rough approximation to "Negro." For the West, it is a hodge-podge. If a two-column presentation is all that can be afforded, let it be "total" and "Negro." The implied residual, non-Negro, will, to be sure, not be homogeneous either, but at least the disturbance will have less nuisance value when buried in the modal racial category. A better solution still, of course, would be to exercise some ingenuity to the end that each minority race with some minimal representation in an areal unit or other statistical category will be shown wherever there is a determination that race or color is a relevant item of tabulation.

We should anticipate that it will indeed be relevant in many places where it has often been disregarded in the past. An example that recently came to my attention is that of detailed occupation by detailed industry. To find such a table for Negroes, one must go back to the Census of 1930. Yet there is every reason to suppose that the removal of barriers to occupational mobility proceeds unevenly by industries, and every reason to want these particular statistics to pinpoint the sectors where business and labor organizations need to be stimulated to assume their proper role in the battle against discrimination. Admittedly, this particular tabulation would require utilization of the full 25 per cent sample, which was used to collect labor force information in 1960 and will presumably be so used again in 1970. Moreover, equally detailed tabulations could perhaps not be justified for the smaller minorities.

Yet we have to keep in mind a cardinal principle that applies in respect not only to minorities defined by race, religion, or ethnic category, but also statistical minorities of any kind--rare occupations, the very poor, single-parent families, or residents of mining communities, for example. The principle is that only a complete canvass of the population can locate enough of the individuals in such minorities to provide an adequate statistical basis for detailed tabulations. If the census does not provide such tabulations, they will not exist. We have known for two decades that most national aggregates can be estimated more reliably from a sample survey than from the decennial count. It is overstating the case only a little to maintain that we need a full census only because we require information on the many small minorities defined by political boundaries, statistical categories, and social groupings.

The issue just raised merely concerns a modification in procedures with respect to content which is already traditional. But I remind you that our concern with equality refers not only to race or color, but also to religion and national origin.

It is true, of course, that we have statistics of a sort on national origin for the segments of the population identifiable by foreign birth or parentage. Since 1930, these statistics have received a progressively diminishing relative emphasis, in view of the declining proportions of the foreign stock in our population and in violation of the principle just enunciated with respect to the identification of minorities as the <u>raison d'être</u> of a census. The foreign stock will still be amongst us in 1970, albeit in reduced proportions, and should receive even more careful attention than in the last two or three censuses. Yet there is a manifest need for data on the national origins of persons born in this country of native parents.

A variety of indications suggest that ethnic or national-cultural differences continue to serve as a significant axis of social structure and to operate, for all we know to the contrary, to limit some kinds of opportunity. Political analysts and professional politicians are convinced that certain nationality groups play distinctive roles in the political process. Measures of "social distance," which sociologists are fond of contriving, reveal impressive stability in comparisons of recent studies with those of several decades ago. Ethnic cultural organizations and voluntary associations continue to lead a vital existence. In many cases the surname, not only among those of Spanish-American extraction, is used by the individual himself and by his associates as an explicit basis of ethnic identification. Despite the general success of the "melting pot" in reducing variance in respect to certain cultural indicators, there remain groups in this country who retain the use of a second language.

Our information on these and other kinds of persisting ethnic identification is spotty and unsystematic, and this situation is not likely to improve in the absence of comprehensive statistics classifying the whole population or a major segment of it in terms of ethnic identification on a somewhat consistent basis. What that basis should be, in specific operational terms, is not entirely clear, since the matter is a complex one. Every obvious suggestion encounters apparent practical difficulties. For example, a question on birthplace of grandparents is subject to the same ambiguities produced by shifting national boundaries that have plagued our traditional statistics on parentage.

Without minimizing the practical problems, I want to take the position--at least for purposes of discussion--that it is less important to have an "ideal" basis of ethnic classification than to have some basis that can at least be replicated. There is much to be said for just a direct question on national origin, after the fashion of the Canadian census. It will be granted immediately that the question will appear ambiguous to some and that not all responses can be expeditiously coded. I would then argue for putting aside such answers as "mixed" or "American" in tabulating by national origin on the assumption that the more straightforward responses are likely to represent reasonably clear ethnic self-conceptions of respondents. If you respond that the census is no place to institute a study in the social psychology of ethnic identity, I would counter that social facts can be nonetheless real, powerful, and important for being inescapably vague. We accustom ourselves to some kinds of vagueness when it becomes apparent that vague concepts may actually have a good deal of predictive validity.

I come finally to the question of religious identification. In this case, there is no longer any doubt that meaningful data can readily be collection.[2] We have firm evidence, moreover, of striking variation over religious groups in socioeconomic status and demographic behavior. The civil rights legislation, as noted above, is fully explicit about the nation's commitment to equality of opportunity for religious as for ethnic and racial groups. The case for a simple question on religious preference of the kind that has already been adequately tested is, as far as I can see, compelling.

Statistics of Movement

We used to think of the census as a sort of periodic stocktaking or social inventory. In fact, during the period when I was learning the rudiments of demography we were frequently exposed to the canonical dictum that the census furnishes the stock data and vital statistics the flows. In fact, the census had long since departed from a strict commitment to the enumeration of stocks, and that concept was only implicit in its procedures anyway. I do not know if anyone has taken the trouble to list all the kinds of census items that represent changes of status or condition over time or the processes by which stocks are accumulated and depleted. The emerging new principle, it appears, is that, where feasible, the census must be pressed into the service of providing information on flows or movements in the event that other statistical mechanisms do not suffice.[3]

I want to interpret the concept of movement broadly, to subsume not only changes of location in geographic space but also alterations of significant social positions, such as occupational role, marital status, or stage in the family life cycle. Apart from the resources of record linkage, mentioned subsequently, there is a fairly severe technical limitation on the ability of the census to supply such data imposed by the fallibility of retrospective reports. There is reason to hope, nevertheless, that a useable level of reliability can be secured for such items as the following (which include the principal innovations in this domain that I wish to urge):

- activity status five years ago (at work, in school, in Armed Forces, other);
- occupation and industry at that time for those at work;
- year of entry into the United States for the foreign born;
- dates of birth of first, last, and next-to-last child;
- dates of first marriage and its termination and of entering current marital status;
- residence classification (farm or nonfarm) of place of birth; and
- place where last attended secondary or elementary school.

Each of these items can be justified by a variety of analytical uses familiar to specialists in relevant fields. Let me, without attempting to summarize these justifications, indicate some general considerations which argue for this substantial expansion of efforts to secure mobility data.

In the specifically demographic field, there is a considerable body of evidence that patterns and shifts in patterns of timing of vital events may be as important for understanding contemporary changes as is the detection of long-run trends in propensity to marry, say, or ultimate size of completed families. Similar kinds of evidence point to hitherto unsuspected cohort effects in regard to geographic mobility. Adequate understanding of the complex phenomena generated by variations in timing requires the juxtaposition of several items of information describing the history of real cohorts. To only a limited degree can this strategy be effected by intercensal comparisons of cross-section data or by accumulation of information on currently registered events. Indeed, it is the questions raised by these partial strategies that clamor for answers which can only be had by assembling longitudinal data on individuals.

A second general consideration is our growing appreciation of the phenomenon of persistence. On the one hand, we are dismayed when pockets of poverty and social pathology persist from decade to decade, sometimes in spite of concerted effort to eliminate them. What we do not know about such persistence is how much it depends on the immobility of the human factor itself, or, by contrast, on the persistence of environmental causes which operate similarly on whatever human material is at hand and despite considerable turnover in that material. I suppose some minimum degree of persistence is implicit in the very concepts of social or spatial structure. But we shall find it hard to understand, let alone modify, the less desirable kinds of persistence until we can separate them into components of mobility and immobility.

Another aspect of the phenomenon of persistence is the stamp of early experience on later fortunes and performance. We know that people can move in social or physical space but yet carry with them propensities engendered in their places of birth or rearing. The unexpected role of farm background in the persistence of group differences in fertility illustrates the potential importance of various sorts of "background" measures that we may hope to obtain--i.e., indicators of earlier experiences that may help to explain current status or condition. To take another example, students of the economics of education are persuaded that the effects of the place in which a person's education was attained persist in the form of differential handicap or advantage--or "returns to education," in their jargon--over long periods of the life cycle. A question on place (it might be difficult to identify "place" more precisely than state or foreign country) where elementary or secondary education was completed, moreover, would be useful not only for this reason but also because it would give us a baseline for measures of geographic mobility preferable in many ways to those provided by place of birth, while the comparison of the two would be very instructive in itself. The reasoning behind the suggestion of year of entry into the United States for the foreign born may also be given in terms of the notion of persistence. Presumably, the modifiable characteristics that immigrants bring with them will actually be modified in some direct relationship with amount of time spent in this country. Hence intergenerational comparisons between characteristics of immigrants and their children should be standardized for the length of time the immigrant generation has lived in the United States. With the item on year of immigration, moreover, one could establish with small error the age at immigration, so that the distinctive problems of those immigrating as mature adults could be studied in comparison with those entering the country as children or adolescents.

A further consideration, which may be offered without pretending to exhaust the reasons for greater emphasis on all kinds of mobility statistics, is that both the private and the public sectors of our society must be increasingly preoccupied with mechanisms of adjustment to short-run changes of major proportions. On the level of manpower analysis, I was impressed by a recent memorandum prepared by George Stolnitz at Resources for the Future which indicated the potential usefulness of a composite inter-industry inter-region labor mobility matrix. Such a matrix could readily be generated by a cross-classification of present residence and economic activity by residence and economic activity five years ago. Three of the four elements in this classification already are standard in the census, and with the addition of activity five years ago we should no longer have to wonder how efficient is geographic mobility in maintaining or improving a match between job skills and job opportunities. Such data would even offer an entering wedge into the issue of distinguishing chronic from transitory poverty, an issue on which public programs seem already to have made an assumption but one, we should hope, which could be modified if improved data so dictated.

There is a quite parallel justification for the suggestion that we enlarge the quantity of retrospective information on fertility, family formation, and family dissolution so as to secure more precise indications of the timing of changes in family cycle patterns. The current uncertainty as to the meaning of recent changes in birth rates is symptomatic of the limitations on our understanding of how the population adjusts its family behavior to short-run cycles and longer swings in the economy. We are, however, beginning to appreciate the distinction between two aspects of family planning--control of ultimate completed size of family, and control of the spacing of births. To the extent that these respond differently to social and economic changes, we shall require information enabling an analytical separation of them if we are to infer their respective causes and trace their effects.

Statistical Systems and Record Linkage

I am aware of two distinct, though related, justifications for a more systematic and comprehensive approach to record linkage than has seemed possible in the past. The first is that non-census record systems may contain the very information we are now attempting to obtain by census questions, and presumably in a more reliable form. For a large number of wage and salary workers, for example, the place of work might be ascertained more accurately and precisely from employer personnel records than from the interview or questionnaire response of the employee himself, and this information might more readily be recorded than the person's own response. The same records could be expected to include a more accurate indication of the occupation of employment than the respondent will be able to report, although there would surely be a problem of reconciling the discrepancies in occupational nomenclature used in different establishments. To take one more example, we suspect that earnings from covered employment and total income are reported more reliably in the Social Security and Internal Revenue Service records, respectively, than in the census. While it may be visionary to suppose that we can foresee the actual dropping of the present census questions on earnings and income, it is not too early to contemplate the use of these sources to provide very significant checks and supplements to the census data.

The second argument for record linkage, of course, is that various record systems contain data which, in principle, are unsuited to collection by interview or self-enumeration. During the present decade, we have seen important examples of record-matching studies of such phenomena as mortality, mental illness, and juvenile delinquency. These could well be regarded as the pilot studies for a new, regularized branch of census operations. The Census Bureau itself, of course, has carried out on its own account various matching studies for purposes of quality checks.

Not to be overlooked in the spectrum of possibilities in this field is that of matching between different sets of information collected by the Bureau. For example, the chance of matching CPS interview records collected in mid-decade with the reports on the same respondents in 1970 should not be missed, if we are serious in making a start on some of the kinds of mobility statistics advocated in the earlier discussion. Looking ahead, if a reliable and relatively inexpensive matching technique can be devised, intercensal record linkage--as, say, between the Census of 1970 and the Census of 1975 (should there be one) or 1980 --could begin to provide an approximation to the population register for which American demographers have often envied some of their European colleagues.

The important possibilities of linkage between systems of vital and civil registration have been convincingly demonstrated in Canada.[4] In this country, we should be well advised to think at the outset of the census as the core resource in such a program.

These proposals, for which I have given only the sketchiest rationale, are obviously much less definite than the recommendations that can now be made for modest innovations within the established tradition of census taking. They clearly require much study and experimentation, but not necessarily a postponement of implementation on that account. Those knowledgeable about the kinds of pilot studies that have already been made can convey more than I about both the promise and the problems of statistical systems generated by record linkage. My impression is that we are close to being reasonably confident of one possible mechanism, which involves inclusion of the social security number on the census schedule as a complete-count identification item, along with name, household relationship, and date of birth. Since this number is already being used in many record systems, and surely will come to appear in many more, the possibilities enumerated above are no longer merely hypothetical. The conversion of the census into a population register may alter it almost beyond recognition, but it has already been argued that such an alteration is a pragmatic response to life in a complex society with its strident demand for more and more elaborate quantitative information.

Obiter Dictum

In summary, I have offered for consideration three major fronts on which we should look forward to an expansion in the scope, detail, and systematic character of the data we can expect from the census. I tried to suggest that the national commitment to the goal of equal opportunity, the need for data on flows and movements, and the desirability of broadening the coverage of the statistical system whose core is the census all represent demands placed upon us by the exigencies of social change and the accumulation of our scientific knowledge and statistical technique. While an indefinitely large number of _ad hoc_ suggestions for changes in the census can be produced, it seems important to have some kind of general rationale for the kinds of changes to receive high priority. Whether such a rationale is implicit in the considerations advanced here is open to discussion, and it is this kind of discussion I would like to see generated by my presentation.

There remains one point on which there may be justifiable fear that such suggestions as are here advanced will founder. In many quarters there is a concern that any expansion, of whatever kind, in the scope of the census threatens civil liberties in the form of what is called an "invasion of privacy." To the extent that this anxiety rests on sheer irrationality, no quantum of information nor cogency of argument can allay it. On the other hand, the argument for such a concern, if advanced by a rational man, can be countered by rational means. In the counter-argument, as I see the matter, two points are cardinal.

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

Second, to the extent that direct relationships of the federal government with the individual threaten the latter's privacy, the invasion has already gone much further in nonstatistical fields than it could conceivably go in a statistical system as such. The resistance to invasion of privacy will be misplaced if it comes to a focus in an attack on statistics instead of the actual places where such invasion occurs. As we all know from personal experience, not only the government, but various private and commercial establishments bear watching in this connection.

We must be eternally vigilant to maintain the safeguards of confidentiality in the statistical system. But we should lend no endorsement to the mistake of reading the intensity of that vigilance as a symptom of any actual threat to civil liberties posed by the kind of statistical system which a modern society must have.

Footnotes

1. Frank Lorimer and Dorothy S. Jones, "The Demographic Characteristics of the Negro Population in the United States," _Journal of Negro Education_, XXII (Summer, 1953), 250.

2. Dorothy Good, "Questions on Religion in the United States Census," _Population Index_, XXV (January, 1959), 3-16.

3. Donald J. Bogue, "The Quantitative Study of Social Dynamics and Social Change," _American Journal of Sociology_, 57 (May, 1952), 565-68.

4. Howard B. Newcombe, _et al_., "Automatic Linkage of Vital Records," _Science_, 130 (1959), 954-59.

ABOUT STATE AND REGIONAL DATA NEEDS*

Werner Z. Hirsch, University of California, Los Angeles

Introduction

Not long ago Charles Schultze, Director of the Bureau of the Budget, pointed to the newly emerging importance of subnational data, stating "New programs initiated to carry out the objectives of the Great Society must be solidly grounded in factual information. The national effort to raise educational levels, to increase employment, to wage war against poverty and crime, to improve transportation and housing facilities—naming only some of our objectives—requires data not now available. It also requires to a much greater extent than ever before, data on a state or local area basis."[1]

In this paper I will discuss some issues involved in furnishing certain data to aid public officials in making state and regional decisions. No attempt is made to identify state and regional data needs for all users. Instead, the main concern is with governmental decisions. I was asked to confine myself to the Census of Population. To a limited extent, some liberties will be taken to go beyond this mandate in discussing ways of collecting new data, including nonpopulation data.

What then is the scope of this paper? At the outset I will explore one way in which state and regional data needs can be identified in a systematic manner. Then the framework will be applied to an example. An examination of ways to supply needed information reveals that the Census of Population already provides much of what is needed; but certain data not now collected would be more readily acquired by integrated recurring surveys than by the current, decennial enumerations. The desirability, scope and feasibility of such survey efforts are then explored.

A Framework for Identifying State and Regional Data Needs

The Bureau of the Census has long shown a deep concern for serving its clients well. Again and again, and especially in preparation for a new decennial census, the Bureau has made an all-out effort to solicit advice on what new and additional information might be needed. A particularly concerted effort was undertaken in the late fifties when the Bureau joined forces with Resources for the Future and sent out questionnaires asking potential users to indicate what new and additional regional information they would want.

It is fair to say that most, if not all, of these efforts have not produced many helpful guides. The main reason is that a shotgun approach was used, and even under the best circumstances, only expressions of highly personalized needs at any one particular moment of time could be obtained. Thus, even if researchers could carefully identify their regional data needs, the Bureau would still face the problem of verifying their validity and ranking their importance and priority. But apparently, in recent years, the Bureau has not been placed in a position where it had to take this second step too often, and there is little evidence that it has a systematic way to cope with the problem. The question therefore arises whether we cannot devise an alternative approach that is more promising.

The approach I suggest is, in a sense, intended to work backwards.[2] It starts by asking, "What are the important decisions to be made with regard to regions and states?" Once these decisions are identified, the next question should be, "What decision rules will be applied to making these decisions?" In a very real sense, the decisions, together with the decision rules, should provide the "demand" for, and establish the character of, the regional and state information needed.

Let us consider this issue in relation to regional public decisions. To decide on relevant data, we must determine in advance what issues they are designed to elucidate and what rules will be used to arrive at a decision. Public bodies need regional information to plan for the future, and to select and operate preferred programs. In planning for the future, governments look for large discrepancies or imbalances that are likely to occur. If, for example, costs are likely to exceed receipts by a wide margin, governments face a problem. If on the other hand, benefits of a program are likely to greatly exceed costs, they face an opportunity. They also look for major changes in magnitudes of variables over time, secular or cyclical, as an indication of major stresses and opportunities.

In formulating programs, i.e., selecting preferred solutions, governments have a set of at least two rules. They are concerned about economic efficiency and therefore want to examine alternative programs in terms of their relative net social benefits. But a government official, particularly an elected official, will not necessarily pick a program because it promises a very large net social benefit. While this might be an essential condition, he insists on another, namely, that the solution be consistent with his political survival. Specifically this means that

---

* Paper prepared for a presentation to the 126th Annual Meeting of the American Statistical Association held at Los Angeles, California on August 15, 1966. The paper benefited from help and advice from Sidney Sonenblum, John M. Kavanagh, Morton J. Marcus and Lawrence H. Revzan.

1. As cited in "Decision-making and Statistical Standards," by James E. Webb, The American Statistician, December 1965.

2. These issues will be discussed in much greater detail in a forthcoming volume, Design of Regional Information Systems.

he will want to know how the losses and gains are likely to be distributed among different interest groups. Detailed information is therefore needed on both social benefits and costs to determine their distribution.

Within such a screening framework, it becomes possible to identify significant data needs and hopefully order them according to their importance. To the extent that this approach succeeds, it should prevent the Census from becoming a data dump when it should be a data bank.

Identifying Some State and Regional Data Needs Within a Decision Framework

I will now apply the framework sketched out in rudimentary form in the previous section to an example, keeping in mind that the Bureau of the Census is mainly concerned with providing raw data and my main task is to reflect on the Census of Population. Therefore I will explore in some detail the nature of population data needed for state decisions on education. Data needed for estimating indirect benefits is not included, and since space is too limited to reproduce the steps that were taken, only the end results are given.

Some of the most important population information I was able to identify as being needed for education decisions, is presented in Table 1, where it is grouped under the following headings: population, households, income, consumer expenditures, leisure and special services. Admittedly this grouping is arbitrary and at one extreme could have been reduced to two categories—population and households, and at the other extreme to a dozen or more categories. In any case, as can be seen particularly from Table 1, there are many cross-classifications.

Most of the population data for education decisions should be available on a city block, census tract, municipal, county, school district and state level.

We also could have produced a table of select population data needed for state and regional water decisions. It would be quite different. For example those who make decisions about water need much less detailed population information than do those facing education problems. Clearly the demand for education is more closely related to age, sex, race, income, etc. than is the demand for water. Furthermore the population data for water decisions should be available on a municipal, county, utility district and state level.

As we compare the information requirements stated in Table 1 with the information in the 1960 Census of Population we find the Census could serve the regional decision maker reasonably well. There are exceptions, however, among them, consumer expenditures, leisure, and specialized services information. It might be possible for the Census to provide some of the additional data needs, but it appears that much of the additional information will require special surveys. For example, I cannot readily imagine the decennial census questionnaire burdened with questions designed to gain information about time spent in recreation participation per week, by age, income group, race and type of recreation. Yet this is crucial information for state and regional recreation decisions. More about the need for special studies will be set forth below.

Another, all too obvious, point should be made. We need up-to-date and recurrent information, often in greater detail, and especially on income, age and education grouping. A further important data gap pertains to detailed migration data as well as intra-area circulation data. Finally, we are often in need of longitudinal data which the Census neglects, except in very rare cases.

I would like to make sure that there is no misunderstanding about the fact that state and regional public decisions require more than population data. Other data needs might be grouped under the headings of manufacturing, commerce and service; policies and attitudes; public service; revenues and expenditures; and property, assets or wealth data.

My assignment excludes a review of the Census of Manufacturers and the Census of Governments. However, I am sure that my earlier discussion about government decisions on education and to a lesser extent on water will prompt readers to ask whether the data identified there are not available in the Census of Governments. First of all let me reply by reminding you that the structure of the Bureau of the Census is such that a given division is concerned with data pertaining to a given phenomenon, i.e., housing, governments, manufacturers, etc. These divisions are not designed mainly to produce information needed by the particular group that is surveyed; and this is not necessarily a bad arrangement. However, to come back to the initial question of whether the present Census of Governments furnishes much of the information needed by public bodies to make state and regional decisions, the answer is an emphatic no. A careful review of the Census of Governments produces the distinct conclusion that much of the information that it presently collects is only in a minor way useful to government decision making. For example, the Census of Governments is mainly looking at administrative budgets which, for decision making, often hide more than they reveal. It would be very useful for the Census of Governments to examine, for example, the desirability of helping to generate programmatic information on governments.

Finally, let me turn to wealth data, not often collected, but found so important by decision makers. The Bureau of the Census could perform a most significant service if it would collect information on the location, physical characteristics, value, utilization, etc. of water, recreation, transportation, education, housing and other physical facilities every 5 or 10 years. For example, in relation to water systems, it would be useful to distinguish between public and private facilities in the following areas: dams, reservoirs, catch and debris basins, aqueducts, drainage facilities, canals, pumping facilities, reclamation plants, desalinization plants, filtration plants, etc.

There is virtue in collecting many of the enumeration data every 5 instead of 10 years since much of the information needed for decision making tends to be of relatively little use if it is old. Decisions have to be based primarily on

Table 1

Select Population Data Needs for State Education Decisions

1. Population
    (a) total numbers
        i. age
            (1) 0-4
            (2) 5-12
            (3) 13-17
            (4) 18-24
            (5) 25-65
            (6) over 65
        ii. sex
        iii. race
            (1) white
            (2) Negro
            (3) other
        iv. income
            (1) 0 - $4,999
            (2) $5,000 - $8,999
            (3) $9,000 - $11,999
            (4) $12,000 - $14,999
            (5) over $15,000
    (b) density rates
        i. central city areas
            (1) age
            (2) race
            (3) income
        ii. suburban areas
            (1) age
            (2) race
            (3) income
        iii. rural areas
            (1) age
            (2) race
            (3) income
    (c) migration rates
        i. age
            (1) 0-4
            (2) 5-12
            (3) 13-17
            (4) 18-24
            (5) 25-65
            (6) over 65
        ii. sex
        iii. race
            (1) white
            (2) Negro
            (3) other
        iv. income
            (1) 0 - $4,999
            (2) $5,000 - $8,999
            (3) $9,000 - $14,999
            (4) over $15,000
2. Households
    (a) total number
        i. families with children ages 1-18
        ii. income class
            (1) 0 - $4,999
            (2) $5,000 - $8,999
            (3) $9,000 - $11,999
            (4) $12,000 - $14,999
            (5) over $15,000
    (b) migration rates
        i. families with children ages 1-18
        ii. income class
            (1) 0 - $4,999
            (2) $5,000 - $8,999
            (3) $9,000 - $11,999
            (4) $12,000 - $14,999
            (5) over $15,000
3. Income
    (a) per capita income
        i. total
        ii. race
            (1) white
            (2) Negro
            (3) other
    (b) household income
        i. total
        ii. race
            (1) white
            (2) Negro
            (3) other
4. Expenditures
    (a) food
    (b) clothing
    (c) total recreation
    (d) total transportation
    (e) total housing
    (f) medical
    (g) utilities
    (h) education
        i. primary
        ii. secondary
        iii. higher
        iv. other
    (i) other expenditures
5. Leisure
    (a) average hours worked per week
    (b) time spent on non-credit educational coursework
    (c) length of school day
        i. primary
        ii. secondary
        iii. higher
        iv. other
    (d) time spent in recreation participation per week
        i. primary school facilities
        ii. secondary school facilities
        iii. higher education facilities
        iv. other
6. Specialized Services
    (a) individuals requiring specialized education

current and projected information.

This leads to my next point: while the Census of Population is doing a good job, subject to certain qualifications, we need a massive new effort to produce key annual or biannual survey information. The collection of certain information, especially if it is to be provided on the basis of longitudinal studies, will involve large-scale efforts. They should not be piecemeal and ad hoc. Instead, they should be carried out by a central unit that integrates and coordinates all efforts so they are comparable and cumulative in terms of the information they provide.

Robert C. Wood, Undersecretary of the Department of Housing and Urban Development, in 1963 called for the establishment of urban observatories. They would make "a common series of investigations under a single research plan which for the first time would provide us with professionally reliable findings simultaneously in a number of areas."3/ Such urban observatories could become instrumental in the development of integrated and coordinated surveys and could provide crucial information to improve decision making.

Is the Bureau of the Census the best agency to carry out this function? The Bureau's long tradition and high competence in collecting data clearly gives it a major claim to this assignment. I would assume that it would want to seriously consider assuming such a responsibility, particularly since the need for new and additional data collection by enumeration appears to be declining. However, the Bureau of the Census would have to acquire competence in an area in which it has relatively little experience. This area concerns the identification of important subnational data needs and the evaluation of their priority. Most likely, for this purpose, the Bureau would have to establish new competence within its own organization. But in addition, it might seriously consider joint ventures in cooperation with research centers in universities and private research institutes engaged in urban and regional research.

Summary

State and regional decision makers have been in need of better and more relevant data for a long time, but this has never been more true than it is today when the nation is so greatly concerned with raising educational and employment levels, fighting crime and poverty, improving transportation and housing facilities, developing and cleaning river basins and recreational facilities, etc. Consequently the demand for data is virtually unlimited. Yet data on a state and regional basis are extremely expensive and before they are collected careful thought must be given to their usefulness. The Bureau of the Census, therefore, might have to concern itself increasingly with carefully identifying the most relevant state and regional data. Much of the information will have to be collected frequently by specialized surveys, which should be well coordinated and integrated.

Since most of the Census information is raw data needed for decisions about the future, it is most important to develop reasonably long internally consistent time series. Therefore whenever possible old definitions should be retained and if they have to be changed, the old series should be continued side by side with the new. In a recent article, Elliott R. Morss complained "Whether the objective has been understanding or prediction, it is extremely difficult to understand the recent emphasis on cross-section analyses. For predictive purposes, time series analyses of individual government units would unquestionably be superior to the cross-section approach."4/ Unfortunately, cross-section data are often all the analyst has to deal with.

In order to identify new and additional regional and state data for which there exists a high priority, a framework is needed within which such a rational identification can take place and the costs and benefits of having or not having the information can be evaluated. With this in mind we have proposed such a framework which helps identify key decisions about regions and states together with appropriate decision rules. An understanding of key decisions together with their decision rules can provide the "demand" for, and establish the character of, the regional and state information needs. This is likely to be a relatively new venture for the Bureau of the Census and it will have to develop a new capability in this respect. In doing so, it might be helped by joining forces with established regional research efforts in universities and private research institutes.

Two other important data sources on the federal level are the Internal Revenue Service, whose personal income tax returns provide a gold mine of statistical information, and the Social Security Administration records. But federal data efforts are not well integrated. As we all know a government-wide data service has been proposed recently by a committee of the Social Science Research Council. This proposal has been examined during the last year by the Office of Statistical Standards of the Bureau of the Budget and a White House appointed committee under the chairmanship of Karl Kaysen. Even if these efforts are brought to fruition, however, they will only go part of the way. Much regional and state information is available in state and local governments and private industry, especially in utility companies. No doubt, state and regional data users would greatly welcome a forceful and well conceived effort to integrate state and regional data collection, projection, and dissemination. The present system is inefficient and expensive. Efforts to bring about coordination in this field, if successful, could greatly enhance the chances of providing more appropriate and better quality state and regional data, and thus improve decision making.

---

3. Robert C. Wood, "Contributions of Political Science to Urban Form," Urban Life and Form, Werner Z. Hirsch, editor, (New York: Holt, Rinehart and Winston, Inc., 1963), p. 123.

4. Elliott R. Morss, "Some Thoughts on the Determinants of State and Local Expenditures," National Tax Journal, Vol. 19, No. 1, March 1966, p. 101.

THE 1970 CENSUS -- METROPOLITAN, LOCAL AND NEIGHBORHOOD NEEDS

Albert Mindlin
Government of the District of Columbia

My task is to try to summarize the changes between the 1960 and 1970 censuses of greatest importance to census users on the metropolitan, county, city, neighborhood, census tract and block levels -- in the truly small area uses of census data. This is clearly a monumental task and cannot possibly be covered fully, or even adequately, in the short time at my disposal. Necessarily I have been selective. This selection is based on the knowledge I have built up in my own work and as chairman of the ASA Census Tract Committee (now known as the ASA Committee on Small Area Statistics). Also the Census Bureau has graciously made available to me various documents -- minutes of the Census Advisory Committee on Small Area Statistics, various speeches made by Bureau staff, notes taken by Bureau staff at the numerous regional meetings held this year, and internal memoranda. I have attempted to boil all of this material down to a single summary paper, focusing on the most important and frequently expressed needs for small area census uses, and discussion of present and proposed tabulations on a small area basis. Some of my remarks will undoubtedly overlap those already made, for many small area needs are the same as state, regional, and national needs. Also much of what I will say is already well-known to the Bureau. My purpose is to establish a public record and expression.

Perhaps the most interesting general statement is that there is considerably less interest in new items, than there is in changes in the handling of items already on the schedule and in other structural changes in the processing and tabulating of the census. I will discuss briefly proposed new items, then go on to suggested definitional changes, suggested improvements, additions and deletions in tables and publication levels, and finally review some important miscellaneous topics. Also to save time I will not attempt elaborate justifications of these proposals. Only the most comprehensive and frequently stated points will be mentioned; and beyond remarks necessary to express the point clearly, this selectivity must suffice for justification. I merely paraphrase a statement made by Mr. Brunsman in a recent internal memorandum, that while in earlier censuses the most articulate users of small area census data were persons involved in health and vital statistics, in recent years and today the users of small area census data have enormously expanded. They include market research, city planning, transportation, urban renewal, civil rights, housing, welfare, anti-poverty. I will add to Mr. Brunsman's statement that indeed the decennial censuses of population and housing constitute the most comprehensive and important single source of the economic, social demographic research on which are based most of the vast urban programs going on today.

So we begin -- new items in no special order of priority.

1. We must obtain place of work on a much more refined basis than the county-central city level of 1960; at least to the census tract level; and indeed, bearing in mind the proposed method of taking the census by mail, which includes elaborate coding tables of address to block-face to census tract, work location can theoretically be handled if obtained as an actual address.

2. We must obtain information on multiple jobs, on the extent of "moonlighting."

3. The concept of residential mobility should be applied to occupations -- what occupation were you in five years ago?

4. Mobility should be expanded to include, if possible, the number of moves over some time period.

5. "Children ever born" should be obtained for all women, not just ever-married women. The present restriction introduces substantial biases in the data, especially in central cities.

6. The family or permanent residence of students, as well as their school residence, should be obtained.

7. There have been several requests for attitudinal questions -- attitudes toward one's job, living quarters, health and recreational facilities, etc. This should be explored.

So much for strictly new items. As we discuss definitions and tables, some recommendations will be equivalent to new items. Let us turn to definitional problems.

8. The present definition of the labor force as anyone who worked at least one hour or had a job or was looking for work in the census week, is badly inadequate for current information on poverty and employment. The hard-core unemployed, or Negro youths who, when asked if they are looking for work, shrug "what for?" are counted out of the labor force exactly the same as suburban housewives. The seriously under-employed, the person who tries to survive by long hours at multiple jobs, the underpaid, these are all difficult or impossible to study for small areas by present census labor force definitions and tabulations.

9. Most people, certainly including Census Bureau staff, recognize the weaknesses of the present subjective housing condition definitions. Furthermore these definitions are totally unfeasible in a mail type survey. New definitions based on objective criteria of facilities, such as shared bath, separate entrance and others, are being studied by the Bureau. I hope this will be presented in some detail this afternoon.

10. Part of the housing condition problem is that many local programs that need this item are focused not on the housing unit but on the structure. Being able to relate condition to structure would be a breakthrough in usefulness of housing information.

11. There have been suggestions to change the definition of gross rent. The concept of gross rent as a definitional effort to make rent comparisons meaningful is very important and must not be lost. The concept may well be improved by changing the definitional components, but it certainly won't be improved by being eliminated.

12. There have been suggestions that the present occupational definition of the census week job be changed to "usual occupation." This hearkens back to pre-depression days and memories of how inadequate the "usual occupation" definition was for depression studies. Yet there is a valid point in the objection. The present definition should be retained, but possibly a new item, "usual occupation" added.

So much for definitions. Let us now take up tables and publication levels.

13. At this point I want to dispose of a curious inconsistency that has apparently crept into census thinking about housing condition. As stated above, the Bureau is developing a research effort to impute housing condition by correlation with facilities. If good correlations can be established, then the publication of facilities on a block level could be reasonable substitute for condition. Well and good. But in other Bureau memoranda the suggestion is made to reduce the facilities questions to a sample. If this is done, then block information on condition by means of facilities will not be possible.

14. There is very general agreement that both the block statistics bulletin and the census tract bulletin must be greatly expanded, with far more cross-tabulations than presently exist. At this point we reach the heart of the census as the most fundamental of all tools for social research in small areas. Let me enumerate the most important of these publication improvements, on at least the census tract level, if not the block level. I emphasize that I am referring here not to tabulation alone, but to publication.

First and foremost, there must be a great expansion of population-housing cross-tabulations. The limited extent of such cross-tabulations was one of the greatest weaknesses in the 1960 census publications for anti-poverty program studies, for housing, school drop-out, relocation, urban renewal and similar programs. For example:

a. Rent or mortgage status by color by income.
b. Housing condition or its correlates, by population characteristics such as color, income, family size and composition, ages of occupants.

15. There must, on at least a census tract level, be greater cross-tabulations, of housing items themselves, for example:

a. Density, i.e. persons per room, by various types of housing -- single, duplex, walk-up, elevator.
b. Condition (or its correlates) and various other housing characteristics by rent or value.

16. The block statistics at a very minimum should include population by age, and a number-of-units-in-structure distribution.

17. The block statistics bulletin should be published for smaller size cities, at least down to the 25,000 level.

18. The census tract bulletin should have deeper cross-tabulations of population items.

a. Income by family size and composition.
b. Characteristics of persons in group quarters.
c. Educational attainment by age, down to age 17 for persons no longer in school.

19. For market research the census tract bulletin should include availability of telephone and whether private or shared.

20. The availability of work-place opens a whole new spectrum of extremely useful census tract (as well as city-wide) tabulations and publication -- characteristics of the work force in a given area: its number, occupational distribution, age, color, distance from home, mode of transportation, etc. For the first time profiles of the work force in a city or a census tract will be possible. This subject should be thoroughly explored and the most important characteristics published.

21. In this connection the present census tract array of residents by industry does not appear to have widespread use.

22. The census tract program should be extended to the smaller cities.

I want to mention now certain additional improvements in published tables, without passing judgment on their need below the city level, but emphasizing them on at least the city level.

23. There should be a substantial expansion of cross-tabulations by color. Studies of discrimination, civil rights, and changes in the status of non-whites in depth have been hampered by the absence of published color cross-tabulations of occupation, industry, wages, educational attainment, age and class of worker.

24. The income upper-end intervals of $10,000 and $15,000 are too low. There should be at least one or two additional breaks.

Finally the following remarks on publication should be made.

25. There has been considerable discussion of the artificiality of political boundaries in

the modern metropolis, and the suggestion that the census publications ignore minor civil divisions and other jurisdictional boundaries. If such a suggestion were adopted in the foreseeable future it would be a disaster. A little reflection on this matter will make this point obvious, and I shan't dwell on it further, other than to point out a simple example of which the Census Bureau is well aware -- relative non-use of data for the urbanized area, which is primarily an academic concept, in contrast to the tremendous use of census data for the cities and counties, which are political jurisdictions. Local public policy, which is heavily dependent on social research, follows political lines. These lines must not be abandoned by the census.

26. One matter under consideration by the Bureau is seriously disturbing. Faced with rising costs and increasing requests for more data, the Bureau has raised the possibility of establishing a hierarchy of publication, in which some items would be tabulated and placed on tapes for sale, but not published in the standard census bulletins; and some items would not be tabulated except on reimbursable request. This of course has been done in the past. What is being contemplated is a great acceleration in the relative decrease of published data. This is a matter of great seriousness on which the Bureau must proceed with extreme caution. The vast bulk of local users of the census have neither the time, the know-how, the mechanical facilities, or the money to obtain and use unpublished data. One of the unique and most precious values of the census is its immediate availability on a reference shelf. To compromise this seriously in order to make more data available could prove monumentally self-defeating. The data, while available in theory, would not be available in fact to the great majority of census users.

There are some aspects of the present publication program that could probably be reduced -- for example tract data by industry. Secondly, in choosing alternatives in publication reduction, the Bureau should consider seriously the great reduction or elimination of the interpretative and graphic material contained in the 1960 volumes. This material has limited value and is expendable. (I do not refer here to the definitions, explanations, schedule displays, and subject reference tables in the Introductions. These are absolutely essential and should be as clear and full as possible.) In addition, most of Volume II and Volume III of the 1960 census publication program have been of little use to local communities. (I do not pretend here to represent national and other large interests. But it is these large interests that are in the best position to pay for special tabulations -- as indeed occurred by, for example, the Department of Labor, the Office of Education, and the Office of Economic Opportunity with the 1960 data.) I strongly urge that a serious review be made by the Bureau of the usefulness of the Volumes II and III programs of the 1960 census before committing its 1970 publication resources. What is needed more than anything else is not these special reports and summaries, but the raw, primeval building block data which are manipulable in a host of ways by a host of users who have only what the Bureau pre-prints for them -- or nothing.

As a final section of this paper I want to turn to some miscellaneous but very important topics.

27. The Census Bureau has most happily demonstrated its responsiveness to the increasingly critical need for information of our rapidly urbanizing society, by seeking out ways of serving local communities over and above the census items as such. One of the most useful ways is a by-product of the proposed mail census of 1970. The creation of the address tapes will enable the Bureau to provide localities, at cost, a variety of address directories. This should be especially noted by those communities that have had difficulty constructing address-census tract directories, or that need block directories. In addition, the Bureau is setting aside a 5-digit blank code field for free use by the community, together with the free address print-outs to expedite use of this code and its entrance onto Bureau tapes. This will provide a computerized look-up table for a variety of city divisions such as voting and police precincts, school and health districts, that will permit publication of census data for any of these city divisions at moderate cost. In addition to this 5-digit local code, the Bureau will have its own block-face code which will be made available to localities for similar use.

I trust that these things will be presented today in more detail by the Bureau speakers. I mention them here for two purposes -- first to express publicly our appreciation for these progressive steps, and second to urge that the Bureau prepare a comprehensive and detailed non-technical brochure explaining these and related matters step-by-step -- the various types of maps, different projection systems, such terms as state plane system, coordinates, the uses of block coordinates, their relation to addresses, why both are needed, how a local community can use block coordinates, what facilities it must have to use them, how it can order material from the Bureau by the codes and coordinates, costs, etc. The tremendously valuable work of the Census Geography Division should be clearly explained to the local communities.

28. There has been much mention of the integration of local and census data, but apparently little clarity of thought that has been reduced to cold print. The much greater accessibility of 1970 census data on a small area basis made possible by the 5-digit code, the block face code, and potentially, the block coordinates are major breakthroughs. But they are not integration with local data. Two proposals have been advanced for genuine integration of local and census data, both of which are strongly urged upon the Bureau. The first is the retention of the individual address tapes with the substantive data for that address, or some equivalent method that permits computer return to the raw data by address at moderate cost. Then a community can send to the Bureau a list of addresses with substantive information for each,

and have the Bureau cross-tabulate this local material with census data. This is true integration of local and census data by address. The second is inclusion on the basic questionnaire of the social security number, and the retention of individual record tapes with this number, or some equivalent method. Then the community can send to the Bureau a list of social security numbers with substantive information for each, and have the Bureau cross-tabulate the local information with census data. This is true integration of local and census data by person. Both of these steps would permit enormous advances in community research, planning and evaluation. If the costs are reasonable, their usefulness can hardly be exaggerated.

I want to close this paper with a few remarks on this very last point, the social security number on the census schedule. This item is very familiar to the Bureau. It has been questioned on the ground of possible public objection, since it obviously is not a substantive item but would be there only for record matching purposes. The prevailing opinion of the Census Advisory Committee on Small Area Data, with which I agree, is that there would probably be little public objection to this item (although there may be some objection from civil liberties groups). As we all know, the social security number is being increasingly required on various documents, and the public is becoming accustomed to providing it. By 1970 it can be reasonably inferred that this item will be taken for granted by most people.

Aside from mechanical and technical problems in processing and using the social security number, I believe the heart of the issue lies elsewhere than the strategic question of public acceptance. It lies within ourselves, as citizens and statisticians. Assuming that we have the power to place the number in the census, what do we ourselves want? There is no question that the horizons of research and planning for social engineering will tremendously widen with the ability to integrate numerous sources by social security number -- census, OASI, IRS, welfare, health, and others. This can, potentially, solve the record matching problem we have wrestled with for years. We are on the historical threshold of comprehensive social data banks, and this item will surely put us across that threshold. Is this what we really want? We live in a democratic society, and presumably confidentiality statutes and regulations are honored. Will this always be so? For the enormous leap forward in social engineering that integrated person records will make possible, do we want to create such governmental files? Or are there ways to, say, create central "look-up" tables by social security number links with other file numbers, without actually integrating the substantive records? Would this suffice? Or is even this too great an invasion of privacy, too great a risk? What is the tipping point? In marching courageously forward to a 1984 utopia, are we not also blindly paving the way for a possible 1984 Big Brother?

This is too big an issue to be covered here. Suffice it to say that the historical moment of truth is rapidly approaching. Already the medicare program is forcing the integration of local health, welfare and vocational rehabilitation records. Whether we like it or not, social data banks, at least primitive ones, are historically upon us, as inevitable as the tide. We must face squarely and solve the problem of how to create effective social data banks and yet preserve confidentiality and personal privacy. Otherwise we may find that the road to a totalitarian hell was partly paved by our liberal good intentions.

I-B

THE 1970 CENSUSES OF POPULATION AND HOUSING: SPECIAL LOCAL STATISTICAL INTERESTS OF THE DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

Chairman, ALBERT MINDLIN, Government of the District of Columbia

COMPATIBILITY OF URBAN LAND UNIT DATA

Robert C. Colwell, Department of Housing and Urban Development

This decade is witnessing a major revolution in data handling as a result of rapid technological progress in the capabilities of computers. Installations of electronic data processing equipment in private industry are being used for a growing list of chores ranging from inventory control to stock market transactions and from airline reservations to material processing equipment. Wherever a job involves assimilation of large quantities of data, it seems that electronic equipment can be devised to do that job faster, more accurately and cheaper. The computer is making significant contributions to the increase in productivity of the U. S. economy.

In Government, computers have found a major place. The Bureau of the Census has long been a pioneer in developing new data processing equipment, and Univac I used in the 1950 Census of Population and Housing was one of the early Federal installations. Today, I suspect that there is hardly a Department or agency of the Federal Government that does not make some use of automatic data processing equipment, directly or indirectly. The Comptroller General recently reported that "the Federal Government is spending directly and indirectly this year nearly $3 billion for the acquisition and use of computer equipment, compared to only minute outlays 15 years ago."

State and local governments are also proceeding to install ADP equipment, finding that burgeoning files of data can be maintained and retrieved more effectively and less costly via the electronic route. Just as electrical bookkeeping equipment began replacing manual records several decades ago, so the computer is replacing EAM in the current decade. The pace is fast and the opportunities are great. Like other technological revolutions, there can be no turning back to older methods. Our economy and our governmental operations have been restructured around the computer. We would drown in a depthless sea of data if these machines ever went on a strike.

The major focus of this paper is a narrow one, but perhaps an important one for the future application of electronic equipment to geographically identifiable data in governmental activities. The Department of Housing and Urban Development, as well as the Bureau of the Census, has a deep involvement in this problem. Increasingly, State and local governments are sharing a similar concern.

The element of geographic location is common to the files in nearly every program that HUD administers. Most of the substantive programs involve land parcels, either separately or in contiguous groupings. Policy analyses and program execution in urban land-use and land-related activities rely on statistical studies of the Department's records as well as on urban and metropolitan universe data, ranging in scale from a block to an SMSA, to the total U.S.A.

Let me illustrate with a few examples:

1. City A proposes a new urban renewal project involving 100 acres. Are the boundaries of the plan properly located for the elimination of slums and blight and the creation of a well-planned neighborhood after renewal? What proportions of clearance and rehabilitation can be expected? Reference to Census block data can throw considerable light on these questions.

2. The XYZ Development Company seeks a commitment to insure mortgages on a 200-house subdivision at prices ranging from $18,000 to $23,000. Is the subdivision too large, and are these prices proper for the proposed location? The records of recent market activity in other subdivisions that FHA has insured will be useful in judging the prospective market absorption, and Census data will help to identify directions of growth in various housing price-groups.

3. Long-range fiscal plans for senior-citizen housing are being formulated. What levels of Federal support are appropriate, and what income groups need to be served? Here, a combination of statistics need to be analyzed, relating demographic and income distribution data to records of available housing and its utilization.

4. A mass transit grant for city K is being considered. Is the proposed corridor the best to serve the most pressing commuter needs? At what level of service will saturation be reached in this corridor? In this case, block and tract data on population and housing characteristics, labor force data, and journey-to-work patterns should be quite helpful.

The list of problems and the variations from place to place are literally never ending. While decisions rest on judgements, the accuracy of the decision is almost invariably better if it rests on empirical data rather than intuitive opinions formed from memory. These informational needs are currently served by three principal data collecting activities. A summary statement followed by a fuller discussion of each may be helpful.

Major Sources of Land-Oriented Data

First, the decennial census of population and housing provides the basic source of information about the housing stock and its occupants, and about the labor force and its characteristics. The 1970 information is expected to cover more items of pertinent information than any prior census, and to be available for more land units than ever before. The wealth of data provided by a decennial census helps to measure the past accomplishments of programs, and by reflecting national needs, furnishes guidance for re-orientation of older programs and the need for new ones. This is the grist for solid analysis on a national scale, as well as for smaller areal units down to the tract and block.

Second, the operating records and files of HUD are a major statistical resource, partly because they show what HUD programs have done, and where they have been used. The cumulative involvement of mortgage insurance, low-income public housing, urban renewal and college housing is quite substantial in many urban areas. Other programs of more recent vintage are already making their mark. The new Department is undertaking to create an integrated information system, pulling together the operating records of the various activities previously conducted by the five constituent agencies of HHFA. This system is needed for both internal management and for broader program analysis.

Third, since computers began to find a place in the offices of local governments, the era of municipal data banks has begun to bud and some blooms are already open. While the first vision of the potential uses of parcel-based data banks may have been seen in the decade of the 1950's, few serious efforts were undertaken until the statistical sixties arrived. Early in 1961, five cities in southwest States submitted an application for a demonstration grant under Section 314, the urban renewal demonstration program, and the Maryland National Capital Planning Commission published a research monograph financed under the urban planning assistance program, Section 701. Concurrently, transportation planners, bogged down in O and D records, also began to turn to land parcel records as a way to anticipate urban traffic rather than see what it had been. From these beginnings, the metropolitan data bank concept has taken root in various forms in a growing list of communities. Where the installation has been able to support comprehensive planning or the community renewal program, assistance from HUD has been available.

In designing automatic data processing systems, many problems have bubbled to the surface and solutions satisfactory to the users have been developed. The record of some of these has been described in the growing volume of reports and articles, but some system designers who have failed to heed the experience of others have paid the price of reliving that experience.

<u>Unit Identification Plans</u>

Every mass data handling system must begin with a unit identification plan. These plans can be distinguished by three major categories in a simple taxonomy.

1. <u>Arbitrary numbers</u>

If the unit is a parcel of land, the identification plan may be based on an arbitrary numbering system unique to the particular set of records. Some property assessment files are so numbered. But where such a unit identification plan is used, there can be no compatibility or comparability with other data files unless a secondary identification which can be recognized geographically is also added to each unit record, or unless a cross-index is compiled.

Arbitrary numbering plans may have the attributes of simplicity and economy. For instance, mortgage loans which are secured by a lien on a parcel of real estate are often identified by the lender and/or the underwriter by means of a serial number containing a prefix or a suffix, or both, which relate to geographical areas such as a State or a branch office jurisdiction. Number assignment in a serial system is simple and automatic. Such a system may also serve as an operations control and a statistical source when there are no number voids or cancellations. But serial numbers are of little use for geographical analysis unless the prefix or suffix is quite sophisticated and detailed. Nevertheless, serial number plans are very attractive to some commercial types of activity, both governmental and private, and probably predominate in the universe of unit identification plans.

2. <u>Geo-political units</u>

Another frequently used plan is that of identification in terms of some geo-political system. Street names and numbers come under this heading, as do census blocks and tracts, tax-roll description, postal zones, wards, precincts, town and city identification, and other such areal elements. Sometimes two or more geo-political units are identified in the heading in order to facilitate the computation of aggregates in terms of such units or to identify cognizance or responsibility. Fire, police, voter registration, insurance, school enrollment and similar data files would be expected to require geo-political identifications.

The geo-political unit was a necessity in the pre-computer era and continues to perform useful functions. Manual records and punched cards often require the notation of larger areal units in order to be economical of time and cost. These limitations are not as compelling when data files are maintained in modern electronic storage.

Boundaries, names, and numbers of geo-political units can be changed by those who create them, and are revised whenever necessary. This risk must be considered when the unit identification plan is designed, recognizing the time and cost involved in file revision as a trade-off against the convenience of ready access to large areal unit location. If the file is expected to be maintained and used for a long time, i.e., for several decades covering the life of a building or the maturity of a long-term lien, file revision may be a necessity and prove costly. But if the file is to be used for a one-time operation or have an expected life of, say, a decade or less, or if the areal unit is expected to be very stable--such as county boundaries--the risk of revision may be small.

Time series analysis based on geo-political unit aggregates also involves a hazard. Boundary revision raises the question of whether absolute location is more significant than areal unit cognizance. For example, the granting of Statehood to Alaska and Hawaii has made time-series analysis of many U. S. statistics somewhat more difficult to compile and certainly more cumbersome to describe.

The continuous process of urban annexation, sometimes of small tracts and other times of large areas, is an ever-present hazard for the urban economist and demographer. This may be illustrated by the population and the area of the city of Los Angeles; both of which have expanded greatly in the 20th Century. For some purposes, it may be meaningful to compare the population living within the city limits in 1900 with the population inside the 1960 boundaries. In other cases, it might be necessary to know how many people in 1960 were living inside the 1900 boundaries or how many in 1900 lived within the city area of 1960.

3. Grid systems

The third method of notation that may be used in a unit identification plan involving geographic location is a grid system with a permanent spatial anchor. Undoubtedly, the most widely used grid locational system is that of latitude and longitude. This method differs from the other two systems in that the intersection of grid coordinates identifies a single point, whereas arbitrary numbers and geo-political units usually refer to a land parcel or area.

If the grid scale is small enough, a single point within a parcel may provide sufficient identification, or several points may be needed to describe boundary lines. Grid point assignments to land parcels involve problems and costs, but once made are unchanging except where parcels are subdivided or merged.

Such plans overcome many of the shortcomings of other methods of identification and are adaptable to computer operations. If used in conjunction with geo-political systems, either as primary or secondary identifiers, grid coordinates provide a ready means of file correction if adjustments are made in either the boundaries or the numbering plan of geo-political units.

We hear that there are some local land parcel data files that employ grid systems other than latitude and longitude. Where these are private installations serving a special purpose, such as an electric power company, an arbitrary grid system scaled for its particular needs may be more suitable and easier to use. In such cases, comparability with other data files may be of no concern. However, where public purposes are involved and the uses of the file may be varied, it would seem that the unit identification plan would be most useful if it carried grid intersects, together with whatever other locational decriptions and unit classes might be helpful.

Recognition of Needs for Comparability

Machine processed files require compatibility if comparative studies are to be made, or if new elements are to be added from other files. We find a growing recognition of this need, as computer applications increase in number and complexity, and as equipment becomes available to do larger and more difficult jobs.

1. Census plans

The Bureau of the Census has stated that it proposes to identify latitude and longitude in degree decimal terms in connection with block-face addresses being developed for the 1970 Census of Population and Housing. Land parcels in Census tracts and in about $1\frac{1}{2}$ million blocks will be so identified. The Census Bureau will discuss this procedure in a paper to be presented this afternoon. This procedure has far-reaching consequences, and opens possibilities for research with 1970 data that have heretofore been impossible.

2. HUD information plans

The new Department of Housing and Urban Development is undertaking to construct an information system to record and describe the activity in each of the many programs it administers. This system will encompass the millions of land parcels securing FHA-insured loans; the manifold characteristics both before and after execution of nearly 2,000 urban renewal projects; low-rent public housing projects going back to 1937; and scores of other programs ranging from college housing loans to open space grants from mass transit loans to code enforcement grants and public facility loans, not to mention FNMA's portfolios.

All of these records are now in various kinds of filing systems, ranging from dockets to manual card records to magnetic tape. Each individual case has some kind of a serial number and some form of geographic location reference, and all have urban land characteristics. The translation of these records into a national information system so that HUD can readily determine what it is doing--and has done--in any and every urban sector is a tremendous undertaking and will be a lengthy and detailed process. But it is essential in order to carry out the broad urban responsibilities which the Congress and the President have now assigned to the Department. The ultimate benefits will far outweigh the initial investment, not only in making program administration more effective but also more efficient.

While HUD has not yet reached the stage of planning the details of its overall information system, it may ultimately adopt the same latitude and longitude land parcel decriptions that the Bureau of the Census plans to use in 1970. These would be carried alongside the case, project, or loan numbers and various geo-political unit identifications. One of the more obvious analytical gains would be the opportunity to aggregate HUD records into Census blocks and tracts for the study of program activity in relation to small-scale universes.

3. Metropolitan data banks

During recent years, HUD and its predecessor agency have given some financial support to the development of metropolitan data banks based on land parcels. This support has come from the urban renewal demonstration program (Section 314), the urban planning assistance program (Section 701), and the community renewal program. Meanwhile, other data banks organized on land parcels have come into being without HUD assistance.

These installations have interesting possibilities both as tools in comprehensive planning and as aids to efficient local governmental operations.

There are now enough of these local data banks in operation to permit a study of their various features as well as the uses to which they have been put. Such a survey is about to be undertaken and is expected to be completed before the end of the current fiscal year. We hope that this appraisal will help the Department in shaping its policies regarding future support of these efforts.

Without seeking to direct the form and structure of local proposals, it would seem reasonable that HUD would probably encourage local agencies to incorporate the same system of land parcel identification that both Census and the Department are expecting to use, viz., a grid coordinate system mathematically translatable into latitude and longitude or into some other plan of coordinates. Inclusion of grid location reference points in local data banks would open up compatibility with Census block and tract data, giving local government the same kind of opportunities that HUD would gain in its own information system.

The Problem of Disclosure

Serious concern is being expressed in various quarters about the potential invasion of privacy that is threatened by computer-based data. This threat arises partly because the capacity of electronic storage makes possible the combining of many data series that previously had to be maintained in separate files under separate jurisdictions. It has been pointed out that the picture described by the totality of several data series presented in concert is different from the view of each by itself. Even though each item could be publicly researched, the mosaic made by the total is one that only a computer could readily produce.

This contention is probably not debatable, for it is a method of analysis that has been successfully used in both financial credit review and in military intelligence to name only two examples. It is important, therefore, that the structure of data files be carefully examined to anticipate whether control of the file by persons who were not benevolent in their motives could be harmful or capricious. It may be helpful to examine the policies of the three principal collectors and repositories of geographically identified data that have been discussed in considering safeguards against the risk of invasion of privacy.

First, the Bureau of the Census operates under very strict laws and regulations regarding disclosure. Its publications and releases are required to show only aggregates which are sufficiently large to preclude identification of individual components. The apprehensions expressed about the invasion of privacy do not relate to Census releases, and the Bureau has never been viewed as a threat because of its laws and its vigorous administration of them. The activities of the Census Bureau relative to this issue are twofold: (a) block and tract data that can be used as benchmarks for other studies; (b) the techniques being developed for translating city street addresses to grid coordinates. Both of these are tools of policy analysis and program planning, but hold no threat to privacy.

Second, some of the programs of HUD deal directly with individuals and private industry, viz., FHA and FNMA, while others deal largely with various kinds of public agencies and local governments. In all of its nonpublic actions, the Department carefully guards the privy relation with its program participants. Its files are treated as confidential as a physician's clinical records of his patients. Where public bodies participate in the Department's program, their actions and the details of the project are usually public information in the offices of the municipality. Public projects often involve open public hearings to assure full disclosure and consideration of local decisions. These actions hold no threat to personal privacy.

Third, local data banks differ from the Federal activities discussed above. They can, and some do, contain a wide range of information, varying from land parcel items to facts about the people who live or work at various addresses. It is the latter type of data that offers the threat to personal privacy.

The recording laws of each of our States require public recording of deeds and liens to establish seniority of interest. This is the essence of land tenure policy in the United States, and our mortgage and real estate industries have developed around this principle. Other local public land records relating to assessments, taxes, zoning, building permits and similar activities are available to anyone who takes the trouble to inspect them. In fact, it is difficult to identify any type of land parcel information in local governments that is not available to the public, except, of course, preliminary planning of future public land acquisition for such things as rights-of-way, public buildings, open space, etc.

In the case of social and demographic data, the picture is different; the threat of invasion of privacy is a potential reality. It would be helpful in the administration of many local social, educational and welfare activities to be able to make small-scale geographical studies of the incidence of crime, truancy, disease, and arson to name only a few items. Also, geographical patterns of income distribution, educational levels and other economic and social data have a use in capital budgeting and many elements of comprehensive planning. But as long as there is public concern about the threat of disclosure of personal information, we should expect to find public resistance to its inclusion in data bank storage.

Possibly there is a way out of this dilemma by following the practices of the Census Bureau. For practical purposes, the needs of local planning and analysis would probably be served just as well by aggregates for small areal units as by maintaining data about persons on an address basis. If the inputs to data banks from files in other municipal departments were limited in detail to a geographic level that protected individuals against possible adverse disclosure, the logic supporting the critics of this analytic tool would vanish.

However, the maintenance of public confidence rests more on faith in the integrity of those intrusted with data than it does on logic. A single indiscretion or an innocent error might destroy the confidence built up over many years in many separate cities. Hence, it will be well for those who design and those who operate local data banks containing personal, social and economic information, to examine the structure of inputs and storage carefully lest outputs impinge on personal rights. Failure to do as much could jeopardize the longer range use of modern equipment that promises help with some of the most difficult kinds of public decisions.

I-C

THE 1970 CENSUSES OF POPULATION AND HOUSING: PLANS FOR THE 1970 CENSUSES

Chairman, CONRAD TAEUBER, U. S. Bureau of the Census

CONTENT PLANNING FOR THE 1970 CENSUS OF POPULATION

Howard G. Brunsman, U.S. Bureau of the Census

The basic decisions which we must reach regarding the 1970 Census include:

1. What subjects should be covered in the Census?
2. What tabulations should be prepared, and how should tabulated results be made available?

We use a number of different approaches to obtain the answers to these questions. Users are continually in touch with us and tell us of their needs for census data. Starting within the past year when we answer requests for data, we ask the requester how he plans to use them. In this way we gain a greater knowledge of the use of census materials. A special unit at the Census Bureau has been set up under Robert Voight to review uses of census data and determine what changes would be desirable to increase such uses.

We also work with a number of advisory committees. Last fall we organized a special Advisory Committee on Population Statistics of which Dr. Philip M. Hauser is chairman. This committee has held four meetings. It has reviewed many aspects of the population statistics program of the Census Bureau but so far has given major attention to the plans for the content of the 1970 Census of Population. We have also organized a Committee on Small Area Data which gives special emphasis to availability and use of data for city blocks, census tracts and other subareas, especially within Standard Metropolitan Statistical Areas.

We also seek guidance from the Census Advisory Committee of the American Statistical Association, the Committee on Population Statistics of the Population Association of America and the special census advisory committees of the American Marketing Association, the American Sociological Association and the American Historical Society.

In 1954 the Watkins' Intensive Review Committee reviewed the operation of the Bureau of the Census. They recommended that Census officials arrange to meet with local users of Census data to learn of their needs. Meetings of this type were held prior to the 1960 Census and proved most useful. We have held 22 similar meetings relating to the 1970 Census within the past six months. These meetings have been sponsored by various groups including the local chapters of the American Statistical Association, the Census Tract Committees, the local Census Tract Key persons as well as the Federal Statistics Users Conference and the local chapters of the American Marketing Association. Each of these local meetings have been attended by between 40 and 200 persons and they displayed enthusiastic interest in Census statistics and in plans for the 1970 Census.

In these meetings with users throughout the country, we have been impressed with the fact that the whole concept of use of Census data is changing. This need has been expressed very eloquently by Dr. Duncan and Mr. Mindlin in the earlier meetings today. We find greater need for statistics to implement local programs in education, poverty, redevelopment, housing, highway and other fields. The Census of Population must become a tool used to find answers to vital problems in local areas and in the country as a whole. We must be prepared to supply special tabulations promptly and at an early date after the census has been enumerated. These special tabulations must be made available in addition to the regular census publications, not in place of them.

A number of additional subjects have been suggested for inclusion in the 1970 Census. Dr. Duncan has referred to the need for additional ethnic detail, information on the year that foreign born persons entered the United States, on child spacing and on religion. He has noted also the desirability of tabulating separately the data for Negroes and other racial groups. Mr. Mindlin has referred to the interest in greater geographic detail on place of work, on information on children even born for never married women and on the time reference of occupation and industry. Both of these men have referred to the interest in including social security number and in activity and occupation five years ago.

In addition to these suggestions, those in which the greatest interest has been expressed include college degree received, field of specialization in college, vocational and technical training, voting and registration to vote and additional detail on sources of income.

When considering changes in content, we take account of cost of the additional items; determination of whether useful information on the subject can be obtained with the approach we propose for the 1970 Census; and the possibility of developing an approach which will obtain the required data in sufficiently accurate form.

The test censuses that Mr. Kaplan described for Louisville in 1964 and Cleveland in 1965 related primarily to procedures. They included some experimentation with content, in particular the exact street address of work place. In the spring of 1966 we conducted small tests of subject matter with the mail-out mail-back approach in Yonkers, New York, and St. Louis Park, Minnesota. In these tests we obtained social security number, ethnic group with two alternate approaches, more detailed description of occupation, address of work place, and activity and occupation six years ago.

We are planning to conduct a special census of the New Haven metropolitan area in April 1967, as a further test of our field procedures. Associated with this test census is a research study designed in cooperation with local and other federal agencies to investigate how census data might be coordinated with local data and to explore the development of packages of standard programs for special tabulations. We wish to take this opportunity to gain experience with new subjects. The content of this test has not yet been determined. Active candidates among new items include the following: whether first marriage ended in divorce; State of birth of parents; college degrees received; field of specialization in college; nursery school attendance; vocational and technical training; greater detail on occupation; address of place of work including street name, number and Zip code and name of employer; greater detail on income sources; activity five years ago and social security number.

During the coming winter we also plan small subject-matter tests similar in approach to the Yonkers test with particular emphasis on new subjects not already tested such as vocational training, education and health. The tests also might include alternate approaches to subjects tested in earlier surveys.

We are planning a dress rehearsal in April 1968 which will cover an entire county or metropolitan area. This will represent a prototype of the 1970 Census and is intended as a final test of procedures and content.

COMPUTER-BASED GEOGRAPHIC CODING FOR THE 1970 CENSUS

William T. Fay and Robert L. Hagan, Bureau of the Census

Almost everyone is aware by this time of the extensive use we plan to make of the mails in the 1970 census. It is possible, however, that many of those who know about it have not considered the amount of planning and preparation necessary to assure its success.

In the time allotted me today I want to tell as much as possible of the part being played by the Census Bureau's Geography Division in this pioneer effort.

I want to make it clear, too, that much of what we are doing now is part of an exploratory phase in which we are mulling over our own ideas and seeking the opinions of others. We want the best possible program and we welcome suggestions from any source.

First of all, there are these reasons for making the change from direct enumeration to mailed questionnaires.

We need a reduction in the time span of the information gathering process to reduce the constant danger of overlooking some people and counting others twice. A speed up in this phase will also contribute to earlier release of data. Publication of most reports on the 1960 census gained 12 to 18 months over the 1950 census but you want even more speed. I need not tell you that the sooner we dispense the information, the more value it has for the Nation.

We want the increased accuracy which will come from making it possible for a family to sit down together to consider the questions, rather than putting the burden of providing the information for the whole family on one person. Quite often the hurried enumerator must question the one member of the family found at home and too often that one person lacks complete information.

We want to improve the quality of census data by diminishing the influence enumerators may have on the answers we receive. Scientific studies have demonstrated that enumerators influence answers to census questions in various ways. While these effects tend to cancel out when the work of a number of enumerators is combined, they can adversely effect the data for individual small areas, each of which is within the territory covered by a single enumerator.

We want to save expense and to reduce the task of recruiting and training the army of enumerators needed in the past. I'm safe in saying that they will never be entirely eliminated but their numbers can be reduced. We plan to use perhaps 100,000 in 1970 as compared with 170,000 in 1960.

A factor which will facilitate a mail census in the increasing drift of our population to urban areas where there are street addresses and where mail is delivered by city carriers. There were 125 million urban dwellers in this country in 1960 and there will be 140 million by 1970.

But if that seems to indicate a simple solution to the need for change in census methods, let's consider some of the complications.

First, there has always been a geographical vagueness about many post office addresses. This is due mostly to the fact that the Giantville post office handles mail for adjacent Dwarftown and even beyond Dwarftown into unincorporated areas.

The Dwarftown citizen usually doesn't mind having a Giantville address as long as he gets his mail promptly. In fact, he sometimes thinks of it as an asset. The folks back home never heard of Dwarftown but they do attach some importance to Giantville and part of that importance tends to rub off on a person who has it as his address.

Sometime the fuzziness about addresses even crosses State lines. For example, the post office at Suitland, Maryland, where the Census Bureau is located, is a branch of the Washington, D. C. post office and we have a Washington address.

In the 1954 Census of Business, the Bureau made an effort to clean up part of that confusion. On the census form a request was printed for a statement of the actual location of the business or plant, regardless of the address. There was nothing complicated about the question but it didn't draw nearly enough accurate answers to solve the problem.

We then decided to see if the post office could help. Our reasoning was that a man trudging the mail route day after day would know the point at which he stepped over the Giantville city limits and entered Dwarftown and that he could make a record of the exact address at which that happened.

He couldn't. Or if he could, he didn't tell us with enough frequency or accuracy to yield us substantial benefit.

Someone had the idea that telephone exchanges were tailored to city boundary lines and that we could identify the locations of business firms from their phone numbers. We found that sometimes this would work but more often it would not. Public utility records didn't help, either.

All of these things foretold some real headaches for anyone attempting a mail census without first clearing up the uncertainties about the relationships between addresses and boundary lines.

We knew it might be possible to blunder along on a mail census by using maps and enumerators in the confused perimeter areas of cities but what we really needed was a master coding file on computer tape which could be used to assign most business locations to States, counties, and cities.

To prepare such a file from scratch would have cost the Bureau far more than we were prepared to pay so we began looking for such a file already in existence.

Finally we investigated a file of punch cards which had been amassed by a directory firm for use in relating census data and sales statistics to dealer areas for purposes of market research and analysis.

That file covered delivery areas of post offices located in cities which had populations of 25,000 or more and it consisted of street names and house number ranges for areas approximating census tracts.

This showed possibilities. It was decided that if the border areas of those cities could be corrected to exact boundaries, the file could be used to assign geographic codes to specific addresses.

In addition, the records could be transferred to computer tapes and programs could be developed which would take an address input, recognize the address elements, and then put them in sequence to facilitate matching against an address reference file.

There were two basic questions to be answered. The first was: "Would a system covering delivery areas of post offices located in cities of 25,000 or more inhabitants provide the base for a solution?"

It was believed that the answer was "Yes," since a major difficulty was the accuracy of coding of establishments located in small cities which had mail delivered from adjacent larger cities. Success, of course, hinged on learning in which community the borderline addresses really belonged.

The other big question concerned the feasibility of developing a computer program which could code a high percentage of addresses accurately.

The file we were studying had been prepared by its owner under a controlled process in which address elements had been standardized to facilitate mechanical coding by use of the card reference file.

Could the Census Bureau use this reference file for coding a large proportion of addresses not prepared under the same controlled conditions?

The answer to this was not so apparent, and so it was decided to put it to a test. Los Angeles County was chosen because it had all of the problems with which we were concerned.

The results were encouraging. About 75 per cent of the addresses from our Census Bureau files were matched to the reference file and coded by computer and the test showed the accuracy level for those addresses to be about 99 per cent.

Following this, a national test was conducted, on a sample basis, with similar results, so the directory firm's entire file for cities of 25,000 or over was obtained as the nucleus of our system. The test procedures were then carried out for the entire United States and the addresses for business firms included in the 1963 Economic Census were coded in this fashion.

The results were satisfactory, that is we did improve accuracy, speed, and cost over past methods. Coverage, however, was not sufficiently extensive, so we are now modifying the basic file to permit coding of business addresses within all city delivery areas in the nation for the 1967 census.

The work described above was a small-scale forerunner of things to come. For the Economic Census the boundaries which had to be recognized in Census publications, and therefore coded by computer or other means, were those for a few thousand urban places, a hundred or so Central Business Districts, a few hundred Major Retail Centers and, of course, the 3,100 counties.

When earnest consideration was given to similar methods for the 1970 Census, the scale changed drastically; then we switched to talk of blocks and block faces with estimated numbers of areas jumping to one and a half and eight million respectively.

Here's where another big problem popped up. Coding of residential addresses to this fine-grained level requires accurate maps with a uniform scale within each urbanized area.

They just didn't exist.

In the 1960 Census, the Bureau needed maps, of course, although not as desperately as now since that was an enumerator census and the periphery problem was not nearly as great. If streets didn't show on maps, or nonexistent ones did, the enumerators were expected to resolve the problems.

The maps used in 1960 were of widely varying sizes and scales. They were barely passable, for the purposes for which we used them, and they resembled an unending series of jigsaw puzzles. In one city, with its suburbs, it was necessary to use 137 different maps for piecing together enumeration districts. There was a constant danger that some areas would be left out or overlapped. That could mean that some people would be overlooked or that some would be approached twice by enumerators of adjoining districts.

For the 1970 Census the problem couldn't be dodged. Something had to be done and the Geography Division decided to tackle it. Here's how it's being done:

We took U. S. Geological Survey 7½ minute quadrangle maps and changed the scale from 1":2,000' to 1":800'. We dropped the topographic detail and updated the street layouts as far as we could with the information we had available.

We knew, of course, that we didn't have enough information and that errors were certain to creep in. This is where we have to lean heavily on local groups, with a much wider knowledge of their own cities than we have. The cooperation we're getting is excellent.

The Bureau can't pay for this help, but we are reciprocating by providing copies of the up-to-date maps being developed, and by creating the capability of providing a vast fund of information which can be blended with local statistics and geared to local governmental units. This is to be facilitated by reserving for local use a five-digit section of the FOSDIC form on which block faces are to be coded. Details of this plan will be given further on in this presentation.

The first step in our map making, the compilation to a scale of 1":800', is done in our Jeffersonville, Indiana, office and it is now more than 40 per cent complete for the urban cores of metropolitan areas. Our goal is to finish the mapping for all of these urban centers by the end of 1968.

Those maps, when finished, will cover 100,000 square miles--about four times the extent covered by the urbanized areas defined in 1960.

When we have developed a map as far as possible in Jeffersonville, we send copies (in 36" x 48" sheets each representing 35 square miles) to a local cooperating agency in the area covered in the map. Usually that agency is a planning group which has agreed to give us the help we vitally need. That group will verify, correct, and update our maps, or they may pass copies on to other local groups in possession of the detailed information necessary.

After the required changes have been noted by local groups, the map is returned to us and we alter our original tracing to conform with the local editing.

Copies are then returned to the local cooperating agency where they are available for local purposes, including the notation of additional changes in streets which should be added to the master maps before the final census deadline.

In the preparation of computerized Address Coding Guides, much the same procedure will be followed.

For this process we'll use FOSDIC worksheets capable of being read electronically. We'll prepare these sheets by printing street names, block face identifications, intersecting streets, and even and odd address range numbers. In doing this, we will use information from a commercial direct mail list and, where available, from directory publishers.

Those partially-completed forms will then be sent to the local cooperating agencies together with copies of our Metropolitan Maps marked with Census area designations. As in the mapping program, the cooperating agencies may themselves verify and complete the worksheets or may farm this process out to other qualified groups within the area.

At this stage, before the forms are returned to us, agencies which are cooperating should determine what use they will make of the five-digit code field provided on the form for local use. We refer to it as the "optional field" because it can be used in various ways, or not used at all if it is not wanted.

You will notice that I say "if it's not wanted." It may be that some communities will not want it; although, if they understand the uses to which it can be put, it seems certain that they will want it. We think the "optional field" is very important, and a little further on I'll speak more of its benefits.

Through the use of address coding guides we will be able, for the first time, to record information for geographic units ranging from one side of a city block to an entire city. The limit to the flexibility of the information available to you after the census will be disclosure rules, computer capacity, and the cost of tabulation. We don't plan to provide this capability for all city delivery areas, but if we can accomplish this for entire urbanized areas and for cities of 25,000 or more inhabitants, the scope of our present plans, far more detailed census data will be available than ever before. A further limitation is the extent of city delivery postal service; beyond these areas we don't plan to code to the block-face level; although, reporting by block is expected to be feasible.

With relatively little added effort a copy of the "Census" Address Coding Guide for an area can be modified locally for broader use by the addition of identification codes for areas such as police precincts, health areas, and so forth. With this accomplished local flexibility is virtually unlimited. Police information, for example, can be matched to the modified coding guide and the police data assigned not only to police precints, but simultaneously to health, school, and other areas, as well as to census tract and block. The same can be done with other local information.

It's hard to imagine better tools for orderly, forward-looking community planning than an accurate address coding guide, but the possibilities are even broader. I'm referring of course to the "optional field" and related methods of securing census data for areas of local interest, rather than being limited to the tabulation areas used by the Bureau.

Let's consider the possibilities in a broad sense and then turn to details.

This tool can be used to secure Census data aggregated to match areas of local interest and then to correlate locally developed statistics with those from the census. As an example, local figures showing a steady growth of juvenile delinquency within a certain police precinct could be related to census population and housing characteristics, coded for the precincts, and from the combined information a clearer picture of the problem and its solution might emerge.

Another situation might bring the question: "How many children between the ages of 6 and 16 live in the 10th school district and what is the racial proportion there?" A special tabulation would put the answer at the school board's finger tips.

Would recreation officials like to know how many children from 5 to 12 live within a 20-block radius of a proposed playground and the ratio of boys to girls? They can get the answer from the Census Bureau computers.

There are four primary ways that we suggest for local consideration in filling in the optional field and a fifth method that can be used separately or in combination with one of the first three. For convenience let me name the methods and then follow with more detailed explanations. They are:

1. Direct Coding
2. Geographic Unit Coding
3. Local Serial Number Coding
4. Census Bureau Coding
5. No Coding

In the first, the Direct Coding method, the procedure is simple, but the possibilities are quite limited. If a block face lies within the 12th police precinct and data for these areas are wanted, the figure 12 is entered in the optional field. If a school district grouping is wanted and the block face is in the 9th school district, the figure 9 is marked in the field. If the block face is in sanitary district 14, the figure 14 may be put into the optional field.

In this method, as in others, when boundaries cut across blocks, rather than following street lines, the severed block face would be treated as two block faces for the purpose of coding.

The two sections would both be placed on the FOSDIC form as separate block faces, the only distinction in coding of the two being in the range of addresses within each section and the code in the optional field. Thus, house numbers 1 to 19 of the block face might be coded as belonging to the 11th police precinct while numbers 21 to 49 might be coded as being part of the 12th police precinct. (The Census Bureau will limit such block face splits to 10 per cent of the total block faces listed.)

The direct coding system has limitations, the three administrative areas cited above and the total of five digits needed to code them into the optional field (12-9-14) exhaust the limits of the field and prohibit coding of the block face into other civic classification on the FOSDIC forms.

In theory, five such area codings could be made for each block face, but in practice such a possibility is rare since it depends on the unlikely supposition that there would be no more than nine (1, 2, 3, 4, 5, 6, 7, 8, 9) each of such administrative areas as districts and precincts and, hence, each would require only one of the five digits in the optional field. A more realistic number of codings would be one or two-three at most, thus drastically limiting the amount of local statistical groupings possible.

Another disadvantage of this system is that once a block face is coded into the optional field as being a part of a local administrative area, it goes into the Census records in that form. If the block face is later shifted, as from one police precinct to another, Census Bureau reference tapes would have to be altered to conform to the change. Otherwise, the value of that particular optional field coding would be lost.

The second method, Geographic Unit Coding, requires preparation of a map on which are shown the boundaries of all the local administrative units for which data will be desired.

The map will look like a hodge-podge of lines, lines which serve to cut the map into the "Geographic Units" referred to in the title of this method. The map will show clusters of blocks, ordinarily, bounded by lines drawn on the map. Each such cluster, or geographic unit, can be described as being entirely within one police precinct, one school attendance area, one traffic zone, and so forth.

On the map, numbers are assigned to each geographic unit and these numbers, or codes, are then inserted in the optional field. One more local action is necessary; a master list prepared to show the combination of codes required to provide data for each area.

The method creates a large number of sub-areas, smaller than tracts but larger than blocks, which can be combined to produce the information wanted for various areas.

The third, or Local Serial Number method, would permit almost unlimited use of the optional field. The flexibility, however, is at least partially offset by its complexity and somewhat greater cost.

In operation, it would require that a serial number be inserted in the optional field for each block face coded on the FOSDIC worksheets. That number would serve as a code link between local records and those of the Census Bureau and would enable the Bureau to extract information from its files to match local areas for which data are desired. One way in which local officials might handle the bookkeeping in this system is shown below.

Giantville / Tract	Local Serial No. of Block Face	Police Precinct	Transportation Zone	Sanitary District	School District	Improvement District
1	1	8	12	7	18	6
1	2	8	12	7	18	6
1	3	9	11	6	17	7
1	4	9	11	6	17	7
1	5	9	10	5	16	8

As can be seen, the block face in Tract 1 which has been identified as No. 1 in the optional field is a part of police precinct 8, transportation zone 12, sanitary district 7, school district 18, and improvement zone 6. The same type of record is shown for block faces 2, 3, 4, and 5 and can be applied to all block faces within a tract or other given area through use of maps and civic records.

The information collected on the tabular forms can be put into machine-readable form to show all of the administrative districts in which each block face lies.

When census information is required for any area for which block faces have been coded, local officials can prepare and send the Bureau a computer tape, or punched cards which include, for each block face, the census tract numbers, the local serial numbers, and the identification of the area for which a tabulation is needed, for example, the school district number. We can then match the tract and serial numbers to our own records and "instruct" our computers to prepare a tabulation for the school districts. The serial number, then, is a common identification, in your records and ours, to permit ready identification of the units we must aggregate to prepare tabulations that you need for local use.

Unlike the direct coding system, block faces can be shifted from one local area to another with no complications other than an adjustment of local records.

Numbers may be assigned in two ways in the Local Serial Number method. In either case they are chosen by the local group for entry in the optional field of the worksheets. The numbers may bear some systematic relationship to a set of locally defined areas or, more frequently, they will be arbitrarily assigned, that is 1, 2, 3, and so forth, as the name suggests. Alternatively numbers will be assigned, upon request, by the Bureau to each block face record as described below.

Marking serial numbers on FOSDIC worksheets can be a tedious and time consuming task with many chances for errors. A computer, however, can do this work rapidly and accurately once appropriate instructions have been written. In recognition of the savings that will result, we are prepared to assign unique identification numbers to each block face, within census tract, if asked to do so.

As FOSDIC worksheets are transferred to computer tape the computer will be programmed to identify all block faces for each block and to assign a two-digit "block face number" to each. These numbers may be changed from time to time. However, we can reproduce these numbers in the optional field on the computer tape and that identification will not be altered even though our "label" changes. For example, block face 12 of block 307 might become block face 03 of block 309 in our part of the record. Nonetheless, if we had entered 30712 in the optional field that identification would remain fixed despite other manipulations, just as if it were a number coded in the optional field initially, and the local participants would have a positive identification of that segment of the address coding guide, once they had received a copy of the product.

Note, further, that this would provide a "structural" code in that the first three digits would be a block number, the last two a block face number. This serves to simplify the preparation of a cross-reference table or dictionary such as the one illustrated above. Specifically, for each block that is not cut by a "local" boundary, and this should include the vast majority of blocks, only one record need be prepared for the block, rather than one for each block face. That record would have a "00," "99," "xx," or other distinctive symbol, yet to be specified, as the last two digits of the equivalent of "Local Serial Number" with a block number as the first three digits. Later we would instruct our computers to recognize that this symbol means "assemble all block faces (within tract) having the specified first three digits."

The fifth method, "No Coding," is included to indicate that "all is not lost" if the optional field is not marked, or if it turns out that a need was overlooked in using this field. In fact most of the benefits of this field can be realized even though it is left blank.

As noted above, we will assign a number to each block face within a block, a number that cannot be considered permanent for the 1970 Census until perhaps four months after the census enumeration. At that time, however, anyone who wishes can secure a copy of the appropriate maps and the address coding guide and can then record the Bureau's identifications of blocks and block faces that correspond to any area of interest. He can then list <u>our</u> identifications to match <u>his</u> areas and we <u>will be</u> able to prepare the desired tabulation.

The disadvantages of the "No Coding" method are two. First, the splitting of a single block face, cut by a local administrative boundary, into two block faces is not feasible. Second, the preparation of local records to relate local areas to Census area identifications cannot be carried to completion until the Bureau's identification of areas is stabilized for the 1970 Census. If, for example, we identify a block face as number 2 within a specific block we must be free to change it to, say, 7 if we wish. If block face 2 <u>is</u> changed to 7, any coding in the optional field for face 2 will be carried over to the new 7 and that link between your records and ours will be just as valid as before. However, any record you prepared based on the fact that we had assigned "2" to a specific block face would be of doubtful value if you could not be sure, and you wouldn't be, that we were not changing identifications to satisfy our internal requirements.

We believe this method will be especially useful as a supplementary aid to those who use the direct coding method and will influence many to use it.

Direct coding is inexpensive, locally, and permits retrieval of census information at minimum cost. The procedures noted above make it feasible, despite the limitations of direct coding, to retrieve other information, when it proves to be desirable to do so. Direct coding, we believe, is the best method for smaller metropolitan areas, especially those that do not now have an ongoing program of computerized data processing.

Geographic Unit Coding is an excellent method, with the exception of the problems that may arise when local area units have boundary changes. It opens up to users the possibility of obtaining census summary data tapes that they may use to regroup our data in various ways of local interest. The problem of preserving confidentiality and the problem of excessive sampling variability rule against the release of detailed census data for individual blocks and block faces and even for block clusters that are small in population. For clusters as large as census tracts, we have in the past provided quite voluminious summary data. We are working hard on the problem of how to present data for various types of small areas without disclosing the characterisitics of any individual or tabulating meaningless numbers. For clusters that contain 1,000 or more persons, I now believe we will be able to provide quite useful summary data tapes.

Local Serial Number coding involves the preparation of an extensive cross-reference table, but is an excellent method of great flexibility and presents no serious problems when local boundaries change, except where a new boundary cuts a block face. That problem, I suggest, is not a serious one.

Census Bureau Coding, in our view, provides all or nearly all of the benefits of Local Serial Number Coding and markedly reduces local effort and costs in "keying in" to census records. If you lean toward the serial number method, we suggest that this alternative is almost certainly preferable.

"No Coding" is an "escape valve" if local groups cannot agree on the use of the optional field, or if an unanticipated need arises.

Further variations are possible through combination of methods shown above. Direct Coding and Small Area Coding may be combined with each other or with Serial Number coding within the five-digit limit.

As an example, the local group may wish to use the first one or two columns in the optional field for small area identification and the last three or four columns for insertion of an arbitrarily chosen code number. Such a combination would result in ready tabulations for the directly-coded areas while at the same time maintaining in census data files the greatest capability for data tabulation for other areas.

I have no doubt that much of this sounds complicated to almost everyone except computer people. With that in mind I assure you that when interested groups with specific problems need further explanation or advice, we will do all we can to help.

Another point of interest; we plan to identify the locations of blocks or block faces by grid coordinates. While the word is "plan," not "promise," I believe we'll carry out this proposal.

Within the areas covered by address coding guides, we expect to have coordinates for block faces; for other parts of urbanized areas, coordinates for blocks. Our coordinates will be recorded in latitude and longitude to four decimal places, that is to 36 feet at most, but those who wish State plane coordinates, rather than latitudes and longitudes, will be able to secure them. The coordinates will be available, at cost of copying, to those who wish them and can be used, within the Bureau, for special tabulations you may desire.

That's all on that topic. More would bore many of you and those who wouldn't be bored can probably imagine potential uses far better than I can describe them.

Somehow we always get around to talking about money; in this case expense involved in local cooperation with the Bureau of the Census. While we believe that local financing of this effort would be successful in many areas, we are not so optimistic as to believe that a substantial proportion of metropolitan area officials will be both willing and able to provide the required funds on short notice. However, the Department of Housing and Urban Development shares our enthusiasm for the new methods and the potential benefits, and will, in about one month, formally announce a program through which eligible agencies can secure "701" grants to cover two-thirds of the costs involved in editing our maps and completing our FOSDIC worksheets.

Further, we are encouraged by the reaction of many top level planners that no metropolitan planning group that really wants a good information base for its work can fail to take advantage of this program.

This paper is a continuation of our efforts to tell local groups of our plans, efforts that will be continuing in the months ahead. We plan to bring our program to the attention of metropolitan area groups and officials of cities with 25,000 or more inhabitants. In this way we hope to secure the local aid we need to change "potential" to "reality" in speaking of the improvements in the 1970 Census.

1970 is still almost four years away, but it is now time to begin planning ways in which this new and different census can best be made to serve the nation at all levels. We'll be pleased to hear your suggestions and comments and, especially, your offers of assistance in our endeavors.

THE "MAIL-OUT/MAIL-BACK" CENSUS RESEARCH PROGRAM

David L. Kaplan, U.S. Bureau of the Census

The first reaction to the title of this paper might properly be "Why?" Do we really need a research program on data-collection techniques for the 1970 Census? After all, the results of the 1960 Census are accepted by the great mass of users as accurate and satisfactory for their purposes. The Census publications are neat and reasonably clear, the columns add up and the rows add across. But for an audience of statisticians, it is really unnecessary to explain that there were certain inadequacies and insufficiencies in the 1960 program which could be remedied at least partially by a more effective system of data collection.

Among the Decennial Census totems to which we statisticians at the Census Bureau pay constant homage are better coverage, improved quality of subject statistics, decreased cost, and speedier preparation of results. To help attain these and related objectives, we carry on a continuing research effort. The mail census program is part of this work.

As concerns coverage--that is, the count of people and their housing units--some techniques for reducing error appear feasible. For example, if we could compress the enumeration period into a shorter time, the likelihood of missing or double-enumerating people on the move could be minimized. Also, a gain could be made if the listing of dwellings was verified more thoroughly than the traditional approach of having the crew leader spot-check the enumerator's canvass. In this connection, a most useful contributor might be the letter carrier since he makes a virtual house-to-house tour every working day of the decade.

To improve the quality of subject statistics within the existing constraints of technical and financial resources, the best hope seems to be through the extension of the use of self-enumeration. Our experience in 1960 indicates that, on the whole, we get better data through self-enumeration than through enumerator interview. It is also true, of course, that the quality of subject statistics would be raised by improving coverage; significant distortions in certain subject distributions undoubtedly result from the differential underenumeration among various population groups and geographic areas.

To cut the cost of data collection, the most promising approaches would appear to be (1) a reduction in the house-to-house canvass by the enumerator and (2) the greater use of telephone to obtain missing information. On the former point, the mailman might well play an important role.

With regard to earlier completion of the results of the census, the data-collection system can contribute in two ways. One is by completing and sending in the census questionnaires in such manner that they can be processed most efficiently. The second way is quite obvious--the sooner the field work gets done, the sooner the processing can begin, and the sooner the results can be compiled.

Another problem which arose in 1960 and which is likely to plague us again in 1970, is the difficulty of recruiting and maintaining a temporary field staff to collect the information. Using the 1960 data-collection procedure again in 1970 would require approximately 200,000 field workers. Considering labor market conditions and the rates of pay which the census can offer, there is likely to be serious difficulty in staffing such a field organization, particularly in the big cities. To the extent that the required number of temporary field workers can be decreased, the problem would, of course, be diminished.

These and related considerations led us early in the 1960's to embark on a research program to test the expanded use of the mails for the 1970 Census. We had had considerable experience with the use of the mails in the 1960 Census and in certain pre-1960 tests. In 1960, an unaddressed Advance Census Form limited to the 100-percent subjects was delivered to all households across the United States by the Post Office. The householder was asked to fill out the form and hold it for the arrival of the enumerator. Also in 1960, for about four-fifths of the country, those households in the 25-percent sample were left with a sample questionnaire booklet which they were requested to fill out and mail back within three days. Our experience with this mail-back approach in 1960, the large-scale experiments conducted during the 1950 Census in the Columbus, Ohio, and Lansing, Michigan, areas, and a substantial test conducted in conjunction with a special census of Memphis city in 1958, all encouraged the belief that a mail-out/mail-back technique might be developed which would yield gains along some and perhaps all of the fronts previously mentioned.

Looked at operationally, there are three vital keys to the success of a mail census. One is the adequacy of the mailing list. The second is the public response which determines the amount and complexity of the follow-up work required. The third is whether the management of a mail census field office can be handled successfully by a temporary organization.

We started the systematic research program in August 1961 with a test in Fort Smith, Arkansas, a city of about 60,000 population. The enumerators made a house-to-house canvass and left a very short questionnaire for householders to mail back. Approximately 86 percent of the households mailed back the questionnaire. Also the list of addresses prepared by the enumerators was used to check an office-generated register of addresses which was based on our 1960 Census records updated through recent building permits and checked by the local Post Office. The results indicated that the corrected office-generated register was at least as complete as the enumerator listings.

In June 1962, we went back to Fort Smith and also went to Skokie, Illinois (another city of about 60,000 population) to test another variant of this approach. We had the Post Office deliver questionnaires addressed from an updated 1960 Census list. A coverage check showed the miss rate for housing units to be about one percent, which compared favorably with the 1960 Census national experience. In each city, close to three-quarters of the householders mailed back their questionnaires.

In April of 1963 we experimented again, this time in Huntington town, Long Island, New York, covering about 150,000 population. For most of the area, we based the mail-out on an updated 1960 Census address register, with a post office check for omissions at the time of mail delivery. In the remaining, essentially rural portion of the area, enumerators made a house-to-house canvass during which they left questionnaires to be mailed in. Also, for the first time in this series of tests, we used both long and short questionnaires. We thus simulated the 1960 (and anticipated 1970) approach whereby most of the population is requested to answer only a limited number of questions and a sample of the population is asked to answer the full range of subjects covered by the census. A coverage check revealed the miss rate for housing units in the address-register portion of the area to be about one percent. The mail return rate was 85 percent for both the short and long questionnaires, a very encouraging sign that the public was willing to cooperate even in answering an extensive questionnaire.

Our experience to date was, thus, quite encouraging on all fronts--public cooperation appeared to be on a satisfactory level, the potentiality for developing an effective mailing list was indicated, and the feasibility of the operation in the field seemed probable. We, therefore, embarked on a larger research program financed by a special appropriation for this purpose. It is noteworthy that officials of the Commerce Department, Budget Bureau, and the Congress recognized so early in the decade that preparatory work on the next Decennial Census should begin and that a feasibility testing program should be financed; more than one million dollars was spent on two large-scale experimental censuses in 1964 and 1965.

In May 1964, we took a special census of the Louisville, Kentucky-Indiana, Standard Metropolitan Statistical Area, covering close to 800,000 population. For mail-out purposes in the city portion of the area, we used the 1960 Census address list, updated with building permit and utility data. Outside the cities, we used a list prepared through a house-to-house canvass by Census personnel about a month prior to Census Day. At mail delivery time, the letter carriers informed the local census office of residential addresses for which they had no questionnaires; the missing addresses were added to the mailing list. Short and long questionnaires were sent out, essentially repeating the subject content of the 1960 Census. The questionnaire return rate was 88 percent, with only a few points difference between the short- and long-form rates. An intensive sample check showed the coverage errors for both occupied housing units and population to be below the 1960 national levels. (The 1960 program did not yield coverage error estimates for individual areas.)

For the second large-scale feasibility test census, conducted in April 1965, we selected the city of Cleveland, Ohio. We chose Cleveland purposely to provide a rigorous test of the mail technique in a large cosmopolitan industrial city in which we had substantial enumeration problems in 1960 and in which we could expect similar problems in 1965 and in 1970.

Some exploration after the Louisville project suggested that we could do better on several accounts by using a commercial mailing list for the register of addresses instead of going through a trasnfer to tape and updating of the addresses on the 1960 Census schedules. While commercial lists do not contain the detailed geographic codes needed for census purposes, they can provide certain important economies and flexibilities over the 1960 materials. The procedure we used was as follows. In December 1964, the commercial list of approximately 300,000 addresses for Cleveland city was printed out on individual labels, each of which was affixed to a card. In January 1965, the cards were given to the local Post Office to check for omissions, revisions, and deletions. The corrections provided by the Post Office were made on the computer tape record. This was done in February, at which time we also inserted the necessary geographic codes, applied the required field control numbers, and designated each fourth household to receive the long-form questionnaire. By early March, individual address labels showing the necessary control designations were printed out and the labels affixed to the appropriate mailing pieces. Also printed out and bound in a separate book was the list of addresses in each assignment area. By March 20, all of the mailing pieces were in the hands of the Cleveland Post Office.

On Monday, March 29, the mailing pieces were delivered by the letter carriers. As they sorted and delivered the census mail, the letter carriers were instructed to make note of any housing unit for which they did not have a mailing piece. These missing addresses were transmitted as rapidly as possible to the local census office, where they were added to the address list for the particular area and mailing pieces sent out.

Householders were requested to send back their completed questionnaires on Census Day--Thursday, April 1. This request was contained in the covering letter on the questionnaire and was also the subject of a very intensive local publicity campaign. The local office checked the mail-back questionnaires for acceptability. Unacceptable returns were followed up by enumerators through telephone or, if necessary, by personal visit. Nonresponse cases, including vacant housing units, were followed up through personal visit by enumerators.

Eighty percent of the householders returned their questionnaires and, as in Louisville, there was very little difference in the return rate as between the short and long forms. Another encouraging aspect was that the great bulk of returns arrived on or shortly after Census Day (also much like the Louisville experience). By April 3, questionnaires had been received from about 70 percent of the households, which meant that the count of more than two-thirds of the population had been accomplished in just a few days, and the chances for error resulting from movement for these people had been minimized. The coverage experience in Cleveland city for occupied housing units and population was on a par with or better than the national averages in 1960 after standardization of the Cleveland results for the proportion of units and population in multi-unit structures (where coverage is lower than in single-unit structures).

Our total experience in all of the research program to date has been gratifying on most all issues. At least in built-up areas which use city-type addresses, and perhaps even in rural areas, the mail approach appears to provide the various benefits we are looking for. There are still many problems to be resolved--for example, how to raise the acceptability rates for the long-form questionnaires, what procedure to use in rural areas, and how to get better coverage in the congested areas of our big cities. We will be working on them intensively in the time remaining before the final 1970 Census materials are locked up. But the decision has been reached to use the mail approach for the larger part of the American population.

Note should be taken of at least two important associated advantages of the mail system. One, it provides greatly increased accuracy in the identification of city blocks, and also permits the identification of individual sides of blocks in the highly-detailed computerized geographic coding system which the Census Bureau is developing for the 1970 Census. Second, the address list, if kept updated after the 1970 Census, will be useful for later sample selection purposes, for the 1980 Census or the proposed 1975 mid-decade census, and for the preparation of local population estimates.

Our next large experimental census is planned for the metropolitan area of New Haven, Connecticut, in April 1967, where we will be refining our techniques and testing subject content. Over the next year, we will also conduct a number of smaller field tests for the same purposes. By the fall of 1967, we will be in earnest preparation for the "dress rehearsal" census scheduled for April 1968. In that project, we will no longer be trying out alternatives; the focus will be on the processes and materials we expect to use in the 1970 Census itself.

MEASURING QUALITY OF HOUSING[1/]

Arthur F. Young, Joseph M. Selove, David A. Koons
U. S. Bureau of the Census

Abstract

Statistics on housing condition produced by the Bureau of the Census have heretofore been based on subjective ratings made by census enumerators. One of the major goals of the 1970 Census of Housing is that of replacing the subjective rating with one based on objective criteria. This paper reports the results of studies evaluating the reliability and accuracy of enumerator ratings. It also reports the results of a multi-variate analysis to measure the association between 1960 condition ratings and objective characteristics. It is concluded that adequate census tract statistics on housing condition could have been produced in 1960 from information supplied by householders. It is proposed that, in 1970, measures of housing quality consist of combinations of objective housing characteristics.

The development of a measure of housing quality which would produce reliable and accurate statistics has been one of the major concerns of the Bureau of the Census, beginning with the first census of housing in 1940. It has also posed one of its most difficult problems.

We have never considered it possible to conceive a single measure that could sum up all the components of housing quality. In 1960 we defined quality in the instructions we gave our census enumerators in these words.

> "Item H6 calls for information about the quality of housing. It tells how many housing units are not providing adequate shelter and are, in their present condition, endangering the health, safety, or well-being of their occupants. We call such units dilapidated. For the units that are not dilapidated, we need to know how many are in good repair and, therefore, sound; and how many are in need of repair and, therefore, deteriorating."

Except for the use of a two-fold, instead of a three-fold classification in 1950, the wording on this point was essentially the same in both censuses.

In the first census of housing the concept underlying our efforts to measure housing quality was the degree of hazard to safety. The enumerator was instructed that a housing unit was to be rated as "Needing major repairs," if the continued neglect of the needed repairs would "seriously impair the soundness of the structure and create a hazard to its safety as a place of residence." However, this rating measured only the physical condition without indicating the level of quality. A tar paper shack could be rated as "Not needing major repairs." In 1950, principally to remedy this defect, units were classified as "Dilapidated" or "Not Dilapidated."

In 1960 a three-way classification was adopted--Sound, Deteriorating, and Dilapidated. Special efforts were taken to make the 1960 definition of Dilapidation identical with that used in 1950. A Sound unit was defined as having no defects or only slight defects which normally would be corrected during the course of regular maintenance. A Deteriorating unit was defined as needing more repair than would be provided during the course of regular maintenance. A Dilapidated unit was defined as one in which the defects were either so critical, or so widespread that the structure should be extensively repaired, rebuilt, or torn down.

The application of the concepts was virtually the same in the last three decennial censuses. Enumerators were instructed to observe each unit and then make an overall judgment according to specified criteria. They were also instructed that their ratings were to reflect only the physical condition of the structure and the unit, and that such factors as neighborhood quality, race or color of inhabitants, for example, were not to be considered. Unlike 1940, in 1950 enumerators were provided with detailed written criteria and instructions as well as photographs depicting levels of condition. In addition, audio-visual techniques were used in training. The 1960 training techniques and instructions were essentially the same as in 1950, except that modifications were made to reflect the three-way classification.

The Census Bureau's condition classification has been combined with availability of plumbing facilities, by the Public Housing Administration, to form a classification which identifies housing as standard or substandard. Although these terms do not appear in census publications, beginning with the 1950 census, tabulations have been provided to which the "substandard" and "not substandard" labels could be directly applied. In terms of the published census categories, a "substandard" unit is:

1/ Based upon an unpublished report, Quality of Housing, by Leon Pritzker and Joseph Selove, April 1966, Bureau of the Census.

1. Dilapidated

    OR

2. lacks one or more of the following facilities:

    a. hot running water in the structure;
    b. flush toilet for private use;
    c. bathtub or shower for private use.

Since 1960 we have devoted considerable time and money to research on methods of improving the measurement of housing and neighborhood quality. The decision that the 1970 census will be conducted by mail has greatly influenced the conduct of the research. Our objective in examining the methods of measurement employed in the 1960 census was to establish their value in providing statistics which could serve as a standard of adequacy. In our judgment, the data provided in the 1960 census fell far short of satisfying the requirements of such a standard.

By adequate, we mean that the statistics should--

1. Provide reliable and accurate data with respect to current levels of quality;

2. be comparable geographically;

3. be built up from data obtained for individual housing units to which individual values should be assigned;

4. be based on methods that distinguish various levels of quality of individual housing units.

Our examination of the data we obtained in 1960 on quality of housing, and the methods we employed for that purpose, has led us to the following conclusions:

1. The 1960 census statistics on condition, that is, whether Sound, Deteriorating, or Dilapidated, are unreliable and inaccurate. On reliability, our best estimate is that if we had sent back a second group of enumerators to rate the housing units of the United States, only about one-third of the units rated as Dilapidated or Deteriorating by the first group would have been rated the same by the second group of enumerators. (See Table 1.)

    With regard to accuracy, it appears that the number of Dilapidated units was understated by at least one-third. The effect of this understatement is to grossly distort estimates of the trend in Dilapidated housing from 1950 to 1960. (See Tables 2 and 3.) The statistics on Dilapidated or Deteriorating housing for blocks appear to be of very low accuracy.

    Statistics for tracts within any given city are adequate, however. This finding is consistent with the others. The random errors of measurement (including enumerator variability) tend to cancel out on the tract level.

    There is little evidence which defines quantitatively the reasons why condition data were poor in quality in the 1950 and 1960 censuses. It is believed that one of the chief reasons is the subjective interpretation of the specific criteria by enumerators when making the overall judgment on condition. Some factors which may influence the enumerators are: (a) the socioeconomic background of the enumerators, or the quality of the neighborhoods and homes in which the enumerators themselves live; (b) the level of instruction given by the supervisory personnel; (c) race or color of the occupants; (d) neighborhood factors such as heavy traffic, noises and odors from commercial establishments, and mixed land usage; (e) general housekeeping habits of occupants, shabbiness of interior and exterior, and artificial fronts of structures. Condition is, at best, a difficult concept to apply, particularly for marginal units. The instructions may have been misinterpreted despite the fact that more time was devoted in training on condition than on any other housing item. Another contributing factor is that the overall determination of condition should be based on observation of the inside and outside of the entire structure. In any decennial census, the enumerator typically sees only one or two rooms or he may conduct the enumeration on the doorstep, or in the public hallway or foyer. In addition, the three-way classification of condition used in 1960 may have contributed to the poorer quality of reporting of condition because it increased the complexity of the rating process and provided the enumerator with an intermediate category in which to place doubtful cases.

2. The Public Housing Administration "Standard - Substandard" classification, although affected adversely by the poor quality of the enumerator ratings, appears to have been a more adequate measure of housing quality in 1960 than the classification based on structural condition alone. (See Table 3.) There is evidence that about one-fourth of the units which could be classified as "substandard" from the findings of one group of enumerators, would have been differently classified from the findings of a second group of enumerators.

    However, the trend appears to have been measured adequately for the decade 1950-1960. The erroneous classifications of structural condition were, in effect, corrected by the plumbing facilities data.

3. We see no feasible method of improving the quality of enumerator condition ratings in a decennial census. This is a consequence of the ambiguities, nonoperational elements, and complexities of the rating process itself, as well as of biasing factors in the environment in which ratings have to be made.

In anticipation of the possibility that the 1970 census would be taken by mail, studies were made (in 1962 and 1963[1]) to test the collection of data on structural condition through self-enumeration. These studies were conducted in the following four areas: the Washington, D.C.-Maryland-Virginia Standard Metropolitan Statistical Area; Philadelphia, Pennsylvania; Fort Smith, Arkansas; and Huntington Township in Suffolk County, New York. Their primary purposes were: (a) to determine how information on condition obtained by mail enumeration would compare with that obtained by direct enumeration in the 1960 census or by ratings from "experts," and (b) to determine the relationship of tenure (whether owner- or renter-occupied) of unit and question wording to respondents' replies.

It was assumed that it would not be feasible to ask the respondents in a mail enumeration if their units were Sound, Deteriorating, or Dilapidated. Therefore, the kinds of questions varied from a set of comparative terms ranging from "excellent" to "very bad," to lists of deficiencies about specific parts of the structure such as the roof, porch, exterior walls, etc. The responses to these questions did not automatically classify units into the three-way classification of condition. It was necessary to devise methods which "translated" the respondents' answers into three categories of condition used in 1960. The results were then compared with those from the 1960 census and ratings made by "experts" who were representatives from the Bureau of the Census, Bureau of the Budget, and the housing agencies. In interpreting the results, it should be recognized that considerable difficulty was experienced in translating the respondents' replies. In addition, the number of cases in these studies was quite small and the only city where a probability sample was used was in Fort Smith, Arkansas.

Data for Fort Smith, Arkansas, (see Table 4) are generally consistent with the findings for the other areas included in this study. In this test, the experts determined condition for the sample units by using the 1960 concepts and instructions. The respondents indicated the condition of their units by checking a set of comparative terms and a list of objective deficiencies. The respondents' replies were then translated into categories which were as close as we could come to the 1960 classification.

[1] "Self-Enumeration of Housing Condition," Housing Division, Bureau of the Census, 1964.

The data in Table 4 indicate that, as it was used in Fort Smith, self-enumeration does not yield condition data comparable to those obtained in direct enumeration, particularly for units at the lower end of the scale. The major problem of measuring condition through self-enumeration under these conditions is that householders rate their units more favorably than "experts," homeowners to a greater extent than renters. While it is recognized that the experts tend to be more critical in their ratings, the same general bias by tenure is found when the results of self-enumeration are compared with those obtained by the 1960 enumerators.

Following these experiments with self-enumeration we turned to regression methods to help identify the characteristics which are related directly to quality of housing. Utilizing data collected in the 1960 census, the analyses related four measures of housing quality to "objective" population and housing characteristics. (See Attachment 2.) The four measures of housing quality were:

1. proportion Deteriorating,
2. proportion Dilapidated,
3. proportion Deteriorating plus Dilapidated,
4. proportion "Substandard."

The results we have obtained to date are based upon block, enumeration district, and tract data from the 1960 census for six cities. Four of these six were chosen at random from cities of population between 100,000 and 200,000 in 1960. The other two, Louisville and Cleveland, were selected because a census had been recently completed in each of these cities.

The independent variables were grouped to provide three separate analyses for each dependent variable and for each type of area. The groups were: "population" variables, "housing" variables, and the combination of the two. The classification of the independent variables as "population" and "housing" variables was arbitrary. The "housing" variables were defined as those variables that could be measured by a census of housing in which no population data at all were collected. Some of the "population" variables, on the other hand, reflect the characteristics both of the housing units and of their occupants (e.g., persons per room).

The computations of coefficients of multiple correlation are based upon census tapes containing data for the 25-percent sample of housing units in 1960.

Our research has indicated the following about the 1960 condition rating and its relation to objective characteristics.

1. There is a gradient in the coefficients of multiple correlation by size of area. The correlations are lowest for

blocks, increasing somewhat for enumeration districts, and are at a maximum for tracts. (See Table 5.) Our analyses thus far indicate that errors in the ratings themselves play an important role in this gradient.

2. At the tract level, the amount of variance explained is, in our judgment, high enough to warrant the use of objective characteristics as substitutes for the enumerator rating of condition. Although the evidence is not as direct as for tracts, our judgment is that objective characteristics will differentiate blocks and enumeration districts, as well as tracts, with respect to quality of housing.

   Comparison of the rank orders of areas based on regressed values and on actual ratings provides one type of operational translation of what these coefficients of correlation signify. Census tracts in Fort Wayne, Indiana, one of the six cities used in our regression study, were ranked by quartiles according to the percent of units which were Deteriorating or Dilapidated in 1960, both according to the actual ratings obtained in 1960, and the values predicted by the regression equations. The tracts which were placed in identical quartiles by both the observed and the predicted values contained 90 percent of the units which were Deteriorating or Dilapidated. (See Table 6.)

3. The estimated multiple correlations for all types of areas--blocks, enumeration districts and tracts, were found to be generally highest where the proportion "substandard" was used as the dependent variable.

   One set of estimates included as an independent variable the percent of units with bathrooms for private use, the other set excluded this variable. We recognize that the standard-substandard classification itself gives heavy weight to the presence of a bathroom, and thus, the inclusion of this variable may appear to involve some type of circular reasoning. However, the presence or absence of a bathroom for private use is an "objective" factor that can be obtained readily in a self-enumerative census. Thus, it seemed appropriate to check on the hypothesis that the presence or absence of a bathroom <u>plus</u> other "objective" factors can substitute effectively for the present definition of the standard-substandard classification. The evidence clearly supports this hypothesis. The bathroom variable produces substantial increases in the correlations.

The statistics based on the "standard-substandard" classification appear to have played a much more important part in the establishment of housing policy and in the execution of housing programs than the statistics on structural condition. A comment is in order therefore about its usefulness as a measure of housing quality in 1970.

One important reason for the adequacy of the statistics based on the "standard-substandard" classification is that the statistics reflect primarily an "objective" characteristic of housing units--the extent of availability of plumbing facilities. In the 1950 census, only about four percent of the units which were classified as "substandard" were Dilapidated units with all plumbing facilities. In the 1960 census this rose to about nine percent. However, in the large cities (cities of 100,000 inhabitants or more) the Dilapidated units with all plumbing facilities made up 15-20 percent of the total "substandard."

Some housing analysts believe that, because of the increased enforcement of housing codes since 1960, there has been widespread installation of plumbing facilities in poor housing. This installation is sufficient to classify low-quality housing as "standard." Thus, the contention is that in 1970 the "standard-substandard" classification will no longer help identify housing of low quality. We would agree that the identification of housing and neighborhood quality by this system will probably be less effective in 1970 than it was in 1960. The question is one of degree, however.

If it is accepted that a modified "standard-substandard" classification is the most useful one, then to overcome this difficulty, the Bureau of the Census must find ways of strengthening this classification by the addition of other directly measurable components of housing quality. We are preparing to test other objective factors for this purpose. This brings me to a discussion of our present plans.

We have considered a number of methods for identifying the quality level of individual units. Three of these start by defining a value on the quality scale underlying each characteristic below which a unit would be classified as inadequate with respect to that characteristic. (For example, a unit could be classified as inadequate if it did not have a complete bathroom for private use.)

The alternatives are:

1. A simple count of inadequacies for each unit.

2. A weighted count of inadequacies where each one is weighted in terms of its importance for a given index of quality of housing.

3. For block tabulations we could identify units without a complete bathroom for exclusive use, and units with a complete

bathroom, but with a number of specified deficiencies. This implies, of course, that the availability of a private bathroom for exclusive use will be taken to indicate inadequate housing, without further qualifications.

We plan to conduct a series of surveys to develop and test measures of housing quality which will consist of combinations of objective characteristics. At present, the list consists of these items:

1. Availability of a complete bathroom, for private use.

   (A complete bathroom includes a flush toilet, a bath or shower with piped hot and cold water, and a lavatory with hot and cold piped water.)

2. Availability of a complete kitchen, for private use.

   (A complete kitchen includes a refrigerator, cook stove, and a sink connected to piped hot and cold water, all in the same room or space.)

3. Kind of heating equipment.

4. Age of structure.

5. Number of closets.

   (This will be expressed as a ratio of closets per room.)

6. Whether there is a telephone in the unit.

7. The rent paid or asked.

8. Value, or price asked.

We have two objectives in these surveys. The first is to determine how well a measure consisting of objective housing characteristics compares with values obtained by an intensive appraisal of housing quality. The second is to determine the sensitivity of our proposed measures to city size and area differences.

To meet the first objective, the survey plan calls for a comparison between measures based on the items I have just described, and values obtained in an intensive appraisal of housing quality. For this purpose we may use some version of the method developed some years ago by the Committee on the Hygiene of Housing of the American Public Health Association. This method is marked by the use of a rating scale on which penalty points are assigned for certain deficiencies in the structure and the unit, as well as those in the physical neighborhood environment. The number of penalty points assigned to a specific item increases with the seriousness of the condition. Regression methods will be used to determine what combinations of housing characteristics are most closely related to the total penalty score, and its subtotals. The APHA penalty scores, and components of those scores will serve as the dependent variables. The characteristics collected on a census schedule supplemented in part by some of the items collected as part of the intensive appraisals will serve as the independent variables.

To meet the second objective, surveys will be conducted in cities selected to provide representation by size of city (population) and geographic region.

Five aspects of this program require emphasis:

First, our measures will be applicable to individual housing units.

Second, the principal purpose of whatever measure we employ, will be to identify housing which gives indications of having a high probability that it is hazardous to health and safety.

Third, by basing our measures of housing quality on objective characteristics, we afford ourselves a greater degree of flexibility than was previously possible in differentiating between levels of quality. Just as mail order catalogues distinguish between good, better, and best merchandise, so, the Bureau of the Census may find it possible in the future to differentiate not only between adequate and inadequate housing, but also to divide the adequate housing into classes of "good," "better," and "best."

Fourth, as we increase the number of objective characteristics, we may also find it possible to develop a measure of neighborhood quality.

Fifth, we will have a much greater degree of assurance than formerly, that the data we will provide on housing quality in one census will be comparable with data which will be collected in censuses to follow. This follows from the fact that measures of housing quality will reflect not judgment, which is subject to a high degree of individual bias, but rather descriptions of the actual facilities and characteristics of the housing unit.

Table 1.--CONDITION OF OCCUPIED HOUSING UNITS; 1960 CONTENT EVALUATION ESTIMATES DISTRIBUTED BY CONDITION RATING IN THE 1960 CENSUS OF HOUSING1/

(In thousands)

Evaluation program rating		Census rating		
		Sound	Deteriorating	Dilapidated
NUMBER				
Total..................	48,853	40,485	6,255	2,113
Sound........................	38,751	35,792	2,703	256
Deteriorating................	6,820	3,928	2,275	617
Dilapidated..................	3,282	765	1,277	1,240

		Total	Sound	Deteriorating	Dilapidated
PERCENT					
Total..................	100.0	100.0	82.9	12.8	4.3
Sound........................	79.3	100.0	92.4	7.0	0.7
Deteriorating................	14.0	100.0	57.6	33.4	9.0
Dilapidated..................	6.7	100.0	23.3	38.9	37.8

1/ Units correctly included in census, and for which condition ratings were obtained or imputed in both the census and the evaluation study.

SOURCE: U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960. Accuracy of Data on Housing Characteristics. Series ER 60, No. 3, Washington, D. C., 1964, table 2A.

Table 2.--ESTIMATES OF NET ERROR IN NUMBERS AND PERCENTAGE DISTRIBUTIONS OF OCCUPIED UNITS CLASSIFIED BY INDICATORS OF QUALITY OF HOUSING, 1960 vs. 1950

Classification	Estimated net error (Census estimates minus evaluation program estimates)					
	Number		Percentage points		Relative error $(\frac{C-EP}{EP} \times 100)$	
	1950 (000)	1960 (000)	1950	1960	1950	1960
Sound...........................	x	1,734	x	3.6	x	4.5
Not Sound.......................	x	-1,734	x	-3.6	x	-17.2
Deteriorating...................	x	-565	x	-1.2	x	-8.3
Not Deteriorating...............	x	565	x	1.2	x	1.3
Dilapidated.....................	343	-1,169	0.9	-2.4	10.3	-35.6
Not Dilapidated.................	-343	1,169	-0.9	2.4	-0.9	2.6
Substandard1/....................	-48	-585	-0.1	-1.2	-0.3	-7.3
Not Substandard1/...............	48	585	0.1	1.2	0.2	1.4

1/ Not an official designation by the Bureau of the Census. Includes units which were Dilapidated, or, if not Dilapidated, lacked one or more of the following: piped hot water in the structure, flush toilet for exclusive use, bathtub or shower for exclusive use.

x--Classification not available for 1950.

SOURCES: (1) 1950 Census data from U.S. Census of Housing: 1960, Volume IV, Components of Inventory Change. Final report HC(4), Part IA, No. 1, table 1.

(2) 1950 data from U.S. Bureau of the Census, Bureau of the Census Technical Paper No. 4, 1960, table 14; and Post-Enumeration Survey, 1950, Results Memorandum #22, December 22, 1953, table 1.

(3) 1960 data from U. S. Bureau of the Census, Accuracy of Data on Housing Characteristics, Series ER 60, No. 31, 1964, tables 2A and 3A.

Table 3.--ESTIMATES OF NET ERRORS IN TRENDS OF OCCUPIED UNITS CLASSIFIED BY CENSUS INDICATORS OF QUALITY OF HOUSING, 1960 vs. 1950 1/

Classification	Differences, 1960-1950 according to: Censuses of Housing in:— Number of units (000)	Censuses of Housing in:— Percent	Evaluation studies in:— Number of units (000)	Evaluation studies in:— Percent
Dilapidated	-1,635	-41.9	-6	-0.2
Not dilapidated	11,690	29.9	9,318	23.2
Substandard	-6,782	-44.5	-6,829	-43.1
Not substandard	16,837	60.8	16,134	57.5

1/ Based on adjusted data. 1950 evaluation program data adjusted for nonresponse. 1960 evaluation program data adjusted for coverage error and nonresponse.

SOURCE: See table 2.

Table 4.--PERCENTAGE AGREEMENT BETWEEN "EXPERT" AND RESPONDENT RATINGS OF CONDITION FOR IDENTICAL UNITS BY TENURE, FOR FORT SMITH, ARKANSAS: 1962

"Expert" rating	Total number of units	Percent	Respondent rating as percent of "expert" rating: Sound	Probably sound	Probably not sound	Not sound
Owner-occupied units:						
Sound	10,788	100	94	4	2	0
Not sound 1/	931	100	44	29	27	0
Renter-occupied units:						
Sound	4,388	100	83	13	2	2
Not sound 1/	1,795	100	46	20	6	28

1/ The category "not sound" is a combination of Deteriorating and Dilapidated as defined for the 1960 Census.

Table 5.--UNWEIGHTED AVERAGE COEFFICIENTS OF MULTIPLE CORRELATION FOR BLOCKS, E.D.'s, AND TRACTS IN SIX CITIES

Classification and variable group	Average of R		
	Blocks	E.D.'s	Tracts
I. HOUSING VARIABLES			
A. Deteriorating	.512	.661	.833
B. Dilapidated	.465	.626	.795
C. Deteriorating and Dilapidated	.587	.709	.856
D. Substandard--bathroom variable excluded 1/	.579	.730	.884
E. Substandard--bathroom variable included 1/	.820	.908	.960
II. POPULATION VARIABLES			
A. Deteriorating	.539	.700	.914
B. Dilapidated	.439	.619	.847
C. Deteriorating and Dilapidated	.596	.725	.922
D. Substandard	.669	.835	.955
III. HOUSING AND POPULATION VARIABLES COMBINED			
A. Deteriorating	.579	.740	.932
B. Dilapidated	.511	.698	.896
C. Deteriorating and Dilapidated	.648	.773	.944
D. Substandard--bathroom variable excluded 1/	.708	.865	.966
E. Substandard--bathroom variable included 1/	.839	.928	.980

1/ Bathroom variable is percent of units with bathroom for private use.

Table 6.--ESTIMATED QUARTILE DISTRIBUTION OF TRACT RANKINGS BY PERCENT OF ALL UNITS WHICH WERE "DETERIORATING" OR "DILAPIDATED" IN 1960, CROSS-TABULATED BY QUARTILE DISTRIBUTION OF TRACTS ACCORDING TO REGRESSED VALUES--FT. WAYNE, INDIANA 1/

Quartile rank of tract according to observed value	Quartile rank according to regressed value			
	1	2	3	4
A. NUMBER OF TRACTS				
1	10			
2		8	1	1
3		2	6	2
4			3	6
B. NUMBER OF DETERIORATING AND DILAPIDATED UNITS				
1	4,548			
2		1,515	265	140
3		157	347	97
4			72	106
C. PERCENT DISTRIBUTION OF DETERIORATING AND DILAPIDATED UNITS				
1	62.8			
2		20.9	3.7	1.9
3		2.2	4.8	1.3
4			1.0	1.5

1/ Based on regressions computed specifically for Ft. Wayne. Input from data contained in the 25-percent sample detailed record of the 1960 Census.

Attachment 2

List of variables used in stepwise regression for six-city analysis

Variable	Numerator	Denominator
1. Median family income	--	--
2. Percent Negro population	Number of Negroes	Total persons
3. Percent nonrelatives of head	Number of nonrelatives	Total persons
4. Percent persons 25+ years of age with less than 5 years of school	Persons 25+ years of age w/less than 5 years of school	Persons 25+ years of age
5. Percent persons 25+ years of age with less than 8 years of school	Persons 25+ years of age w/less than 8 years of school	Persons 25+ years of age
6. Percent of families with less than $3,000 of income	Families with less than $3,000 of income	Total families
7. Percent of persons unemployed	Unemployed persons in civilian labor force	Persons in civilian labor force
8. Percent of persons employed as household workers, service employees, or laborers	Persons employed as household workers, service employees, laborers	Employed persons
9. Percent of females employed as household workers	Females employed as household workers	Employed females
10. Percent of owner-occupied units	Owner-occupied units	Occupied units
11. Percent of units occupied by nonwhites	Units occupied by nonwhites	Occupied units
12. Percent of units classified "deteriorating"	"Deteriorating" units	All units
13. Percent of units classified "dilapidated"	"Dilapidated" units	All units
14. Percent of units classified "deteriorating" or "dilapidated"	"Deteriorating" or "dilapidated" units	All units
15. Percent of units "substandard"	"Substandard" units	All units
16. Percent of units with bathroom for exclusive use	Units with bathroom for exclusive use	All units
17. Percent of units in structures containing 3 or more units	Units in structures with 3+ units	All units
18. Percent of units in structures built 1939 or earlier	Units in structures built 1939 or earlier	All units
19. Percent of units heated by "other means with flue," "other means without flue," or "not heated"	Units heated by "other means with flue," "other means without flue," or "Not heated"	All units
20. Percent of units heated by "other means without flue" or "not heated"	Units heated by "other means without flue" or "not heated"	All units
21. Percent of occupied units with one person occupancy	Occupied units with one person occupancy	Occupied units
22. Percent of occupied units with 1.01 or more persons	Occupied units with 1.01+ persons per room	Occupied units

Variable	Numerator	Denominator
23. Percent of occupied units occupied by recent movers	Occupied units occupied by recent movers	Occupied units
24. Percent of owner-occupied units valued under \$10,000 or renter-occupied units with gross rent less than \$80	Owner-occupied units valued under \$10,000 and renter-occupied units with gross rent less than \$80	Occupied units except renter-occupied with "no cash rent"
25. Percent of owner-occupied units valued under \$10,000 or renter-occupied units with gross rent less than \$60	Owner-occupied units valued under \$10,000 and renter-occupied units with gross rent less than \$60	Occupied units except renter-occupied with "no cash rent

BIBLIOGRAPHY

Post-Census Evaluation of Housing Quality Statistics:

1. U. S. Bureau of the Census, The Post-Enumeration Survey: 1950. Bureau of the Census, Technical Paper No. 4, Washington, D. C., 1960. table 14.
2. U. S. Bureau of the Census, Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: Accuracy of Data on Housing Characteristics, Series ER 60, No. 3, Washington, D. C., 1964. tables 2A, 2B, 2C, 3A, 3B, 3C, 10.

A Critical View of Census Statistics on Housing Quality:

3. Wurster, Catherine Bauer. Housing and the Future of Cities in the San Francisco Bay Area. Institute of Governmental Studies, University of California, Berkeley, 1963. pp. 13-15
4. Duncan and Hauser. Housing A Metropolis--Chicago. The Free Press of Glencoe, Illinois, 1960. Chapter 2.

Suggested Criteria for Defining Housing Quality:

5. Dewhurst, J. Frederic and Associates. America's Needs and Resources. The Twentieth Century Fund, New York, 1947. pp. 157-164.
6. Dewhurst, J. Frederic and Associates. America's Needs and Resources, A New Survey. The Twentieth Century Fund, New York, 1955. pp. 219-222.

Some Intangible Measures of Housing Quality:

7. Beyer, Glenn H., Housing and Society. The Macmillan Company, New York, 1965. Chapter 8.

An Intensive Appraisal Method of Housing and Neighborhood Quality

8. American Public Health Association, Committee on the Hygiene of Housing. An Appraisal Method for Measuring the Quality of Housing: A Yardstick for Health Officers, Housing Officials and Planners. New York, 1946.

A Study of the Relationship Between the Components of the APHA Dwelling Appraisal to Aspects of Health, Safety, and Welfare:

9. Wilner, Walkley, Pinkerton, and Tayback. Housing Environment and Family Life; A Longitudinal Study of the Effects of Housing on Morbidity and Mental Health. The Johns Hopkins Press. Baltimore, 1962.

The Usefulness of the "Standard-Substandard" Measure of Housing Quality:

10. Fisher, Robert Moore. 20 Years of Public Housing. New York, Harper Bros., 1959, p. 30.

II

STAGNATION IN DECLINE OF AMERICAN DEATH RATE

Chairman, CLYDE KISER, Milbank Memorial Fund

CHANGE IN MORTALITY TREND IN THE UNITED STATES AND CERTAIN OTHER COUNTRIES

Iwao M. Moriyama, National Center for Health Statistics

After a long period of rapid decline in mortality, the downward trend lost its momentum during the 1950 decade in the United States (See Figure 1). The death rate began to level off about 1950 and has been stagnant for the past 15 years.

A number of countries under enemy occupation experienced excess mortality during the war years, but after the cessation of hostilities there appeared to be a resumption of the downward trend of the prewar period. However, in the 1950 decade a marked slowing down in the rate of decline began to take place. The same phenomenon has occurred elsewhere. There has been either a marked change in the rate of decline or relatively little change in the death rates since 1950 in the following countries: Sweden, Finland, Norway, Denmark, the Netherlands, England and Wales, Austria, Belgium, Czechoslovakia, France, Italy, Chile, Japan, Australia and New Zealand. The movement of the death rate in recent years has been distinctly upward in Norway, the Netherlands and Denmark.

In the United States, the change in trend has occurred in the death rates for both males and females, and for whites and nonwhites. The death rate has leveled off for residents of almost every State. In a number of States, the crude death rate is now rising.

The leveling off effect may be seen in the death rate for each age group up to 55 years (See Figure 2). The effect is much more marked in the younger ages where the slope of the trend line has been steeper.

Of particular interest and significance is the trend of the infant mortality rate (See Figure 3). During the period 1933 to 1949, the infant mortality rate for the United States decreased about 4 percent each year. Beginning about 1950, the rate of decrease in infant mortality dropped to 1 percent per annum. For nonwhite infants the deceleration in the downward trend is even more marked. During the years 1933 to 1949, the mortality rate for nonwhite infants decreased 4.6 percent per annum. Between 1950 and 1964, the rate of decline slowed down to 0.6 percent per year. As a consequence, the gap between the rate for white and nonwhite infants has widened during the past decade.

The trend of neonatal mortality rates as well as that for the postneonatal period has been affected. The infant mortality rate for residents of urban places has leveled off as has the rate for rural residents. The change in infant mortality trend has not been limited to any particular section of the country, although the degree of leveling off of the rates has not been uniform.

Because of the change in the infant mortality trend in the United States, the present situation in this country is less favorable relative to other countries than was the case 25 to 30 years ago. The infant mortality rate for the United States is currently significantly higher than the rates for the various Scandinavian countries, the Netherlands, Switzerland, England and Wales, Australia, New Zealand and several other countries such as Japan, Czechoslovakia and France. Although there are indications of changes in trend for other countries such as Australia, England and Wales, New Zealand, Norway and Sweden, these changes are not nearly as great as those experienced in the United States. On the other hand, the rates for other countries of low mortality continue to decline. Therefore, the United States is steadily losing ground with respect to its international ranking of infant mortality rates.

There are a number of possible reasons for the change in the mortality trends. For example, it is possible for the death rate to level off as a result of changes in completeness of death registration, definitions or in statistical practices. It may be due to errors in intercensal population estimates, or in the statement of age on death certificates or in the census enumeration. However, there does not seem to be any evidence to support the view that the changes in trend are statistical artifacts.

The studies of mortality trends in the United States, England and Wales and in Chile indicate that the leveling off of the death rates can be accounted for by a combination of two sets of factors. The first is the dramatic drop in the death rates for diseases of infectious origin with the successive introduction and application of pneumonia serum therapy, the sulfa drugs and the antibiotics. The accelerated decline started about 1938 and then lost its impetus in the 1950's. By that time, the mortality from diseases of infectious origin had reached a level where it no longer contributed in a major way to the total number of deaths. Even if the trend of the death rates for the infective and parasitic diseases, including pneumonia and influenza, had continued downward without interruption, this would not have accounted for all of the leveling off of the total death rate.

The long-term decline in mortality from the infectious diseases resulted in a major realignment of the principal causes of death which uncovered a second set of factors. These factors involve the trends of mortality from the presently numerically important causes of death, namely, malignant neoplasms and cardiovascular-renal diseases at all ages, congenital malformations through the childhood years, accidents and other violence from childhood through middle age, cirrhosis of the liver in middle age, and diabetes mellitus from middle age into old age. Also, new problems are emerging. The dramatic upward trend of the chronic bronchopulmonary disease mortality from middle age onward seems particularly significant. The combined effect of these various trends is to slow down the rate of decline of the total death rate.

It is difficult to determine precisely the roles played by the different factors in affecting the general mortality trend. The indications are that the leveling off of the death rates for the infective and parasitic diseases, influenza and pneumonia, and the other diseases of the respiratory system did not account for a large part of the deceleration of the rate of decline of the death rates for all causes for most age groups. Much more important seem to be the trends for the chronic diseases such as cardiovascular-renal diseases and malignant neoplasms in the adult population, malignant neoplasms and accidents and other violence in the younger population, and congenital malformations among children. These diseases and conditions constitute the hard core for which prevention of deaths is more difficult.

An interesting and important question is why the rate of decline of the pneumonia death rate changed during the past decade after a period of impressive decrease. These changes in the rate for age groups in which the death rate has reached a low level are understandable. However, the pneumonia death rates are still high for the older age groups, and for infants and preschool children. Similarly, the rates for nonwhites are still higher than those for whites and the rates for males are higher than those for females. There is considerable spread in the level of the death rates for the various age, color, and sex groups, but the leveling off of the pneumonia death rate for all groups has taken place about the same period. This parallelism in the trends by sex, color, and age appears too regular to be true. It seems strange that these changes should have occurred at about the same time. For some segments of the population the pneumonia death rates are so low that for all practical purposes, the notion of an irreducible minimum can be accepted. For the other subgroups of the population, it would seem that there should be further prospects of reduction in the death rate. Actually, the pneumonia death rate is now increasing for the older age groups, indicating the possibility of antibiotic-resistant organisms playing an increasing role in older pneumonia patients.

With regard to the total death rate, there may be some question as to whether the irreducible minimum has been reached. The examination of the death rates by cause of death indicates that further declines are possible. Also, comparison of death rates by age and sex for the various countries of low mortality for the years 1959 and 1960, depending upon their availability, shows that the structure of the death rate for the United States is far from the lowest. If the lowest age-sex specific death rate achieved by any country of low mortality in 1959 and 1960 had been obtained in the United States in 1960, there would have been about 397,000 fewer deaths in the United States. This means that the crude death rate for the United States would have been 7.3 per 1,000 population as compared with the recorded death rate of 9.5 per 1,000 population. For males, the expected death rate would have been 7.8 as compared with the recorded rate of 11.0 per 1,000 population. For females the corresponding rates would have been 6.9 as compared with 8.1 per 1,000 population.

It is difficult to say what the biological irreducible minimum is. However, it is obvious that the death rates in the United States are far above the low levels established in the Scandinavian countries and in the Netherlands. Even for these countries, the rates cannot be considered the lowest possible figures because lower mortality rates than those recorded in 1959 or 1960 have been observed for some age groups in the years prior to 1959.

The recent change in the rate of decline of the death rates in the United States does not appear to be a temporary phenomenon. The provisional death rate of 9.4 per 1,000 population in 1964 and 1965 are low relative to past rates, but they are far from the continuation of the downward trend established between 1938 and 1954. Further decreases may be recorded in the future, but it seems unlikely that the death rate for the United States will soon approach the levels already attained by various other countries.

Summary and Conclusions

After a long period of rapid and substantial decline, the death rate for the United States has reached the point where further decreases as experienced in the past cannot be anticipated. Similar changes in the trend of the crude death rate may be observed in the data for a number of other countries.

In the United States, the mortality trends for every age group from infancy to late middle age or old age--whites as well as nonwhites, and females as well as males--have changed in recent years.

Further declines in mortality are possible but they are likely to be modest ones. In view of the mortality experience for the past decade, it does not seem likely that the death rate for the United States will soon approach the levels already attained by various other countries.

FIGURE I. CRUDE DEATH RATES: UNITED STATES, 1900-1965

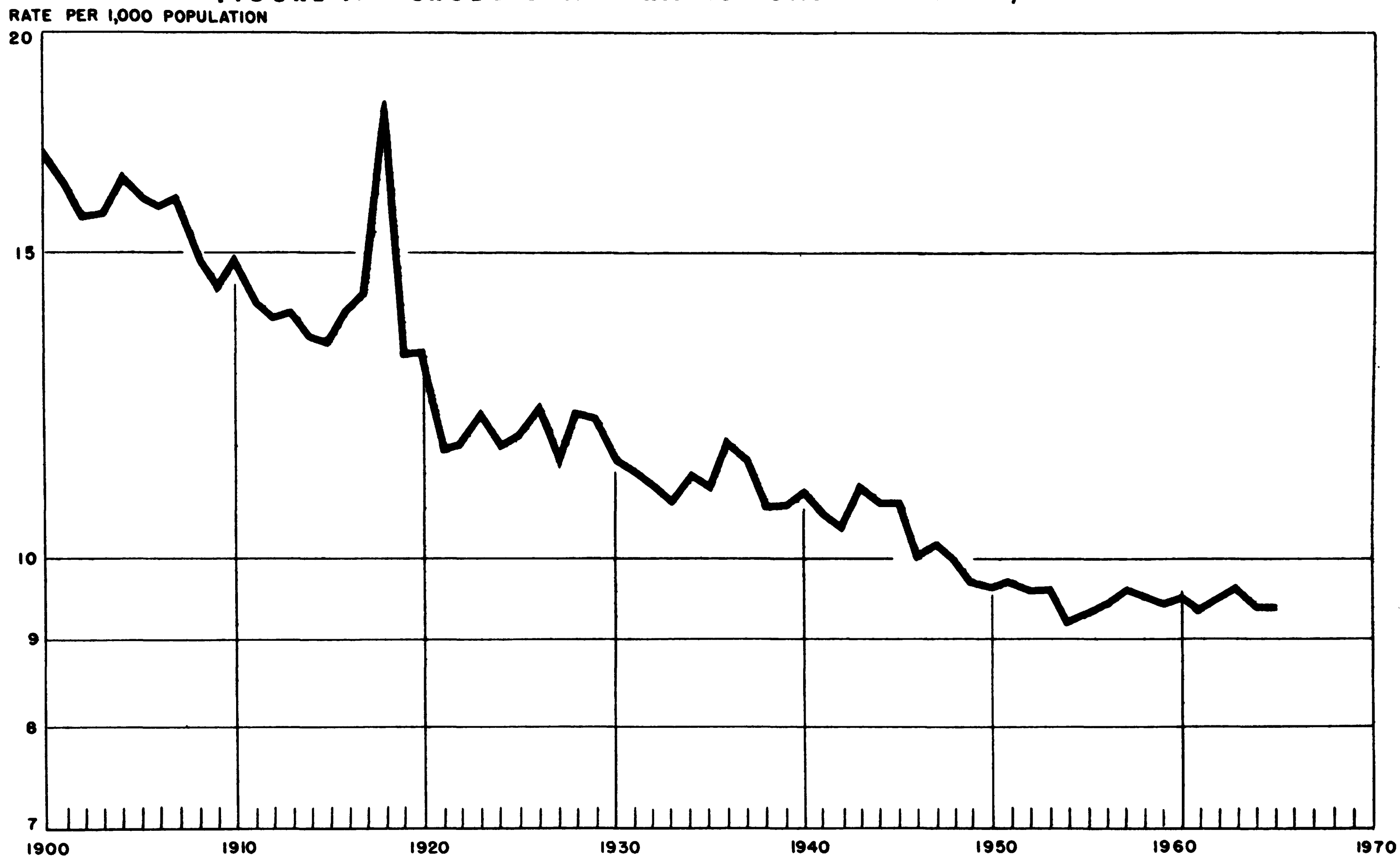


FIGURE 2A. DEATH RATES PER 100,000 POPULATION: UNITED STATES, 1930-60

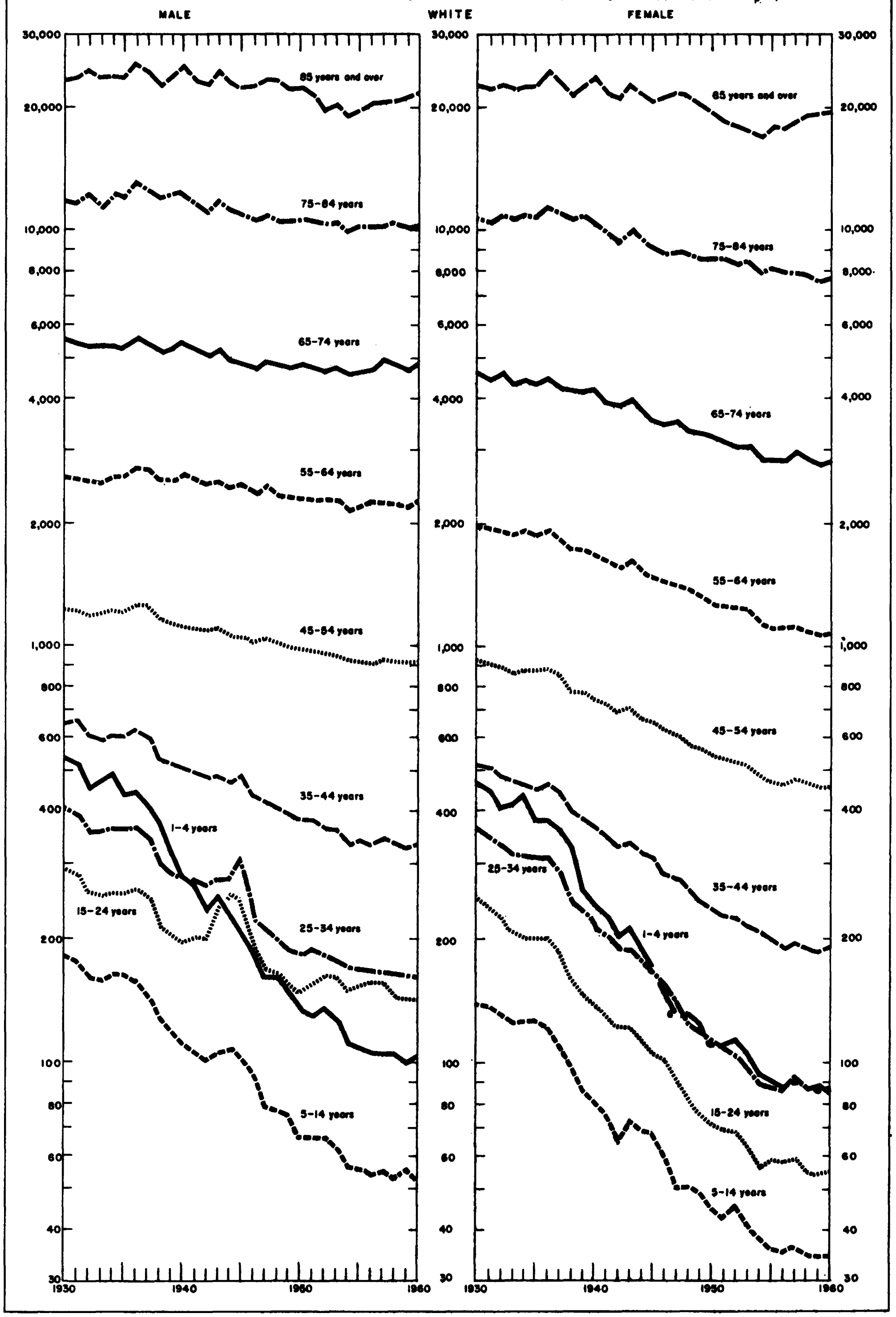


FIGURE 2B. DEATH RATES PER 100,000 POPULATION: UNITED STATES, 1930-60

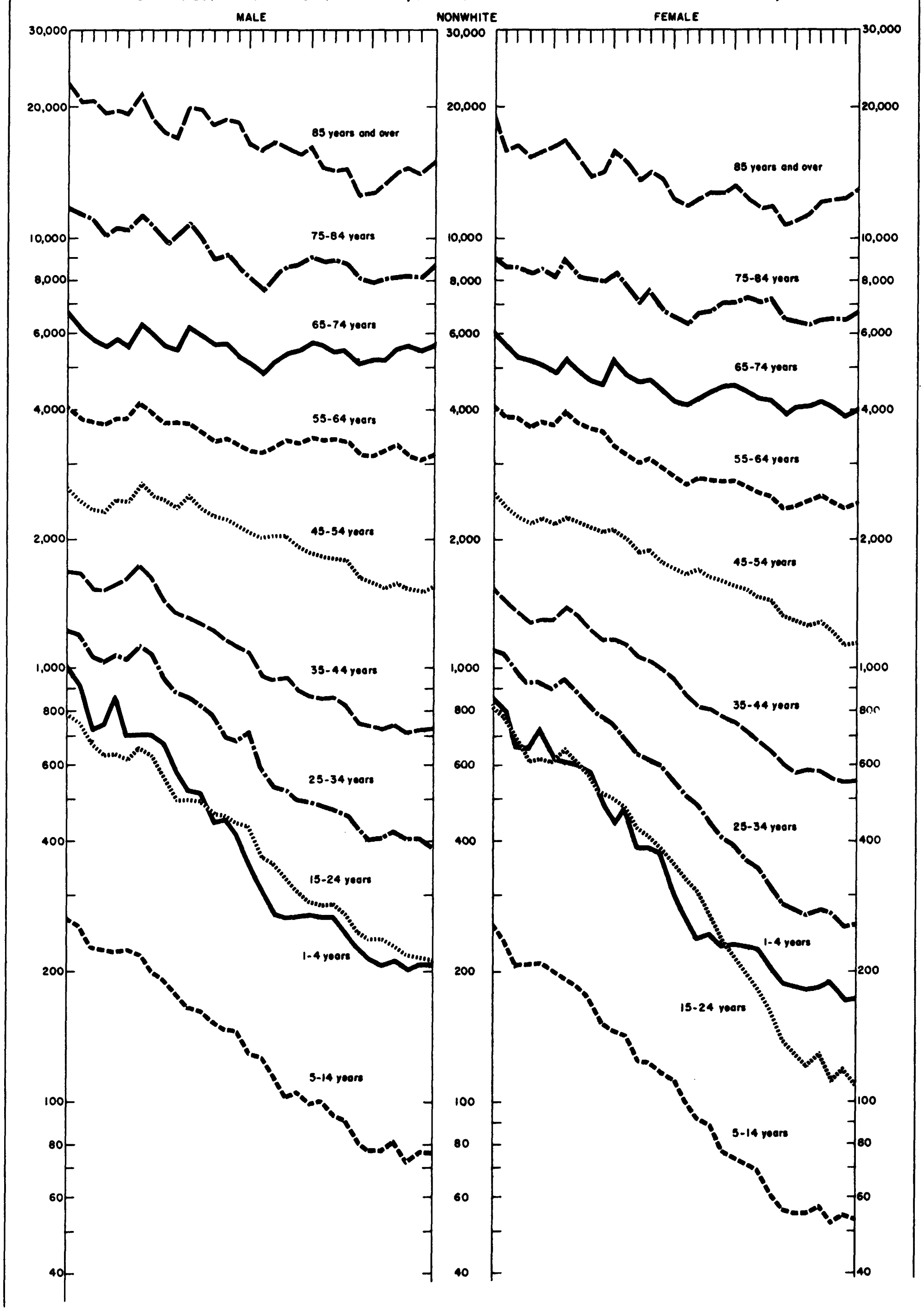
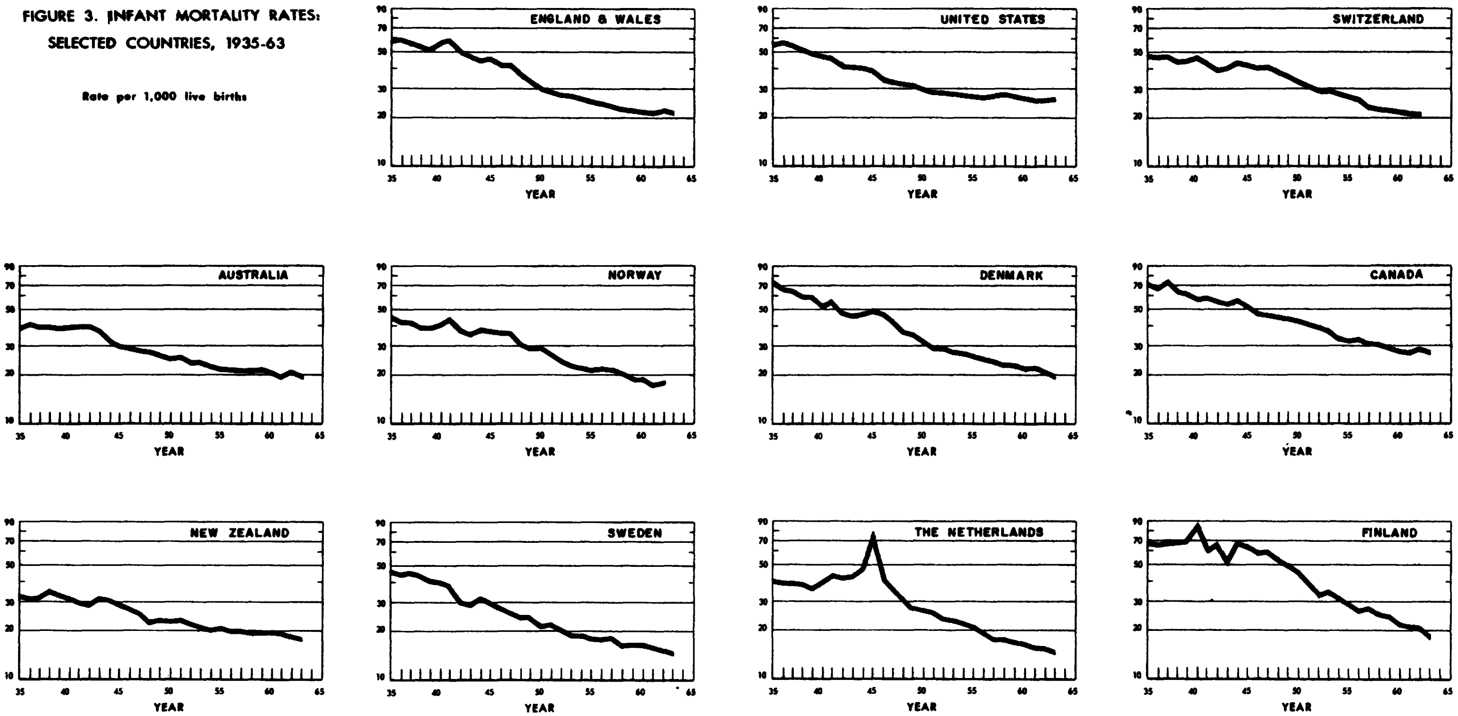


FIGURE 3. INFANT MORTALITY RATES: SELECTED COUNTRIES, 1935-63

Rate per 1,000 live births



INTERNATIONAL RESEARCH ON MORTALITY[1]

Ruth R. Puffer[2] and G. Wynne Griffith,[3] Pan American Health Organization

The marked decline in death rates during this century in the United States as well as in other countries of the Americas has been due principally to the measures taken for the prevention of communicable diseases, and the decline has been much greater for mortality in infancy and childhood than in adult life. Although the decreases in the deaths due to malaria and to tuberculosis have also affected the death rates in adult life, further reduction in death rates in the United States, as in other countries, requires greater understanding of the causation of mortality in adults, especially that due to cardiovascular diseases and cancer. The slowing down in the decline in mortality in the last decade in the United States (1) and in other countries indicates the need for a continuing analysis of these problems.

The fact that death rates from diseases of the heart, particularly coronary heart disease, were much higher in the United States than in many other countries of the world was the initial observation (2) that led the Pan American Health Organization to undertake the Inter-American Investigation of Mortality. It was thought that clues to the causative factor or factors could be obtained if the wide differences in the observed mortality were verified, that is, if the differences were shown not to be due to variations in medical practice, in terminology, and in the classification of the underlying causes. Thus, this Investigation was designed to study mortality in adult life in the 60-year age span from 15 to 74 years with particular attention to cardiovascular diseases and cancer. In this collaborative research project, which involved the intensive study of deaths in 12 widely separated cities (two English-speaking, two Portuguese-speaking, and eight Spanish-speaking), great care has been taken to obtain complete data to ensure the uniform assignment of the underlying cause of death. The results of this large-scale study of 43,298 deaths are now being analyzed and will be published in full later this year.

A previous publication (3) has described the events which led to this Investigation in the period 1957 to 1961, the planning conferences held in 1961 and 1962, the selection of the 12 participating cities and the principal collaborators in these cities, and the method of data collection to assemble all available clinical, laboratory, and autopsy findings on approximately 2,000 deaths per year for two years in each city. The field work was carried out from 1962 to 1964 and all records were received for processing and analysis in 1965. The underlying cause of death has been classified, in accordance with international procedures, with the assistance of two outstanding medical referees: Dr. Percy Stocks, formerly of the General Register Office of England and Wales and former Director of the WHO Center for International Classification of Diseases; and Dr. Dario Curiel, former Director of the Latin American Center for Classification of Diseases in Venezuela.

The results have revealed remarkable differences in mortality in adult life. Much more information on causes of death is available from hospital and autopsy reports than is provided on official death certificates. In several cities death rates for preventable diseases and conditions in early adult life are excessive. The Investigation indicates that it is time for statisticians to take a new look at mortality statistics and to introduce needed changes so that mortality statistics will again become (as in the past) an important research tool.

At this time some of the results are presented in order to show the nature of the ma-

---

1. This research project was made possible by a research grant (GM-08682) of the National Institute of General Medical Sciences, U. S. Public Health Service. This paper was prepared for the Meeting of the Social Statistics Section of the American Statistical Association on 15 August 1966, Los Angeles, California.
2. Chief, Health Statistics Branch, Pan American Sanitary Bureau, executive organ of the Pan American Health Organization, Washington, D. C. and Principal Investigator, Inter-American Investigation of Mortality.
3. Epidemiologist-in-Charge, Inter-American Investigation of Mortality, Pan American Sanitary Bureau, Washington, D. C.

terial, the need for prevention, and possible directions for further research. Data for five of the cities are used--San Francisco in the United States; Bristol in England; Guatemala City in Central America; Caracas, Venezuela, in northern South America; and Santiago, Chile, at the other end of the continent. The principal collaborators in these cities are as follows:

Dr. Ellis D. Sox, Director of Public Health, City and County of San Francisco, California

Professor R. C. Wofinden, Medical Officer of Health, City and County of Bristol, and Professor of Public Health, University of Bristol School of Medicine, England

Dr. J. Romeo de León, Jr., Medical Officer, Epidemiology Branch, Division of Public Health, Institute of Nutrition of Central America and Panama (INCAP), Guatemala City, Guatemala

Dr. Carlos L. González, Technical Adviser, Ministry of Health and Social Welfare, and Professor of Preventive Medicine, Vargas Medical School, Central University of Venezuela, Caracas

Dr. Adela Legarreta, Professor of Biostatistics, School of Public Health, University of Chile, Santiago

Death Rates by Age for Cities and Corresponding Countries

Differences in the age-specific death rates for the five cities in the 60-year age span are clearly evident, the death rates in the three 10-year age groups from 15-44 years being two or three times higher in one city than in another (Table 1 and Figure 1). The death rate of San Francisco is twice as high as the rate of Bristol in the age group 35-44 years, and about 50 per cent higher for age groups 25-34 years and 45-54 years.

Figure 2 was prepared in order to illustrate the possible similarity between these death rates of cities and those of the corresponding countries as a whole, and to point out any equally wide variation in the rates in those countries. For the national rates, the figure includes the data for two additional age groups, under 15 years and 75 years and over. The data for the cities have been confined to the 60 years for which accurate death rates for residents were

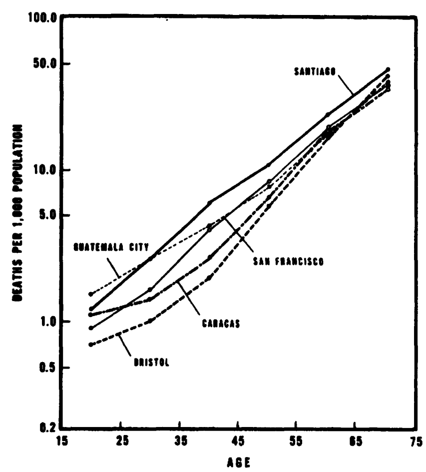
Table 1

Annual Death Rates per 1,000 Population for 10-Year Age Groups, 15-74 Years, in Five Cities, Inter-American Investigation of Mortality, 1962-1964

Age group	Bristol, England	Caracas, Venezuela	Guatemala City, Guatemala	San Francisco, U.S.A.	Santiago, Chile
15-74 years					
Crude	8.9	4.1	5.4	10.5	7.7
Age-adjusted	4.6	5.0	5.8	5.5	7.3
15-24 years	0.7	1.1	1.5	0.9	1.2
25-34 years	1.0	1.4	2.6	1.6	2.6
35-44 years	1.9	2.6	4.2	4.0	6.0
45-54 years	5.7	6.7	7.8	8.3	10.6
55-64 years	16.2	18.3	16.7	18.8	23.2
65-74 years	41.7	34.6	37.1	35.7	45.0

Figure 1

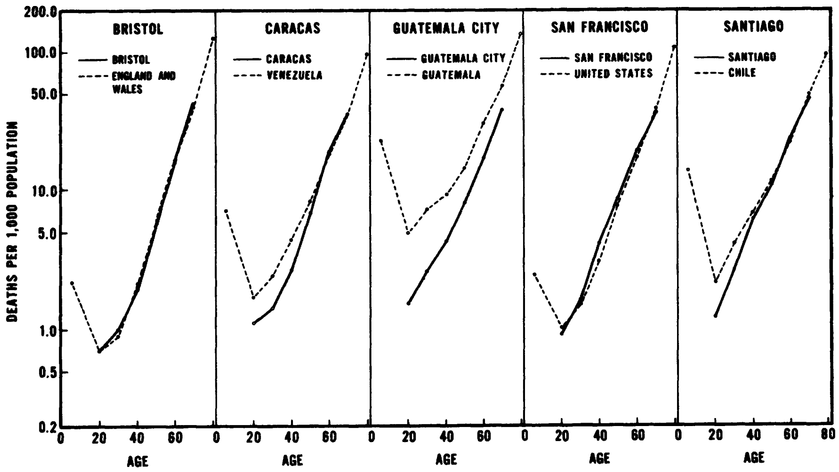
Annual Deaths per 1,000 Population for Adults 15-74 Years of Age, in Five Cities, Inter-American Investigation of Mortality, 1962-1964



obtained in the Investigation.

As was pointed out by the principal collaborator in Bristol, the death rate in that city is practically the same as for England and Wales as a whole, and has been so for many years. In Venezuela the situation is entirely different, the rate for the country being much higher than in the capital city, especially for the age period 15-54 years. Since the regis-

Figure 2
Annual Deaths per 1,000 Population for Adults 15-74 Years of Age in Five Cities, and for All Ages in Corresponding Countries, Inter-American Investigation of Mortality, 1962-1964



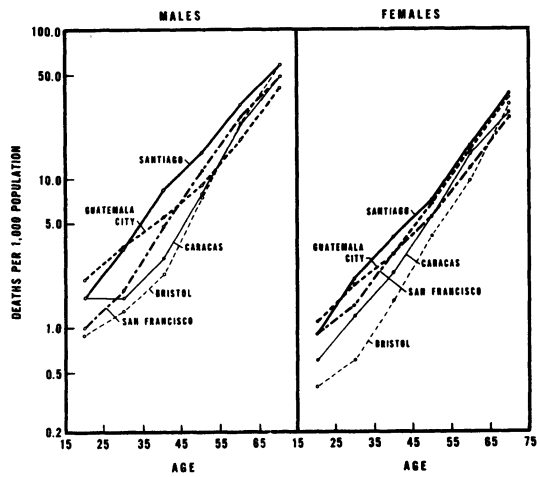
tration of deaths in rural areas is known to be incomplete, the urban-rural differences are probably even greater. In Guatemala City likewise the death rates for the city are well below those for the country as a whole, while in Chile the death rates for the country also exceeded those in Santiago in the three 10-year age groups from 15-44 years. Many factors probably contribute to the lower death rates in the capital cities in these Latin American countries--such as, for example, the concentration of health and medical facilities in cities.

In San Francisco mortality appears to follow a different pattern, the death rates in that city being definitely in excess of those in the United States for the age span 35 to 64 years.

The data presented in these two figures give a clear indication of wide variations in death rates in the 30-year age span 15-44 years in these five cities and between the capital cities and Latin American countries as a whole. Indeed, mortality in rural areas in Latin America may be many times--six to eight--higher than the lowest of the rates found in one of these five cities. The excessive mortality, not only in certain capital cities, but even more so outside them is a challenge to all concerned with the health of the population. Also the differences between mortality in cities and that in the corresponding countries certainly indicate the need for analysis of resident data by causes in urban and rural areas, as a basis for measuring health problems in Latin American countries as well as in the United States.

Death rates by sex also show marked variations, the mortality of males being consistently higher than that of females in the same age groups in the same city (Figure 3 and Table 2). The age-adjusted death rate* of males

Figure 3
Annual Deaths per 1,000 Population for Males and Females 15-74 Years of Age in Five Cities, Inter-American Investigation of Mortality, 1962-1964



* Death rates have been adjusted by the direct method to a standard population, the composite population of the cities included in the Investigation.

Table 2

Annual Death Rates per 1,000 Population, by Sex, for 10-Year Age Groups, 15-74 Years, in Five Cities, Inter-American Investigation of Mortality, 1962-1964

Age group	Male					Female				
	Bristol	Caracas	Guatemala City	San Francisco	Santiago	Bristol	Caracas	Guatemala City	San Francisco	Santiago
15-74 Years										
Crude	11.1	4.5	6.4	13.8	10.0	6.9	3.6	4.6	7.5	5.8
Age-adjusted	6.3	6.3	7.0	7.2	9.8	3.2	4.0	4.9	4.0	5.4
15-24 years	0.9	1.6	2.2	1.0	1.6	0.4	0.6	1.1	0.9	0.9
25-34 years	1.3	1.6	3.5	1.8	3.3	0.6	1.2	1.9	1.4	2.1
35-44 years	2.3	2.9	5.5	4.8	8.4	1.5	2.3	3.1	3.3	4.0
45-54 years	7.3	7.9	9.1	11.3	15.0	4.1	5.5	6.8	5.6	7.0
55-64 years	23.8	23.1	18.6	26.1	31.9	9.6	14.6	15.3	11.8	16.6
65-74 years	58.0	47.7	40.8	47.5	57.3	31.4	27.6	34.6	25.7	37.1

for the 60-year age span in Bristol (6.3 per per 1,000 population) was nearly twice the rate of females (3.2). In San Francisco the corresponding age-adjusted rates were higher for both males and females, namely, 7.2 and 4.0 per 1,000 population. For the other three cities, they were as follows: Caracas, male 6.3, female 4.0; Guatemala City, male 7.0, female 4.9; Santiago, male 9.8, female 5.4.

The data from these five cities were of good quality: of the 18,869 deaths studied, 51.2 per cent had occurred in hospitals, autopsies had been performed on 44.2 per cent, and 62.7 per cent of the deceased had been hospitalized during their last year of life.

Important Causes of Death in Forty-Year Age Span

In these populations mortality in young adult life shows much more variation than in the older age groups. An analysis of the causes of death is therefore presented for the 40-year age span 15-54 years with division into two 20-year age groups, 15-34 years and 35-54 years, and with adjustment of the rates to the same standard populations (Tables 3 and 4). Because of the higher rates in males than in females, the material is presented for the sexes separately. In Figure 4 and Table 3 the death rates from selected important causes of death in males are shown for the two 20-year age spans. In early adult life, 15-34 years, external causes (accidents and violence) caused more than half the male death rate in four of the five cities. Malignant neoplasms and diseases of the heart were responsible for from 7 to 27 per cent of the age-adjusted death rates. Bristol had the most favorable death rate and practically all the excess mortality in Caracas and San Francisco, as compared with Bristol, was due to accidents, suicide, and homicide. In the other two cities, Guatemala City and Santiago, mortality from accidents and violence was also very high; in addition the group comprising alcoholism, alcoholic psychosis, and cirrhosis of the liver (treated as a single group

Figure 4
Annual Age-Adjusted Death Rates from Important Causes of Death per 100,000 Population for Males 15-34 Years and 35-54 Years in Five Cities, Inter-American Investigation of Mortality, 1962-1964

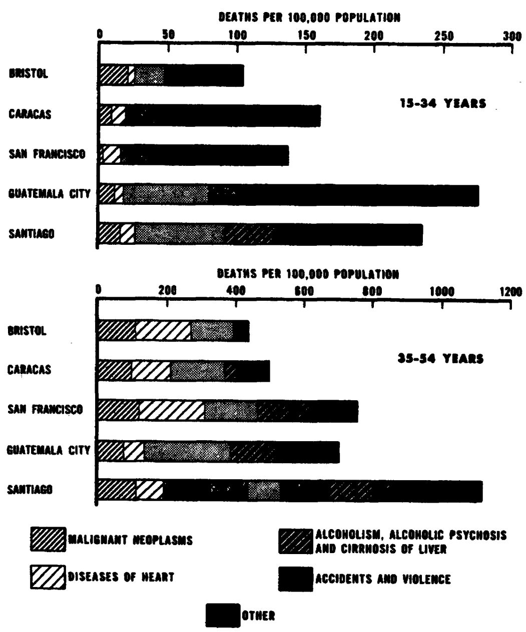


Table 3

Annual Age-Adjusted Death Rates from Important Causes of Death per 100, 000 Population, for Males 15-34 Years and 35-54 Years of Age, in Five Cities, Inter-American Investigation of Mortality, 1962-1964

Cause of Death	15-34 Years					35-54 Years				
	Bristol	Caracas	San Francisco	Guatemala City	Santiago	Bristol	Caracas	San Francisco	Guatemala City	Santiago
All causes	104.8	160.6	137.3	275.3	234.7	436.7	499.8	756.0	702.4	1116.9
Tuberculosis (001-019)	-	5.8	-	20.7	20.4	7.6	13.8	11.0	52.4	157.8
Malignant neoplasms (140-205)	20.8	8.8	2.9	11.1	15.5	107.2	98.5	121.1	76.1	110.7
Diseases of heart (400-443)	4.8	10.1	13.0	6.4	10.7	163.8	113.7	189.3	58.8	79.2
Alcoholism, alcoholic psychosis, and cirrhosis of liver (307, 322, 581)	-	1.2	5.3	26.4	37.5	3.2	34.5	160.9	132.8	291.6
Vascular lesions affecting central nervous system (330-334)	1.2	1.6	1.9	4.7	3.1	25.6	24.4	17.6	16.3	35.1
External causes (E800-E999)	57.3	111.6	94.8	147.9	105.6	38.4	93.6	130.0	181.2	198.9
Motor vehicle accidents (E810-E825, E830-E835)	33.1	17.2	38.3	39.1	36.1	11.4	19.4	24.8	65.1	68.2
All other accidents (E800-E802, E840-E962)	12.4	17.5	29.6	30.4	29.0	8.7	22.1	27.5	47.1	58.2
Suicide (E963, E970-E979)	10.8	15.0	16.0	30.6	19.4	18.3	17.4	63.1	27.0	34.7
Homicide (E964, E965, E980-E999)	1.0	61.9	10.9	47.8	21.1	-	34.7	14.6	42.0	37.8
All other causes	20.7	21.5	19.4	58.1	41.9	90.9	121.3	126.1	184.8	243.6

since 84.2 per cent of all cirrhosis deaths at these ages in these cities were known to be associated with alcoholism) appeared as a significant cause. Tuberculosis also contributed to the excessive mortality.

In the next 20-year age span (35-54 years) malignant neoplasms and diseases of the heart caused a major proportion of the male death rate in four of the cities. The death rate from diseases of the heart in San Francisco was much higher than in the three Latin American cities. This high rate is principally accounted for by coronary heart disease. The death rate from this cause was also high in Bristol. As pointed out recently by Reid (4) the recorded mortality from arteriosclerotic heart disease among U.S. white males is substantially higher than among British males. The difference in rates in these two cities, however, was not significant. Table 5 shows for each city two sets of age-adjusted death rates from arteriosclerotic heart disease, the first based on the death certificates and the second on the final assignments of cause made by the medical referees after review of all available information. In the age span here considered negligible changes occur in Bristol for both sexes, but in San Francisco the death rates, particularly of males, were reduced following review. Thus the differences between these two cities became less marked. Although in Guatemala City and Santiago the rates for both sexes increased slightly on final assignment, the disparity between them and the two English-speaking cities remains considerable, a fact which confirms one of the observations that led to this Investigation.

In the age group 35-54 years violence and the various forms of alcoholism were responsible for excessive death rates in San Francisco, Guatemala City, and Santiago and together they accounted for much of the excess over the rate in Bristol. The importance of alcoholism has been recognized in San Francisco and Santiago; however, the additional data collected in this Investigation resulted in many more deaths being attributed to alcoholism and alcoholic cirrhosis of the liver. Of the deaths certified as due to cirrhosis of the liver, alcoholism had been mentioned in only 33.3 per cent, whereas after the Investigation, of those classed to this cause 84.2 per cent were associated with alcoholism. Also, the excessive death rates from external causes in these cities of the Americas could

Table 4

Annual Age-Adjusted Death Rates from Important Causes of Death per 100,000 Population, for Females 15-34 Years and 35-54 Years of Age, in Five Cities, Inter-American Investigation of Mortality, 1962-1964

Cause of Death	15-34 Years					35-54 Years				
	Bristol	Caracas	San Francisco	Guatemala City	Santiago	Bristol	Caracas	San Francisco	Guatemala City	Santiago
All causes	52.3	88.2	109.1	143.2	143.7	258.6	363.5	425.1	462.8	528.5
Tuberculosis (001-019)	-	4.3	4.1	23.9	14.2	4.4	12.1	1.1	35.5	31.1
Malignant neoplasms (140-205)	8.5	10.5	15.2	10.8	14.8	117.0	135.0	119.1	133.5	131.4
Diseases of heart (400-443)	0.7	4.9	3.1	9.3	5.8	50.5	61.6	45.9	46.8	49.9
Alcoholism, alcoholic psychosis and cirrhosis of liver (307, 322, 581)	1.2	0.8	13.2	4.7	5.0	0.9	4.7	104.0	21.8	81.2
Vascular lesions affecting central nervous system (330-334)	-	1.1	9.8	0.6	3.1	21.2	24.2	26.8	23.1	37.9
Complications of pregnancy, delivery and puerperium (640-689)	1.2	17.7	-	24.9	38.3	1.3	6.5	1.8	15.9	27.8
External causes (E800-E999)	24.2	24.6	46.6	13.8	26.3	20.9	20.3	65.7	16.1	27.6
Motor vehicle accidents (E810-E825, E830-E835)	9.8	3.6	11.8	2.8	5.4	8.1	8.7	10.3	5.9	6.8
All other accidents (E800-E802, E840-E962)	6.1	4.5	7.2	2.6	7.8	1.5	2.8	21.3	2.6	8.1
Suicide (E963, E970-E979)	7.3	11.5	13.6	2.4	11.3	10.3	6.1	25.5	3.8	9.8
Homicide (E964, E965, E980-E999)	1.0	5.0	14.0	6.0	1.8	1.0	2.7	8.6	3.8	2.9
All other causes	16.5	24.3	17.1	55.2	36.2	42.4	99.1	60.7	170.1	141.6

not have been as clearly defined on the basis of the information on the death certificate. In some Latin American cities, the external cause of injury is not routinely recorded on death certificates and thus the high death rates from motor vehicle accidents have not been generally recognized or publicized. These excessive death rates from accidents, suicide, and homicide indicate serious problems in several of those cities. San Francisco appeared to have an unusually high suicide death rate in the age period 35-54 years. Political disturbances in 1962 adversely affected the homicide rate in Caracas.

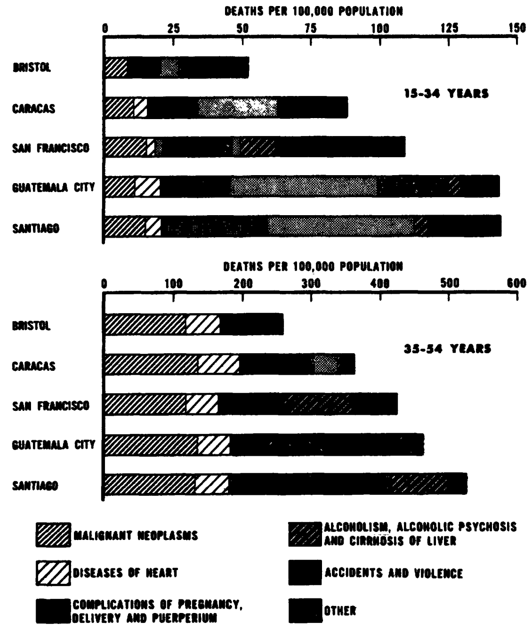
Similar data on mortality in females in the two 20-year age spans are provided in Figure 5 and Table 4. External causes did not result in the high death rates which were noted for males in four of the cities. However, in the 20-year age span 35-54 years the group of deaths due principally to alcoholism contributed significantly to mortality in both San Francisco and Santiago. The death rates in the five cities from malignant neoplasms and heart disease did not vary significantly. In the three Latin American cities death rates from complications of pregnancy, delivery, and the puerperium were many times in excess of those in the two English-speaking cities; in Santiago half of these deaths were the result of abortions. Tuberculosis also played a role in the higher death rates of females in Guatemala City and Santiago.

Each city has its own distinct pattern of mortality. The Bristol experience is the most favorable for both males and females in early and in middle adult life. The major problems would seem to be motor vehicle accidents in younger, and coronary disease in older men. The second most favorable experience is that of Caracas, and were it not for the high homicide rate due in large part to the political disturbances which coincided with the period covered by the Investigation, male mortality in this city would be similar to that in Bristol in both the younger and the middle adult years. In San Francisco external causes accounted for

Table 5
Age-Adjusted Death Rates per 100,000 Population at Ages 15-54 Years, by Sex, from Arteriosclerotic Heart Disease, according to Death Certificate and Final Assignment in Five Cities, Inter-American Investigation of Mortality, 1962-1964

City	Males		Females	
	Death certificate	Final assignment	Death certificate	Final assignment
Bristol	49.1	48.7	9.0	8.9
Caracas	30.3	29.2	11.3	9.0
Guatemala City	7.8	9.2	2.1	2.5
San Francisco	72.5	65.1	11.0	10.0
Santiago	13.0	14.7	4.9	6.4

Figure 5
Annual Age-Adjusted Death Rates from Important Causes of Death per 100,000 Population for Females 15-34 Years and 35-54 Years in Five Cities, Inter-American Investigation of Mortality 1962-1964



70 per cent of the male mortality from 15-34 years, while in the following two decades coronary disease and alcoholism in its various forms contributed heavily to the death rate of males. Among females external causes in both age groups and the causes associated with alcoholism in the second period are major problems. The pattern in Guatemala City is characterized by relatively high male rates from tuberculosis, alcoholism and cirrhosis of the liver, and external causes in both age groups, while mortality from cancer and heart disease is comparatively low. Females also have high tuberculosis death rates and maternal mortality is excessive. The notable features in Santiago are the high rates from alcoholism and cirrhosis of the liver, seen in males in both age groups and in females of middle age. Tuberculosis takes a heavy toll, of middle aged men particularly, and maternal mortality is high.

Summary

Death rates in adult life show wide variations in the five cities included in this study, which reports part of the findings of the Inter-American Investigation of Mortality, undertaken by the Pan American Health Organization. Of the five cities, four (Caracas, Guatemala City, San Francisco, and Santiago) are widely separated in the Americas and the other (Bristol) is in England. The death rates in the three Latin American cities were much lower than those for the corresponding country as a whole in young adult life. The death rate in Bristol was similar to that in England and Wales, while the rate in San Francisco for the age span 35-54 years was significantly higher than that in the United States as a whole.

Analysis by causes in two 20-year age spans indicated excessive mortality in these cities among young males 15-34 years of age due principally to external causes (motor vehicle accidents, other accidents, suicides, and homicides). In all cities except Santiago over half of the death rate was due to these external causes. Malignant neoplasms and diseases of the heart were responsible for only a small portion of the death rate. In the next 20-year age span, in addition to external causes, the group comprising alcoholism, alcoholic psychosis, and cirrhosis of the liver (principally alcoholic) was responsible for a high proportion of male deaths in three cities--San Francisco, Guatemala City, and Santiago. Diseases of the heart, mainly coronary heart disease, caused higher death rates in San Francisco and in Bristol than elsewhere. This difference between the English-speaking and the Spanish-speaking cities confirmed one of the observations that led to the Inter-American Investigation of Mortality. Although the differences are even greater at older ages (55-74 years), the disparity is already evident in this age period. By contrast, in the 40-year age span 15-54 years the female death rates from heart disease in all the cities were of approximately the same size.

The causes of mortality in these populations have become more clearly defined through this Investigation, by ensuring that the information in hospital and autopsy records was utilized in the assignment of the cause of death. A system is recommended whereby hospital records on deaths and supplementary reports on autopsies would be combined with the causes stated on death certificates so as to clarify the cause and make the data more useful for preventive action as well as for research.

The Inter-American Investigation of Mortality has shown that collaborative research utilizing standard definitions and procedures is feasible on an international scale. The time has come to utilize new tools combining information from various sources and to make greater progress in the analysis of mortality, which for many years to come will continue to be the most valuable basic data for the geographic study of diseases on a world basis.

REFERENCES

(1) The Change in Mortality Trend in the United States, National Center for Health Statistics, U. S. Public Health Service, Series 3, No. 1, 1964.

(2) Puffer, R. R. and L. J. Verhoestraete: Mortality from Cardiovascular Diseases in Various Countries, with Special Reference to Atherosclerotic Heart Disease, World Health Organization, Bulletin, Vol. 19, 1958.

(3) Puffer, R. R., G. W. Griffith, D. Curiel and P. Stocks: International Collaborative Research on Mortality, World Health Organization, Public Health Papers No. 27, 1965.

(4) Reid, D. D.: Studies of Diseases among Migrants and Native Populations in Great Britain, Norway and the United States. I. Background and Design, included in National Cancer Institute Monograph 19, U. S. Public Health Service, 1966.

DISCUSSION

Carl L. Erhardt, New York City Department of Health

The problem discussed by Dr. Moriyama has concerned health authorities now for a decade. We had become so used to declining death rates that resistance to further decreases is difficult to accept. We must keep in mind that man is mortal, that the population will continue to have increasing numbers of old people and that dramatic new therapies have not appeared in the sixties. Nevertheless, as Moriyama has pointed out, such explanations are inadequate, since lower age-specific mortality rates do exist in other countries. The white-nonwhite differentials in the United States are alone sufficient evidence that improvement in the gross rates must be possible.

There is little I can add to Moriyama's discussion, but the change in rate of decline he notes does seem most evident (Figure 2 A) among white males at ages 15-24 and 25-34 and to have occurred earlier than at other ages--about 1948, I should judge from the chart. This date is immediately after World War II with its high peak of mortality at these ages. Could these war losses, affecting the healthiest young men, have been, in part at least, responsible for the earlier change in trend at these ages--and also for the later stabilization (about 1955) at ages 35-44, as this younger cohort grew older? Perhaps some birth cohort analyses might shed light on this point. However, the same course is not visible among nonwhite males and could not be expected among females of either age group. Nor, if true, can it explain the later stabilization at younger or older ages.

I wonder if our emphasis on age or cause-of-death information will lead us to solutions of the apparent enigma in the course of U. S. mortality rates. Let us consider infant mortality, for example. Is it not possible that social factors that do not necessarily reveal themselves directly in cause-of-death statistics are the important influences? Marriage has been occurring at younger ages; divorce rates have been high; the proportion of working mothers has increased. Hence satisfactory family life has been disturbed. What about the possible effect of determined efforts to maintain the life of poor risk infants (i.e., those of low birthweight and/or short gestation)? Studies of their later progress are few and results are not uniformly clear that they do as well as mature infants.

We pride ourselves on the levels of inoculation of infants against infectious diseases in most areas and among most population subgroups. But is there less concern for babies, if not neglect, in the disruptions of family life, in the striving for financial success, in the efforts to provide the so-called benefits of suburban living? It seems to me we may need more research into the sociology of mortality rather than the medical aspects.

Moriyama also raises a question about the change in rate of decline of pneumonia mortality during the past decade. To some extent I suspect this may reflect an artifact of reporting, but reporting changes would not alter the lack of change in crude rates; they would only magnify upward trends for chronic diseases. There appears to be need for bacteriologic work that has been neglected because of availability of broad-spectrum antibiotics. However, as with infant mortality, I believe we have to seek other factors that are not immediately evident no matter how closely we look at age- or cause-specific mortality. Here again, I urge study of the sociology of mortality.

We have become an effete society. I recall an old Harold Lloyd film where his chauffeur simply made a U-turn to drive him to visit his fiancee who lived directly across the street. The scene would not be as hilarious nowadays when the most muscular, agile adult apparently cannot walk a block or two to get a pack of cigarettes or his bottle of bourbon. If home, he has to take the car. On arrival at his destination, he must double park lest he have to walk twenty paces from the empty parking space a short distance away. It is an appalling thought that the 16-year old might take a bus (or even, horribile dictu, walk) to school. Lack of exercise, poor diet (largesse de richesse), smoking, excess alcohol, possible effects of air pollution or food additives; there appear to be many possibilities of social (and environmental) factors that only after many years are showing their effects in terms of mortality.

These considerations are pointed up in reviewing the paper by Puffer and Griffith based on their work with their several collaborators. The paper indicates that differences in diagnosis or classification are not the explanation of international intercity variations in mortality of substantial magnitude. The authors have described these intercity variations in gross, by sex, by age and by cause of death and have also delineated the patterns peculiar to each city. They are to be congratulated, with their colleagues, on this accomplishment. Perhaps the most important contribution, from my point of view at this time, is the relatively considerable proportion of the mortality rates in these cities that is contributed by conditions associated with alcoholism, by tuberculosis and by external causes.

The exclusion of these conditions among males at 15-34 years of age (Table 3) reduces the rates to a range of 37.2 (San Francisco) to 80.3 (Guatemala City) instead of 104.8 (Bristol) to 275.3 (Guatemala City). There is not much difference in the ratios of these high rates to the lows (2.2 and 2.6). But now the major contribution rests in the residual, "all other causes". At 35-54, however, the same exclusions reduce the rates to a range of 336.0 (Guatemala City) to 468.6 (Santiago), with a ratio between high and low of 1.4 instead of 2.6 (between Bristol and Santiago).

Again, it seems to me that social action is indicated. It is obvious with tuberculosis, but alcoholism, accidents, suicides, homicides are also social problems. These observations, too, confirm my opinion that the sociology of mortality needs to be studied.

It is perhaps particularly fitting that these two papers are read as a pair--and before the Social Statistics Section of ASA. Until medicine discovers or devises major advances in cancer and heart disease prevention and/or cure, influencing the social body seems to provide the greatest promise for further reduction in death rates. I also pose the question whether the present stagnation in mortality rates in the United States has a social rather than a purely medical component underlying it.

III

EVALUATION OF 1960 CENSUS COUNT

Chairman, MILOS MACURA, United Nations

EVALUATION OF COVERAGE IN THE 1960 CENSUS OF POPULATION THROUGH CASE-BY-CASE CHECKING

Eli S. Marks, University of Pennsylvania
Joseph Waksberg, Bureau of the Census

I. INTRODUCTION

The population count was the original purpose for taking the census and still remains by far its most important statistic. In addition to providing the basis for allocating Congressmen, population counts are important to public planning, the allocation of funds, and the projection and analysis of other statistics. Moreover, since the amount of error in the population count is likely to vary from group to group within the total population, any coverage loss can affect all census data.

Coverage thus occupies a unique position in census planning. Attempts to insure high coverage and evaluating the success of these attempts have become central considerations in developing the methods to be used in recent censuses. The resources used for these purposes are significant portions of the total resources available for a census.

There are essentially two methods of evaluating census data. One is by case-by-case analysis of a sample of census returns, using whatever means are available to uncover errors in the census. The other is by analysis of the statistics themselves, comparing them with other related information (on births, deaths, previous census counts, etc.) and examining problems of internal consistency. This report and the companion paper by Messrs. Siegel and Zelnik, describe the use of these methods in evaluating the coverage of the 1960 Population Census, and provide alternative estimates of undercounts. The present paper is restricted to the results of the case-by-case studies of 1960 coverage. This includes reinterviews and matching the census against samples selected from various independent lists. The Siegel-Zelnik paper describes methods used in analyzing the statistics and presents estimates of undercounts which combine the best features of both systems.

It should be noted that, for some purposes, what appear to be coverage errors are really the joint effect of coverage and content errors. For example, if one is analyzing the population counts by age and sex, the number of persons reported in a given category is affected not only by the degree to which they are missed in the census, but also errors in age reporting and even in reporting or tabulating sex. Because of the pervasive role of age, sex, and color as bases of cross-classifications in the census and because many of the evaluation methods are tied to analyses of the population by these items, we are likely to think of these items as virtually defining the population. Thus, analyses of coverage tend to involve age, sex and color distributions of the population.

Since this paper is concerned with techniques of evaluation that involve examining the census records to ascertain whether specific individuals have been correctly counted, it is possible to keep "pure" coverage errors separate from age-reporting differences. For reasons to be described later, only "pure" coverage errors will be discussed in this paper, although much of the statistical data distribute under-enumeration by age, sex and color. The Siegel-Zelnik paper will contain a discussion of the joint effect of coverage and other reporting errors on the population counts.

Our definition of coverage should be made clear, by pointing out an important distinction. In any census there is always a small proportion of households with no one ever at home when the enumerator calls, and such households are, therefore, not personally enumerated. In 1960, as in previous censuses, the enumerators were instructed to obtain and report counts of the people in these households from neighbors, and if possible, a minimal amount of other information, i.e., sex, age, color, relationship and marital status. These cases are not treated as undercoverage, although frequently no information other than the count of individuals could be obtained. They were treated as simply contributing to the persons for whom individual items of information were not reported. In general, computer methods were used to impute data for such not-reported cases and most census tables show the estimated distributions for the entire enumerated population. (Appendix tables show the amount of such imputations.) Undercoverage describes persons whose existence was not reported to census enumerators, and who are therefore not represented in any way in the census.

II. COVERAGE ANALYSES PRIOR TO 1960.

It might be useful to start off with a brief resume of estimates of undercoverage in the 1950 Census and various actions taken to improve coverage in the 1960 Census.

Estimates of undercoverage in the 1950 Census of Population and their uncertainties.--At least three estimates are available of net underenumeration of the total resident population of the United States:

1. Post-Enumeration Survey (PES) of the 1950 Census: 1/
   2.1 million persons (1.4 percent of the enumerated population).
2. The analytic method developed by Ansley J. Coale: 2/
   5.4 million persons (3.6 percent of the enumerated population).
3. The Bureau of the Census "minimum reasonable estimate." 1/
   3.7 million persons (2.5 percent of the enumerated population).

The PES estimate.--There are a number of reasons for believing that the PES estimates of omitted persons, and thus of the net underenumeration in the 1950 Census, are substantial understatements:

1. The PES check for missed persons did not cover transient hotels and motels, nor other "quasi" households where 35 or more persons lived. The enumerated population excluded from the scope of the PES missed-persons check is estimated at 4.1 million or a little less than 3 percent of the enumerated population. However, there is reason to believe that persons in such quarters are missed at a disproportionately high rate. This exclusion therefore probably resulted in an understatement of more than 3 percent for the missed population.

2. There was a time lag between the Census and PES of between 4 and 5 months. Thus, it was obviously difficult for the PES interviewer to obtain reliable information about persons living at specified sample locations at the time of the census who left before the time of the PES interview (almost 10 percent of the population of the United States).

3. The PES made use of substantially the same enumerative methods that were employed in the census and thus apparently had the same difficulties in identifying persons with tenuous or temporary attachments to households. Evidence for the soundness of this view comes from various demographic analyses performed after the 1950 Census that indicated that the PES estimates were just about as deficient as the census estimates for those groups of the population subject to the greatest risk of underenumeration, for example, males aged 15-39.

1/ U.S. Bureau of the Census. "The Post-Enumeration Survey: 1950," Technical Paper No. 4.

2/ Coale, Ansley J. "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, Vol. 50, March 1955, pp. 16-54.

The Coale estimate.--The procedure was based on the assumption that the rates of underenumeration in the 1930 Census by age, sex, and color were equal to the smaller of the error rates in 1940 and 1950. On the basis of this assumption an iterative process was applied that provided estimates for the population 15 years old and over in 1950. For children aged 0-14 in the 1950 Census, the estimate was derived from Birth Registration Tests of 1940 and 1950. Coale, himself, recognized that his procedure had some limitations, especially for the older ages. Indeed, for the age group 65 and over, he rejected the estimates based on the iterative procedure and accepted in their place the PES estimates.

The Census "minimum reasonable" estimate.--This method made use of data on births adjusted for underregistration on the basis of the 1940 and 1950 Birth Registration Tests to produce estimates of net undercounts of children under 15 years of age. The estimated net undercount turned out to be about 1.5 million, while the PES estimate for the same age group was about 0.7 million. There was an implicit assumption that the estimates of the population based on adjusted births had no errors. For the age group 15-39, the method employed sex ratios derived from the Coale analysis. These ratios were applied to the 1950 Census count of females as adjusted by the estimate of female underenumeration obtained from the PES. Although the sex ratios employed would appear much more reasonable to most demographers than the sex ratios of the PES, there was no external validation and indeed the ratios are subject to the limitations inherent in Coale's method. For the population 40 years old and over the PES estimates of underenumeration were used without further adjustment.

Studies of methods of improving coverage prior to 1960.--A variety of techniques and procedures were tested during the 1950's to determine their feasibility and effectiveness for improving coverage in the 1960 Census of Population. These included such devices as: matching of census rolls with names on lists of persons considered hard to enumerate (e.g., welfare records), missed persons forms distributed to all school children, use of local neighborhood and block leaders as enumerators, use of post office resources to improve coverage, improved quality control and enumerator training methods. The results of some of these tests are described in Census Working Paper No. 19. 3/

Some of the changes in methodology used in the 1960 Census were based on the results of these studies. The tests did not provide any encouragement for use of matching special lists as a coverage improvement program. In retrospect, we think that further investigation of this is warranted. The earlier tests were conducted on a small scale and it is possible that better methods of list construction and matching can be developed.

3/ U.S. Bureau of the Census. "Tests of Use of Post Office Resources to Improve Coverage of Censuses," Working Paper No. 19 (1965).

Methods used to improve coverage in the 1960 Census.--One major aim of the enumeration procedures for the 1960 Censuses was to reduce the number of functions performed by the enumerators at one time, because it appeared that the great complexity of their jobs in the earlier censuses had led to insufficient emphasis on some of the more important aspects of coverage and content of the censuses. By and large, there was more than one end in view for the adoption of each of the new enumerative procedures in 1960. Those procedures and devices that had as one of their objectives the improvement of coverage were the following: 4/

1. The Advance Census Report Form.--This was the form delivered by mail to each household in advance of the census. Each household was asked to list its members together with a minimum amount of demographic data. Enumerators were required to pick up the forms and check them for completeness. The belief was that, to the extent that the public filled out this form and listed the members of each household on or about April 1, the errors of coverage which arise from a protracted period of enumeration--i.e., the failure to count some persons who moved during the enumeration period and the double-counting of others--would be reduced.

2. Separation of collection of nonsample data (Stage I of the Census) from the collection of sample data (Stage II of the Census).--The first stage was designed to concentrate primarily on coverage, with the goal of providing improved counts of people and housing units.

3. Use of a separate listing book.--This was done to provide better control on the callbacks at households where no one was at home on the first visit, and to help insure that the enumerators would list households in the appropriate order of canvass, for purposes of improving coverage.

4. Quality control of field work.--It provided supervisory personnel with definite procedures for detecting and, when necessary, rejecting unacceptable work.

The problems of slum-area enumeration.--Some of the techniques and devices tested during the 1950's were designed especially to deal with coverage problems in slum areas. None of the new developments actually adopted in 1960, however, were especially focussed on slum areas; they were designed to improve the enumeration in all areas. In retrospect, it now seems that the scope and intensity of the procedures used in 1960 were not sufficient to deal with the extraordinary difficulties that were encountered in the slum areas.

III. METHODS OF EVALUATION.

Case-by-case evaluation through reinterview.--The process consists of obtaining in reinterviews a list of a probability sample of the population whose coverage in the census is to be checked. This list is then matched with the census list, name by name, to uncover cases of both underenumeration and overenumeration. The list is obtained in the first instance by specially selected and specially trained interviewers who are given more adequate compensation than census enumerators. Every effort is made to produce a "quality job." Also the matching operation is undertaken with great care.

There are two major advantages in the use of this method: First, the method identifies and provides estimates of the components of the net coverage error, i.e., the underenumeration and the overenumeration. Second, it provides an opportunity to learn about reasons for coverage error and factors associated with coverage error in individual cases.

There are, on the other hand, two limitations: First, the method is an enumerative method and thus is subject to the weakness inherent in the use of such a method to the extent to which people are not identified in a question and answer procedure and to the extent to which people are not reached by an enumerator. No matter how well trained or motivated a reinterviewer is, there will be deficiencies in the coverage of the reinterview method. The second limitation is that of matching. There are always uncertainties in a process where reliance has to be placed on two reports of the name of the individual.

Various techniques have been used to overcome the matching problem. The technique used in 1960 defines both over and underenumeration relative to a specified small area, say, an enumeration district, that contains the usual residence of the person being checked. This eases the matching problem considerably and produces in principle an unbiased estimate of the net underenumeration. As a result of this technique, an individual enumerated in the wrong place gets counted both as a missed person (where he should have been enumerated) and as an overenumerated person (in the wrong place). Relative to a larger area, say the United States as a whole - such individuals are counted once and only once, and therefore should not be included as part of coverage error. Thus, coverage evaluation on the basis of small areas can result, on a national basis, in substantial overestimates of both gross misses and gross overenumerations.

4/ U.S. Bureau of the Census. "Procedural Report on the 1960 Censuses of Population and Housing," Working Paper No. 16 (1963).

Case-by-case evaluation through record checks.--In broad terms, the method consists of defining a frame, e.g., a list of names of the entire population of the United States if that is the population to be checked. This frame is defined independently of the census enumeration to be evaluated. A sample is drawn from the frame and an attempt is made to determine the usual residence on the census date of the persons in the sample. The final step in the procedure is to match a sample of names on the frame against the names appearing in the census and thus to obtain an estimate of gross underenumeration in the census.

There are a number of limitations. First, the frame we have been able to develop is not quite complete. Second, there is the problem of determining the usual residence of a sample of persons whose names are shown on the frame, especially when the frame is several years old. Third, there is the matching problem. Fourth, there is the fact that the method does not produce an estimate of net underenumeration but only of omitted persons.

The principal advantage in the method lies in the fact that there is much greater "independence" between the frame and the census enumeration than is obtained by reinterview methods. In the United States, however, the principal component of the frame, in the absence of a population register, was the previous census of population together with missed persons detected in the 1950 PES. To the extent to which persons who were missed in the 1950 PES (possibly half of the total missed persons) are omitted in successive censuses there is a lack of independence. There is, however, considerable variation in the completeness of enumeration by age. Thus, there is reason to believe that the correlation between errors of omission in successive censuses is not very great.

IV. RESULTS OF 1960 EVALUATION THROUGH REINTERVIEWS.

The 1960 coverage evaluation through reinterviews involved two studies:

a. EP-8: This was a re-enumeration of housing units based on an area sample of 2,500 segments containing about 25,000 housing units. Each segment was re-enumerated in a search for housing units omitted from the census and for listings in the census that did not represent real housing units. The 2,500 segments were a subsample of segments previously canvassed for the Survey of Change and Residential Finance (SCARF), and the coverage check enumerators had available both the SCARF and 1960 Census information. This redundancy was presumed to improve the chance of identifying missed units.

   This study provided only an estimate of the number of missed or over enumerated housing units and the persons in such units. No information on their characteristics was obtained.

b. EP-9: The second study was a re-enumeration of persons and housing units based on a list sample of about 15,000 living quarters enumerated in the census. Its primary purpose was to check on the quality of census coverage of persons in enumerated units. A secondary purpose was to make an additional check for missed housing units and missed persons in them; this check was done by determining whether housing units adjacent to the sample enumerated unit were also enumerated. Some characteristics of persons missed in the census were obtained in this study.

   An analysis of the EP-9 data indicates that this study was apparently not successful in fulfilling its second purpose. According to the EP-9 survey, the census missed about 1,140,000 persons in 460,000 missed occupied housing units. The comparable figures from the area sample approach used in EP-8 were 2,850,000 persons in 1,290,000 occupied housing units. On the basis of other information available and earlier experience with area sampling devices, the Bureau's judgment is that the larger EP-8 figure is much closer to the truth.

   It is probable that the EP-9 estimate of missed and over-enumerated persons in partially enumerated households is also seriously deficient. However, it is not possible to construct any alternative estimates based on reinterview methods. We believe that most of the deficiencies occur for males. Females appear to have been covered reasonably well.

   In any case, we have prepared what we believe is the best estimate of coverage error from the reinterview surveys by combining the results of the EP-8 and EP-9 studies. For the most part EP-8 has been used for the components arising from missed and overenumerated living quarters and EP-9 for errors in partially enumerated households. However, a small part of the coverage errors arose because of processing errors rather than field enumeration. The evidence on this was mostly from EP-9 and these figures are reflected in our final estimates.

Table 1 contains the result of this combination of EP-8 and EP-9 results. It shows a gross undercount of 5,653,000 persons, a gross overenumeration of 2,325,000 and a net undercoverage of 3,328,000 persons or 1.9 percent of the census count. If one were concerned about the errors in enumeration alone, the net undercoverage would increase to about 3,700,000. The table also contains additional information on the sources of coverage errors.

Reasons for underenumeration.--In the 1950 Census, of the 3.4 million persons estimated by the PES to have been missed in the census, about 2.5 million lived in dwellings that were themselves missed. The judgment is that this heavy concentration of coverage error in missed dwellings is more of a reflection of the inadequacy of the 1950 PES rather than of the true pattern of coverage error in the census. And, indeed, the Evaluation Program of the 1960 Census indicates that of the 5.7 million people estimated to have been missed, 3.1 million lived in dwellings that were themselves missed. The 1960 Census Evaluation Program estimates that slightly more than one-half of the missed persons lived in dwellings that were missed, as compared to about three-fourths in the 1950 PES.

We believe this is a reflection of the fact that a reinterview technique is reasonably satisfactory in detecting living quarters (and their occupants) that the original enumerators missed but does not adequately measure missed individuals in partially enumerated living quarters. Estimates of undercoverage shown in these tables for females are quite close to those prepared independently (as will be indicated in the Siegel-Zelnik paper). However, they seem to be low for males, and are seriously low for nonwhite males. We suspect that to a great extent this arises from a failure of the reinterview to discover persons who have a loose attachment to existing households.

The 1960 evaluation program does not provide much information about reasons for coverage error. The 1950 PES, however, does provide a few clues as to why dwellings were missed. For example, about two-thirds of the missed dwellings were in buildings that were completely overlooked by census enumerators.

Age-Sex-Color Distributions.--A modification in this procedure, however, is necessary to produce age-sex-color distributions. Since the EP-8 study did not obtain any information on the characteristics of missed persons, age-sex-color distributions for missed persons in missed units have been prepared by using the EP-9 percentage distributions for these items and applying them to the EP-8 counts, within a number of subgroups of the population. This, of course, is a valid procedure only if one assumes that the missed units uncovered in EP-9 are similar to all units missed. It is unlikely that this is exactly the case, but no other alternative exists.

Resulting estimates of undercoverage by age, sex, and color are shown in table 2. As mentioned earlier, these take into account pure coverage error only. They are therefore not strictly comparable to the corresponding distributions arising from the demographic analyses described in the Siegel-Zelnik paper, which reflect the effect of both coverage and age-reporting error. A separate study of the census evaluation program focussed on the content of the information reported in the census, including age, and in theory it would be possible to combine the results of the two studies to produce the net effect of both coverage and content errors. Unfortunately, the measures of age misreporting are probably rather poor estimates of the bias in age reports. For this reason, in presenting estimates of error by age, sex, and color, we have used the reinterview studies as a measure of coverage errors only.

In comparing the reinterview results with those from demographic analyses, it can be seen that there is a moderately good correspondence for white females. We take this fact as evidence that these estimates are fairly reliable. There are very large discrepancies, however, for white males, nonwhite females and nonwhite males. With respect to the differences in age distributions, there are reasons to feel that the reinterview results are better in the upper ranges (65 and over). In the lower ranges (under age 15) the figures based on demographic analysis are undoubtedly superior. The reinterview results are also deficient for nonwhite adult males, and this group is probably the one most seriously understated in the reinterview approach. For the other population groups, there is considerable uncertainty as to which source is more accurate.

V. USE OF RECORD CHECKS FOR COVERAGE EVALUATION

Record checks are defined as studies in which samples of persons from independent lists or administrative records are checked for completeness of coverage in the census. The 1960 record check evaluation of coverage of the total population has been published in Series ER-60 No. 2. A brief summary of the study follows:

The record checks are based on sample studies of four population groups which together make up the total population, with some trivial omissions:

(a) persons enumerated in the 1950 Census;
(b) children born during the intercensal period;
(c) persons missed in the 1950 Census but detected by the 1950 PES; and
(d) aliens who registered with the Immigration and Naturalization Service in January 1960.

Their combined representation is believed to be 98 percent or more of the entire population.

The four population samples totalled 7,612 persons of whom 425 were found to be "out of scope" (persons deceased, outside the United States, or erroneously included in the sample), resulting in a working sample of 7,187. Definite information about enumeration status was available for 6,003 sample persons; of these 1.3 percent were identified as having been missed in the 1960 Census. Major limitations in the ability to arrive at precise estimates of omissions arise from a failure to account for 16.5 percent of the working sample because of "noninterviews," mostly caused by the inability to obtain 1960 addresses for 932 sample persons and because of a "probably missed" group for whom a precise determination about inclusion in the 1960 Census could not be made.

Because noninterview and probably missed cases were believed to involve more underenumeration than the 6,003 cases for whom definite enumeration information was obtained, it did not appear reasonable to apply to the problem group the 1.3 percent missed rate established for the 6,003 sample persons of known enumeration status. Hence various assumptions were made about the enumeration status of the problem groups leading to a range of estimates of underenumeration. For this reason, the record check results must be viewed as providing estimates of minimum and maximum estimates of underenumeration, arising from rather extreme assumptions about the noninterview cases.

Using the record checks in this way leads to a minimum estimate of underenumeration of 4.7 million persons, and a maximum of 8.5 million, or 2.6 to 4.7 percent of the 1960 enumerated population. Further details are shown in table 3. These studies yield measures of undercoverage only. Net error in coverage can be estimated by using estimates of erroneous inclusion in the census (overcoverage) from the reenumerative studies. Using these figures results in an estimate of net undercoverage ranging from 1.3 to 3.4 percent. This range encompasses both the reinterview estimate and the one arrived at by analytic means.

Actually, we suspect we were too conservative in arriving at estimates that could be reasonably considered as minimum and maximum. Table 3 reflects ranges of assumptions that looked so broad that there would be practically no disagreement that the true value would lie in the range. If one wishes to take a somewhat bolder attitude and some small risk of being wrong, it is possible to narrow the range considerably, but still produce a minimum and maximum that appear to encompass the true figure with a high degree of probability.

This kind of approach produces an estimate of missed persons somewhere in the interval of 3.8 percent to 4.4 percent, and a net coverage error of from 2.5 to 3.1 percent. As the Siegel-Zelnik paper will indicate, this appears to be a reasonable range for the net coverage error.

The small sample sizes and the uncertainties arising from the effect of the noninterviews prevent these studies from providing usable estimates of undercoverage by sex-age-color. An examination of the missed person rates for each of the four population groups separately shows the kind of results that would be expected. Although persons enumerated in the 1950 Census comprised about three-fourths of the population in 1960, they accounted for only two-thirds of the persons missed. Under-enumeration in the other three groups all accounted for more than their share of the population.

In addition to the study of coverage of the total population described above, there were two other record-check studies concentrated on population groups which were believed to represent special coverage problems arising from uncertainties in "usual place of residence" - the elderly and college students. For the elderly, a sample was selected from the list of Social Security recipients. For college students, a sample of colleges was designated and lists of students enrolled at these colleges obtained and sampled.

In the case of the elderly, only underenumeration rates were obtained. The college student study, however, included data on both under - and over enumeration. This was done by a preliminary questionnaire to the sample students inquiring about the location of all residences at which they might have been enumerated, e.g., dormitory, parent's home, whether on vacation during the census period, etc.

The fairly current addresses obtained for these two population groups resulted in a much higher rate of success in determining enumeration status than in the more general study--94.5 percent for the Social Security recipents and 92.9 percent for the college students. Again, with some risk of being wrong, we can state the sample estimate in terms of a fairly narrow range.

Exclusive of sampling error, the gross number of missed OASI beneficiaries appears to be between 512,000 and 573,000 persons, or 5.1 to 5.7 percent of the beneficiaries estimated as enumerated. Similarly, our evaluation sample places the net undercount of college students between 66,000 and 71,000 or around 2.5 to 2.7 percent of the correctly enumerated count.

Table 1.--REINTERVIEW ESTIMATES OF TOTAL COVERAGE ERROR IN THE 1960 CENSUS OF POPULATION, BY TYPE OF ERROR

Item	Estimate	
	Number	Percent
Census Count	179,323,000	100.0
Undercoverage in ED's 1/	5,653,000	3.2
In Missed Living Quarters	3,143,000	1.8
In Partially Enumerated Living Quarters	2,510,000	1.4
Overenumeration in ED's 1/	2,325,000	1.3
In Overenumerated Living Quarters	244,000	0.1
In Properly Included Living Quarters	2,081,000	1.2
Net Undercoverage	3,328,000	1.9
Gross Coverage Errors 1/	7,978,000	4.4

Note: This table includes both enumeration and central office processing coverage errors. The data are based on the EP-8 and EP-9 studies. The estimates represent simple inflation of the sample data.

1/ A person counted only once, but in the wrong Enumeration District, is treated both as a missed person (where he should have been enumerated) and as an overenumerated person (in the wrong Enumeration District). Relative to the United States, this evaluation procedure can result in a substantial overestimate of gross coverage errors. In principle, this procedure does not affect the estimate of net undercoverage.

Table 2.--REINTERVIEW ESTIMATES OF NET COVERAGE ERROR IN THE 1960 CENSUS OF POPULATION, BY AGE, SEX AND COLOR

Color & Age	No. of Persons (in thousands)		Percent of Enumerated Population (including Armed Forces overseas)	
	Male	Female	Male	Female
White				
All Ages	1,219	1,332	1.5	1.7
Under 5	118	145	1.3	1.7
5-14	100	246	0.6	1.6
15-29	189	253	1.2	1.6
30-44	206	268	1.3	1.6
45-64	446	315	2.8	1.9
65 and over	160	105	2.3	1.3
Nonwhite				
All Ages	417	361	4.2	3.4
Under 5	39	27	2.6	1.8
5-14	109	76	4.6	3.3
15-29	3	55	0.1	2.5
30-44	105	87	5.7	4.2
45-64	121	103	7.3	5.9
65 and over	40	13	6.7	1.8

Table 3.--RECORD CHECK MAXIMUM AND MINIMUM ESTIMATES OF MISSED PERSONS AND STANDARD ERRORS OF THESE ESTIMATES, BY POPULATION GROUPS SAMPLED

(in thousands. Excludes estimates for persons not covered by record check studies.)

Item	Total	Population Group Sampled			
		Persons enumer-ated in 1950 Census	Birth reg-istrations from April 1, 1950 to April 1, 1960	Persons estimated as missed in 1950 Census by PES	Aliens regis-tered in Jan. 1960
Maximum Missed Rate					
Missed Persons	8,329	5,612	2,035	295	386
Standard error	816	691	185	69	79
Missed persons, percentage	4.8	4.3	5.5	10.5	15.4
Standard error	0.5	0.5	0.5	2.5	3.1
Minimum Missed Rate					
Missed Persons	4,737	3,193	1,180	167	197
Standard error	641	535	140	52	56
Missed persons, percentage	2.7	2.4	3.1	5.7	7.3
Standard error	0.4	0.4	0.4	1.8	2.1

AN EVALUATION OF COVERAGE IN THE 1960 CENSUS OF POPULATION BY TECHNIQUES OF DEMOGRAPHIC ANALYSIS AND BY COMPOSITE METHODS

Jacob S. Siegel and Melvin Zelnik*
U.S. Bureau of the Census

Method of Demographic Analysis

Introduction.--This paper presents (1) the results of studies using methods of demographic analysis to evaluate the 1960 Census counts and (2) several sets of composite estimates which combine (a) the results derived by various analytic techniques or (b) the results derived by analytic techniques and the case-by-case checking techniques involving reinterviews and matching against independent lists discussed in the companion paper by Messrs. Marks and Waksberg. Because of the close relation between coverage of the total population and the accuracy of the data by age, sex, and color, we are concerned here both with overall underenumeration and with net undercounts (or overcounts) by age, sex, and color.

There are a variety of specific techniques of evaluation that may be classified as techniques of demographic analysis. These techniques make possible the comparison of census counts with some expected result or standard usually derived by the manipulation of such demographic data as census counts and birth, death, and immigration data. There are anumber limitations to these techniques. First, the expected results or standards may be defective, either because of errors in the data underlying them or because of oversimplified assumptions in their construction. Second, these techniques serve best to provide estimates of census error which are relative to a previous census or to other categories in the same census, rather than absolute estimates of error. Third, these techniques provide measures of net census error only--i.e., they cannot distinguish between content and coverage error or between compensating overcoverage and undercoverage. Thus, for example, in dealing with the population of the United States classified by age, sex, and color, the method cannot distinguish between coverage error and errors in classifying persons by age, sex, and color.

The advantages of these techniques are, basically, that they deal with an entire universe or subuniverse and are, for the most part, not handicapped by sampling error or the problems of matching; they focus on levels of error which may be more effectively measured, e.g., net census error or deviations from expected ratios; and the defects of the standards or expected results may be small. In addition, the techniques of demographic analysis often provide a strong basis for judging the demographic reasonableness of census results and of other methods of evaluation.

Although the analytic techniques cannot identify the sources of error, it is still advisable to maintain a conceptual distinction vis-a-vis these sources. Thus, when the focus of our analysis is on the total population, the estimated net errors are estimates of coverage error only. In this context we shall use the term "net underenumeration." When the focus of the analysis is on some segment of the total population, e.g., a specific age-sex-color group, the net error actually refers to the joint effect of both errors of coverage and errors of classification. In this context, we shall use the term "net undercount (or overcount)."

Intercensal estimating equation.--An estimate of the accuracy of the 1960 Census count relative to the 1950 Census count can be arrived at by comparing the difference between the two census figures, on the one hand, and the algebraic sum of the estimates of the components of change during the decade, on the other. If the former figure is larger than the latter figure, then the absolute amount of net census underenumeration has decreased; if the latter is larger, then the absolute amount of net census underenumeration has increased. This assumes that the estimate of net change based on components is without error. Several reports and papers have presented the results of comparisons of this kind.1/ The latest Census Bureau's report giving intercensal population estimates for 1950 to 1960 implies that the amount of net underenumeration was almost exactly the same (difference of 3,000) in 1960 as in 1950. However, in their study published in Demography, Taeuber and Hansen gave an estimate of 277,000 increase in coverage between 1950 and 1960. The difference in these two estimates is a result of different assumptions concerning the amount of net migration of U.S. citizens (exclusive of those moving between Puerto Rico and the United States).2/

Taeuber and Hansen note another element of uncertainty in the determination of the relative levels of coverage in the two censuses, namely the possibility of overenumeration in the 1960 Census figures due to over-imputation of persons. The authors state that the range of this overenumeration could reasonably be from 100,000 to 400,000. However, they do not make an allowance for over-imputation in their estimates.

---

* Dr. Zelnik has been associated with the School of Hygiene and Public Health, Johns Hopkins University, since August 1966.

The authors wish to make special acknowledgement of the contribution of Leon Pritzker of the U.S. Bureau of the Census in connection with the conceptual development of the composite estimates.

If we make maximum and minimum allowances for over-imputation and net movement of U.S. citizens, we can generate a range of estimates for the change in coverage between 1950 and 1960. Thus, assuming 400,000 over-imputations and net in-migration of 280,000 U.S. citizens gives a reduced coverage of 403,000 in 1960. Alternatively, assuming no over-imputations and net outmigration of 172,000 U.S. citizens results in an increased coverage of 449,000 in 1960.

Table 1 shows, for various estimates of the percent net underenumeration in 1950, the percent of net underenumeration in 1960 assuming the following changes in absolute coverage between 1950 and 1960: a) no change; b) an increase of 277,000; c) a decrease of 403,000; and d) an increase of 449,000. If the net underenumeration in 1950 was 1.4 percent, as indicated by the 1950 Post-Enumeration Survey, and if the absolute decrease in coverage between 1950 and 1960 was 403,000, then the percent net underenumeration was the same in 1960 as in 1950. All other comparisons show a smaller percent underenumeration in 1960 than in 1950.

Estimates by age, sex, and color.--Estimates have also been obtained of the relative consistency of the 1950 and 1960 Census counts by age, sex, and color. The residuals derived by comparing the expected population is 1960, based on the 1950 Census counts and data on births, deaths, and net immigration, with the census counts in 1960 represent also the differences between the net undercounts in 1960 and 1950 for age cohorts, assuming that the estimates of intercensal change based on component data are without error. The estimates of intercensal change used to bring the 1950 Census figures forward are consistent with an overall estimate of no change (3,000) in absolute coverage between 1950 and 1960. The estimates of intercensal change and the resulting residual estimates have been set forth in Current Population Reports, Series P-25, No. 310.3/ Such estimates are of quite limited usefulness, however, in establishing the net undercounts in the later census since it is almost impossible to eliminate from the residuals the contribution of errors in the earlier census. They may, however, direct attention to possible anomalies in one or the other census, as for example, the apparent large net overcount of persons 65 and over in 1960, especially of nonwhites (13 percent for each sex), an anomaly which has now appeared in three successive censuses.

Some analytic studies carried out at the Census Bureau have yielded estimates of absolute coverage of population by age, sex, and color in 1960. In Current Population Reports, Series P-25, No. 310, adjusted census data for 1950 and 1960 were employed in the process of preparing intercensal estimates of the population from 1950 to 1960. The adjustments in 1960 were derived by carrying adjusted census figures for 1950 forward to 1960 by estimates of intercensal change and comparing the results with the 1960 Census counts. In the case of whites, the estimates of net census undercounts for ages 15 and over for 1950 were those developed by Coale and Zelnik for native whites.4/

The Coale-Zelnik estimates of net undercounts for native whites 15 and over are based on estimates of births for 1855-1934, which in turn result from the backward projection of females 15-29 in the eight censuses from 1880 to 1950 and the assumption of a uniform level of net undercount of these females amounting to 1.4 percent. Coale and Zelnik found that the figures for white female births obtained by "reviving" native white females enumerated in the age span 15-29 were consistently higher than comparable estimates of the same birth cohorts derived by "reviving" females enumerated for ages below 15 and above age 30; i.e., females 15-29 appeared to be the most fully enumerated group. The estimate of 1.4 percent is offered as the minimum net undercount implied by the available evidence and is associated in part with a net undercount of 1.0 percent for white women 15-54 years of age shown by the 1950 Post-Enumeration Survey. The assumption of a net undercount of females 15-29 in recent censuses, combined with the fact that births estimated from females 15-29 in one census are (approximately) equal to births estimated from females 15-29 in the preceding and following censuses, led to the assumption of uniform net undercounts over time. In order to derive estimates of male births, female births were inflated by a constant sex ratio at birth; this was equivalent to increasing births estimated from males by 3 percent, on the average.

The estimates of net undercounts for nonwhites 15 years old and over in 1950 were derived by a variation of Coale's iterative technique using an assumption that the percent of net undercount in 1940 for each 5-year age group below age 35 and each 10-year age group 35 and over was the same as the average of the percents of net undercount for the same age groups in 1950 and 1960.5/ In the case of both whites and nonwhites, estimates of net undercounts under 15 years of age in 1950 were determined by use of birth statistics adjusted for underregistration, brought forward with allowance for changes due to death and migration. Tables 2 (col. 2) and 3 (col. 4) give the resulting percent net undercounts by age, sex, and color, for the resident population of the United States in 1960. 6/

Recent work suggests that the pattern of net undercounts by age in 1960 is markedly different from the patterns in 1950, 1940, and 1930, even though the age patterns in the three earlier censuses are not too dissimilar from one another.7/ As a result of this difference, the iterative technique linking 1950 and 1960 is probably not a suitable technique for estimating net undercounts in 1960, as was done for the nonwhite estimates in Series P-25, No. 310. We have, therefore, prepared alternative estimates of net undercounts in 1960 for the nonwhite population carrying forward the original Coale estimates of adjusted nonwhite population in 1950 (i.e., those published in 1955) with our estimates of intercensal changes.8/ The Coale estimates of 1950 were derived by an iterative technique on the general hypothesis that the age-patterns of net undercounts were similar in the 1930, 1940, and 1950 Censuses, and on the specific conservative hypothesis that the percent errors in 1930 were equal to those of 1940 or 1950, whichever is less. The least reliable results of this method are at the older ages. Accordingly, the Post-Enumeration Survey results were substituted for persons 65 and over. The 1950 Coale estimates extended to 1960 are offered here as no less reasonable than, and possibly superior to, the other estimates available (see table 3, col. 5).

Estimates based on adjusted births.--In all of the estimates which have been presented up to this point, the expected populations under age 15 in 1950 and under age 25 in 1960 are based on registered births adjusted for underregistration, registered deaths, and estimates of net migration. Since the number of births in any period of time is considerably larger than the number of deaths and net migrants, errors in the completeness of birth registration are of greater consequence for estimates of net census underenumeration and net census undercounts than errors in the other components.

Two tests of the completeness of registration of births have been conducted--one in conjunction with the 1940 Census, the other in conjunction with the 1950 Census. The percent completeness by color and occurrence in hospitals, according to these tests, is as follows:

Year and color	Total	In hospital	Not in hospital
1940			
All classes....	92.5	98.5	86.1
White........	94.0	98.6	88.2
Nonwhite.....	82.0	96.3	77.2
1950			
All classes....	97.9	99.4	88.2
White........	98.6	99.5	88.2
Nonwhite.....	93.5	98.2	88.2

Estimates of the completeness of registration for the intercensal years 1940 to 1950 were derived by interpolating between the 1940 and 1950 test results. Specifically, it was assumed that the change in percent completeness followed a linear trend between the decennial years with respect to hospital and non-hospital births, for the white and nonwhite groups separately, for each State.

No test of birth registration completeness was carried out in 1960. As a result, estimates of the completeness of birth registration for years subsequent to 1950 were based on the results of the 1950 test, on the assumption that percent completeness by occurrence in- and not-in-hospital, by color and by State, was the same as in 1950. (In effect, an estimated change in registration for each color group comes about from changes in the proportion utilizing hospitals for childbirth.) Similarly, the results of the 1940 test were used to derive estimates of the completeness of birth registration for the years 1935-39.

Although there were some differences in the designs of the two Birth Registration Tests, they consisted, essentially, of matching birth records covering some specified period of time immediately preceding the census with cards prepared for infants born during that period and enumerated in the census. The major source of error in this procedure, ignoring problems of matching, involves infants who were not enumerated. If, among these persons, the proportion whose birth was not registered was the same as among those enumerated, then their omission from the census would not affect the estimate of completeness of birth registration. However, it is quite likely that among the not enumerated, the proportion whose birth was not registered was higher than among those enumerated. This would mean that estimates of the completeness of birth registration were too high, thereby introducing a downward bias in estimates of population based on adjusted births.

Chandra Sekar and Deming have examined the effect on the estimates of completeness of registration, of the omission of infants in the census 9/ and have suggested a method for estimating the bias. The basic objective of the method is to subdivide an area (either geographically or by a combination of characteristics) into sub-groups each of which is highly homogeneous with respect to enumeration completeness (a completely homogeneous population is defined as one in which each individual has an equal probability of being enumerated). Within such sub-groups, the correlation between unregistered and unenumerated events would be very low. An estimate of the total number of births (registered and unregistered) in the area could then be derived by cumulating the "total number of births" corrected for underregistration. Relating the figure for registered birth to this total would then give a figure approximating the unbiased estimate of completeness of registration. This method was applied by the National Office of Vital Statistics to the results of the 1940 and 1950 tests--in the former case to all States and in the latter to the State with the lowest registration completeness in 1950. The results indicated that underenumeration had little effect on estimates of registration completeness.

In a further attempt to explore this question, we have used the national results of the 1950 Birth Registration Test and the 1950

Infant Enumeration Study in conjunction with two assumptions of dependence between unregistered and unenumerated events to generate two alternative estimates of birth registration completeness in 1950 for total United States. The assumptions we employed were: a) the degree of "not registered" was zero for the "not enumerated" group, and b) the degree of "not registered" for the "not enumerated" was ten times as great as the degree of "not registered" for the "enumerated." These two assumptions led to estimates of completeness of registration of 98.0 percent and 96.9 percent, respectively, as contrasted to the actual estimate of 97.9 percent. Thus, rather extreme assumptions had only very slight effects on the level of the completeness of registration.

Aside from the issue just discussed, it has been suggested that the results of the Birth Registration Tests are biased upward, especially in the case of nonwhites, because of the difficulties of establishing matches.10/ Since it is not possible at this date to examine the original documents, we have employed a fairly extreme assumption to measure the sensitivity of the estimates of net census undercount to an overestimate of the incompleteness of registration. We have assumed for illustrative purposes a reduction of one-third in the annual estimates of the incompleteness of registration of births, by color. Table 4 compares the estimates of net census undercount, for the population under 25 years of age, by age, sex, and color, based on the official estimates of the completeness of registration of births, with estimates employing the assumption stated. We want to emphasize that the results do not provide any information on the accuracy or inaccuracy of the official estimates of the completeness of birth registration. They merely indicate what the effect would be on estimates of net census undercount if the estimates of completeness of registration were in error by as large a margin as we have assumed. The effect is especially noticeable for nonwhites in the age groups 15-19 and 20-24. The estimate of net undercount of nonwhite males 20-24 is reduced from 21 to 11 percent; the corresponding figure for females is reduced from 11 percent to 2 percent.

We are inclined to accept the official estimates of completeness of birth registration, even though these estimates lead to large estimates of net census undercount for nonwhites, especially males. The estimates for nonwhite males aged 20-24 in 1960 arrived at through the use of the official birth registration figures are not very different from the undercounts estimated for nonwhite males aged 20-24 in 1950 by Coale, using an iterative technique,11/ and these estimates are, in turn, similar to the undercount estimated for nonwhite males in 1940 by Price,12/ using Selective Service data. While consistency in the level of net census undercount over three censuses is not proof of the accuracy or validity of any one or all three of the estimates, it does suggest the reasonableness of the estimates. Confidence in the estimates is increased by the fact that the undercounts for the 1940 and 1960 censuses are based on two quite different methods and bodies of data.13/ In our opinion the high undercounts for nonwhite males in 1960 and concern with the possibility of inadequate matching in the tests do not provide sufficient reason for rejecting the results.

Method of expected sex ratios.--We have also employed another analytic technique for estimating net undercounts by age, sex, and color--namely, the application of "expected sex ratios." There are two problems involved in the use of this technique if it is to provide absolute estimates of adjusted population for both sexes. First, it requires an acceptable, independently determined, set of estimates of net undercounts, by age, for one sex. Second, the estimation of expected sex ratios involves a number of approximations which may lead to varying degrees of error. Both of these problems are more difficult to resolve in the case of nonwhites and the older population. We have completed the preparation of a set of expected sex ratios for the resident population, by age and color, which take account of the observed or estimated "actual" sex ratios of births, changing mortality by sex (represented by sex ratios of survival rates from various official life tables combined as quasi-generation life tables), excess war deaths, and the cumulative effects to 1960 of net civilian and military movement to and from the United States by sex. These sex ratios are offered as more realistic than those serving the same purpose which are computed from a single sex ratio of births and a conventional life table for 1960.

The method assumes that the underregistration of births does not vary by sex and that the sex ratios of survival rates are not seriously affected by errors in the basic data used in constructing official life tables. In order to measure illustratively the effect of net census undercounts on the level of the sex ratios of survival rates, and hence on the level of expected sex ratios, the 1900 and 1960 life tables for Negroes or nonwhites were recalculated on the basis of the percent net undercounts for 1960 and the resulting sex ratios of survival rates were compared with similar ratios based on unadjusted life tables. These computations indicate that the expected sex ratios would tend to be higher at most ages if the life tables were adjusted and that they are particularly sensitive to the level of the net undercounts at the older ages, where mortality rates are high.

A comparison of the expected sex ratios and the "enumerated" ratios in 1960 indicates ratios lower than expected at all ages below 50, for whites and nonwhites separately, especially at ages 20 to 49 for nonwhites (ee table 5 for summary results). At ages above 55 the "enumerated" ratios are usually higher than expected, especially for nonwhites. (Adjustment of the expected sex ratios at these ages for net census undercount does not bring them up to the level of the "enumerated" ratios, however.) The expected sex ratios agree quite closely with the sex ratios of the estimated population under 25 years of age, by age and color, based on births adjusted for underregistration, deaths, and net immigration. Estimates of adjusted male population 25 and over in 1960, employing the sets of expected sex ratios, were derived from the analytic estimates for females and are shown in table 2 (col. 3) and table 3 (col. 6).

Estimates of Negro population.--Another analytic set of estimates of net undercounts is that prepared by Bogue and his associates for the Negro population by age and sex (table 3, last col.).14/ We have given only limited attention to these estimates in our evaluation studies because, as the published critique by Zelnik indicates,15/ they suffer from a number of deficiencies. The methodology includes the use of an adjustment for age heaping to allow for net age-misreporting of grouped data (although the former type of adjustment is not particularly relevant to the latter problem), an (arbitrary) 2-percent estimate of net undercoverage by age, an additional (duplicate) allowance for net undercount of children under 10 based on births adjusted for underregistration, and use of a synthetic life table for 1960 to determine the sex ratios by age. The estimates of net undercounts arrived at by this procedure are generally lower than the other analytic estimates, with the outstanding exception at ages under 10. Coale has pointed out that the Bogue's study has the highly doubtful implication that birth registration has been deteriorating.

Preferred analytic composite.--Using the analytic techniques described so far, a set of estimates was defined representing a "preferred composite based on demographic analysis" (table 6, cols. 3 and 4). The percents of adjustment under age 25 by age and sex were derived from births adjusted for underregistration, carried forward with deaths and net immigration. The percents for the white population 25 and over by age and sex were based on extensions to 1960 of the Coale-Zelnik estimates for 1950. The estimates for nonwhite females 25 and over by age were based on extensions to 1960 of the Coale estimates for 1950. The figures for nonwhite males were obtained by applying expected sex ratios to the nonwhite female population.

Results by age, sex, and color.--The differences among the alternative sets of estimates of net undercounts derived by demographic analysis, excluding the Bogue-Misra-Dandekar estimates, are small. Thus, the estimated net underenumeration for the total population is 3.1 or 3.2 percent, depending on the specific combination made of estimated net undercounts for the sex-color groups. For males, the estimates range from 3.8 to 4.0 percent and for females from 2.3 to 2.4 percent. These alternative sets of estimates yield the same estimate of net undercount for whites, 2.2 percent. For nonwhites, the estimates vary from 10.2 to 10.6 percent. All of these undercounts assume no change in the overall coverage of the 1950 and 1960 Censuses (i.e., a net immigration of 280,000 civilian citizens between 1950 and 1960).

The estimated net undercount for white females is 1.6 percent and for white males 2.8 or 2.9 percent. The differences between the two sexes are most pronounced from ages 15 to 49. At ages beyond this, females appear to be no better enumerated than males. The estimated undercounts for nonwhite males vary from 12.2 to 12.7 percent and for nonwhite females from 7.8 to 8.8 percent. The nonwhites show approximately the same pattern as the whites--smaller undercounts for females through the young adult ages, with smaller undercounts for males at the older ages.

There is wide variation in the estimates of net undercount for persons 65 and over, particularly for white and nonwhite males.

Synthesis of Methods of Evaluation

Comparison of results.--We have described several sets of purely analytic estimates of net undercounts in the 1960 Census and the Marks-Waksberg paper has described the estimates of coverage error derived from the record-matching studies and the reinterview studies. We should now like to consider these in relation to one another.

The results of the various methods for females, and particularly white females, are close, but the results for the other sex-color groups are quite different. (For this discussion, the analytic series principally referred to is the series identified as "preferred composite based on demographic analysis.") The comparative estimates of missed females are 1.8 percent for the reinterview method and 2.4 percent for the analytic method, and of missed males are 1.8 percent for the reinterview procedure and 3.9 percent for the analytic method. Both for males and females the discrepancy is much greater for nonwhites than for whites, although the figures for white males differ significantly too. For nonwhites as a whole, the respective figures are 3.8 percent and 10.5 percent. The estimate of net underenumeration of white females from the reinterview studies and the demographic analysis is virtually identical, 1.7 and 1.6 percent.

The relationships are much more erratic for individual age groups. Figures from the reinterview studies are affected not only by coverage errors but also by age-reporting errors and other problems of estimation, including sampling errors. Although differences between the analytic estimates and the reinterview estimates (whether total net census error or net coverage error), considered in terms of broad age groups, are relatively small for white female, they are particularly great for white males in the age groups 15-29, 30-44, and 65 and over, for nonwhite males in all age groups except 5 to 14, and for nonwhite females under 5, 15-29, 45 to 64, and 65 and over. In these cases, with the exception of the age group 65 and over, the analytic method shows the larger net undercounts.

Relative limitations of methods.--Each of the three methods of evaluation is subject to various limitations and varying degrees of error. Reference was made to the limitations of the reinterview and record check studies in the paper by Marks and Waksberg. In brief, the record check studies provide an impracticably wide range of estimates of the extent of gross underenumeration. Allowing for gross overenumeration of 1.3 percent, the figures on net undercoverage range from 1.3 percent to 3.4 percent, depending on the assumptions made with regard to the coverage of persons for whom a definite determination about inclusion in the 1960 Census could not be made (16.5 percent of the sample). The figures encompass the available estimates, although some narrowing of the range may be possible.

The analytic studies strongly suggest that the reinterview studies understate the overall coverage error for males and nonwhites. This deficiency applies in the case of all sex-color groups except white females. The understatement of the error for nonwhite males by the reinterview studies is especially apparent in comparison with the "preferred analytic" estimates, but it remains evident even when the "analytic" estimates for males are derived by applying expected sex ratios to female population adjusted by the net coverage error from the reinterview studies. The 4 percent net underenumeration of nonwhite males from the reinterview studies compares with 12 percent and 7 percent from the analytic estimates referred to (table 6, cols. 1, 3, and 5). Some of the estimates of net coverage error by age are unreasonably low, as for example, 0.1 percent for nonwhite males aged 15-29 and 1.2 percent for white males aged 15-29, or unreasonably high, as for example, 6.7 percent for nonwhite males 65 and over. Furthermore, the population sex ratios by age implied by the net coverage errors from the reinterview studies tend to be too low, particularly for nonwhites, although there are some striking "errors" in the opposite direction (table 5, col. 5).

Marks and Waksberg caution against the use of the theoretically more appropriate estimates of net census errors (net coverage error combined with net age-misreporting error) from the reinterview studies on the ground that the age-misreporting-error component is subject to very large sampling errors and response biases, too large to add any information to that afforded by the net coverage error. This component is also affected by the assumptions of the estimating method. Accordingly, we must generally rely on the net coverage error from the reinterview studies to represent the net census error; yet the net coverage error may substantially understate or overstate the net census error if there is a pronounced bias in age reporting. The reliability of the net census errors should be greater for whites and broader age groups.

The analytical approach also has its limitations. There is considerable dependence of the estimates of net undercounts for persons under 25 on the results of the Birth Registration Tests in 1940 and 1950 and yet there is some uncertainty as to the accuracy of these tests. Difficulties in matching of the census records with birth certificates would tend to cause an overstatement of the underregistration of births; and, as indicated earlier, the estimates of net undercounts are quite sensitive to any errors in the correction for underregistration. On the basis of the 1940 test results indicating that 18 percent of nonwhite births were not registered, there is no doubt that there was substantial underregistration of nonwhite births in 1940, but there is a real question about the precise extent of underregistration.

Next, the estimates of net undercounts for nonwhites above age 25 (i.e., 1950 Coale estimates extended to 1960) depend heavily on the assumption of similarity of the pattern of net undercounts at successive recent censuses (1930, 1940, and 1950 in the Coale estimates) and on the estimates of net undercounts for children in these censuses based on births adjusted for underregistration (under 15 in the 1950 Coale estimates). The iterative technique has a tendency to accumulate errors as one goes up the age scale, so that the estimates for the older ages, particularly 65 and over, may be defective. Coale himself rejected his original estimates for 65 and over.

Furthermore, the Census Bureau extension of the 1950 Coale-Zelnik estimates of net undercounts for whites above age 25 in 1960 may be questioned on a number of grounds. There is, first, the acceptability of the fundamental assumptions by which the basic estimates were derived in 1950, particularly the assumption of a common level of net undercount for females 15-29 in each prior census. Further questions relate to the procedure of estimating population by age in 1950 after the historical series of birth estimates was determined, the procedure of extending the estimates for native whites in 1950 to include the foreign-born, and, as in general, the adequacy of the estimates of intercensal change by which the adjusted population in 1950 was carried to 1960.

Composite estimates based on reinterview studies and demographic analysis.--Despite these limitations, we want to take advantage of these various estimates to develop a set or sets of estimates of population for April 1, 1960, by age, sex, and color, that would be significantly more accurate than the 1960 Census statistics and which could be recommended for general use.

A number of possible criteria for such estimates may be identified. A single set of "best" estimates of net undercounts may be sought. On the other hand, it may be preferable to try to establish a range, giving high and low estimates, or minimum and maximum estimates, of net undercounts. These could be developed in combination with, or independent of, a set of "best" estimates. The range would suggest the degree of uncertainty associated with the estimates of net undercount, although no specific mathematical probabilities could be assigned to the high and low figures. The calculation of alternative estimates has certain limitations and certain advantages. Offering alternative estimates presents certain practical difficulties to many users, who prefer a single set of figures; on the other hand, the availability of a set of high and low figures makes possible choice by the user in conformity with a cost analysis of his problem, which may call for a high or a low figure. He will often prefer the high series, particularly if this is consistent with maximum costs. On the other hand, the high estimates involve the risk of deviating from the true figure more than the census counts do, i.e., they involve the risk of serious overstatement. In fact, it may be considered desirable to avoid overstatement altogether and to develop a set of adjustments which may be regarded with a high degree of certainty as being understatements of the errors in the census counts and yet the largest acceptable estimates of error. As lower bounds of the true figures or "minimum reasonable" estimates, such figures may be described as representing a highly conservative choice of a single set of best estimates.

A few experimental sets of estimates of this kind have been prepared. The starting point for the first set is the estimate of net coverage error for white females obtained from the reinterview studies and the demographic analyses. As noted, the estimate is virtually identical in both sources, 1.7 or 1.6 percent of the census count. The estimated errors for white females differ somewhat by age group in the two series, however. The estimates of net coverage error from the reinterview studies, which are remarkably constant up to age 45 and approximately so throughout the age distribution, were adopted. Thus, estimates of the total number of white females in the United States by age were obtained from the reinterview studies. Since we are attempting to understate the net error, the estimates of nonwhite females derived from the reinterview studies were used in the same way (3.4 percent at all ages as compared with 8.8 percent from the demographic analyses). To derive estimates of the adjusted male population, expected sex ratios were applied to the estimates of adjusted female population. The results are shown in broad age groups in table 6 (set 1). The resulting net errors are: for whites males 3.0 percent, for nonwhite males 6.7 percent, and for the total population 2.6 percent. The use of expected sex ratios gives a net undercount for white males 1.3 percentage points greater than for white females and 3.3 percentage points greater for nonwhite males than for nonwhite females. The resulting figures for males exceed the net coverage errors of the reinterview studies in most age groups and hence are not minimal in relation to the available estimates. However, the net coverage errors for males from the reinterview studies are, for the most part, untenably low now only in relation to the figures for males obtained from the adjusted figures for females and expected sex ratios but also in relation to the net coverage errors for females.

The second set of composite estimates of net undercounts (whites only) also makes use of the fact that the overall net coverage error for white females obtained in the reinterview studies is about the same as the figure shown by the analytic studies. In these calculations, estimates of the population under 25 years of age based on adjusted births, deaths, and net immigration were combined with estimates for females 25 and over consistent with an all-ages coverage error of 1.6 percent for females, and estimates for males 25 and over were then derived from the estimates for females by use of expected sex ratios. These calculations happen to imply a net coverage error of 1.6 percent for the white female population 25 and over also. Females were assigned the net census errors (net coverage error plus net age-reporting error) from the reinterview studies for very broad age groups (Series A, ages 25-44, 45-64, and 65 and over; Series B, ages 25-34, 35-44, 45-54, and 55 and over), and the figures so adjusted were distributed into smaller age groups on the basis of the "demographic" estimates. The resulting estimates of error are shown in table 6 (set 2). The net error for white males is 2.9 percent.

Still another approximation to "conservative best" estimates or "minimum reasonable" estimates are given as set 3 in table 6. In this set the demographic estimates of net undercounts based on adjusted births, deaths, and net immigration were accepted only for ages under 15 (under 5 and 5 to 14).16/ This choice implies sufficient confidence in the results of the 1950 Birth Registration Test to accept the estimates based on births since 1945, but not in the results of the 1940 Test to accept the estimates based on earlier births.17/ For the next three older groups of females, 15-29, 30-44, and 45-64, we selected the smaller figure as between the estimates of net undercount from the demographic analyses and the net coverage error from the reinterview studies. Accordingly, all figures came from the reinterview studies except that for white females 30-44 years of age. For this group, the figure selected was quite small, 0.1 percent, but it agrees with the net census error from the reinterview studies. Estimates of net undercounts for males were derived by applying expected sex ratios to the adjusted figures for females. To complete this set of estimates, the population 65 and over, for whites and nonwhites, was assumed to have no net error since we have been unable to establish whether the census counts overstated or understated the population. The census counts were then divided by sex on the basis of expected sex ratios.

Estimates for males so calculated are often well above the coverage errors for males from the reinterview studies. This set of figures shows net errors of 2.4 percent for white males, 1.1 percent for white females, 8.0 percent for nonwhite males, and 4.7 percent for nonwhite females (table 6, set 3). These levels of net error are somewhat lower for white males and females, and somewhat higher for nonwhite males and females, than the levels indicated by the composite estimates based on both demographic analysis and reinterview studies, previously computed (i.e., sets 1 and 2 of table 6). The overall level of net underenumeration in 1960, 2.2 percent, is roughly the same as the figure expected on the basis of the Census Bureau's "minimum reasonable" estimate of 2.5 percent for 1950.

A logical weakness of this procedure is that it is unlikely that the net error for all nonwhites 65 and over is zero while nonwhite males are substantially overstated (6 percent) and nonwhite females are substantially understated (5 percent). If, on the other hand, the census counts for both sex groups are taken without adjustment, the implied sex ratio for nonwhites is quite unreasonable (90 as compared with an expected value of 80).

Of course, many other sets of composite estimates of net undercounts, designed to represent conservative estimates, are possible. One could accept the "preferred composite based on demographic analysis" as the best estimates, and reduce these by a fixed proportion, say one third or one-quarter, to derive a set of conservative best estimates. However, this procedure would give estimates for white females below those from the reinterview studies and estimates for ages under 25 below those based on adjusted births.

The various composite series described here would not necessarily increase the accuracy of the relative distribution by age, for the census date. Furthermore, the changes by age since the census date, implied by current estimates of population adjusted on the basis of the composite estimates of net census undercounts, would not necessarily be more realistic than if the data had not been adjusted. These limitations result from the fact that the proportion of the actual net undercounts allowed for in these composite estimates varies from age to age. On the other hand, the absolute level of the adjusted census counts or current estimates at each age (except possibly 65 and over) would be closer to the theoretical truth.

Differential coverage.--One of the main reasons for the concern about the extent of coverage error in censuses stems from the fact that it varies widely among groups of the population (age, sex, color, socioeconomic status, etc.) and geographic areas in the country (States, cities, counties, urban-rural, etc.). If the level of coverage error were the same among all population groups and geographic areas, there would be less need for and concern about the availability of estimates of coverage error.

In general, we believe that coverage error contributes more heavily to the anomalies of the census counts by age, sex, and color than errors of age reporting. Therefore, we are less concerned about age-reporting errors than undercoverage for most of the age distribution; possibly at ages over 50, age-reporting errors become relatively important and may dominate.

Our studies of the quality of the 1960 Census data have indicated the following differences in coverage:

1. The enumeration of males is less complete than that of females, at least up until 45 or 50.

2. The enumeration of nonwhites is substantially less complete than that of whites, probably at all ages but certainly up until about age 60 or 65.

3. The enumeration of males at ages 15 through 44, especially for nonwhites, is less complete than at other ages or the average level over all ages. The population under 5 is no longer to be singled out as a group with especially bad coverage.

In addition, by inference from the 1960 reinterview studies and from the results of the 1950 Post-Enumeration Survey, we may conclude that:

1. There are important geographic variations in the completeness of enumeration. Coverage is probably poorer in the central cities of our metropolitan areas than in the suburban counties and probably poorer in the South than in the rest of the United States. Coverage is probably poorest in the slum areas of our big cities, but we do not have evidence from interview or other studies to support this conclusion.

2. The underenumeration of young children (children under 5 years of age) is very probably closely related to the underenumeration of their parents. According to the Infant Enumeration Study of 1950, in 80 percent of the cases where infants were missed their parents were also missed.18/

Age-reporting errors.--Because of the limitations of the record-matching studies and the reinterview studies, we have little basis for describing the pattern of age-reporting errors in the 1960 Census. Two estimates are available: (1) estimates based directly on the reinterview studies and (2) estimates derived by taking the difference between the net undercounts based on demographic analysis and the net coverage errors from the reinterview studies. Reference has already been made to the inadequacies of the former estimates arising from sampling error, response biases, and the assumptions in the estimating method. The lack of comparability of the components of the second estimate and the limitations of each component have also been noted. The estimates for white females alone may be informative. In short, we do not have any solid facts about the age-reporting errors in the 1960 Census.

There is some evidence, although far from conclusive (for example, the residual estimates for the population 65 and over for the 1950-60 decade), to suggest that the 1960 Census may contain a net overstatement of persons 65 years old and over. This overstatement is accompanied by an understatement, arising from age-reporting errors, in the age groups that immediately precede the 65-and-over group. However, as of now, we do not know the extent, or even the direction, of error in the census count 65 and over. Further research may contribute to a clarification of this question.

Conclusion.--In conclusion, we know little in a formal manner regarding the reasons for underenumeration or the geographic variations in coverage error, and have only rough or approximate measures of net census errors by age, sex, and color. We continue to have considerable concern about the validity of the differences by color and by age shown by the available estimates of net coverage error or net census error. We have greater confidence in the validity of our estimates of the differences by sex.

We have developed a number of sets of estimates of net undercounts by age, sex, and color, some involving a synthesis of methods and techniques, but we have so far been unable to arrive at a single set of figures which we feel we can recommend for general use. Efforts along these lines will continue with the hope of achieving this goal.

Footnotes

1. See U.S. Bureau of the Census, Current Population Reports, Series P-25, Nos. 331 and 310; Donald S. Akers, "Estimating Net Census Undercount in 1960 Using Analytical Techniques," paper presented at the annual meeting of the Population Association of America, Madison, Wisconsin, May 1962; and Conrad Taeuber and Morris H. Hansen, "A Preliminary Evaluation of the 1960 Censuses of Population and Housing," Demography, Vol. I, No. 1, 1964.

2. On the basis of the two available estimates of net movement of this group (-172,000 and +280,000), Taeuber and Hansen assumed zero net movement while the other studies had adopted the figure of +280,000.

3. U.S. Bureau of the Census, "Estimates of the Population of the United States and Components of Change, by Age, Color, and Sex, 1950 to 1960," Current Population Reports, Series P-25, No. 310, by J. S. Siegel, D. S. Akers, and W. D. Jones, June 30, 1965.

4. Ansley J. Coale and Melvin Zelnik, New Estimates of Fertility and Population in the United States, Princeton University Press, Princeton, N. J., 1963.

Other estimates of net undercounts for (native) whites in 1960 than those described in the text are shown in: M. Zelnik, "Errors in the 1960 Census Enumeration of Native Whites," Journal of the American Statistical Association, Vol. 59, June 1964, pp. 437-459.

5. The results were adjusted by the application of sex ratios based on quasi-generation life tables. The estimates for 10-year groups in 1950 were then distributed into 5-year groups by a graduation method. See, also, Akers, op cit.

6. The estimates of intercensal change used to bring the 1950 adjusted census figures forward are consistent with an estimate of no change (3,000) in overall coverage between 1950 and 1960; the figures are those shown in table 12 of Current Population Reports, Series P-25, No. 310. The estimates of net undercounts in 1960, shown in tables C-2 of that report, refer to the total population of conterminous United States (excluding Alaska and Hawaii), including members of the Armed Forces overseas, and are based on the population adjusted for net undercounts.

7. Melvin Zelnik, "An Examination of Alternative Estimates of Net Census Undercount, by Age, Sex, and Color: 1950 and 1960," paper contributed to the annual meeting of the Population Association of America, New York, N.Y., April 1966.

8. A. J. Coale, "The Population of the United States...A Revision of Census Figures," op. cit; M. Zelnik, "An Examination of Alternative Estimates...," op. cit.

9. C. Chandra Sekar and W. E. Deming, "On a Method of Estimating Birth and Death Rates and the Extent of Registration," Journal of the American Statistical Association, Vol. 44, No. 245, March 1949.

10. D. J. Bogue, B. D. Misra, and D. P. Dandekar, "A New Estimate of the Negro Population and Negro Vital Rates in the United States, 1930-1960," Demography, Vol. I, 1964.

11. A. J. Coale, op. cit.

12. Daniel O. Price, "A Check on Underenumeration in the 1940 Census," American Sociological Review, Vol. XII, Feb. 1947, pp. 44-49.

13. Other grounds for accepting the official registration figures have been offered by A. J. Coale. He points out that, if failure to match had represented a considerable part of what was considered underregistration, some of this would have occurred in hospital births, which were in fact nearly completely registered. He has also indicated that a study by J. T. Yamaguchi of the Princeton Office of Population Research, comparing the population as enumerated in 1960 by age, by geographic division of birth, with corresponding estimates for the census date based on births in each geographic division adjusted for underregistration, tends to confirm the high level of underregistration of births in 1940.

14. Bogue, Misra, and Dandekar, op. cit.

15. Melvin Zelnik, "An Evaluation of New Estimates of the Negro Population," Demography, Volume 2, 1965, pp. 630-639.

16. For whites these were generally lower than the estimates of net coverage error; for nonwhites they were higher.

17. This is an arbitrary choice since 1945 does not represent a turning point in the improvement of birth registration; rather, improvement was gradual during the forties.

18. U.S. Bureau of the Census, Infant Enumeration Study: 1950, Procedural Studies of the 1950 Censuses, No. 1, 1953.

Table 1. Estimated Percent of Net Underenumeration in 1960, for Various Percents of Net Underenumeration in 1950 and Various Absolute Amounts of Change in Coverage Between 1950 and 1960

(Base of percent is census count of resident population)

1950		1960 percent net underenumeration according to 1950-60 change in coverage			
Source	Percent net underenumeration	No change in coverage	Coverage increase of 277,000	Coverage decrease of 403,000	Coverage increase of 449,000
PES estimate 1/...........	1.4	1.2	1.0	1.4	0.9
Minimum reasonable estimate 1/..................	2.5	2.1	1.9	2.3	1.8
Arbitrary 3 percent.......	3.0	2.5	2.4	2.8	2.3
Coale estimate 2/.........	3.6	3.0	2.9	3.3	2.8

1/ U.S. Bureau of the Census, The Post-Enumeration Survey: 1950, Technical Paper No. 4, 1960.

2/ 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, Vol. 50, March 1955, pp. 16-54.

Table 2.--ESTIMATES OF PERCENT NET CENSUS UNDERCOUNT OF THE WHITE POPULATION, BASED ON ANALYTIC AND COMPOSITE METHODS, BY AGE AND SEX: 1950 AND 1960

(Base of percent is census count of resident population; minus sign denotes net census overcount)

Sex and age	1950	1960			
	Series P-25, No. 310[1]	Series P-25, No. 310[2]		Composite estimate[5]	
		A[3]	B[4]	A	B
Male, all ages	3.2	2.8	2.9	2.9	2.9
Under 5	4.5	2.0	2.0	2.0	2.0
5-9	3.1	2.5	2.5	2.5	2.5
10-14	1.0	2.6	2.6	2.6	2.6
15-19	4.2	4.0	4.0	4.0	4.0
20-24	5.9	4.5	4.5	4.5	4.5
25-29	5.2	4.9	4.4	4.9	4.7
30-34	4.5	5.1	3.2	3.7	3.5
35-39	2.1	4.8	2.6	3.1	3.3
40-44	3.4	3.5	1.9	2.4	2.6
45-49	2.2	2.5	1.6	0.7	-0.1
50-54	2.2	4.4	3.7	2.7	2.0
55-59	5.3	0.7	0.4	-0.6	0.7
60-64	3.6	3.6	3.1	2.1	3.4
65 and over	-2.0	-3.4	4.0	4.7	4.4
Female, all ages	2.1	1.6		1.6	1.6
Under 5	3.8	1.2		1.2	1.2
5-9	2.5	1.6		1.6	1.6
10-14	1.1	1.5		1.5	1.5
15-19	1.8	2.5		2.5	2.5
20-24	1.6	2.5		2.5	2.5
25-29	0.3	1.4		1.9	1.7
30-34	0.1	0.6		1.1	0.9
35-39	-1.4	-0.2		0.2	0.4
40-44	2.0	-0.2		0.3	0.4
45-49	1.5	0.7		-0.3	-1.0
50-54	2.7	4.4		3.4	2.7
55-59	7.6	1.6		0.7	2.0
60-64	7.3	4.4		3.4	4.7
65 and over	2.3	2.1		2.8	2.5

[1] U.S. Bureau of the Census, Current Population Reports, Series P-25, No. 310, 1965. Figures relate to resident population of the United States. Estimates under 15: Based on adjusted births, deaths, and net immigration data. Estimates 15 and over: Based on Coale-Zelnik estimates for native whites.

[2] Estimates under 25: Based on adjusted births, deaths, and net immigration data.

[3] Estimates 25 and over for males and females are extensions to 1960 of Coale-Zelnik estimates for 1950.

[4] Males 25 and over: Expected sex ratios applied to adjusted female population.

[5] Males and females under 25: Based on adjusted births, deaths, and net immigration data. Males 25 and over: Expected sex ratios applied to adjusted female population. Females 25 and over: Adjusted for net census error in broad age groups, as indicated by reinterview studies, consistent with an all-ages net coverage error of 1.6 percent and a uniform net coverage error of 1.6 percent for ages 25 and over by age. Percent redistribution of females into 5-year age groups in alternative series based on:

A. Distribution within age groups 25-44 and 45-64 from P-25, No. 310.

B. Distribution within age groups 25-34, 35-44, 45-54, and 55 and over from P-25, No. 310.

Table 3.--ESTIMATES OF PERCENT NET CENSUS UNDERCOUNT OF THE NONWHITE POPULATION, BASED ON ANALYTIC AND COMPOSITE METHODS, BY AGE AND SEX: 1950 AND 1960

(Base of percent is census count of resident population; minus sign denotes net census overcount)

Sex and age	1950			1960			
	P-25, No. 310[1]	Coale[2]	Bogue et al.[3]	P-25, No. 310[1]	1950 Coale estimates extended to 1960[4]		Bogue et al.[3]
					A[5]	B[6]	
Male, all ages	14.8	15.0	8.8	12.7	12.6	12.2	8.9
Under 5	11.1	11.0	8.9	8.4	8.4	8.4	11.0
5-9	11.9	12.0	2.8	6.0	6.0	6.0	8.9
10-14	6.6	7.0	-	5.5	5.5	5.5	3.0
15-19	15.3	18.0	11.7	14.3	14.3	14.3	4.7
20-24	19.0	19.0	22.6	21.2	21.2	21.2	16.8
25-29	24.9	20.0	19.5	18.9	21.7	24.5	17.7
30-34	34.4	20.0	19.2	16.0	16.0	22.0	18.7
35-39	19.6	12.0	6.3	22.7	17.8	16.9	16.9
40-44	13.6	20.0	5.3	28.8	14.4	14.6	12.8
45-49	16.4	13.0	4.8	21.8	13.5	13.1	7.4
50-54	14.8	11.0	8.3	15.5	22.8	21.7	6.2
55-59	18.0	17.0	17.3	13.3	9.3	6.3	-2.0
60-64	17.1	24.0	20.2	18.5	13.5	10.6	10.6
65 and over	-13.7	12.0	-15.7	-9.2	12.2	1.9	-10.1
Female, all ages	9.5	11.0	3.5	7.8	8.8		3.8
Under 5	10.3	10.0	7.6	6.8	6.8		9.8
5-9	9.7	10.0	0.3	5.1	5.1		8.1
10-14	7.0	7.0	-1.7	4.4	4.4		2.1
15-19	8.7	12.0	3.4	11.2	11.2		1.3
20-24	2.6	8.0	3.3	10.7	10.7		2.7
25-29	6.7	8.0	4.3	6.4	9.6		2.2
30-34	8.6	9.0	4.3	1.0	6.3		1.5
35-39	2.2	3.0	-5.2	5.4	6.7		3.3
40-44	5.6	18.0	1.1	6.4	6.8		1.3
45-49	12.3	12.0	2.8	8.3	9.2		-
50-54	19.6	16.0	12.4	8.3	22.3		2.7
55-59	35.0	30.0	30.9	11.5	11.1		-1.3
60-64	36.4	36.0	36.4	20.9	16.5		10.8
65 and over	6.5	5.0	-14.4	17.2	14.0		0.8

- Represents zero or rounds to zero.
[1] U.S. Bureau of the Census, Current Population Reports, Series P-25, No. 310, 1965. Figures relate to resident population of the United States.
[2] A. J. Coale, Journal of the American Statistical Association, 1955. Figures relate to resident population of conterminous United States.
[3] D. J. Bogue et al., Demography, 1964. Figures relate to resident Negro population of conterminous United States.
[4] Estimates under 25: Based on adjusted births, deaths, and net immigration data.
[5] Males and females 25 and over: 1950 Coale estimates carried forward to 1960 by estimates of intercensal change given in Series P-25, No. 310.
[6] Males 25 and over: Expected sex ratios applied to adjusted female population for 1960.

Table 4.--ESTIMATED PERCENTS OF NET CENSUS UNDERCOUNT FOR THE POPULATION UNDER 25 YEARS OF AGE, BY AGE, COLOR, AND SEX, BASED ON ADJUSTED BIRTHS: 1960

(Base of percent is census count of resident population)

Color and age	Official factors for underregistration of births		Assuming one-third reduction in underregistration of births	
	Male	Female	Male	Female
WHITE				
Under 5	2.0	1.2	1.7	0.9
5-9	2.5	1.6	2.1	1.3
10-14	2.6	1.5	1.9	0.9
15-19	4.0	2.5	2.0	0.6
20-24	4.5	2.5	1.5	0.1
NONWHITE				
Under 5	8.4	6.8	6.9	5.2
5-9	6.0	5.1	3.9	3.1
10-14	5.5	4.4	1.7	0.8
15-19	14.3	11.2	6.3	3.8
20-24	21.2	10.7	10.6	1.8

Table 5.--COMPARISON OF "ENUMERATED" AND EXPECTED SEX RATIOS WITH SEX RATIOS OF ADJUSTED POPULATIONS, BY BROAD AGE GROUPS AND COLOR: 1960

(Males per 100 females in resident population)

Age and color	"Enumerated" sex ratios	Expected sex ratios	Sex ratios based on populations adjusted by--[1]				
			Net coverage error from reinterview studies	Preferred composite based on demographic analysis	Composite based on reinterview studies and demographic analysis		
					Set 1	Set 2	Set 3
WHITE							
All ages	97.4	98.6	97.3	98.5	98.6	98.6	98.7
Under 5	104.0	104.8	103.6	104.8	104.8	104.8	104.8
5-14	103.9	104.9	102.9	104.9	104.9	104.9	104.9
15-29	98.5	100.8	98.2	100.7	100.8	100.6	100.8
30-44	96.4	98.8	96.1	100.6	98.8	98.8	98.8
45-64	95.6	95.2	96.5	95.8	95.2	95.2	95.2
65 and over	82.3	83.8	83.1	77.8	83.8	83.8	83.8
NONWHITE							
All ages	94.7	97.6	95.4	97.6	97.6	(X)	97.7
Under 5	99.9	101.4	100.7	101.5	101.4	(X)	101.5
5-14	100.0	100.9	101.3	101.0	100.9	(X)	101.0
15-29	91.6	98.7	89.6	99.1	98.7	(X)	98.7
30-44	88.4	97.9	89.7	97.8	97.9	(X)	97.9
45-64	95.8	95.0	97.1	94.9	95.0	(X)	95.0
65 and over	90.1	80.5	94.4	80.5	80.5	(X)	80.5

X Not applicable.
[1] See footnotes of tables 2, 3, and 6 for an explanation of the basis of the adjusted figures.

Table 6.--ESTIMATED PERCENTS OF NET COVERAGE ERROR AND OF NET CENSUS UNDERCOUNTS BASED ON ANALYTIC AND COMPOSITE METHODS, BY SEX, COLOR, AND BROAD AGE GROUPS: 1960

(Base of percent is census count of resident population; minus sign denotes net census overcount)

Age and color	Net coverage error from reinterview studies[1]		Preferred composite based on demographic analysis[2]		Composite based on reinterview studies and demographic analysis[3]					
					Set 1[4]		Set 2[5]		Set 3[6]	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
WHITE										
All ages	1.6	1.7	2.8	1.6	3.0	1.7	2.9	1.6	2.4	1.1
Under 5	1.3	1.7	2.0	1.2	2.5	1.7	2.0	1.2	2.0	1.2
5-14	0.6	1.6	2.5	1.6	2.5	1.6	2.5	1.6	2.5	1.6
15-29	1.2	1.6	4.5	2.2	3.9	1.6	4.4	2.3	3.9	1.6
30-44	1.3	1.6	4.5	0.1	4.2	1.6	3.1	0.6	2.6	0.1
45-64	2.8	1.9	2.8	2.6	1.4	1.9	1.4	1.8	1.4	1.9
65 and over	2.3	1.3	-3.4	2.1	3.1	1.3	4.4	2.5	1.0	-0.8
NONWHITE										
All ages	4.2	3.4	12.2	8.8	6.7	3.4	(X)	(X)	8.0	4.7
Under 5	2.6	1.8	8.4	6.8	3.3	1.8	(X)	(X)	8.4	6.8
5-14	4.6	3.3	5.8	4.8	4.1	3.3	(X)	(X)	5.8	4.8
15-29	0.1	2.5	19.5	10.6	10.3	2.5	(X)	(X)	10.4	2.5
30-44	5.8	4.2	18.0	6.6	15.3	4.2	(X)	(X)	15.4	4.2
45-64	7.3	5.9	13.4	14.4	5.0	5.9	(X)	(X)	5.0	5.9
65 and over	6.7	1.8	1.9	14.0	-8.9	1.8	(X)	(X)	-5.9	5.3

X Not applicable.

[1] Based on reinterview studies EP-8 and EP-9.

[2] Whites and nonwhites under 25: Consistent with percents of net undercount in Series P-25, No. 310 (based on adjusted births, deaths, and net immigration data). Whites 25 and over: Consistent with percents of net undercount in Current Population Reports, Series P-25, No. 310 (Coale-Zelnik estimates for 1950 extended to 1960). Nonwhite females 25 and over: Coale estimates for 1950 extended to 1960. Nonwhite males 25 and over: Expected sex ratios applied to the adjusted female population.

[3] Lower bounds of true errors or "minimum reasonable" estimates.

[4] Females: Net coverage error from the reinterview studies. Males: Expected sex ratios applied to the adjusted female population.

[5] Males and females under 25: Based on adjusted births, deaths, and net immigration data. Females 25 and over: Adjusted for net census error in broad age groups as indicated by reinterview studies, consistent with an all-ages net coverage error of 1.6 percent and a uniform net coverage error of 1.6 percent for ages 25 and over by age. Males 25 and over: Expected sex ratios applied to adjusted female population. Estimates correspond to detailed estimates labeled "Composite estimates B," table 2.

[6] Males and females under 15: Based on adjusted births, deaths, and net immigration data. Females 15-64: Net coverage error from reinterview studies or preferred composite based on demographic analysis, whichever is lower. Males 15-64: Expected sex ratios applied to adjusted female population. Population 65 and over: Census counts by color, distributed by sex on the basis of expected sex ratios.

DISCUSSION

Joseph Steinberg, Social Security Administration

To users of census data, the two papers presented this morning are welcome additions to knowledge. We are a good deal closer today to some worthwhile goals as a result of this work--the goals of improved postcensal estimates of population, by age and sex--and of knowledge where errors in the count occur.

The Marks-Waksberg paper summarizes the results of a large volume of case-by-case evaluative work on coverage of the 1960 Census. It is, probably, as close as anyone can come to a final summary of substantial direct evaluative efforts. It is interesting to note that the final reinterview estimate of net undercoverage in the 1960 Census is 1.9 percent of the population; and that record checks give a range (with a "bolder attitude") of 2.5 to 3.1 percent. The Siegel-Zelnik paper takes account of recent evaluations of the assumptions in some analytic techniques and presents the results of those analytic methods now considered as preferable for the 1960 Census situation. The estimated net underenumeration by demographic analysis is 3.1 or 3.2 percent.

In working toward the goal of a "corrected" distribution of the population by age, sex, and color, the Siegel-Zelnik paper presents three experimental sets of estimates. They state that "the absolute level of the adjusted census or current estimates at each age (except possibly 65 and over) would be closer to the theoretical truth." But, they conclude that they have so far been unable to arrive at a set of figures to be recommended for general use.

Users of decennial census data have had available "some estimates of net undercounts for 1960 by age, sex, and color" in the United States Summary volume of the 1960 Census of Population. The analytic results in the Siegel-Zelnik paper are quite different for both white males and females 65 and over, (and by age group) for nonwhite females, and to some lesser degree, for nonwhite males. These substantial changes reflect the differences between the analytic methods used in the current presentation and the results in the Summary volume of the 1960 Census. Users have also had results in the Current Population Reports, Series P-25, No. 310, which also differ from the currently presented analytic results. As we examine the Marks-Waksberg paper and related matters, a few additional observations seem pertinent.

1. Many users might agree that not only is the most important statistic from the Census the population count, but so, too, in the postcensal period is this true--and by age and sex.

2. Useful information is available from the case-by-case approach on the relative contributions to coverage errors from field work and imputation procedures. About 10 percent of the undercount and 40 percent of the overcount arose in the processing operations. Net undercoverage due solely to field work is estimated as 1.9 percent.

3. The paper states that the reinterview results on the measures of age reporting are rather poor estimates of the bias in age reports. The user hopefully might look forward to some final summarization and greater elaboration suggesting the basis for this statement--and also similarly in regard to other items such as income, etc.

4. While data are not yet published, one might look toward the results of the recent registration for Medicare as an additional basis for evaluating the estimates of net undercoverage for the equivalent age groups in the 1960 Census. Likely, one will find some undercoverage. The counts of aged registrants would provide a lower bound to the estimates. The next few years will indicate the extent of underregistration for Medicare--as nonregistered aged persons apply for hospitalization benefits, the validity of their age allegations are examined and substantiated through appropriate means, and a range of estimates of underregistration are derived. At that point, a more reasonable estimate of undercoverage will be possible.

As we examine the Siegel-Zelnik paper, a few additional observations also seem pertinent.

1. One must agree that the techniques of demographic analysis often provide a strong basis for judging the demographic reasonableness of census results or other methods of evaluation. However, as the authors point out, there are a number of limitations to these techniques (albeit there are limitations--whether smaller or greater--in other evaluative techniques). One might, therefore, assume that evaluative analytic techniques (and various derived results) may continue to be presented at intervals based upon alternatives not as yet considered. One would hope that despite the important research considerations, that one of the existing results (modified as necessary) can be recommended for use.

2. It is interesting to observe, that the several levels of estimates of relative undercoverage, indicate that efforts to improve coverage in 1960 by various means do appear to have worked to some degree--percentage-wise up to 0.7 or 0.8 percent better than 1950.

3. The authors mention an anomaly between the Census of 1950 and 1960--an apparent large net overcount of persons 65 and over in 1960--an anomaly which has now appeared in three successive censuses. For 1960, could this more likely reflect the better counting of persons at 65 and over than errors in age reporting? For 1960, the unpublished evidence from the Medicare registration in 1965-1966 may perhaps support this hypothesis.

4. The finding of Coale and Zelnik that "females 15-29 appeared to be the most fully enumerated group," in 1950 is strikingly inconsistent with the 1960 preferred composite result based on demographic analysis. And, the 30-44 group among white females is now best! Did the method produce the result, or are we to assume that the same cohort was enumerated well in 1950 and 1960.

5. If one places credence in reinterview net coverage error results for nonwhite females under 5, 5-9, and 10-14, what would the rate of underregistration of births be that would be consistent? Would this offer some added possibility anent nonwhite males?

6. The present analytic technique involves the use of expected sex ratios. The authors state that their research indicates that these expected sex ratios are rather sensitive to the level of net undercounts at the higher ages. Possibly the authors may wish to suggest in later discussion the implications of this comment with respect to the levels of estimates that may be derived for age sex groups--and the reasonable bounds on their estimates of the degree of sensitivity in their present use of the technique at such higher ages.

7. One can certainly join in their concern about the need for measurement of differential coverage error. For overall evaluation, there is need for adequate estimates with respect to geographic variation. There would also seem to be substantial need for measurement and reporting by socio-economic classifications. Users of census counts (and census data classified by a variety of characteristics) often assume them to be accurate. As noted later, major policy perhaps may have somewhat inadequate factual underpinning if unadjusted census results (whether unadjusted for undercoverage or biases in, say, age reporting, are the basis.

8. This user would prefer that a single "best" estimate be recommended together with a technical note on the range of undertainty. While the range would not be a probability confidence interval, it could nevertheless furnish useful information. In this way, if a user chose to use the upper limit for cost analysis or for a target, it would of necessity be so identified. This would still leave the needs of the general audience untouched and uniformly served.

Evaluative work of the type reported in these papers is important to the producers and the users of data.

To the producers, the challenge is:

1. To find improved techniques for doing a better job next time.

2. To find improved techniques for doing a better evaluative job.

3. To take advantage of the synthesized knowledge for providing the user with "better" populatiòn estimates. Would that this third challenge could provide the user, at an earlier date, with uniformly available, authoritatively backed, results.

The Census staff continues to examine methods for improved census taking. Problems of census taking and survey operations in slum areas have been identified since, at least, the 1950 Census as of a much higher order than in many other situations. Would that the properly motivated local people presently pressing for national recognition of the need for improvement in the lot of the poor could be effectively challenged to understand the need for and to help work toward improved data on coverage and content in their local areas.

The thrust toward improved evaluative methods has been suggested by Marks-Waksberg as coming perhaps through greater use of matching of special lists. This effort is one which deserves support--with substantial resources made available (perhaps some as early as this fiscal year) so that as many as possible of the identified limitations in the use of this technique (i.e., size of sample, or failure to get current addresses) can be overcome.

Perhaps, consideration should also be given to a method currently in use in another federal agency--record checks after an evaluation reinterview.

The suggestion to provide "best estimates" of postcensal population by age and sex is made with full realization that this creates a major problem for the provider. But certainly one might agree that the absence of such authoritative estimates leads to chaotic multiple estimates or the use of "comparable to Census" estimates when, in fact, the evidence we have heard today is that significant underestimates (in program decisions) may be the result of use of unadjusted census data. The relevant current data for many classes of decisions involve in large part the accuracy of current estimates of population by age, sex, and color. The number of persons who are unemployed and employed, as estimated from the Current Population Survey, are now and have virtually always been tied to "comparable to Census" estimates of population by age, sex, and color. As we examine the estimates of undercoverage in the census, shouldn't we be concerned as to the impact on this use, especially when decisions based upon whether we have full employment or the likely current volume of unemployment, are involved? (Besides the CPS survey results are already being adjusted for undercoverage relative to "comparable to Census" estimates.) Similarly, estimates of the current size and change in the number of the poor and their characteristics, and the possible levels of cost of various alternatives under proposals for a negative income tax are likely to be changed. Other estimates such as the number of persons eligible for Government programs such as Medicare; the number of persons living in substandard housing or eligible for rent subsidies, etc., get changed. It would appear useful for statisticians to face up to the question, and to answer affirmatively, that there is need for publication of a single set of official authoritative figures on "best estimates" of the population by age and sex, not only for the cnesus period, but also on a postcensal basis--and that these be used in the preparation of all important series.

DISCUSSION

W. Parker Mauldin, The Population Council

I should like to commend the Bureau of the Census for the serious efforts it has made and continues to make to evaluate Census data. They have used varied approaches, have undertaken detailed analyses and the magnitude of the effort is very great indeed. They are to be commended for the high standards they set and for the pioneering work that they are undertaking.

At the same time I should like to commiserate with them because their difficulties are enormous. There is a tremendous movement of people each year and this makes it difficult for them to carry out an effective re-interview in order to determine how good the census itself was. For example, as is pointed out in the Marx-Waxberg paper, almost 10% of the people had moved between the census and the PES, although the PES (Post Enumeration Survey) was conducted within five months of the Census.

I was struck with the fact that the enumeration status of college students could not be determined for 7% of those in the sample; and the enumeration status could not be determined for 5.5% of Social Security recipients. Both of these are groups for whom current addresses should be excellent, at least relative to many other groups in our population. In both cases the Bureau of the Census had what appeared to be current addresses shortly before the census, but even so they were unable to determine the enumeration status of significant proportions of both groups. If these groups are so mobile, it seems likely that other groups are even more mobile.

There are a few points on which I would like clarification. In the use of record checks for coverage evaluation four population groups were included:

1. Persons enumerated in the 1950 Census.
2. Persons missed by the 1950 Census, but picked up by the Post Enumeration Survey.
3. Children born during the intercensal period.
4. Aliens who registered with the Immigration and Naturalization Service in January, 1960.

In the Marx-Waxberg paper, it is stated "their combined representation is believed to be 98% or more of the entire population." In 1950 Coale estimated that 3.6% of the population was missed; but the Post Enumeration Survey accounted for only 1.4% or 2.2% less than the Coale estimate. The minimum reasonable estimate prepared by the Bureau of the Census was 2.5% missed or 1.1% more than was accounted for by the PES. It seems to me that 2.5% or possibly more of the population was not included in the record check for coverage evaluation. I would prefer, therefore, that the emphasis be on 98% or less rather than 98% or more. This may seem a trivial point but I think that it reflects a philosophy that runs throughout the papers, namely, that those preparing estimates of a corrected count tend to use minimum figures wherever possible rather than maximum or even expected or medium estimates.

Using the record checks for coverage evaluation, one estimates an under-enumeration of 2.6 to 4.7%. But this represents undercoverage only. Net error is estimated by subtracting from the underenumeration the overenumeration, estimated to be 1.3%.

We have been told that "the technique used in 1960 defines both over- and underenumeration relative to a specified small area, say, an enumeration district that contains the usual residence of the person being checked." The estimated gross overenumeration of 2,325,000 persons or 1.3% of the total would appear to have been based on this small area concept. This concept is a perfectly good one if one is deriving the estimate of underenumeration in the same way; however, when the record check is used one is not using the small area concept but rather the entire United States is the basic area. That is to say, a list sample has been prepared and the Bureau of the Census searched every enumeration district in which they thought it might be likely to find the individual. Thus the individuals on the list were counted as being correctly enumerated if they were found in any one of several enumeration districts. Overenumeration, however, was estimated on the small area basis, which overstates overenumeration in the entire United States. I am not entirely sure of the procedures used, but in reading the papers it seems to me that this was the procedure and it is for this reason that I ask for clarification on this point.

One of the puzzling features of the composite estimate is that more males than females were missed in every age group up to 45 years of age. The really surprising thing about this is why male babies and young male children in the age group 0-4 and 5-9, for example, would be missed more than females. The re-interview data do not show that more males are missed. Thus the apparent undercount of males in the 0-4 group, for example, is the result of the estimate of the expected number of males. The estimating procedure includes an allowance for under-registration of births and possibly this leads to an inflation in the number of males. Basically the procedure is to take the number of females in the specified age group, apply an expected sex ratio to this number and thus estimate the number of males. It is quite possible that the expected sex ratio is incorrect. The magnitude of the sex ratios would appear to be all right, but when they lead to an inconsistent result of this type one wonders whether the basic assumption is correct or if the re-interview data should be trusted more for this age group than is the case.

I am also puzzled that the sex ratio for the 5-9 age group is larger -- trivially larger, but still larger -- than for the 0-4 age group among whites.

In reading the paper one finds such terms as the "conservative best estimate" and the "minimum reasonable estimate." For some purposes it is desirable to derive minimum figures and in any case one would be inflating the actual census count. Nonetheless, the philosophy behind this bothers me in that we need a best estimate rather than a "conservative best" or "minimum reasonable" -- why not simply a "reasonable estimate?" I know that for administrative purposes it would be useful to have a single set of figures but, at the same time, it seems to me that there are too many uncertainties in the estimating procedures to permit us to afford the luxury of a single set of corrected figures.

There are two minor points in the Segal-Zelnick paper that I should like to comment on. On page 6A the statement is made that "The assumption of the net undercount of females 15-29 in recent censuses, combined with the fact that births estimated from females 15-29 in one census are approximately equal to births estimated from females 15-29 in the preceding and following censuses, led to the assumption of the uniform net undercounts over time. Literally, this seems unlikely; they probably are talking about ratios being equal rather than the numbers and this should be indicated.

Segal and Zelnick also state that since the number of births in any period of time is considerably larger than the number of deaths and net migrants, errors in the completeness of birth registration are of greater consequence for estimates of net census underenumeration and net census undercounts than errors in the other components. This is true if the magnitude of the errors is the same in each of the components but is not necessarily true if errors are larger in one component than in another and, indeed, in this instance I think that errors certainly are larger in some components than in others.

IV

EDUCATIONAL DATA FROM LARGE SCALE TESTING PROGRAMS

Chairman, MORRIS B. ULLMAN, U. S. Office of Statistical Standards

THE MENTAL TEST QUALIFICATION OF AMERICAN YOUTHS FOR MILITARY SERVICE AND ITS RELATIONSHIP TO EDUCATIONAL ATTAINMENT*

Bernard D. Karpinos, Medical Statistics Agency,
Office of The Surgeon General, Department of the Army

I. Introduction

Should you happen to be in the office of a local board, within the Selective Service System, when a group of youths had just returned from an Armed Forces examining station, and should you, out of curiosity, ask any one from among those who were disqualified for military service what happened to him, his most likely answer would be: "I flunked the physical." This would have been overwhelmingly correct in World War I; it would have been also to a great extent the proper answer in World War II, but it is not so now. At present, it is highly probable that the youth was disqualified because he failed to meet the mental, rather than the medical, requirements, as his reply would imply; he could be a perfect physical specimen. On a nationwide basis, there is now a fifty-fifty chance of this being so. In giving that particular reason why he was "turned down" (outside the fact that he may be more embarrassed to admit that he "flunked" the mental tests), the youth resorted to a common expression ("flunked," or "failed the physical") in regard to disqualifications for military service--be these for medical, mental, or moral reasons. In fact, this common usage of the term had led to many misinterpretations of the disqualification data, especially, when such misinterpretations favored predetermined promotional purposes--quite unfortunate even if these purposes as such are socially desirable (1).

The number of youths disqualified for failing the mental tests is now as large as that of youths disqualified for medical reasons. During the recent 1964-1965 period, out of somewhat over 2 million draftees, namely, registrants forwarded within this period by the local boards to the Armed Forces Examining and Induction Stations (AFEIS) for preinduction examination (the number of such draftees had greatly increased in these years because of the intensified Vietnam crisis), somewhat less than half-a-million examinees were disqualified for medical reasons, and about the same number failed the mental tests. (These figures include 40,000 draftees who failed simultaneously both the mental and medical requirements, and consequently are added to each of these two groups. See "Numerical Summary," Table 1, bottom.) These were predominantly youths around 20 years of age (2,3). One could hardly doubt that this represents a great loss of manpower, from a military and a social viewpoint.

Grave a problem as it is on a national basis, its gravity is much accentuated by the wide geographic and ethnic differences revealed by these data, specifically, with respect to the disqualifications for mental test failures.

II. Geographic and Ethnic Differentials in The Mental Test Disqualifications of Draftees

Close to one-fourth of the examined draftees were thus disqualified during this 1964-1965 period because they could not meet the current mental requirements (more exactly, 23.4 percent: 21.4 percent of the examinees failed the mental tests, only, and 2.0 percent failed the mental tests and were simultaneously disqualified for medical reasons--the latter constituting an overlapping group; see Table 1.) (See (3), also subsequent section VI, for a discussion of the current mental requirements.) However, while of the white (actually, non-Negro) draftees 16.0 percent (including the overlapping group) failed the mental tests on a nationwide basis, of the Negro draftees 63.3 percent (including the overlapping group) were so disqualified, showing thus a disqualification rate for the Negro draftees about 4 times as high as for the white draftees.

Wider variations in the mental disqualifications, than those indicated on the nationwide basis, are revealed when these data are both geographically and ethnically differentiated--and the smaller the geographic area the more conspicuous these variations.

By geographic region--without ethnic differentiation--the disqualification rates for failing the mental tests ranged during this period from 14.4 percent in North Central to 36.5 percent in the South. Greater differences were found when differentiated by geographic division: the range in this case was from 11.2 percent in West North Central to 37.1 percent in South Atlantic (Table 1). Still greater differences were disclosed by State: their range was from 6.4 percent (Iowa) to 59.7 percent (Mississippi)(Table 2).

When further differentiated by ethnic group, the ranges in the disqualification rates for failing the mental tests were:

> White: a. Geographic region, from 10.9 percent in North Central to 22.4 percent in the South; b. Geographic division, from 9.4 percent in West North Central to 28.3 percent in East South Central; c. State, from 6.3 percent (Iowa) to 31.8 percent (Tennessee)(Table 2).

*Abstracted from a projected monograph by the author: "Mental Qualification of American Youths for Military Service."

Negro: a. Geographic region, from 42.6 percent in the West to 71.2 percent in the South; b. Geographic division, from 41.5 percent for the Pacific to 75.0 percent in East South Central; and c. State, from 27.7 percent (Washington) to 83.9 percent (Mississippi)(Table 2).

Obviously, the differences are both geographic and ethnic.

In order to visualize more clearly these geographic and ethnic variations, the States have been arrayed (ranked), from the lowest to the highest disqualification rates for mental reasons, and grouped in terms of 5-percent intervals, by race, as follows:

Percent Disqualified	A. Total
5-10	Iowa, Washington, Utah, Minnesota, Montana, Oregon, Idaho;
10-15	Nebraska, Wyoming, North Dakota, Wisconsin, Kansas, Rhode Island, South Dakota, New Hampshire, Vermont, Indiana, Alaska, Colorado;
15-20	Michigan, Ohio, Massachusetts, Pennsylvania, California, Oklahoma, Nevada, Connecticut, Illinois, Missouri
20-25	Arizona, New Jersey, Maine, New York Hawaii;
25-30	Delaware, Texas, Maryland, Florida, New Mexico;
30-35	West Virginia, Kentucky, Arkansas, Virginia;
35-40	Tennessee, District of Columbia;
40-45	Alabama, Louisiana, Georgia;
45-50	North Carolina;
50-55	South Carolina;
55-60	Mississippi.

Percent Disqualified	B. White
5-10	Iowa, Washington, Utah, Minnesota, Oregon, Montana, Nebraska, Kansas, Wisconsin, Idaho, Wyoming;

Percent Disqualified	B. White (continued)
10-15	North Dakota, Rhode Island, Nevada, Michigan, Indiana, Ohio, South Dakota, Illinois, New Hampshire, Pennsylvania, Vermont, Colorado, California, District of Columbia, Florida, Missouri, Oklahoma, Connecticut, Alaska, New Jersey, Massachusetts;
15-20	Delaware, Maryland, Texas, Louisiana, Arkansas, Arizona, South Carolina, New York;
20-25	Alabama, Georgia, Maine, Virginia, Hawaii;
25-30	Mississippi, New Mexico, North Carolina;
30-35	Kentucky, West Virginia, Tennessee.

Percent Disqualified	C. Negro
25-30	Washington;
40-45	Massachusetts, Oregon, California, Minnesota, Indiana, Utah, Michigan;
45-50	Ohio, Rhode Island, Arizona, Pennsylvania, District of Columbia;
50-55	Kansas, Nebraska, New York, Kentucky, Colorado, Oklahoma, West Virginia;
55-60	Missouri, Illinois, Connecticut, Delaware, Wisconsin, Maryland, New Mexico, New Jersey;
60-65	Nevada, Virginia, Texas;
65-70	Tennessee, Florida;
70-75	Arkansas, Alabama, Louisiana;
75-80	Georgia, North Carolina;
80-85	South Carolina, Mississippi.

There is no intention to underplay here the equal seriousness of the medical problem. Also, if much emphasis is put on the mental test failures, it is not because it is the topic of the present discussion--there is always a tendency to hyperbolize the problem dealt with at a particular time. It is done out of sincere conviction that more can be done for those who failed the mental tests, than for the medically disqualified youths, in two directions: a. In rehabilitating the youths who had been disqualified, or are about to be disqualified, for these reasons; and b. In the preventive sense, by sincere attempts to preclude, or to minimize as far as possible, such future results through proper education and other social programs. (It should be mentioned here that a certain rehabilitation program is carried out by the United States Public Health Service for youths disqualified for medical reasons; a parallel rehabilitation program has been initiated by the Department of Labor for those disqualified for failing the mental tests, and similar programs are planned by the Department of Defense.

It is not merely a military problem. To quote President Kennedy in this connection: "A young man who does not have what it takes to perform military service is not likely to have what it takes to make a living. Today's military rejects include tomorrow's hardcore unemployed" (5). Special investigations indicated that those who could not qualify for mental reasons were principally youths who lacked sufficient education--dropouts, and in general, youths of low quantitative and qualitative levels of education; youths who came from poor economic groups, and who were mainly "job seekers," having a relatively high rate of unemployment. These findings reflect obviously to a high degree the potential civilian productivity of these youths, as well as their future adjustment in civilian life (5,6).

III. Geographic and Ethnic Differentials in the Mental Qualification of Draftees Found Acceptable

Of no less importance than the wide geographic and ethnic variations in the disqualification of draftees for failing the mental tests are the corresponding variations in the mental qualification of those who met the minimum mental requirements. These geographic and ethnic variations are revealed in Table 3, by geographic region and division, and in Table 4, by State.

The mental qualification of draftees found acceptable (met the current mental, medical, and moral military requirements) are expressed in terms of four mental groups: mental groups I and II--the upper mental groups, and III and IV--the lower. It may suffice to state here that these mental groupings are derived from the scores made by the draftees on the mental tests, the upper groups indicating higher scores. (The meaning of these groupings is fully discussed in the subsequent section VI,- Mental Testing.)

For evaluative purposes, the percent of the draftees found acceptable falling within the mental groups I and II have been taken as an index. So appraised, the range in term of the percent of draftees found acceptable in mental groups I and II, were found to be:

Total: a. Geographic region, from 53.0 percent in the West to 31.4 percent in the South; b. Geographic division, from 53.5 percent in the Pacific to 24.8 percent in East South Central; c. State, from 59.9 percent (Oregon) to 20.5 percent (Mississippi).

White: a. Geographic region, from 54.5 percent in the West to 35.5 percent in the South; b. Geographic division, from 55.2 percent in the Pacific to 27.8 percent in East South Central; c. State, from 59.9 percent (Oregon) to 26.0 percent (Kentucky and Mississippi).

Negro: a. Geographic region, from 13.5 percent in the West to 4.6 percent in the South; b. Geographic division, from 13.9 percent in the Pacific to 3.5 percent in West South Central; c. State, from 19.2 percent (New Mexico) to 1.7 percent (South Carolina).

To facilitate visualization of these geographic and ethnic variations, the States have been arrayed (as done above for the disqualifications rates) by their percent in mental groups I and II, from the highest to the lowest percent, grouped in 5-percent intervals, as follows:

Percent Qualified in Mental Groups I and II	A. Total
60-55	Oregon, Alaska, Montana, Utah, Washington;
55-50	Iowa, Minnesota, Idaho, Colorado, California, Nebraska, Kansas, Nevada, Wyoming;
50-45	Wisconsin, Arizona, Connecticut, New Hampshire, Rhode Island, Ohio, South Dakota, Illinois, Hawaii;
45-40	Maine, Massachusetts, Michigan, Indiana, Vermont, Missouri, New Mexico, Pennsylvania, Oklahoma, New Jersey, New York, Delaware, Maryland;
40-35	Florida, Texas;

Percent Qualified in Mental Groups I and II	A. Total (continued)
35-30	Virginia;
30-25	Georgia, West Virginia, Tennessee, Louisiana, North Dakota, North Carolina, District of Columbia, Arkansas, Kentucky;
25-20	Alabama, South Carolina, Mississippi.

Percent Qualified in Mental Groups I and II	B. White
65-60	Oregon;
60-55	Alaska, Washington, Utah, Montana;
55-50	Colorado, Idaho, Nebraska, Kansas, Wyoming;
50-45	Connecticut, Wisconsin, District of Columbia, Illinois, Ohio, Arizona, Rhode Island, New Hampshire, Michigan, South Dakota, Delaware, Maryland, Indiana, Florida, Massachusetts, Hawaii;
45-40	Maine, Missouri, New Jersey, Pennsylvania, Oklahoma, New York, Vermont, New Mexico, Texas, Virginia;
35-30	Georgia, Louisiana;
30-25	North Dakota, North Carolina, Tennessee, Arkansas, Alabama, West Virginia, South Carolina, Mississippi, Kentucky.

Percent Qualified in Mental Groups I and II	C. Negro
20-15	New Mexico, Rhode Island, Nebraska, Minnesota;
15-10	New York, Colorado, California, Washington, Ohio, Michigan, Wisconsin, District of Columbia, West Virginia, Massachusetts, Indiana, Maryland;

Percent Qualified in Mental Groups I and II	C. Negro (continued)
10-5	Kansas, Missouri, Pennsylvania, Illinois, Oregon, New Jersey, Connecticut, Delaware, Oklahoma, Arizona, Virginia, Nevada, Kentuc Kentucky;
5 or less	Utah, Tennessee, Georgia, Texas, Mississippi, North Dakota, North Carolina, Louisiana, Arkansas, South Carolina.

IV. Underlying Causes

Naturally, the foremost question is: What are the underlying causes for such wide and unquestionably very much socially-disturbing geographic and ethnic variations in the mental qualification of the draftees (qualified and disqualified) for military service? Due to the very nature of the mental testing, the immediate response to such an inquiry would be, of course, a study of the relationship between the mental test scores of the examinees and their educational achievement. Indeed, past studies indicate definite positive association between the examinee's mental score and his educational attainment (7-10). However, those studies were carried out on a nationwide basis for the sole purpose of generally comparing Army mental test scores with education. There had been no attempt to evaluate such variations in the mental qualifications, as found here, in terms of educational achievement. It is toward such a differential evaluation that this study was principally directed. In its final analysis, the present study becomes to a convincing extent an assessment of the quality of our educational system, geographically and ethnically defined. Surely, it is fully realized that the youth's mental test score does not depend solely on the educational opportunities he is exposed to, but in addition--and significantly so--on the knowledge he gained from his formal educational training, which depends, of course on his cultural and economic environment outside the school--a problem much wider in scope than the educational opportunities as such (2,7).

To establish the relationship between the mental test scores and educational level, a series of tabulations was undertaken in which the mental test scores of the youths examined for military service were cross-tabulated by educational attainment. However, prior to discussing the extent of these tabulations, the manner in which they were accomplished and consolidated, and the form in which the distributions are presented in final form, it is most important, for proper interpretation of these data, that the meaning of both the educational attainment and the mental testing be clearly understood.

V. Educational Achievement

In relating the mental qualification of the examined youths for military service to their educational attainment, the latter presents a rather simple index. It is expressed in terms of the highest grade, or the highest year of schooling, completed. This information is supplied by the youth at the time he registers with his local board. It is copied from the registrant's classification questionnaire (SSS Form 100) onto the DD Form 47 ("Record of Induction") which the draftee brings with him when sent to an Armed Forces examining station. In the case of applicants for enlistment, this information is usually supplied directly to the recruiting station.

VI. Mental Testing

Armed Forces Qualification Test (AFQT).

General. The mental qualification of the examined youths for military service, determined at the Armed Forces Examining Stations (AFEIS), is by far a more complex index. From July 1950, when the Selective Service processing of youths for military service was reestablished under the provision of the 1950 Selective Service Act (no induction processing took place between January 1949 and July 1950), the Armed Forces Qualification Test (AFQT) has been the primary mental test used for determining the examinee's mental qualification for military service (1-3).

Development. Structurally, the AFQT, as originally developed, was a modified version of the Army General Classification Test (AGCT) used in World War II. The AFQT was equated with the AGCT. Stated schematically, this comparative relationship between the AFQT and the AGCT was established in the following manner: a. The AGCT was administered in 1949 to representative samples of groups of youths at various recruiting stations, b. on the basis of their scores on the AGCT, a further sample was selected from these representative groups in such a manner that its distribution by AGCT scores corresponded to that of the World War II "mobilization population," used in standardizing the AGCT scores; c. the matched sample was tested with AFQT, and d. its AGCT and AFQT distributions were then correlated and equated. Obviously, the current AFQT scores, as expressed in percentiles, relate to that World War II mobilization population, which presumably represented the expected distributions by AGCT of the civilian manpower pool available then for military service. Actually, the AGCT was standardized on the military personnel, officers and enlisted men, of all military services, as of the end of 1944. Since the exemptions and deferments from military service were then at a minimum, it was assumed that this military population may be taken as "unbiased representation of the civilian manpower pool," with respect to age, education, occupational status, and geographic distribution (11,12). The current mobilization population might differ in its distribution from that of World War II. Inasmuch, however, as all AFQT scores were standardized on the same basis, the present comparative analysis is not affected by it.

Contents. Like the AGCT, the AFQT consisted originally of the following time-honored content areas: vocabulary--ability to handle words and understand verbal concepts; arithmetic--ability to reason with numbers and solve simple mathematical problems; spatial relations--ability to distinguish forms and patterns. In 1953, mechanical ability, namely, ability to interrelate tools and equipment, was added as a content area. This analysis is based on this (1953) revised version of the test. It contains since then 100 questions equally distributed among those four mentioned content areas. It is arranged in cycles of increasing difficulty in each of the content areas. It is a self-administered spiral omnibus-type test; it emphasizes power, rather than speed (1-3, 13).

Objectives. Functionally, the AFQT differs from the AGCT in that the AFQT was delegated a dual objective: a. To measure the examinee's general mental ability to absorb military training within a reasonable length of time, so as to eliminate those who do not possess such ability--a qualification device; and b. to provide a uniform measure of the examinee's potential general usefulness in the service, if qualified on the tests--a classification device. It was hence specifically intended to predict potential success in general military training and performance ("military Trainability"). It has been validated for that purpose.

The AFQT thus combines a mental prescreening function, almost non-existent in World War II, and the function of the Army General Classification Test (AGCT) administered in World War II at the reception centers for assessing the mental ability of those who entered the Army for assignment purposes. With the introduction of the AFQT, the latter function was thus transferred from the Army reception centers to the examination stations. As originally conceived, the main purpose of this transfer was to provide a basis for an equitable qualitative distribution of manpower among the Armed Services--a purpose now defunct.

Administration and Scoring. The test is given to every youth when initially examined for military service, except to Spanish-speaking registrants in Puerto Rico. (This latter group is tested with an equivalent test in Spanish.) The test was developed through the joint efforts of all military services.

The test questions are of the usual multiple-choice type. The examinee's score ("raw score") during this period, as now, has been derived from the number of correct "net" answers, computed by subtracting from the number of the test questions answered by the examinee correctly one-third of the number of questions answered by him incorrectly--a procedure adopted to compensate for "successful guessing."

For comparative analysis, the examinee's raw score is converted to percentile score, and is so recorded on the individual's reports of examination (DD Form 47 and SF 88). This is done on the basis of specific conversion tables applicable to the particular version of the test. As stated before, these percentile scores were standardized on the "mobilization population" of World War II.

In order to provide a more workable classification of the examinee's degree of trainability, the percentile scores on the AFQT have been condensed into five mental groups. The mental groups, the current corresponding required "net" correct answers, the percentile score on the AFQT, and the percentages within each mental group of the standard population are as follows:

Mental Group	Required Net Correct Answers	Percentile Score	Percent within Standard Population of Each Group
I	89-100	93-100	7
II	74- 88	65- 92	28
III	53- 73	31- 64	34
IV	25- 52	10- 30	21
V	24 or less	9 or below	10

The mental groupings--from mental group V to I--obviously indicate progressive gradation of trainability. The individual's mental group is recorded on his examination reports, alongside his percentile score (as indicated above), providing thus a mutual check.

The Aptitude Area Tests

From the time it was initiated (1950) until August 1958, the AFQT was the only mental test (except for the previously-mentioned Spanish test in Puerto Rico) used for determining the examinee's mental qualification for military service. The mental qualification standards were established by the Universal Military Training and Service (UMTS) Act, requiring a minimum percentile score of 10 on the AFQT. Examinees scoring below the 10 percentile (mental group V) have been classed as mentally unfit for military service, unless found administratively acceptable ("administrative acceptees"), on the basis of "terminal screening" (3).

In August 1958, additional mental tests ("aptitude area" tests) and additional minimum mental requirements based on these tests were introduced. At first, the Army Classification Battery (ACB) was used, replaced late in 1961 by the Army Qualification Battery (AQB). These tests were designed to determine the person's potential usefulness in particular kinds of military jobs or assignments ("aptitude areas"); specifically, in the major occupational categories into which the jobs for enlisted men have been grouped. Only examinees in mental group IV on the AFQT have been subject to these tests.

Since August 1958, there have been, hence, two types of mental test failures: a. Examinees who failed the AFQT (below 10 percentile; mental group V); plus, b. those in mental group IV on the AFQT who failed the minimum requirements on the "aptitude area" tests (ACB or AQB). (The latter group is referred to as "Trainability Limited" (3).) The 1964-1965 data, relating to the mental qualifications of draftees for military service (Tables 1-4), are based on both the AFQT and the current AQB requirements. (For more detailed discussion, see 3-4.)

VII. AFQT Scores as Related to Educational Level

Extent of Study

The evaluation on the relationship between the mental test scores and educational achievement is based on the AFQT scores, alone. It comprises the experience of 5 years and 7 months--from January 1953 through July 1958 (prior to the introduction of the "aptitude area" tests.) During this period the same mental requirements obtained, namely, a minimum of 10 percentile score on the AFQT, only. Also, the contents of the AFQT and the manner of scoring were the same during the entire period.

Some three-quarter of a million medical examination forms of youths examined for military service during this period were cross-tabulated, representing a fifty-percent sample. These forms included the following groups of youths: a. disqualified draftees; b. inducted draftees; and c. disqualified applicants for enlistment. Each of these groups were cross-tabulated separately. In addition, there were available the distributions of enlistees during this period by mental group.

All data, from these various groups, were weighted and combined, in accord with available general military data. (The procedures applied in combining these data are explained in the projected monograph; see footnote to the legend of this paper.)

Cross-distributions were thus derived for all youths examined during this period for military service, by geographic division and race.

Overall Evaluation

Since it seemed desirable to obtain an evaluation of all youths in the population, not limited to youths examined for military service, the 19-21 age group were selected for this purpose as reported in the 1960 Census. The cross-distributions obtained from the youths examined for military service were therefore adjusted (or standardized) according to the educational achievement of these groups as derived from the Census data.

The cross-distributions of the mental test scores by educational attainment are presented in Table 5, by geographic division and race. The spread by educational attainment is as tabulated; that by AFQT score is in terms of mental groups and their corresponding percentile intervals. Note, that the population bases for all youths and for the white youths are 100,000, and for the Negro youths 10,000--smaller bases because of the smaller number of Negro youths involved. This should be clearly kept in mind in reading these tables. For instance, on the basis of their educational attainment, it is expected that our of 100,000 white youths, age 19-21, in Continental United States, 4,863 (4.9 percent) would be in the lowest mental group V; out of 10,000 Negro youths it is expected that 3,076 (30.8 percent) would be in mental group V (Table 5).

Similarly, according to these tables, for instance, 9,783 out of 100,000 white youths (9.8 percent) in the Continental United States would be in mental group I, whereas only 44 out of 10,000 Negro youths (0.4 percent) in the Continental United States would score as high.

The last rows in the presented distributions are thus the actual distributions of the 19-21 age groups by educational attainment according to the Census. Each column represents a distribution of the particular educational level by percentile score--a qualitative evaluation. The last column depends, of course, on both the level of education and its quality.

Median Years of School Completed versus Expected Median Percentile Scores on the AFQT

It seems quite needless to elaborate what can be readily seen from the tables (Table 5) with respect to the geographic and the ethnic variations in the expected AFQT scores. However, in order to form a clearer judgment as to the relationship between educational attainment and the AFQT scores, Table 6 was prepared in which the medians of educational attainment are contrasted with those expected on the AFQT. In comparing these medians on an ethnic basis only, there seems to be no very significant difference in those relating to educational attainment: a median of 12.4 years of school completed for the white youths, compared with a median of 11.0 years of school completed for Negro youths. But the expected medians on the AFQT are 55 percentile for the white youths, versus 15 percentile for the Negro youths.

More striking are the differences when these medians are geographically and ethnically differentiated: for the white youths the range in the medians of educational attainment is from 12.5 years of school completed (West North Central) to 12.2 years (East South Central)--a negligible range, indeed; but the range in the corresponding AFQT medians is from 63 to 47 percentile. For the Negro youths, the range in the medians of educational attainment is from 12.3 years of school completed (Pacific) to 10.2 years (East South Central), whereas the corresponding range in the expected AFQT medians is from 25 percentile--very low, to an even lower median of 11 percentile.

Reemphasizing, the determining factors here are not only the level of formal education and its quality, but the interrelated socio-economic factors outside the school.

VIII. References

(1) Karpinos, Bernard D.: Qualification of American Youths for Military Service, Medical Statistics Division, Office of The Surgeon General, Department of the Army, Washington, D. C., 1962.

(2) ________________: "Results of the examination of youths for military service, 1964," Supplement to Health of the Army, May, 1965. Medical Statistics Agency, Office of The Surgeon General, Washington, D. C. 20315.

(3) ________________: "Results of the examination of youths for military service, 1965," Supplement to Health of the Army, July, 1966. Medical Statistics Agency, Office of The Surgeon General, Washington, D. C. 20315.

(4) ________________: "Results of the examination of youths for military service, 1965 (Complementary Analysis)," Supplement to Health of the Army, September, 1966. Medical Statistics Agency, Office of The Surgeon General, Washington, D. C. 20315.

(5) U. S. Department of Labor: One-Third of a Nation: A Report of Young Men Found Unqualified for Military Service, January 1, 1964.

(6) ________________: Young Workers: Their Special Training Needs, Manpower Research, Bulletin No. 3, May, 1963.

(7) Department of the Army: Marginal Man and Military Service; A Review. December, 1965.

(8) Bingham, Walter V.: Personnel Classification Testing in the Army, Science, 100:275-280, September, 1944.

(9) Bingham, Walter V.: Inequalities in Adult Capacity from Military Data, Science, 104:147-152, August 16, 1946.

(10) Fuchs, Edmund F.: Army Test Scores vs. Education Since World War I (Abstract), American Psychologist, 12:466, 1957.

(11) Department of the Army: Development of Armed Forces Qualification Test and Predecessor Army Screening Tests, 1946-1950, PRS Report 976, November 1952.

(12) Department of the Army: Follow-Up of the Standardization of the Armed Forces Qualification Test, PRS 959, June 1952.

(13) ________________: Development of the Armed Forces Qualification Test, Forms 3 and 4, PRB 1078, October, 1953.

TABLE 1. RESULTS OF PREINDUCTION EXAMINATION OF DRAFTEES FOR MILITARY SERVICE, BY GEOGRAPHIC REGION AND DIVISION, AND BY RACE

1964 - 1965

Geographic Region and Division[1]	Percent disqualified, by cause				
	Total	Adminis-trative reasons	Failed Mental Tests	Medically Disqualified	Failed Mental Tests and Medically Disqualified
TOTAL					
UNITED STATES	47.0	1.4	21.4	22.2	2.0
NORTHEAST	46.4	1.4	17.8	25.2	2.0
New England	48.4	2.0	13.8	30.4	2.2
Middle Atlantic	45.9	1.3	18.9	23.8	1.9
NORTH CENTRAL	38.2	1.1	13.1	22.7	1.3
East North Central	39.4	1.1	14.3	22.7	1.3
West North Central	35.1	1.3	10.1	22.6	1.1
SOUTH	55.2	0.9	33.6	17.8	2.9
South Atlantic	55.6	1.1	33.9	17.4	3.2
East South Central	59.2	0.4	38.7	17.4	2.7
West South Central	50.4	1.1	28.0	19.1	2.2
WEST	44.2	2.6	13.7	26.7	1.2
Mountain	42.7	1.3	14.3	25.7	1.4
Pacific	44.6	2.9	13.5	27.1	1.1
WHITE[2]					
UNITED STATES	42.1	1.4	14.5	24.7	1.5
NORTHEAST	43.9	1.3	14.4	26.5	1.7
New England	47.8	2.0	12.8	30.9	2.1
Middle Atlantic	42.8	1.1	15.0	25.1	1.6
NORTH CENTRAL	35.8	1.1	9.8	23.8	1.1
East North Central	36.6	1.0	10.5	24.0	1.1
West North Central	33.8	1.3	8.4	23.1	1.0
SOUTH	45.5	1.0	20.3	22.1	2.1
South Atlantic	44.9	1.2	19.4	22.0	2.3
East South Central	50.4	0.4	26.0	21.7	2.3
West South Central	41.9	1.1	16.6	22.7	1.5
WEST	43.1	2.6	12.2	27.2	1.1
Mountain	42.1	1.2	13.5	26.1	1.3
Pacific	43.3	2.9	11.8	27.6	1.0

TABLE 1. (Continued)

Geographic Region and Division[1]	Percent disqualified, by cause				
	Total	Adminis-trative reasons	Failed Mental Tests	Medically Disqualified	Failed Mental Tests and Medically Disqualified
NEGRO					
UNITED STATES	74.4	1.2	59.3	9.6	4.3
NORTHEAST	69.3	2.1	48.6	14.1	4.5
New England	67.0	1.4	46.2	14.2	5.2
Middle Atlantic	69.5	2.2	48.8	14.1	4.4
NORTH CENTRAL	63.4	1.3	46.7	12.3	3.1
East North Central	63.0	1.3	46.1	12.5	3.1
West North Central	66.3	1.5	51.0	10.4	3.4
SOUTH	79.0	0.8	66.5	7.0	4.7
South Atlantic	78.5	1.0	65.3	7.0	5.2
East South Central	81.8	0.3	71.3	6.5	3.7
West South Central	77.3	0.8	64.3	7.6	4.6
WEST	63.9	3.0	40.2	18.3	2.4
Mountain	67.9	1.8	50.0	12.1	4.0
Pacific	63.5	3.1	39.2	18.9	2.3

[1]States within each geographic division: New England--Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut; Middle Atlantic--New York, New Jersey, Pennsylvania; East North Central--Ohio, Indiana, Illinois, Michigan, Wisconsin; West North Central--Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas; South Atlantic--Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida; East South Central—Kentucky, Tennessee, Alabama, Mississippi; West South Central--Arkansas, Louisiana, Oklahoma, Texas; Mountain--Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada; Pacific--Alaska, California, Hawaii, Oregon, Washington (according to the Bureau of The Census).

[2]White: Non-Negro

NUMERICAL SUMMARY, BY RACE			
Results of Examinations	Total	White	Negro
Number Examined	2,025,435	1,714,285	311,350
Number Disqualified, by Cause	952,128	720,784	231,344
Administeative reasons	27,659	23,900	3,759
Failed mental test, only	432,708	248,394	184,314
Medically disqualified, only	451,780	421,994	29,786
Failed mental tests and medically disqualified	39,981	26,496	13,485

TABLE 2. DISQUALIFICATION RATES (PERCENT) OF DRAFTEES FOR MENTAL REASONS, BY STATE AND RACE

1964 - 1965[1]

State	Total	White (Non-Negro)	Negro	State	Total	White (Non-Negro)	Negro
UNITED STATES	23.4	16.0	63.6	Missouri	19.1	14.0	55.6
				Montana	7.7	7.7	*
Alabama	40.7	20.2	73.0	Nebraska	10.0	8.8	51.3
Alaska	14.2	14.2	*	Nevada	17.7	11.2	61.9
Arizona	20.1	19.1	47.4	New Hampshire	12.5	12.5	*
Arkansas	33.0	19.0	71.2				
California	16.0	13.9	41.9	New Jersey	20.7	14.3	59.6
				New Mexico	28.8	28.5	58.9
Colorado	14.4	13.4	53.6	New York	23.9	19.9	52.6
Connecticut	17.9	14.1	57.5	North Carolina	45.4	29.7	79.1
Delaware	25.1	16.1	58.1	North Dakota	10.8	10.8	*
District of Columbia	38.3	13.8	48.4				
Florida	28.6	13.9	67.6	Ohio	15.4	12.0	45.4
				Oklahoma	17.3	14.0	53.6
Georgia	42.0	20.7	76.4	Oregon	8.0	7.6	40.7
Hawaii	24.6	24.5	*	Pennsylvania	15.7	12.5	48.2
Idaho	9.6	9.6	*	Rhode Island	11.6	11.0	46.3
Illinois	18.7	12.5	55.9				
Indiana	13.5	11.4	42.1	South Carolina	51.5	19.8	80.6
				South Dakota	11.9	12.1	*
Iowa	6.4	6.3	*	Tennessee	37.8	31.8	66.3
Kansas	11.2	9.4	50.6	Texas	26.8	18.6	64.0
Kentucky	31.7	30.2	52.9	Utah	7.4	7.3	43.4
Louisiana	41.4	18.7	74.0				
Maine	21.2	21.2	*	Vermont	13.2	13.1	*
				Virginia	33.2	21.5	63.8
Maryland	27.5	17.9	58.6	Washington	7.2	6.9	27.7
Massachusetts	15.5	14.9	40.0	West Virginia	30.8	30.2	54.5
Michigan	15.3	11.3	43.5	Wisconsin	10.9	9.5	58.3
Minnesota	7.5	7.4	41.9				
Mississippi	59.7	25.6	83.9	Wyoming	10.1	9.8	*

[1]Includes a. Failed mental test, only, and b. Failed mental test and medically disqualified; see Table 1.
*Number examined too small for reliable rates.

TABLE 3. PERCENT DISTRIBUTION OF DRAFTEES FOUND ACCEPTABLE FOR MILITARY SERVICE, BY MENTAL GROUP, GEOGRAPHIC REGION AND DIVISION, AND BY RACE[1]

1964 - 1965

Geographic Region and Division	Percent found acceptable, by mental group[2]				
	I	II	III	IV	Administrative Acceptees
TOTAL					
UNITED STATES	7.6	34.7	43.1	13.4	1.2
NORTHEAST	7.4	35.1	45.2	11.3	1.0
New England	8.9	37.1	43.0	9.6	1.4
Middle Atlantic	7.0	34.6	45.7	11.8	0.9
NORTH CENTRAL	8.6	38.7	41.3	10.9	0.5
East North Central	8.1	37.9	41.8	11.7	0.5
West North Central	9.9	40.4	40.2	9.0	0.5
SOUTH	4.8	26.6	45.9	20.0	2.7
South Atlantic	5.0	27.5	46.4	18.8	2.3
East South Central	3.3	21.5	48.7	22.9	3.6
West South Central	5.6	29.0	42.7	19.9	2.8
WEST	11.5	41.5	37.0	9.6	0.4
Mountain	11.0	40.5	37.5	10.0	1.0
Pacific	11.6	41.9	36.8	9.5	0.2
WHITE					
UNITED STATES	8.2	36.9	42.9	11.2	0.8
NORTHEAST	7.8	36.6	45.0	9.9	0.7
New England	9.1	37.7	43.1	8.9	1.2
Middle Atlantic	7.5	36.3	45.5	10.2	0.5
NORTH CENTRAL	9.1	40.2	41.0	9.3	0.4
East North Central	8.6	39.9	41.5	9.6	0.4
West North Central	10.1	41.0	39.7	8.7	0.5
SOUTH	5.5	30.0	46.8	16.1	1.6
South Atlantic	5.8	31.6	46.5	14.8	1.3
East South Central	3.7	24.1	50.8	19.4	2.0
West South Central	6.2	32.2	43.9	15.8	1.9
WEST	11.9	42.6	36.6	8.6	0.3
Mountain	11.2	41.0	37.4	9.5	0.9
Pacific	12.1	43.1	36.3	8.3	0.2

TABLE 3. (Continued)[1]

Geographic Region and Division	Percent found acceptable, by mental group[2] I	II	III	IV	Administrative Acceptees
	NEGRO				
UNITED STATES	0.4	7.5	43.9	41.3	6.9
NORTHEAST	0.7	10.6	48.0	34.8	5.9
New England	0.8	8.5	40.5	42.6	7.6
Middle Atlantic	0.7	10.8	48.6	34.2	5.7
NORTH CENTRAL	0.5	10.1	47.4	39.7	2.3
East North Central	0.5	10.1	45.6	41.7	2.1
West North Central	0.5	10.1	62.3	24.0	3.1
SOUTH	0.2	4.4	40.6	45.0	9.8
South Atlantic	0.4	5.2	45.6	40.9	7.9
East South Central	0.1	3.5	34.1	47.3	15.0
West South Central	0.1	3.4	33.6	53.1	9.8
WEST	0.7	12.8	47.6	37.1	1.8
Mountain	0.2	8.8	38.6	45.6	6.8
Pacific	0.8	13.1	48.4	36.3	1.4

[1]See footnotes 1 and 2, Table 1.

[2]See text "Mental Testing," and also (2) on the meaning of mental groups. "Administrative acceptees" refers to examinees who failed the mental tests but were declared administratively acceptable on the basis of personal interviews.

TABLE 4. PERCENT OF DRAFTEES FOUND ACCEPTABLE FOR MILITARY SERVICE IN MENTAL GROUPS I AND II, BY STATE AND RACE

1964 - 1965

State	Total	White (Non-Negro)	Negro	State	Total	White (Non-Negro)	Negro
UNITED STATES	42.3	45.1	7.9	Missouri	42.3	44.6	9.7
				Montana	57.2	57.2	*
Alabama	23.6	28.8	2.0	Nebraska	52.3	53.0	17.2
Alaska	58.5	58.9	*	Nevada	50.8	53.9	5.6
Arizona	48.0	49.2	5.7	New Hampshire	47.6	47.6	*
Arkansas	25.9	29.2	2.0				
California	52.6	54.6	14.0	New Jersey	41.9	44.5	7.9
				New Mexico	42.1	42.2	19.2
Colorado	53.1	53.7	14.3	New York	41.3	43.3	14.3
Connecticut	47.8	49.7	7.9	North Carolina	26.6	29.7	3.6
Delaware	41.3	46.1	7.7	North Dakota	46.4	46.4	*
District of Columbia	26.4	49.4	11.5				
Florida	39.8	45.7	2.3	Ohio	46.8	49.2	11.9
				Oklahoma	41.9	43.8	7.4
Georgia	29.4	34.4	4.3	Oregon	59.9	60.3	7.9
Hawaii	45.1	45.1	*	Pennsylvania	41.9	44.0	9.1
Idaho	53.1	53.1	*	Rhode Island	47.5	47.9	17.2
Illinois	46.1	49.3	8.9				
Indiana	44.3	45.8	10.7	South Carolina	21.7	27.4	1.7
				South Dakota	46.4	46.4	*
Iowa	54.2	54.2	*	Tennessee	27.2	29.4	4.6
Kansas	51.9	52.9	9.8	Texas	37.6	40.9	4.0
Kentucky	25.0	26.0	5.5	Utah	57.0	57.2	4.8
Louisiana	27.1	32.7	2.9				
Maine	45.0	45.0	*	Vermont	42.8	42.8	*
				Virginia	35.0	40.3	5.6
Maryland	40.8	45.8	10.3	Washington	56.7	57.2	12.7
Massachusetts	44.7	45.2	10.8	West Virginia	28.1	28.4	11.4
Michigan	44.5	47.5	11.7	Wisconsin	49.2	49.6	11.6
Minnesota	53.9	54.0	16.2				
Mississippi	20.5	26.0	3.7	Wyoming	50.6	50.7	*

*Number examined too small for reliable distribution.

TABLE 5. EXPECTED DISTRIBUTION OF MALE YOUTHS, AGE: 19-21, BY QUALIFICATION (MENTAL GROUP) ON THE ARMED FORCES QUALIFICATION TEST (AFQT), ACCORDING TO THEIR EDUCATIONAL ATTAINMENT, BY GEOGRAPHIC DIVISION AND RACE, 1960*

Mental Qualification		Educational Attainment: Years of School Completed									
Mental Group	Percentile Score	None	Elementary School			High School		College			Total
			1-4	5-6	7-8	1-2	3-4	1-2	3-4	5 or more	
CONTINENTAL UNITED STATES:					Total				BASE: 100,000		
I	100-93	-	2	1	28	202	3,204	3,928	1,301	20	8,686
II	92-65	5	6	14	378	1,698	13,876	8,317	1,778	22	26,094
III	64-31	23	47	147	2,170	5,934	21,551	6,022	925	10	36,829
IV	30-10	58	305	920	4,465	5,795	7,972	890	93	1	20,499
V	9 or below	528	1,382	1,486	2,303	1,567	610	14	2	-	7,892
	Total	614	1,742	2,568	9,344	15,196	47,213	19,171	4,099	53	100,000
					White				BASE: 100,000		
I	100-93	-	2	1	31	228	3,608	4,420	1,469	24	9,783
II	92-65	6	6	18	423	1,892	15,449	9,252	1,976	24	29,046
III	64-31	25	44	155	2,355	6,317	22,445	6,270	941	9	38,561
IV	30-10	56	292	883	4,331	5,029	6,363	721	72	-	17,747
V	9 or below	483	961	937	1,428	757	280	14	3	-	4,863
	Total	570	1,305	1,994	8,568	14,223	48,145	20,677	4,461	57	100,000
					Negro				BASE: 10,000		
I	100-93	-	-	-	-	1	16	22	5	-	44
II	92-65	-	-	-	3	23	203	128	28	1	386
III	64-31	-	5	9	78	304	1,481	416	79	1	2,373
IV	30-10	7	41	119	547	1,157	2,009	216	25	-	4,121
V	9 or below	88	458	561	890	768	310	1	-	-	3,076
	Total	95	504	689	1,518	2,253	4,019	783	137	2	10,000
NEW ENGLAND:					Total				BASE: 100,000		
I	100-93	-	4	3	45	246	3,065	4,483	1,650	15	9,511
II	92-65	-	7	21	477	2,029	14,136	9,022	2,698	20	28,410
III	64-31	48	47	133	2,674	6,345	20,758	7,407	1,555	12	38,979
IV	30-10	90	244	912	4,512	5,110	6,623	927	148	-	18,566
V	9 or below	388	521	696	1,708	918	296	5	2	-	4,534
	Total	526	823	1,765	9,416	14,648	44,878	21,844	6,053	47	100,000
					White				BASE: 100,000		
I	100-93	-	4	3	47	254	3,150	4,565	1,700	15	9,738
II	92-65	-	8	22	491	2,082	14,420	9,284	2,760	21	29,088
III	64-31	50	43	125	2,706	6,435	20,872	7,511	1,563	12	39,317
IV	30-10	93	243	885	4,501	4,857	6,225	919	126	-	17,849
V	9 or below	396	500	639	1,505	723	238	5	2	-	4,008
	Total	539	798	1,674	9,250	14,351	44,905	22,284	6,151	48	100,000
					Negro				BASE: 10,000		
I	100-93	-	-	-	-	-	27	182	-	-	209
II	92-65	-	-	-	-	34	477	45	66	-	622
III	64-31	-	18	40	159	336	1,705	396	132	-	2,786
IV	30-10	-	23	177	490	1,335	1,972	121	86	-	4,204
V	9 or below	9	122	255	838	734	221	-	-	-	2,179
	Total	9	163	472	1,487	2,439	4,402	744	284	-	10,000

TABLE 5. (Continued)

Mental Qualification: Mental Group	Mental Qualification: Percentile Score	Educational Attainment: Years of School Completed: None	Elementary School: 1-4	Elementary School: 5-6	Elementary School: 7-8	High School: 1-2	High School: 3-4	College: 1-2	College: 3-4	College: 5 or more	Total
MIDDLE ATLANTIC:						Total			BASE: 100,000		
I	100-93	-	4	3	30	154	2,528	3,720	1,716	33	8,188
II	92-65	27	9	17	268	1,525	12,751	8,889	2,644	39	26,169
III	64-31	38	66	118	1,911	5,846	21,453	6,842	1,466	16	37,756
IV	30-10	98	309	685	4,269	6,272	8,176	929	121	2	20,861
V	9 or below	459	904	1,217	2,304	1,530	577	31	4	-	7,026
	Total	622	1,292	2,040	8,782	15,327	45,485	20,411	5,951	90	100,000
						White			BASE: 100,000		
I	100-93	-	3	4	33	166	2,769	4,083	1,888	36	8,982
II	92-65	30	8	19	291	1,652	13,728	9,668	2,892	43	28,331
III	64-31	42	64	124	2,026	5,980	21,664	7,254	1,548	17	38,719
IV	30-10	88	305	643	4,044	5,346	6,898	889	115	1	18,329
V	9 or below	429	832	1,067	1,816	1,026	432	33	4	-	5,639
	Total	589	1,212	1,857	8,210	14,170	45,491	21,927	6,447	97	100,000
						Negro			BASE: 10,000		
I	100-93	-	1	-	-	4	25	25	7	-	62
II	92-65	-	1	-	4	33	347	149	30	-	564
III	64-31	-	7	6	83	457	1,946	294	70	2	2,865
IV	30-10	19	36	108	642	1,506	2,030	131	17	-	4,489
V	9 or below	75	160	263	693	631	196	2	-	-	2,020
	Total	94	205	377	1,422	2,631	4,544	601	124	2	10,000
EAST NORTH CENTRAL:						Total			BASE: 100,000		
I	100-93	-	3	1	31	235	3,999	4,609	1,661	18	10,557
II	92-65	-	7	8	445	1,910	16,250	9,300	1,932	16	29,868
III	64-31	28	48	78	2,138	6,253	21,903	5,494	758	6	36,706
IV	30-10	48	257	512	4,080	5,573	6,948	675	66	1	18,160
V	9 or below	437	695	625	1,495	1,063	382	10	2	-	4,709
	Total	513	1,010	1,224	8,189	15,034	49,482	20,088	4,419	41	100,000
						White			BASE: 100,000		
I	100-93	-	3	1	33	252	4,315	4,975	1,794	19	11,392
II	92-65	-	8	8	468	2,039	17,355	9,922	2,060	16	31,876
III	64-31	30	41	79	2,230	6,380	22,054	5,584	757	6	37,161
IV	30-10	43	240	495	3,962	4,819	5,742	570	53	-	15,924
V	9 or below	425	604	550	1,204	637	213	12	2	-	3,647
	Total	498	896	1,133	7,897	14,127	49,679	21,063	4,666	41	100,000
						Negro			BASE: 10,000		
I	100-93	-	-	-	1	3	20	21	9	-	54
II	92-65	-	-	-	15	37	302	183	39	1	577
III	64-31	-	9	5	103	474	2,011	444	77	2	3,125
IV	30-10	11	46	73	554	1,458	2,138	193	21	1	4,495
V	9 or below	58	183	154	496	617	240	-	1	-	1,749
	Total	69	238	232	1,169	2,589	4,711	841	147	4	10,000

TABLE 5. (Continued)

Mental Qualification: Mental Group	Mental Qualification: Percentile Score	Educational Attainment: Years of School Completed: None	Elementary School 1-4	Elementary School 5-6	Elementary School 7-8	High School 1-2	High School 3-4	College 1-2	College 3-4	College 5 or more	Total
WEST NORTH CENTRAL: Total										BASE: 100,000	
I	100-93	-	-	-	62	231	5,167	5,869	1,975	20	13,324
II	92-65	-	15	12	705	1,744	18,458	10,729	1,986	19	33,668
III	64-31	-	28	78	2,694	4,739	21,647	5,862	744	7	35,799
IV	30-10	23	200	392	3,471	3,701	5,837	655	52	-	14,331
V	9 or below	470	514	369	845	469	199	11	1	-	2,878
	Total	493	757	851	7,777	10,884	51,308	23,126	4,758	46	100,000
White										BASE: 100,000	
I	100-93	-	-	-	64	242	5,404	6,115	2,065	20	13,910
II	92-65	-	16	12	733	1,809	19,160	11,135	2,049	20	34,934
III	64-31	-	31	77	2,782	4,818	21,746	5,860	711	6	36,031
IV	30-10	24	190	380	3,391	3,183	5,064	632	40	-	12,904
V	9 or below	452	446	294	602	299	115	12	1	-	2,221
	Total	476	683	763	7,572	10,351	51,489	23,754	4,866	46	100,000
Negro										BASE: 10,000	
I	100-93	-	-	-	-	-	11	55	4	-	70
II	92-65	-	-	-	10	35	338	199	65	1	648
III	64-31	--	-	10	81	302	1,949	590	146	3	3,081
IV	30-10	-	40	66	522	1,483	2,241	117	30	-	4,499
V	9 or below	87	194	200	607	415	199	-	-	-	1,702
	Total	87	234	276	1,220	2,235	4,738	961	245	4	10,000
SOUTH ATLANTIC: Total										BASE: 100,000	
I	100-93	-	-	-	16	144	2,058	2,136	637	13	5,004
II	92-65	-	3	21	322	1,423	10,807	5,442	1,156	23	19,197
III	64-31	4	37	250	2,424	6,445	22,294	5,147	809	15	37,425
IV	30-10	17	266	1,595	6,117	7,113	9,097	852	95	1	25,153
V	9 or below	618	2,529	2,932	3,886	2,426	823	7	-	-	13,221
	Total	639	2,835	4,798	12,765	17,551	45,079	13,584	2,697	52	100,000
White										BASE: 100,000	
I	100-93	-	-	-	17	184	2,623	2,726	810	17	6,377
II	92-65	-	2	24	405	1,776	13,536	6,807	1,423	29	24,002
III	64-31	5	32	290	2,879	7,448	25,005	5,590	825	13	42,087
IV	30-10	18	260	1,596	6,101	6,130	6,671	622	53	1	21,452
V	9 or below	466	1,274	1,421	1,887	840	191	3	-	-	6,082
	Total	489	1,568	3,331	11,289	16,378	48,026	15,748	3,111	60	100,000
Negro										BASE: 10,000	
I	100-93	-	-	-	1	-	7	5	3	-	16
II	92-65	-	-	1	3	18	120	64	21	-	227
III	64-31	-	6	10	82	292	1,276	360	75	2	2,103
IV	30-10	1	28	159	618	1,057	1,764	167	25	-	3,819
V	9 or below	116	695	826	1,092	801	304	1	-	-	3,835
	Total	117	729	996	1,796	2,168	3,471	597	124	2	10,000

TABLE 5. (Continued)

Mental Qualification: Mental Group	Mental Qualification: Percentile Score	None	Elementary School 1-4	Elementary School 5-6	Elementary School 7-8	High School 1-2	High School 3-4	College 1-2	College 3-4	College 5 or more	Total
EAST SOUTH CENTRAL:						Total			BASE: 100,000		
I	100-93	-	-	1	14	127	1,496	1,746	569	12	3,965
II	92-65	15	7	27	328	1,150	8,746	5,604	1,212	13	17,102
III	64-31	24	48	258	2,771	5,471	20,671	5,852	831	5	35,931
IV	30-10	34	424	1,732	6,829	6,690	9,461	1,028	93	-	26,291
V	9 or below	707	3,316	3,730	5,041	2,964	950	3	-	-	16,711
	Total	780	3,795	5,748	14,983	16,402	41,324	14,233	2,705	30	100,000
						White			BASE: 100,000		
I	100-93	-	-	1	18	161	1,883	2,197	719	13	4,992
II	92-65	20	9	33	410	1,436	10,873	6,970	1,492	15	21,258
III	64-31	31	43	299	3,356	6,359	23,289	6,433	879	9	40,698
IV	30-10	41	440	1,913	7,425	5,920	7,182	822	67	1	23,811
V	9 or below	792	2,248	2,195	2,810	953	240	3	-	-	9,241
	Total	884	2,740	4,441	14,019	14,829	43,467	16,425	3,157	38	100,000
						Negro			BASE: 10,000		
I	100-93	-	-	-	-	-	6	7	2	-	15
II	92-65	-	-	-	2	8	86	52	18	-	166
III	64-31	-	6	10	60	218	1,096	372	64	-	1,826
IV	30-10	1	36	106	463	954	1,790	179	19	-	3,548
V	9 or below	38	729	944	1,331	1,044	359	-	-	-	4,445
	Total	39	771	1,060	1,856	2,224	3,337	610	103	-	10,000
WEST SOUTH CENTRAL:						Total			BASE: 100,000		
I	100-93	-	-	2	19	183	2,410	3,329	976	17	6,936
II	92-65	-	6	19	307	1,389	11,986	7,977	1,505	17	23,206
III	64-31	52	69	254	2,119	5,231	20,661	6,782	939	9	36,116
IV	30-10	107	525	1,520	4,728	5,798	9,327	1,181	129	-	23,315
V	9 or below	594	1,965	2,101	2,661	2,137	954	12	3	-	10,427
	Total	753	2,565	3,896	9,834	14,738	45,338	19,281	3,552	43	100,000
						White			BASE: 100,000		
I	100-93	-	-	3	22	218	2,864	3,966	1,163	20	8,256
II	92-65	-	8	23	363	1,635	14,136	9,373	1,770	21	27,329
III	64-31	63	78	282	2,429	5,927	22,581	7,245	1,000	7	39,612
IV	30-10	115	534	1,573	4,675	4,871	6,578	812	91	-	19,249
V	9 or below	504	1,517	1,409	1,270	635	206	10	3	-	5,554
	Total	682	2,137	3,290	8,759	13,286	46,365	21,406	4,027	48	100,000
						Negro			BASE: 10,000		
I	100-93	-	-	-	-	-	8	6	1	--	15
II	92-65	-	-	1	2	11	96	81	15	-	206
III	64-31	-	3	11	54	166	1,081	441	64	1	1,821
IV	30-10	7	48	124	500	1,057	2,343	308	31	-	4,418
V	9 or below	105	425	565	979	985	479	2	-	-	3,540
	Total	112	476	701	1,535	2,219	4,007	838	111	1	10,000

TABLE 5. (Continued)

Mental Qualification		Educational Attainment: Years of School Completed									
Mental Group	Percentile Score	None	Elementary School			High School		College			Total
			1-4	5-6	7-8	1-2	3-4	1-2	3-4	5 or more	
MOUNTAIN:						Total			BASE: 100,000		
I	100-93	-	5	3	22	251	4,533	5,408	1,349	13	11,584
II	92-65	-	4	11	475	2,113	16,885	9,890	1,302	13	30,693
III	64-31	-	25	83	2,141	5,617	21,266	6,152	545	13	35,842
IV	30-10	70	429	638	3,694	4,795	6,629	586	34	-	16,875
V	9 or below	729	1,101	898	1,235	768	261	13	1	-	5,006
	Total	799	1,564	1,633	7,567	13,544	49,574	22,049	3,231	39	100,000
						White			BASE: 100,000		
I	100-93	-	6	3	23	267	4,825	5,753	1,427	15	12,319
II	92-65	-	4	12	506	2,201	17,785	10,386	1,371	13	32,278
III	64-31	-	27	88	2,178	5,790	21,579	6,364	550	8	36,584
IV	30-10	74	304	574	3,461	4,405	5,922	441	29	-	15,210
V	9 or below	525	783	547	949	597	194	13	1	-	3,609
	Total	599	1,124	1,224	7,117	13,260	50,305	22,957	3,378	36	100,000
						Negro			BASE: 10,000		
I	100-93	-	-	-	-	-	-	-	13	-	13
II	92-65	-	-	-	-	73	285	219	22	-	599
III	64-31	-	-	-	158	292	1,639	285	45	10	2,429
IV	30-10	-	236	164	733	1,088	1,763	286	14	-	4,284
V	9 or below	392	605	637	569	343	129	-	-	-	2,675
	Total	392	841	801	1,460	1,796	3,816	790	94	10	10,000
PACIFIC:						Total			BASE: 100,000		
I	100-93	-	5	1	29	304	4,535	5,495	1,259	27	11,655
II	92-65	-	2	10	295	2,260	16,672	9,632	1,458	21	30,350
III	64-31	5	32	55	1,291	6,447	21,801	5,842	622	10	36,105
IV	30-10	61	229	328	2,168	5,167	8,114	1,125	85	1	17,278
V	9 or below	496	810	493	908	1,132	746	22	5	-	4,612
	Total	562	1,078	887	4,691	15,310	51,868	22,116	3,429	59	100,000
						White			BASE: 100,000		
I	100-93	-	6	1	31	328	4,834	5,833	1,338	31	12,402
II	92-65	-	2	11	317	2,407	17,607	10,004	1,518	21	31,887
III	64-31	5	34	57	1,343	6,748	22,017	5,658	564	8	36,434
IV	30-10	40	219	319	2,138	4,873	6,762	794	70	-	15,215
V	9 or below	519	811	484	814	873	533	23	5	-	4,062
	Total	564	1,072	872	4,643	15,229	51,753	22,312	3,495	60	100,000
						Negro			BASE: 10,000		
I	100-93	-	-	-	-	-	65	110	24	-	199
II	92-65	-	-	-	-	36	452	478	67	2	1,035
III	64-31	-	-	-	65	251	1,902	825	139	2	3,184
IV	30-10	31	36	46	253	900	2,566	544	28	1	4,405
V	9 or below	23	79	61	213	450	351	-	-	-	1,177
	Total	54	115	107	531	1,637	5,336	1,957	258	5	10,000

*The distributions by mental qualification were standardized according to the distributions of male youths age: 19-21, by educational attainment, derived from United States Census of Population, Volume 1, Characteristics of the Population (Tables 101 and 102), 1960.

TABLE 6. MEDIAN YEARS OF SCHOOL COMPLETED BY MALE YOUTHS, (19-21) YEARS OF AGE, AND THEIR EXPECTED MEDIAN SCORES OF THE ARMED FORCES QUALIFICATION TEST (AFQT), BY GEOGRAPHIC DIVISION AND RACE, 1960*

Geographic Division	Median Years of School Completed			Expected Median Percentile Score on AFQT		
	Total	White	Non-White	Total	White (Non-Negro)	Negro
UNITED STATES	12.3	12.4	11.0	50	55	15
New England	12.4	12.4	11.4	54	56	21
Middle Atlantic	12.3	12.4	11.2	51	55	20
East North Central	12.4	12.4	11.4	56	58	22
West North Central	12.5	12.5	11.6	62	63	21
South Atlantic	12.0	12.2	10.3	40	51	13
East South Central	11.8	12.2	10.2	36	47	11
West South Central	12.2	12.3	11.0	46	51	13
Mountain	12.4	12.4	10.7	57	60	17
Pacific	12.4	12.4	12.3	57	60	25

*See footnote to Table 5.

EDUCATIONAL DATA FROM LARGE SCALE TESTING PROGRAMS: RESULTS AVAILABLE FROM COLLEGE LEVEL TESTING PROGRAMS

W. B. Schrader, Educational Testing Service

Since 1950, a number of salient changes have occurred in American higher education. A steadily increasing proportion of high school graduates has sought admission to colleges and universities. This trend, coupled with the marked increase in the birth-rate following World War II, has brought unprecedented demands for college education in recent years. As pressures for college admission increased, problems of college choice and needs for guidance became severe. Questions of "talent loss" and of needs for highly educated persons were widely discussed. Along a somewhat different line, there were many significant efforts to modify the high school curriculum led by outstanding scientists and scholars. As a result, the academic preparation of candidates for college admission has also been undergoing change. Finally, testing programs dealing with the transition from high school to college and from college to graduate and professional school have grown rapidly during this period.

This report will be concerned with data developed by college level testing programs and relevant to current questions in higher education. College level testing will be considered to include tests administered in the junior and senior years of high school if they are oriented toward planning for college as well as tests administered during the college years. It will be assumed that a testing program includes the development or selection of tests, their administration under standard conditions, scoring and reporting, and the development of materials (often statistical) to aid in score interpretation. In particular, the preparation of norms will be considered an integral part of a testing program, even if this involves <u>special</u> test administrations not part of the basic program operations.

Essentially, this report will be concerned with comparisons of various educationally significant groups with respect to performance on widely-used tests of academic ability and achievement. Brief treatment will be given to each of the following questions:

- -- How different in ability are college freshmen from high school seniors?
- -- How do major fields of study differ with respect to the ability of students that they attract?
- -- What indications of trends can be observed on the basis of year-to-year comparisons of student groups?

In addition, I plan to describe briefly a study now under way to learn more about the current state of subject-matter preparation of candidates for admission to college.

The relation of test performance to college attendance was studied on a state-wide basis by Learned and Wood (1938) in their classic Pennsylvania study of the 1930's. Other early state-wide studies by Toops (1940) in Ohio, Phearman (1949) in Iowa and Berdie (1954) in Minnesota supported the generalization that a large proportion of the highly able students were not attending college. Wolfle (1954) in his influential book <u>America's Resources of Specialized Talent</u> developed a national picture by judicious compilation of data from a variety of sources. Shortly afterward, a large national survey involving a brief test and a questionnaire and bearing on this question was initiated. This was the 1955 National Study of High School Students and Their Plans conducted by ETS as part of a broader study of scientific talent done by the College Board with the support of the National Science Foundation. Major reports emerging from this study include: Cole (1956), Stice, Mollenkopf, and Torgerson (1956), and Educational Testing Service (1957). More recently, Project Talent has developed a great quantity of information on this and many other topics. Among the many publications, Flanagan and others (1964) and Flanagan and Cooley (1966) are particularly relevant. Darley (1962) and a National Science Foundation report prepared by Bridgman have also made important contributions to the question of how ability is related to college-going. Like Wolfle, they used existing data to arrive at national estimates.

The utilization of data from testing programs to estimate the ability level of college freshmen is seriously handicapped by the fact that different colleges use different tests, and that some colleges presumably do not test their freshmen. Moreover, since only students who are considering enrollment in college are likely to take the College Board or the American College Testing Program tests, comparison of college freshmen with a cross-section of high school seniors cannot be obtained from operational data. One solution to this problem, clearly, is to administer a widely-used test to a representative sample of high school seniors and then follow them up to find out which ones enter college. This is essentially the NSF and Project Talent design except that it calls for using an operational test rather than specially developed tests. Such a study has been done with the support of the College Entrance Examination Board to provide data to aid in score interpretation. This study utilized the Preliminary Scholastic Aptitude Test (PSAT) which is somewhat shorter than the Scholastic Aptitude Test (SAT) but is parallel in content. This parallelism is important because the SAT has been extensively validated as a predictor of college grades.

The initial purpose of the study was to develop high school norms for the PSAT and, indirectly, for the SAT. In drawing the sample of schools for these norms, a comprehensive list, maintained by ETS, of public and private secondary schools was used. Each school was classified on the basis of the nine census regions and as public or private. The number of schools to be drawn from each of the eighteen subsets was made proportional to the number of schools in the subset. Two hundred schools were selected for the base sample. Usable senior test data were obtained for 147 schools and 9,745 students. Although the proportion of schools cooperating was smaller than would be needed for definitive results, it was judged that the data were adequate to aid in score interpretation. Testing was conducted in October, 1960. The study is described in Chandler and Schrader (1966).

In the follow-up phase of the study, conducted by Seibel (1965), only a subsample of the 9,745 students was used. Using a composite score giving a weight of 2 to PSAT-Verbal and a weight of 1 to PSAT-Mathematical, all students in the top 5%, half of the students in the next 25% and one-tenth of the students in the bottom 70% were selected for follow-up. This reduced the number to 2,423. In the analysis, results for each student were weighted in accordance with the sampling ratio for his group. Some of Seibel's results are shown in Table 1. I must emphasize that these results describe not freshmen in general but those freshmen who entered college within a year after graduation from high school. There is good reason to believe that the means are somewhat higher than they would be for an entire freshman class. Subject to this limitation, however, they show the relative test performance of four educationally significant groups on a widely-used test. Rank-in-class as shown in Table 1 was converted to a percentile rank, and the resulting percentage expressed as a normal deviate on a scale having a mean of 13 and a standard deviation of 4. I might add that Seibel's data on rank-in-class were obtained from the high school and that the enrollment of each student in college was confirmed by the college. Seibel's figures for percent going to college within a year after high school graduation are: boys, 41%, girls, 32%, and total 36%. Project Talent results indicated that among the high school seniors in 1960, 46% of the boys, 33% of the girls, and 39% of the total reported that they had attended a two-year or four-year college since high school. However, 49% of the boys, 35% of the girls and 42% of the high school graduates stated that they had attended a recognized college, as reported by Flanagan and others (1964). The Project Talent results seem reasonably consistent with Seibel's findings.

From Seibel's report, it may also be estimated that, when PSAT 2V + M is used as the measure of ability;

Of boys in the top 5% in ability, 96% went directly to college

Of boys in the top 40% in ability, 74% went directly to college

Of girls in the top 4% in ability, 91% went directly to college

Of girls in the top 36% in ability, 66% went directly to college

It should be noted that the follow-up involved only 507 cases in the top 5%, and 1,240 cases in the next 25%, so that sampling error as well as sample bias must be considered in evaluating the percentages given. Nevertheless, they tend to support the view advanced by Berdie and Hood (1965) in their study of college plans of Minnesota high school seniors in 1960 that progress is being made in increasing the proportion of high ability youth going to college.

Scores earned in testing programs may yield useful information on ability differences among major fields of study and on career choices. An early example of relevant data is provided by Chauncey's (1952) article which included, along with extensive data on many aspects of the 1951 Selective Service College Qualification Test, data on the performance of students majoring in, or planning to major in, various subjects. These results indicated that, insofar as the sample was representative, a student enrolled in engineering, physical science or mathematics could expect to rank lower in ability relative to his fellow majors than a student of equal ability enrolled in, say, business or education. It should be clear, of course, that in any one college the relative ability levels might differ from the national pattern.

It should be noted that comparison of education majors with majors in other fields does not directly answer the question of the relative ability level of prospective teachers. In a study of college sophomores conducted in 1963 to provide norms for the Comprehensive College tests, it was possible to examine this question. In this study, as reported by Haven (1964), it turned out that of 1,227 prospective teachers who listed a major field, only 472 gave education as their major field. If prospective college teachers are excluded, 464 out of 1,046 prospective teachers, or less than half, gave education as their major field. When prospective teachers were classified by level of school, secondary school teachers were quite close to, though below, the general average in each of five tests covering various aspects of college work. In Natural Sciences and in Mathematics, they were about one-fifth of a standard deviation below the general average. These results are based on about 150 cases.

Table 2 presents data drawn from the Graduate Examinations National Program with respect to Verbal and Quantitative scores earned by candidates in various fields. It may be useful in interpreting Table 2 to note that the scales for each test were established in 1952 so that a sample of college seniors would have a mean of 500 and a standard deviation of 100. All comparisons across fields are complicated, of course, by varying practices with respect to requirement of the tests by various departments and by national selection programs. Certainly no definitive conclusions can be drawn from them. They do illustrate, however, the possibility of developing information about major fields as a by-product of operational testing programs.

Test program data should be useful for making year-to-year comparisons of the ability levels of total candidate groups and of particular candidate subgroups. This statement assumes that a continuing score scale is maintained by appropriate equating methods. Most large-scale testing programs do have continuing scales to facilitate the use of scores. The utility of year-to-year comparisons should be greater if the candidate group represents a well-defined and stable population.

In my judgment, the Medical College Admission Test has a particularly good set of statistics for year-to-year comparisons. The score scale was set in 1951 to yield a mean of 500 and a standard deviation of 100. Since then the scales have been maintained by equating. During this period, virtually all candidates for medical school admission took the test. Only one of the four tests has changed sufficiently to require a major change in title: "Understanding Modern Society" was replaced by "General Information" beginning in May 1962. A series of articles in the Journal of Medical Education has reported annual studies of the characteristics of medical applicants and accepted students. Thus a valuable body of data has been built up over the years. Data on number of applicants and the mean scores of applicants and accepted students are especially relevant. The complete series up to 1965 entrants can be found in two articles: Hutchins and Gee (1962) and Johnson (1965). On the Science test a consistent upward trend is apparent for classes admitted since 1957. The mean for applicants has risen from 482 to 513 and for accepted students, the mean has risen from 516 to 556 over the same period. It is more difficult to characterize the pattern in the other tests, except to note that the General Information test mean has been rising fairly noticeably since it was introduced and that mean scores in Verbal and Quantitative have been somewhat higher during the past five years than they were during the preceding five.

The Law School Admission Test, since it was instituted in 1948, has undergone a rapid growth both in the number of law schools requiring it and the number of candidates taking it. On the basis of unpublished data covering the testing years 1958-59 to 1964-65, there has been a consistent upward trend in scores from 1959-60 to 1964-65 from a mean of 477 to a mean of 510. (The standard deviation is about 100.) During this same period the candidate group nearly doubled in size -- from 20,735 to 39,162.

With respect to the College Board Scholastic Aptitude Test, the very rapid growth of this test both in the number of colleges requiring it and in the number of candidates taking it would seem to make year-to-year comparisons of scores difficult to interpret. From a different viewpoint, however, what is remarkable is the relatively small change in mean scores despite the marked change in the group tested. The mean Verbal score was 483 for seniors tested in the period December 1956 through March 1957 and 471 for seniors tested in the period December 1965 through March 1966. The corresponding means for Mathematical scores were 504 and 494. (For both tests the standard deviation for the more recent scores is about 110.) The number of candidates is more than four times as large for 1965-66 as for 1956-57. With respect to the verbal test, Stewart (1966) recently completed a study of the long-range stability of the score scale. She found that check equatings using December 1963 data agreed satisfactorily with original equatings done in March 1953 and in February 1957.

It may also be worth mentioning that boys scored 67 scaled score points higher on the mathematical test than did girls for the 1956-57 program data but only 46 points higher in the current year's data.

A somewhat different way in which data collected as an integral part of program operations may have some general usefulness is illustrated by the extensive questionnaire study of high school curricula now being conducted by ETS for the College Board. In this study, the aim is to describe the academic preparation of candidates currently choosing the various College Board achievement tests rather than of high school seniors or prospective college students generally.

The sample for this study included in all about 38,500 candidates. No candidate was asked to complete more than one questionnaire. The sampling plan allowed for slight overlap between fields, but a candidate drawn for two fields was assigned to the field having fewer candidates.

Ten questionnaires were prepared with the help of the examining committees in the various fields. The questionnaires cover English, Mathematics, History and Social Studies, three sciences -- Biology, Chemistry, and Physics -- and four languages -- Latin, French, German, and Spanish. Part of each questionnaire is concerned with the amount of work the student has taken in each subject, and with course titles in the particular field covered by his questionnaire, and in closely-related fields. Each questionnaire includes many questions about learning activities and course content by topics. For example, the mathematics questionnaire asks the student whether or not he has studied the normal curve, statistical sampling, and statistical inference.

In conclusion, data obtained in connection with operational testing programs may provide information relevant to broad educational questions. Test data collected in operational programs have two distinct advantages: students are likely to be highly motivated and relatively large samples can be obtained economically. For some purposes, however, the interpretation of results based on program data is difficult or impossible, particularly because the sample is self-selected. For such purposes, only a specially designed study can yield the information needed.

REFERENCES

Berdie, Ralph F., After High School -- What? Minneapolis, Minnesota: The University of Minnesota Press, 1954.

Berdie, Ralph F. & Hood, Albert B., Decisions for Tomorrow: Plans of High School Seniors for After Graduation. Minneapolis, Minnesota: University of Minnesota Press, 1965.

Chandler, Marjorie Olsen & Schrader, William B., Preliminary Scholastic Aptitude Test 1960 National Norms. Statistical Report 66-3. Princeton, New Jersey: Educational Testing Service, 1966.

Chauncey, Henry, "The Use of the Selective Service College Qualification Test in the Deferment of College Students." Science, 1952, 116, 73-79.

Cole, Charles C., Encouraging Scientific Talent. New York: College Entrance Examination Board, 1956.

Darley, John G., Promise and Performance: A Study of Ability and Achievement in Higher Education. Berkeley, California: Center for the Study of Higher Education, University of California, 1962.

Educational Testing Service, Background Factors Relating to College Plans and College Enrollment Among Public High School Students. Princeton, New Jersey: Educational Testing Service, 1957.

Flanagan, John C. & Cooley, William W., Project Talent: One-Year Follow-Up Studies. Cooperative Research Project No. 2333. Pittsburgh, Pennsylvania: School of Education, University of Pittsburgh, 1966.

Flanagan, John C. and others, The American High-School Student. Cooperative Research Project No. 635. Pittsburgh, Pennsylvania: Project Talent Office, University of Pittsburgh, 1964.

Harvey, Philip R. & Marco, Gary L., Aptitude and Advanced Test Scores of 1963-1964 National Program Candidates by Undergraduate Major Fields. Graduate Records Examination Special Report No. 65-3. Princeton, New Jersey: Educational Testing Service, 1965.

Haven, Elizabeth W., The Sophomore Norming Sample for the General Examinations of the Comprehensive College Tests. Princeton, New Jersey: Educational Testing Service, 1964.

Hutchins, Edwin B. & Gee, Helen Hofer, "The Study of Applicants, 1960-61." The Journal of Medical Education, 1962, 37, 1203-1212.

Johnson, Doris G., "The Study of Applicants, 1964-65." The Journal of Medical Education, 1965, 40, 1017-1030.

Learned, William S., & Wood, Ben D., The Student and His Knowledge. Carnegie Foundation for the Advancement of Teaching Bulletin, No. 29. New York: The Foundation, 1938.

National Science Foundation, The Duration of Formal Education for High-Ability Youth. Report NSF 61-36.

Phearman, L. J., "Comparisons of High School Graduates Who Go to College with Those Who Do Not." Journal of Educational Psychology, 1949, 40, 405-414.

Seibel, Dean W., Follow-Up Study of a National Sample of High School Seniors: Phase 2 - One Year After Graduation. College Entrance Examination Board Research and Development Report, RDR 65-6, No. 1. Princeton, New Jersey: Educational Testing Service, 1965.

Stewart, E. Elizabeth, The Stability of the SAT-Verbal Score Scale. Research Bulletin 66-37. Princeton, New Jersey: Educational Testing Service, 1966.

Stice, Glen, Mollenkopf, Wm. G., & Torgerson, Warren S., Background Factors and College-Going Plans Among Public High School Seniors. Princeton, New Jersey: Educational Testing Service, 1956.

Toops, Herbert A., "The Prediction of College-Going." School and Society, 1940, 51, 257-261.

Waite, Annette C. & Harvey, Philip R., An Analysis of Graduate Record Examinations Scores by the Undergraduate Field of Study, 1962-63. Graduate Record Examinations Special Report 64-2. Princeton, New Jersey: Educational Testing Service, 1964.

Wolfle, Dael, America's Resources of Specialized Talent: A Current Appraisal and a Look Ahead. New York: Harper and Brothers, 1954.

TABLE 1

Test Performance and Rank in Class for High School Seniors, College Entrants, Students Who Completed One Year of College, and Students Who Completed the Year in Good Standing[a]

Group	PSAT-Verbal Mean	PSAT-Verbal S.D.	PSAT-Verbal N[b]	PSAT-Mathematical Mean	PSAT-Mathematical S.D.	PSAT-Mathematical N[b]	PSAT: 2V+M Mean	PSAT: 2V+M S.D.	PSAT: 2V+M N[b]	Rank In Class Mean	Rank In Class S.D.	Rank In Class N[b]
BOYS												
High School Seniors	36.3	11.3	4585	43.2	11.5	4585	115.8	32.3	4585	12.4	3.8	4199
Entered College	43.2	11.4	1878	50.3	11.1	1878	136.8	31.7	1878	14.5	3.5	1825
Completed One Year	44.1	11.4	1599	50.9	11.2	1599	139.1	31.8	1599	14.8	3.5	1558
Completed Year in Good Standing	46.1	10.9	1263	52.4	11.0	1263	144.6	30.5	1263	15.3	3.5	1246
GIRLS												
High School Seniors	37.0	11.8	5162	38.5	10.3	5162	112.5	32.2	5162	14.3	3.6	4799
Entered College	46.0	12.2	1659	45.9	10.5	1659	137.9	32.6	1659	16.3	3.4	1609
Completed One Year	46.6	11.8	1506	46.4	10.5	1506	139.6	31.8	1506	16.4	3.4	1466
Completed Year in Good Standing	47.3	11.7	1377	47.0	10.3	1377	141.7	31.2	1377	16.6	3.3	1349
TOTAL												
High School Seniors	36.7	11.6	9747	40.7	11.1	9747	114.0	32.3	9947	13.4	3.8	8998
Entered College	44.5	11.8	3537	48.3	11.1	3537	137.3	32.1	3537	15.3	3.6	3434
Completed One Year	45.3	11.7	3105	48.7	11.1	3105	139.3	31.8	3105	15.6	3.5	3024
Completed Year in Good Standing	46.7	11.3	2640	49.6	11.0	2640	143.1	30.9	2640	16.0	3.5	2595

[a]Source: Seibel, D. W. Follow-up Study of a National Sample of High School Seniors: Phase 2 - One Year After Graduation. College Entrance Examination Board Research and Development Reports, RDR 65-6, No. 1, 1965.

[b]All N's are weighted total N's.

TABLE 2

Mean Score on Graduate Record Examinations
Aptitude Test for National Program Candidates by
Major Field of Study[a]

(Fields having 500 or more candidates in 1963-64)

	Mean Score				Number of Candidates[b]	
	1962-63		1963-64			
Major Field	Verbal	Quanti-itative	Verbal	Quanti-itative	1962-63	1963-64
Agriculture	419	507	421	505	553	785
Biology	524	530	518	525	2086	3096
Business	463	518	461	515	1775	2898
Chemistry	554	643	551	640	2596	3557
Economics	553	588	546	581	1334	2241
Education	453	439	448	427	6266	8807
Engineering	520	680	516	675	5432	7754
English	599	493	600	492	4422	6953
Fine Arts	529	455	524	450	415	645
French	583	488	580	486	544	970
Geology	542	592	549	599	604	717
History	567	498	569	495	3206	5367
Home Economics	438	421	432	410	317	551
Humanities	603	529	609	530	521	703
Mathematics	562	670	559	668	3153	4778
Music	497	454	492	457	752	1086
Natural Science	524	583	523	576	568	625
Philosophy	624	562	621	565	751	1017
Physical Education	402	417	399	411	773	1223
Physics	589	697	589	695	2827	3472
Political Science	584	531	582	528	1599	2476
Psychology	565	531	565	529	3822	5696
Religion	540	491	537	487	430	573
Social Science	528	497	522	476	830	1252
Sociology	533	486	532	477	952	1606
Spanish	533	457	533	446	383	595
Speech	507	440	501	433	372	551
Zoology	542	548	546	560	771	1110

[a]Sources: Waite, Annette C. and Harvey, Philip R., An Analysis of Graduate Record Examinations Scores by the Undergraduate Field of Study, 1962-63. Graduate Record Examinations Special Report 64-2, 1964.

Harvey, Philip R. and Marco, Gary L., Aptitude and Advanced Test Scores of 1963-64 National Program Candidates by Undergraduate Major Fields. Graduate Record Examinations Special Report 65-3, 1966.

[b]Numbers of cases given are for the Verbal Score. Numbers for the Quantitative Score may differ slightly.

THE EDUCATIONAL OPPORTUNITIES SURVEY

Frederic D. Weinfeld, U.S. Office of Education

Last July 2nd the U. S. Commissioner of Education presented to the President and the Congress, a report on the Equality of Educational Opportunity. This report was in compliance with Section 402 of the Civil Rights Act of 1964, which required that "The Commissioner shall conduct a survey and make a report to the President and the Congress, within two years of the enactment of this title concerning the lack of availability of equal educational opportunities for individuals by reason of race, color, religion or national origin in public educational institutions at all levels in the United States, its territories and possessions, and the District of Columbia".

The Educational Opportunities Survey was carried out by the National Center for Educational Statistics of the U.S. Office of Education, directed by Alexander M. Mood. In addition to its own staff, the Center used the services of outside consultants and contractors. James C. Coleman of Johns Hopkins University had major responsibility for the design, administration and analysis of the survey. Ernest Q. Campbell of Vanderbilt University shared this responsibility, and particularly had major responsibility for the college surveys, while I had the fortune to be Project Officer for the Survey.

Commissioner Harold Howe II described the survey with these words in his letter of transmittal:

> Stated in broadest terms, the survey addressed itself to four major questions.
>
> The first is the extent to which the racial and ethnic groups are segregated from one another in the public schools.
>
> The second question is whether the schools offer equal educational opportunities in terms of a number of other criteria which are regarded as good indicators of educational quality.
>
> . . . .
>
> Only partial information about equality or inequality of opportunity for education can be obtained by looking at characteristics, which might be termed the schools' input. It is necessary to look also at their output--the results they produce. The third major question, then, is addressed to how much the students learn as measured by their performance on standardized achievement tests.
>
> Four is the attempt to discern possible relationships between students' achievement, on the one hand, and the kinds of schools they attend on the other.

Work was started on the Survey in the Spring of 1965 and plans were made for administration and testing in late September of 1965. The plans called for the testing and surveying of about 800,000 students in some 5,000 schools throughout the country in grades 1, 3, 6, 9 and 12 together with their teachers, principals and superintendents.

The Educational Testing Service of Princeton, New Jersey, was awarded the contract, on the basis of competitive bids, for conducting the Educational Opportunity Survey, including test administration, test scoring, data processing and data analysis. They also consulted on various aspects of the Survey and convened an Advisory Panel to aid in the design and analysis of the study.

I might add at this point that in addition to the Survey, which made up the major body of the report to the Congress, there were also several small contracts let for specific projects and studies. Among these was the study directed by Charles Nam of Florida State University using the Current Population Survey of the U.S. Bureau of the Census to collect additional data, especially about school drop-outs who would not be picked up by the main survey.

The Survey sample consisted of a 5% sample of schools. This was a two-stage, self-weighing, stratified, cluster sample with counties and SMSA's being the Primary Sampling Units (PSU's) in the first stage and with high schools being the PSU's in the second stage. When a high school was drawn in the sample the elementary schools feeding into that school were automatically included in the sample also. Since the Educational Opportunities Survey was primarily concerned with the children of minority groups, and since these groups constituted only about 10% of the total school population, the schools were stratified according to the percentage of non-white students. Strata with higher percentage of these students were given larger sampling ratios and were sampled more heavily, with the final results that over 40% of the students in the Survey were from minority groups.

The instruments for the Survey were designed to collect the data needed for analysis along the lines of the four aforementioned questions concerning the availability of equal educational opportunity. The Survey depicts evidence of inequalities in educational opportunity by developing comparative statistical information for items and resources that educators agree are relevant to school quality. Comparison was made of the exposure to these relevant items of school quality by children of minority groups and by children of the majority group. To obtain these data, questionnaires were devised and administered to the teachers, the school principals, and the superintendents of the participating sample schools.

The Teacher Questionnaire contained some 72 questions including: personal data, professional training, type of college attended, teaching experience, type of school and student preferred, job satisfaction, opinions on issues and problems of integration, such as bussing and compensatory programs for the disadvantaged, and problems existing in their school. The final part of the Teacher Questionnaire consisted of a voluntary test of 30 contextual vocabulary items. The purpose of this test was to get a measure of the verbal facility of the teachers.

The 100 item Principal Questionnaire was the main source of information about the school. The questions covered school facilities, characteristics, staff, programs, racial composition, problems, curricula, extra-curricular activities, etc. There were also questions on the personal background and training of the principal and on his opinions on problems of integration.

The Superintendent Questionnaire consisted of 41 questions dealing with administrative information about the school system, selected statistics about the school system and its expenditures, attitudes towards current school issues, and personal information about the Superintendent.

Detailed factual and attitudinal data about the students were also obtained by questionnaires. Included were items of home background information so that the student data could later be controlled on these items of socio-economic status, family background, family interest in education, etc. Different questionnaires, appropriate to each of the grade levels were used.

The 12th Grade Student Questionnaire for example, was comprised of some 116 items. In addition to the questions on home background and the usual personal and school data there were questions on the students' attitude towards school, race relations, and the world, such as: "How good a student do you want to be in school?" "If you could be in the school you wanted, how many of the students would you want to be white?" and "Good Luck is more important than hard work for success. (Agree or Disagree)".

Tests of the various school skills were to be the yardsticks for measuring the detrimental effects of poor school facilities and characteristics upon student learning. The Test Battery was designed as an integral part of the entire research design. The object was to obtain as much data as possible within the limitations of time and available resources. Two of the basic skills chosen were reading comprehension and mathematics ability. These two areas are common in all school curricula and are taught in all schools at all grade levels. Another area deemed of importance was that of the general level of knowledge gained by the students either from their school courses or from experiences in the outside world. A test of general information was therefore included in the battery in order to measure this type of learning. Two other ability tests, were used to measure the students' skills in the verbal and reasoning areas. The two tests of this type included in the battery were the verbal and non-verbal ability tests.

One major limitation in the design in the test battery was the time required for test administration. It was desirable and considered administratively feasible to have the test battery and the questionnaires completed in no more than one school day. The lower grades had to have a shorter battery because of the limited attention span of the younger children. Therefore the testing time increased in the various test batteries until it reached its maximum length in the 12th grade.

Since the lead time before the administration of the Survey in September 1965 was too short to develop specific tests in the above areas existing standardized tests were used. However, because full length standardized tests usually require more time than would have been available, it was decided to use shortened, or half-length, forms of these tests rather than to omit tests in any area. Another requirement was that the various tests be interlocked through as many grades as possible so that scores on the same type of tests administered at different grade levels could be compared. The scaling allowed us to have a comparable measure of growth between the different grades.

The law required that the Survey be made at "all levels" and so it was decided to administer the tests to selected grades at spaced intervals. This would give us a good picture of what was going on in the schools without having to test at grade level. The grades chosen were Grades 1, 3, 6, 9 and 12.

The tests were of the multiple-choice answer type and were provided with machine scorable answer sheets. These answer sheets were scored and processed by a machine which scanned the penciled responses optically and put the results directly onto magnetic tape. For Grades 1 and 3 an accordion type answer booklet was used. The pupils marked their responses directly onto the booklet without using a separate answer sheet. This procedure eliminated errors that might have been caused by young children in transcribing their responses onto a separate answer sheet. The use of machine scorable tests for Grades 1 and 3 is a relatively new procedure which cuts down costs and errors considerably by eliminating hand scoring.

This then is the basic data which we have from the survey. It is comprehensive data in the sense that we have collected for each individual student as much information possible about him, his teachers, and his school. Because of this collection of integrated data, related aspects of the global educational situation can now be investigated. The data was processed for the report on IBM 7090 computers and the collated data is now on magnetic tape. For each student in the sample, the tape record now consists of his student questionnaire responses, his test scores, the average questionnaire responses and average test scores of the students in his school, the average of his teachers' questionnaire responses, his principal's questionnaire responses, his school superintendent's questionnaire responses, and the appropriate sampling weight for the student. The data are grouped by geographical regions. There are 5 SMSA regions and 3 non-SMSA regions. The sampling design did not allow for any smaller breakdowns by States or by counties.

The following three reports have been published: a 33 page summary of the report, the complete report which was presented to the Congress, a 737 page document, and a Supplemental Appendix to the Survey which contains basic correlation matrices for samples of 1,000 students from various regional, racial, and grade groupings. Copies of these reports are available upon request.

I would like to mention briefly some of the preliminary analyses of the data which we have conducted for the report. The extent of the segregation in the public schools of racial and ethnic groups - Commissioner Howe's first question is shown in figures 1 through 4 on pages 4, 5, 6 and 7. The second question, whether the schools offer equal educational opportunity, is partially answered in tables 1 through 4 starting on page 10. Here the exposure of various groups to many of the tangible school facilities, characteristics, and relevant items of school quality is compared. Teacher and principal characteristics are also compared in this way in tables 5 and 6. As I mentioned previously there is also on tape the aggregate values of the responses of the student for each school. These peer group characteristics are compared in tables 7 and 8.

The third question - comparative student performance on standardized achievement test - is sketchily presented in table 9 on page 20. Much more detailed data about all these questions is presented in later sections of the report.

The fourth question, the relationship between student achievement and the kinds of schools they attend, is really the interesting one. Using samples of 1,000 students from various regional, racial, and grade groups it was found that most of the variability in school achievement resided in the within school variance and only a lesser percentage was accounted for by the between school variance. This was after the socio-economic home background of the students was partialled out. For most minority groups their achievement was found to be more highly related to the type of school they attend than the majority students.

Using techniques of multiple-partial regression, again partialling out student home background, it was found that variation in school facilities and characteristics account for relatively little of the variance in student achievement insofar as this is measured by the standardized tests used as criteria. Of all the school variables, the quality of teachers showed a stronger relationship to pupil achievement, it was progressively greater at the higher grades, indicating a cumulative impact of the quality of the teachers in a school or pupil's achievement. Again, teacher quality is more important for minority pupil achievement than for that of the majority.

Besides these regression studies there were investigations of the attitudes and aspirations of the students, future teachers of minority groups, educational opportunity at public institutions of higher education, non-enrollment or school drop outs, project Headstart pupils, disadvantage associated with foreign language in the home, guidance counselors and vocational education. All these studies are contained in the main Report.

The National Center for Educational Statistics is now conducting a continuing program of analysis of this data. Investigations currently planned are: the development of indices of student socio-economic background, student educational background, teacher characteristics, and school quality; an examination of the relationship of school expenditures to achievement school quality and efficiency; a study of students and schools in Appalachia, the effect of social and regional teacher mobility; validation of previous regression equations; and various specific educational problems such as the effect of watching television, etc.

The data from the Educational Opportunities Survey are all in the public domain and will later be made available by listings, cards, or tape to interested educational researchers who may wish to use the data and analyze it for their own specific purposes and fields of interest. The confidentiality of the data will, of course, have to be maintained. The data, it must be remembered, is cross-sectional data. The Educational Opportunities Survey was a one-day, one-shot, survey with complete anonymity of all participants so that there can be no follow-up studies made as has been done with data from Project TALENT or that can possibly be done with data from various college testing programs.

As Dr. Mood stated in the report:

> In view of the fundamental significance of educational opportunity to many important social issues today, Congress requested the survey of educational opportunity reported

in this document. The survey is, of course, only one small part of extensive and varied activities which numerous institutions and persons are pursuing in an effort to understand the critical factors relating to the education of minority children and hence to build a sound basis for recommendations for improving their education.

Probably the main contribution of the survey to this large and long range effort will be in the fact that for the first time there is made available a comprehensive collection of school data (at selected grade levels) gathered on consistent specifications throughout the whole Nation.

DISCUSSION*

John W. Tukey, Princeton University and Bell Telephone Laboratories

Dr. Karpinos has been careful to apply the then-available techniques to persuade the data to answer each question of immediate concern rather than a mixture of that question with others. In particular he has made much use of standardized rates. When such studies are made in the future, however, there may be a real advantage to going further and comparing superstandardized rates. (A description of their use in a particular context will appear in Appendix 2 to Chapter IV-6 of the forthcoming NAS-NRC report on the National Halothane Study.)

The Equality of Educational Opportunity Survey involved many problems, some narrowly statistical, others broad problems of interpretation. The extent to which differences were associated with schools rather than individuals was unplanned for, so that the classical problem of "the correct error term" (so familiar in agricultural experiment) arose at a relatively late stage of the analysis.

It would be desirable, perhaps essential, to know how to interpret the results found for the effects ascribable to "teacher quality". Do these effects come from differences in teachers? Or, by some other route, from those basic differences in community attitudes toward the importance of education that causes some communities to expend money and effort in getting and keeping "better" teachers? This is but one of many problems that the great increase in factual knowledge gained from the survey has dumped in our laps.

Dr. Schrader's discussion of % going to college in various ability ranges provides a helpful and stimulating summary of the data, at least at a tabular level. We are now ready to try to bring sufficient order out of the numbers that we have a reasonable chance of noticing changes in the ongoing process, if any exist. To this end, we clearly ought to change our focus from raw % to some measure that more adequately allows for the greater difficulty of altering extreme fractions by 1%. I would urge an examination of the logit of the fraction going to college as a function of ability and epoch. The time rates of change at various abilities will be very much more alike, and much more likely to reveal whether or not we have a smoothly running process, subject to easy extrapolation.

*Prepared in part in connection with research at Princeton University sponsored by the Army Research Office (Durham).

V

METHODOLOGY OF PROGRAM EVALUATION

Chairman, FREDERICK MOSTELLER, Harvard University

Page

ASSESSING THE PROGRESS OF EDUCATION

Ralph W. Tyler, Center for Advanced Study in the Behavioral Sciences

Education today is of great concern to all Americans. Without education our young people cannot get jobs, are unable to participate intelligently and responsibly in civic and social life and fail to achieve individual self-realization. Education is increasingly recognized as the servant of all our purposes.

Because of its primary importance, our people are seeking information to guide their thinking and action in support of education. They are asking many questions, such as, are we making progress in raising educational levels? Are there areas or fields in which progress is lagging, where more support and effort should be focussed? Is progress more pronounced in certain sectors of the population such as urban, rural, central city, suburban, lower socio-economic levels, upper socio-economic levels? As schools increase their efforts to solve particular problems, questions will be raised about the progress thus achieved. As time goes on, school people and laymen alike will be seeking to understand more fully the relation between the various "inputs" into our schools and the progress of education.

The need for information of this sort by teachers, administrators, school boards, legislators, community leaders and the public generally, is a legitimate one. The great educational tasks we now face require many more resources than have thus far been available, and these resources must be wisely used to produce the maximum effect in extending educational opportunity and raising the level of education. To make these decisions, dependable information about the progress of education is essential, otherwise we scatter our efforts too widely and fail to achieve our goals. Yet, we do not now have the necessary comprehensive and dependable data. We have reports on numbers of schools, buildings, teachers and pupils and about the moneys expended. But we do not have sound and adequate information on educational results. Because dependable data are not available, personal views, distorted reports and journalistic impressions are the sources of public opinion and the schools are frequently attacked and less frequently defended without having necessary evidence to support either claim. This situation will be corrected only by a careful, consistent effort to obtain valid data to provide sound evidence about the progress of American education.

The need for data on progress has been recognized in other spheres of American life. During the depression, the lack of dependable information about the progress of the economy was a serious handicap in focussing efforts and in assessing them. Out of this need grew the index of production, the Gross National Product, which has been of great value in guiding economic development. Correspondingly, the Consumer Price Index was developed as a useful measure of the changes in cost of living and inflation. Mortality and morbidity indices are important bases for indicating needed public health measures. Facing the need for massive efforts to extend and improve education, the demand for valid information to support the requests and to guide the allocation of resources must be met.

In recognition of this need, Carnegie Corporation of New York, a private foundation, in 1964 appointed an Exploratory Committee on Assessing the Progress of Education. I was asked to serve as Chairman. Dr. Jack Merwin of the University of Minnesota is the Staff Director. The Committee's assignment is to confer with teachers, administrators, school board members and others concerned with education to get advice on the way in which such a project may be constructively helpful to the schools and avoid possible injuries. The Committee is also charged with the development and try-out of instruments and procedures for assessing the progress of education. The Committee has been working on these assignments for nearly two years. Recently, the Fund for the Advancement of Education joined in supporting the project.

The discussions with administrators, curriculum specialists, teachers and school board members clearly recommended that the initial assessment include more than the 3 R's and that it ultimately cover the range of important educational tasks of the modern school. In harmony with this suggestion, instruments are now being constructed by four leading test development agencies in the fields of reading and the language arts, science, mathematics, social studies, citizenship, fine arts and vocational education. In subsequent years, other important areas will be included.

Because the purpose of the assessment is to provide helpful information about the progress of education that can be understood and accepted by public-spirited lay citizens, some new procedures are being developed. In each field, scholars, teachers and curriculum specialists have formulated statements of the objectives which they believe faithfully reflect the contributions of that field and which the schools are seriously seeking to attain. For each of these major objectives, prototype exercises have been constructed which, in the opinion of scholars and teachers, give students an opportunity to demonstrate the behavior implied by that objective. These lists of objectives and prototype exercises which help to define them have been reviewed by a series of panels of public-spirited citizens living in various parts of the country in cities, towns and villages. Each panel spent two days reviewing the material and making a judgment about each objective in terms of the questions: "Is this something important for people to learn today? Is it something I would like to have my children learn?" This process resulted in very few revisions of the original listing of objectives. The procedure was designed to insure that every objective being assessed is: (1) considered important by scholars, (2) accepted as an educational task by the school, and (3) deemed desirable by leading lay citizens. This should help to eliminate the criticism frequently encountered by current tests in which some item is attacked by the scholar as representing shoddy scholarship or criticized by school people as something not in the curriculum or by prominent laymen as being unimportant or technical trivia.

A national assessment to identify kinds of progress being made in education, and problems and difficulties arising, will not be very meaningful unless separate measures are obtained for populations within the total country which vary among themselves and thus present different degrees and kinds of progress and different problems to be solved. The particular populations that need to be treated separately may change over the years ahead, but for some time age, sex, socio-economic status, geographic location and rural-urban-suburban differences will probably be significant. Hence, the present plan is to assess a probability sample for each of 192 populations defined by the following subdivisions: boys and girls, four geographic regions, four age groups (nine, thirteen, seventeen, and adult), three divisions by urban, suburban, rural classifications, and two socio-economic levels.

The fact that populations are to be assessed and not individuals makes it possible to extend the sampling of exercises far beyond that of an individual test in which each person takes all of it. It may be that a comprehensive assessment would require so many exercises that if it were to be taken by one person he would need ten hours or more to complete them. With a population sample, 20 persons, each spending 30 minutes, would together take all the exercises. In this case, a population of 10,000 persons would furnish a sample of 500 for each of the assessment exercises and no one would have given more than 30 minutes of his time. Assuming that an assessment would be made every 3 to 5 years, in order to ascertain the kinds of progress taking place, it is very unlikely that many of those individuals who participated in the earlier assessments would be involved in any of the subsequent ones. Hence, from the point of view of the child or adult, no serious demand would be made on his time. Furthermore, it is unlikely that the children taking the exercises in later years would be drawn from the same classrooms as the earlier ones. Therefore, the demands made upon a teacher in releasing a child for half-an-hour will be minimal. The assessment, though costly, should be feasible and involve little or no inconvenience to individuals or to schools.

Since the assessment does not require that all participants be in classes, the exercises to be used are not limited to the usual test items. Interviews and observational procedures are also to be employed to furnish information about interests, habits, and practices that have been learned. Because school objectives commonly include these areas, it is necessary to see that some assessment is made of the levels of attainment.

The assessment exercises will differ from current achievement tests in another important respect. An achievement test seeks to measure individual differences among pupils taking the test. Hence, the items of the test are concentrated on those which differentiate among the children. Exercises which all or nearly all can do, as well as those which only a very few can do, are eliminated because these do not give much discrimination. But, for the purposes of assessing the progress of education, we need to know what all or almost all of the children are learning and what the most advanced are learning, as well as what is being learned by the middle or "average" children. To construct exercises of this sort is a new venture for test constructors. Under the contract they are to develop exercises at each age level in which approximately one-third represent achievements characteristic of most of those at that age level, one-third represent achievements characteristic of about half of those at that age level, and one-third which represent the achievements characteristic of the most advanced, that is, the top ten percent, of that age level.

To summarize the educational attainments of these several populations it is not necessary to compute test scores. Instead, the following sorts of things would be reported:

> For the sample of seventeen-year-old boys of higher socio-economic status from rural and small town areas of the Midwest region, it was found that:
>
> 93% could read a typical newspaper paragraph like the following.
>
> 76% could write an acceptable letter ordering several items from a store like the following.
>
> 52% took a responsible part in working with other youth in the playground and community activities like the following.

24% had occupational skills required for initial employment.

It is anticipated that the assessment would be the responsibility of a commission of highly respected citizens. They and the commission staff would prepare reports of the findings of the assessment, much as we now obtain reports of the findings of the decennial census. These reports would be available to all people interested in education, providing them in this way with significant and helpful information on what has been learned by each of the 129 populations. In subsequent years, the progress made by each of these populations since the preceding assessment would also be reported.

The contractors expect to have the assessment exercises completed by the end of the summer or early fall. The official try-outs will then be held during the Fall and Winter so that the exercises can be revised and the completed assessment instruments presented to the Carnegie Corporation by late spring, 1967. This should make it possible for a commission to institute the first official assessment during the school year, 1967-68.

The technical advisory committee for the program has been of very great help in establishing guiding principles and in proposing solutions to the problems arising during the progress of the study. Its membership is Robert Abelson, Lee Cronbach, Lyle Jones and John Tukey, Chairman. The Committee is also responsible for the design of the official try-outs to assure adequate range in difficulty of the exercises and comparability among the various sections.

Since many of those whose educational achievements will be assessed are not available in schools, the Committee is seeking the help of leading survey research centers in conducting assessment interviews with adults, seventeen-year-olds who are out of school, and younger children who may not be available for testing in the schools. This should also prove to be an interesting pioneering task.

There has been a good deal of misunderstanding of this project because it has been confused with a national achievement testing program in which a pupil takes a complete test, his score is reported and the mean scores for classrooms, schools and school systems are made public. Many of the critics of the assessment project assume that it would be handled similarly and thus put undue pressure on pupils and teachers who wish to show up well on the tests.

Other critics view any national assessment with alarm because they have visions of a giant Federal government using national programs of any sort as a means for getting control of the local schools. These seem to be fears that have no direct relation to this project for the Committee's plans have been shaped by the advice and criticism given us by teachers, administrators, school board members, and public-spirited laymen. Through the various conferences and many meetings, the Committee has been able to identify concerns and problems that such an assessment must deal with. With the help of the counsel we have received, the plans are being drawn so as to carry the project through in a way that will not injure our schools but will provide greatly needed information. We believe that the assessment of the progress of education will make a modest but constructive contribution to the improvement of American education by providing data useful in guiding the efficient allocation of resources.

BOOBYTRAPS AND PITFALLS IN THE EVALUATION OF SOCIAL ACTION PROGRAMS

Peter H. Rossi
National Opinion Research Center, University of Chicago

I: Introduction:

If one were to measure success by the popularity of evaluation research, then empirical social research has certainly arrived. Perhaps, the best example of this popularity lies in the legislation authorizing the present War on Poverty in which the agencies involved are specifically directed to set aside funds for evaluation research. Other ameliorative programs may not give as much formal recognition to such activity, but nevertheless seek social researchers to add to their staffs for this purpose or attempt to get social research centers to provide evaluations of their programs.

There are other measures of success besides popularity. If one were to measure success by the proportion of evaluation researches which are conducted with powerful enough designs to render unequivocal evaluation statements, then empirical social research does not appear to be a smashing success. For a variety of reasons -- some substantive, others related to the present state of development of research methodology, and still others concerned with the "politics" of evaluation -- there are very few evaluation researches which have the elegance of design and clarity of execution which would achieve widespread admiration among social researchers.

The purpose of this paper is to explore some of the main reasons why evaluation research is hard to do well and to suggest some ways in which these difficulties can be overcome. Providing much of the materials on which this paper has been based have been the experiences with such research of the National Opinion Research Center over the past few years. However, I venture that the experiences of other research centers and of individual researchers has not been very different: At least my informal, but undoubtedly highly biased, survey would indicate strong similarities between our experiences and theirs.

In principle, the evaluation of action programs appears to be most appropriately undertaken through the use of experimental designs. All the elements which would strongly recommend such research designs are usually present: The program involved is something which is added to the ongoing social scene by purposive social action as opposed to events which are not under the control of some individual or agency. Because an action program is under someone's control, the construction of experimental and control groups is, in principle, possible. Furthermore, the program is usually not designed to cover an entire population, but only some portion of it so that some of a target population would not be covered, making it possible to think in terms of control groups. Thus, in principle, it is not difficult to design an extremely elegant program of experiments to evaluate the effectiveness of the usual action program. Controlled experiments, however, are not frequently used in evaluation research. For example, there is not a single evaluation research being carried out on the major programs of the War on Poverty which follows closely the model of the controlled experiment.

II: Action Programs and the Contemporary Scene:

There can be little doubt that the present historical period is one in which there is considerable groping for new and presumably more effective treatments for a variety of presumed ills. We have rediscovered the poor, suddenly become intensely aware that Negroes are an incredibly disadvantaged group, become worried over the plight of the aged, and concerned about a presumed wasteage of brainpower. We also have enough national income to allocate some part of our resources to new programs designed to correct some of the obvious faults in our society.

However, there is an ironic twist to developing a heavy conscience in this historical period. This is because we cannot ordinarily expect that the new treatments we can devise will produce massive results. It appears as if we are in much the same position in the treatment of diseases. The introduction of modern medicine and modern sanitation procedures into a country which has had neither can very dramatically reduce morbidity and mortality, as experiences in some of the emerging nations indicate. But, in the United States of today, each new gain in morbidity and mortality can be expected to be smaller and more difficult to achieve. Providing potable water is much easier to achieve, and more dramatic in its impact on morbidity and mortality, than any attempt we can make to lower the incidence of lung cancer, especially if we try it through lowering levels of smoking in individuals.

Similarly with respect to our social ills. Dramatic effects on illiteracy can be achieved by providing schools and teachers to all children: Achieving a universally high enough level of literacy and knowledge, so that everyone capable of learning can find a good spot in our modern labor force, is a lot more difficult. Hence, the more we have done in the past to lower unemployment rates, to provide social services, etc., the more difficult it is to add to the benefits derived from past programs by the addition of new ones. Partly, this is because we have acheived so much with the past programs and partly this is because the massive efforts of the past have not dealt with individual motivation as much as with benefits to aggregates of individuals.

In part, the concern of contemporary practitioners in the applied fields with evaluation arises out of their increased methodological sophistication. But, in even larger measure, it arises out of the expectation -- held at some level or other -- that massive effects are not to be expected from new programs and the new treatments aregoing to be increasingly expensive in terms of time and money. The problem of evaluation in this historical period is that the new treatments can be expected to yield marginal improvements over present treatments and that cost-to-benefit ratios can be expected to rise dramatically. Hence, there is considerable interest in research but considerable apprehension over what it will show concerning the effects of programs.

To illustrate, let us consider the case of Project Headstart: We have apparently wrung most of the benefits we can out of the traditional school system. Although everyone would agree that universal schooling for children up to approximately age sixteen has been a huge success, as opposed to a system of no schooling or of schooling mainly for those to pay for it themselves, there still remains considerable room for improvement, especially in the education of the poor and otherwise disadvantaged. A supplementary pre-school program bringing such children more into parity with those better off because of family background sounds like an excellent program. But, it is hardly likely to produce as much benefit as the introduction of universal elementary schooling did, especially since it is designed to do the job that a full-time institution, the family, neglected to do for one reason or another.

Effective new treatments which produce more than equivocal results are expensive. For example, each trainee at a Job Corps camp costs somewhere between five and ten thousand dollars a year (depending on which estimates you hear), as compared to considerably less than one thousand dollars per year in the usual public high school. Yet a year in a Job Corps Training Center is not going to be five to ten times more effective than a year in a public high school.

Paradoxically, the costs of evaluation are also expensive for these new programs. If effects can be expected to be small, greater precision is needed in research to demonstrate such effects unequivocally. This is another reason why I stressed the controlled experiment as the ideal evaluation research design: Its ability to detect effects is quite powerful compared to alternative methods.

Although as social scientists we can expect the new social programs to show marginal effects, the practitioner does not ordinarily share our pessimism -- at least, not when he faces the Congressional Appropriating Committee. Hence, the claims made in public for the programs are ordinarily pitched much higher, in terms of expectation of benefits, than we could realistically expect with the worst of research and much better than we could expect with the best of research. Thus it turns out that one of the major obstacles to evaluation research is the interests in the maintenance of a program held by its administrators. Their ambivalence is born of a two horned dilemma: On the one hand, research is needed to demonstrate that the program has an effect; on the other hand, research might find that effects are negligible or non-existent.

III: <u>Commitment to Evaluation:</u>

The will to believe that their programs are effective is understandably strong among the practitioners who administer them. After all, they are committing their energies, careers and ideologies to programs of action and it is difficult, under such circumstances, to take a tentative position concerning outcomes. Hence, most evaluation researches which are undertaken at the behest of the administrators of the programs involved are expected to come out with results indicating that the program is effective. As long as the results are positive (or at least not negative) relationships between practitioners and researchers are cordial and sometimes even effusively friendly. But, what happens when it comes out the other way?

A few years ago, the National Opinion Research Center undertook research with the best of sponsorships on the effect of fellowships and scholarships on graduate study in the arts and sciences fields. It was the sincere conviction, on the part of the learned societies which sponsored the research, that such fellowships and scholarships were an immense aid to graduate students in the pursuit of their studies and that heavily supported fields were thereby able to attract better students than fields which were not well supported. The results of the study were quite equivocal: First, it did not appear that financial support had much to do with selection of a field for graduate study. Secondly, it did not appear that graduate students of high quality were being held back from the completion of their graduate programs by the lack of fellowships or scholarships: Those who were committed found some way to get their Ph.D's, often relying on their spouses to make a capital investment in their graduate training. The equivocal nature of the results was quite disappointing to the sponsors whose first reaction was to question the adequacy of the study's methodology, leading to the coining of a National Opinion Research Center aphorism that the first defense of an outraged sponsor was methodological criticism. The findings affected policy not one whit: The sponsoring groups are still adamantly claiming more and more in the way of financial support for graduate students from the federal government on the grounds that such support materially affects the numbers of talented students who will go to graduate study beyond the B.A., and, furthermore, materially affects the distribution of talent among various fields of study.

Relations between the sponsoring learned societies and our researchers have been cool (if not distant) ever since. The learned societies believe their problem has been badly researched, and the researchers believe that their results have been badly ignored.

Sometimes both the researcher and the practitioner suffer from the will to believe leading to evaluation research containing the most lame sets of qualified results imaginable. Perhaps the best example can be gleaned from the long history of research on the effects of class size on learning. The earliest researches on this topic go back to the beginnings of empirical research in educational psychology and sociology in the early twenties. Since that time there is scarcely a year in which there has not been several dissertations and theses on this topic, not to mention larger researches done by more mature scholars. The researches have used a variety of designs ranging from the controlled experiment to correlational studies, the latest in the series being the results on this score obtained by James Coleman in his nationwide study of schools conducted for the Office of Education under the Civil Rights Act of 1964. The results of these studies are extremely easy to summarize: By and large, class size has no effect on the learning of students, with the possible exception of classes in the language arts. But, the net results of more than two hundred researches on educational ideology and policy has been virtually nil. Every proposal for the betterment of education calls for reductions in the size of classes, despite the fact that there is no evidence that class size affects anything except possibly the job satisfaction of teachers. Even the researchers in presenting their results tend to present them apologetically, indicating the ways in which defects in their research designs may have produced negative findings as artifacts.

In fact, I do not know of any action program that has been put out of business by evaluation research, unless evaluation itself was used as the hatchet to begin with. Why is this the case? Why do negative results have so little impact? The main reason lies in the fact that the practitioners, first of all (and sometimes the researchers), never seriously entertained in advance the possibility that results would come out negative or insignificant. Without committment to the bet, one or both of the gamblers usually welch.

The ways by which welching is accomplished are myriad. It is easy to attack the methodology of any study: Methodological unsophisticates suddenly become experts in sampling, questionnaire construction, experimental design, and statistical analysis, or borrow experts for the occasion. Further replication is called for. But, most often it is discovered that the goals of the program in terms of which it was evaluated are not the "real" goals after all. Thus, the important goals of school systems are not higher scores on multiple choice achievement tests, but better attitudes toward learning, a matter which the researcher neglected to evaluate. Or, the goals of a community organization in an urban renewal area were not really to affect the planning process but to produce a committment to the neighborhood on the part of its residents while the planning took place.

Perhaps the best example of how "real" goals are discovered after goals that were evaluated were found to be poorly attained can be found in the work of a very prominent school administration group. This group, fully committed to the educational modernities of the forties and fifties, found to its surprise that whether or not a school system adopted its programs had little to do with the learning that students achieved. Hence, they dropped achievement tests as a criterion of the goodness of a school or school system and substituted instead a measure of how flexible the administration was in adopting new ideas in curriculum, producing an evaluation instrument which, in effect, states that a school system is good to the extent that it adopts policies that were currently being advocated by the group in question.

IV: Assuring Positive Results:

Given unlimited resources, it is possible to make some sort of dent in almost any problem. Even the most sodden wretch on skid row can be brought to a semblance of respectability for some period of time (provided that he is not too physically deteriorated) by intense, and expensive, handling. But, to make an impact on the denizens of all the skid rows in all of our great cities requires methods that are not intensive and are not expensive case by case. There is not sufficient manpower or resources to lead each single skid row inhabitant back to respectability, if only for a short period.

Yet, many action programs, particularly of the "demonstration" variety, resemble the intensive treatment model. They are bound to produce results if only because they maximize the operation of the Hawthorne and Rosenthal effects, but cannot be put into large scale operation because either manpower or resources are not available. Hence, programs which work well on the initial run on a small scale with dedicated personnel can be expected to show more positive results than the production runs of such programs with personnel not as committed to the program in question.

The distinction I want to make in this connection is that between "impact" and "coverage." The impact of a technique may be said to be its ability to produce changes in each situation to which it is applied, while the coverage of a technique is its ability to be applied to a large number of cases. Thus, face-to-face persuasion is a technique which has high impact as a means of getting people to come in for physical examinations, but its coverage is relatively slight. In contrast, bus and subway posters may have low impact in the sense of producing a desired effect each time someone is exposed, but large coverage in the sense that many people can be exposed to bus and subway posters very easily.

An extremely effective technique for the amelioration of a social problem is one which has both high impact and high coverage. Perhaps the best example of such techniques can be found in medicine whose immunizing vaccines are inexpensive, easy to administer and very effective in reducing the incidence of certain diseases. It does not seem likely that we will find vaccines, or measures resembling them in impact and coverage, for the ills to which action programs in the social field are directed. It is more likely that we will have action programs which have either high impact or high coverage, but not both. The point I want to emphasize here is that it is a mistake to discard out of hand programs which have low impact but the potentiality of high coverage. Hence, programs which show small positive results on evaluation and which can be generalized to reach large numbers of people can, in the long run, have an extremely significant cumulative effect.

Examples of such programs in the social action field do not easily come to mind. But perhaps an illustration from the field of public health can be cited appropriately: Over the past few decades public health information specialists have been plagued by the fact that their most effective techniques have low coverage and their best mass techniques have little impact. Evaluation research after evaluation research has indicated that it is possible to raise the level of an individual's health knowledge and utilization of health facilities if you can get him to come to a course of lectures on the topic. In contrast, public health information campaigns utilizing the mass media have been shown to have minute effects. Yet, the information of the American population concerning health matters has appreciably increased over the past two decades. It is apparently the case that while no one campaign was particularly effective, their cumulative effects were considerable.

V: The Control Group Problem:

The key feature of the controlled experiment lies in the control exercised by the experimenter over the processes by which subjects are allocated to experimental and control groups. In a well-designed experiment, such allocations are made in an unbiased fashion. But, there are many ways in which a well thought out plan can go awry.

Perhaps the major obstacle to the use of controlled experiments in evaluation research is a political one. The political problem is simply that practitioners are extremely reluctant to allow the experimenters to exercise proper controls over the allocation of clients to experimental and control groups. For example, the proper evaluation of a manpower retraining program requires that potential trainees be separated into experimental and control groups with a contrast being made between the two groups at a later time. This obviously means that some potential clients, who are otherwise qualified, are barred arbitrarily from training - an act which public agencies are extremely reluctant to authorize.

In part, the political problem arises because researchers have not thought through sufficiently the problem of what constitutes a control or non-experimental experience. The logic of experimental design does not require that the experimental group not undergo some sort of treatment, it merely requires that the experimental group not be given the treatment which is being evaluated. In short, we have not been ingenious enough in inventing placebo treatments which are realistic enough to give the public official the feeling that he is not slighting some individuals at random. For example, a placebo treatment for a job retraining program may be conceived of as some treatment designed to help men get jobs but which does not involve retraining and, over which the training program should demonstrate some advantage. Perhaps testing and intensive counseling might be an acceptable placebo for a control group in an experimental evaluation of job training. Or, a placebo treatment for the evaluation of a community mental health center might be referrals to general practitioners for the kinds of treatment they either administer themselves or provide referrals to.

Even in the best circumstances and with the best of sponsors, the carrying out of controlled experiments can run into a number of boobytraps. There is, for example, the case of an evaluation research all set to go and well designed but whose program did not generate enough volunteers to fill up either the experimental or the control groups. Under these circumstances, the administrator opted to fill up the experimental groups abandoning all attempts at segregating the volunteers into experimental and control groups.

Or, there is the example of a well designed research on the effectiveness of certain means of reaching low income families with birth control information whose design was contaminated by the City Health Department setting up birth control clinics in areas which had been designated as controls!

Or, there is the risk that is run in long range experimental designs that the world may provide experiences to control, which would duplicate in some essential fashion, the experimental treatment. Thus, Wilner et al., in the evaluation of the effects of public housing unfortunately undertook their research in a period when the quality of the general housing stock in Baltimore was being improved at so fast a rate that the contrast in housing conditions between experimental and control groups had greatly diminished by the end of the observational period.

In sum, it is not easy either to obtain sufficient consent to undertake properly controlled experiments or to carry them out when such consent is obtained.

VI: A Strategy For Evaluation Research:

There are a number of lessons to be drawn from the various sections of this paper which hopefully could go some distance toward devising a strategy for the conduct of evaluation research. While it is true that in a Panglossian best of all possible worlds, the best of all possible research designs can be employed, in a compromised real world, full of evil as it is, it is necessary to make do with what is possible within the limits of time and resources. The problem that faces us then is how can we set up the conditions for doing as best a job we can and produce research which is as relevant as possible to the judgment of the effectiveness of social policy programs.

Although the idea of evaluation research has gained wide acceptance, we are a long way from a full commitment to the outcomes of evaluation research. It is part of the researcher's responsibility to bring to the practitioner's attention that in most cases the effects of action programs are slight and that there is more than an off-chance possibility that evaluation will produce non-positive results. The policy implications of such findings have to be worked out in advance; otherwise the conduct of evaluation research may turn out to be a fatuous exercise.

Secondly, we have a long way to go in devising ways of applying controlled experiments to problems of evaluation. Political obstacles to the use of controls often make it hard to get acceptance of such designs, and the difficulty of maintaining controls in a non-sterile world make full-fledged experimental designs relatively rare in use.

Earlier in this paper, I suggested that we take a lesson from medical research and search for the social analogues of placebos to be administrered to our control groups. There are other directions in which experimental designs should go: For example, considering the high likelihood that treatments have small effects, we need very powerful designs to demonstrate positive results. But because power costs money, it is worthwhile considering research designs which evaluate several types of experimental treatments simultaneously so that the outcomes will be more useful to the setting of program policy. To illustrate: it is considerably more worthwhile to have the results of an experimental evaluation which provides results on several types of Job Corps camps than on job corps camps in general. Looking at the differential effectiveness of several job corps camps provides more detailed and better information for the improvement of job corps programs than would a gross evaluation of the program all told.

This paper has stressed the model of the controlled experiment as the desired one for evaluation research. But, it is abundantly clear that for a variety of reasons, controlled experiments are rarely employed as evaluational devices and that they are difficult to employ. Most frequent are some sort of quasi-experiments in which the control groups are constructed by methods which allow some biases to operate and correlational designs in which persons subjected to some sort of treatment are contrasted with persons who have not been treated, controlling statistically for relevant characteristics.

The important question which faces the evaluation researchers is how bad are such "soft" evaluational techniques, particularly correlational designs? Under what circumstances can they be employed with some confidence in their outcomes?

First of all, it seems to me that when it is massive effects that are expected and desired, "soft" techniques are almost as good as subtle and precise ones. To illustrate, if what is desired as the outcome of a particular treatment is complete remission of all symptoms in each and every individual subject to treatment, then it is hardly necessary to have a control group. Thus if a birth control technique is to be judged effective if and only if it completely eliminates the chance of conception in an experimental group, then the research design is vastly simplified. The question is not whether those who use the method have less children than those who do not, but whether they have any children at all, a question which can be easily decided by administering the technique to a group and counting births (or conceptions) thereafter.

The obverse of the above also holds. If a treatment which is to be tested shows no effects using a soft method of evaluation, then it is highly unlikely that a very precise method of evaluation is going to show more than very slight effects. The existence of complex and large interaction effects which suppress large differences between a group subject to a treatment and statistical control groups seems highly unlikely. Thus if children participating in a Head Start program show no gain in learning ability compared to those who did not participate in the program, holding initial level of learning constant, then it is not likely that a controlled experiment in which children are randomly assigned to experimental and controlled groups is going to show dramatic effects from Head Start programs.

Of course, if a correlational design does show some program effects, then it is never clear whether selection biases or the program itself produce the effects shown.

This means that it is worthwhile to consider soft methods as the first stage in evaluation research, discarding treatments which show no effects and retaining those with opposite characteristics to be tested with more powerful designs of the controlled experimental kind.

Although ex post facto designs of a correlational variety have obvious holes in them through which may creep the most insidious of biases, such designs are extremely useful in the investigation of effects which are postulated to be the results of long acting treatments. Despite the fact that it is possible that cigarettes cause cancer, the evidence from ex post facto studies of the correlation between cigarette smoking and lung cancer can hardly be ignored, even though the evidence is not pure from the viewpoint of a purist. Similarly, NORC's study of the effects of Catholic education on adults, despite all our efforts to hold constant relevant factors, can still be easily produced by self selection biases that were too subtle for our blunt instruments to detect. We have nevertheless gained a great deal of knowledge concerning the order of effects that can be expected, were a controlled experiment extending over a generation conducted. The net differences between parochial school Catholics and public school Catholics are so slight that we now know that this institution is not very effective as a device for maintaining religiosity and that furthermore the effects we found are quite likely to have been generated by selection biases.

From these considerations a strategy for evaluation research is beginning to emerge. It seems to me to be useful to consider evaluation research in two stages--a Reconnaissance Phase in which the soft correlational designs are used to screen out those programs it is worthwhile to investigate further; and an Experimental Phase in which powerful controlled experimental designs are used to evaluate the differential effectiveness of a variety of programs which showed up as having sizable effects in the first phase.

AN ECONOMIC EVALUATION OF THE RETRAINING PROGRAM IN MICHIGAN: METHODOLOGICAL PROBLEMS OF RESEARCH

Einar Hardin and Michael E. Borus, Michigan State University

A research project on the economics of the retraining program is currently being conducted at Michigan State University under a contract with the U. S. Department of Labor. The project has three main aims. It seeks to estimate (1) the economic benefits and costs to society as a whole which result from occupationally oriented institutional retraining courses undertaken in Michigan under the Manpower Development and Training Act and under the Area Redevelopment Act, (2) the economic benefits and costs which accrue to persons enrolling in retraining courses and (3) the impact of retraining courses upon the tax revenues, expenditures, and transfer payments of the government. Special efforts are being made to estimate how the impact of retraining varies with the characteristics of the training courses, the occupations for which training was intended, the labor markets, and the persons enrolled.

Of the many methodological problems which have arisen in our study and which may be common to a variety of benefit-cost studies of human resource programs, we shall deal with four: (1) estimating the social product gains on the basis of gains in trainee earnings, (2) use of a control group design for estimating the impact of retraining upon earnings of trainees, (3) accessibility and quality of earnings data, and (4) measuring the marginal social cost of retraining courses.

Estimating the Social Product Gains

Society may undertake retraining activities with many economic aims in mind: to alleviate unemployment and low earnings, to alleviate shortages of particular goods, or to expand the overall output of the nation. These objectives may be ends in themselves, and they may be means to broader, less clearly economic goals. For the main analysis, we interpret social economic benefits to mean the contribution that retraining courses make to aggregate national output over and beyond the direct effect of increased government expenditures on the retraining activities. Subsidiary analyses consider the impact of retraining upon employment status, welfare assistance status, occupational level, and other aspects having possible interest to policymakers.

There seems to be no way of observing directly the contribution of retraining courses to national output. It must instead be assessed by inference. The generally accepted method of inference is based on microeconomic theory, in particular the marginal productivity principle, as applied to perfectly competitive markets. According to this theory, the rate of remuneration earned by a resource will equal the contribution to output which was made by the unit of that resource last added to production. It follows from this proposition that if retraining induces a person to transfer from one occupation to another, the change in aggregate output resulting from retraining will equal the difference between the person's earnings in the two occupations. Similarly, if retraining increases the person's efficiency in a particular occupation, aggregate output and trainee earnings will both increase and by approximately the same amount. If this line of reasoning is carried further, one could infer that the gains in aggregate output attributable to the whole retraining program are roughly equal to the sum of the gains in earnings of the trainees.

The equality of gains in output and earnings is strictly true only for marginal adjustments. If retraining courses were to affect a significant share of the labor force, a calculation based on discrete instead of marginal changes might become necessary. We do not think this is a major problem at the present time.

A much more important problem which must be overcome in measuring the gain in social output is that of non-market externalities. This problem has long been discussed. In the present context, the analysis should recognize that any reduction in unemployment, crime, and delinquency which might result from retraining would lower the demands placed on the public employment service and the welfare and social service agencies. The release of resources from these services to other uses in the economy will not be reflected in the wages that an employer pays a retrainee. No employer can be expected to count as a material benefit to himself the amount of resources which his action in hiring the unemployed saves society. As more becomes known about the effects of unemployment and low earnings upon the demand for social services, increasingly accurate adjustments for these external benefits should prove possible. At this time, however, benefit-cost analysis can at best make only tentative allowance for these factors.

A second type of analytical problem of even greater importance than that of non-market externalities occurs within the labor market. Retraining may enable an unemployed person to obtain a job which otherwise would have been filled by someone else. The trainee by securing a job increases his earnings. Although this increase equals the increment in output which society obtains by keeping the job filled instead of vacant, the job might not have remained vacant in the absence of retraining. Retraining would then merely cause one person to be displaced by another without adding to social product. In the course of time, the increased supply of persons available for a particular job would tend to reduce the relative wage rate for the job and would probably increase aggregate employment and output. But the extent and speed of this offset to the displacement effect are matters of faith rather than of knowledge.

Alternatively, retraining may enable a trainee to fill a vacancy which would otherwise have remained unfilled, and the trainee may have been recruited from an occupation where his former position was easily filled from among the unemployed. In this instance, the increment in social product attributable to retraining approximates the total earnings of the trainee in his new occupation. It is equivalent to the gain in earnings of the trainee plus the gain in earnings of the person who would have been unemployed but who instead filled the trainee's former position. Retraining then not only upgrades one person, the trainee, but also creates a vacuum into which an unemployed person may move. A vacuum effect will also occur when the labor shortage occupation for which training takes place has important complementarities with other occupations which are easily filled from among the unemployed.

The vacuum effect can also appear in an economy experiencing full employment. If retraining of an individual for a shortage occupation induces another person to move up into the trainee's former occupation, the aggregate benefits to society equal the sum of the gains in earnings of the trainee and his replacement. If filling the shortage creates a demand for manpower in complementary occupations and if these vacancies are also filled by transfers from lower paid occupations, the aggregate benefits to society are enlarged further, but this additional benefit is not reflected in the gains of the trainee.

Fortunately for benefit-cost analysis these are extreme cases. Displacement is not likely to be important, if full employment prevails or if retraining is directed towards shortage occupations, as required by the Manpower Development and Training Act. The vacuum effect will be limited, if the trainees are recruited directly from among the unemployed, who are given priority in selection for the training courses, or if they are recruited from other shortage occupations. These mixed tendencies perhaps reduce the average bias in estimating the social product gains from private earnings. They also leave an uncomfortably wide margin for interpretation and qualification of findings.

Control Group Design

There are also problems in measuring how the retraining program affects the private earnings of the trainees themselves. To measure the private benefits of trainees, one must estimate how they would have fared in the labor market had they not become enrolled. A simple before-and-after comparison of the earnings of the trainees is not useful, because the earnings of a particular group of persons normally change over the course of time in response to cyclical and seasonal variation and to continuing displacement, attrition, and replacement hiring. Therefore, it is necessary to set up a control group of non-trainees for each course, who are as nearly similar to the trainees in their qualifications for retraining as it is possible to find, and to compare the earnings of the two groups for the same period. Such a control group might be composed of persons who have been considered for enrollment in the course, who possess the basic qualifications for enrolling, who have expressed an interest in enrolling, but who in fact have not enrolled.

In addition, it is necessary to obtain work history and other information for the nontrainees as well as for the trainees. By using personal background data from the period before training as independent variables in the regression analyses, one can hope to eliminate statistically the major remaining differences between trainees and nontrainees as to initial skills, motivation, work habits, appearance, and other marketability factors.

Although the control group design might hold constant the basic marketability of individuals, it does not fully eliminate the danger of self-selection particularly when the earnings are measured for a rather short period of time after the end of the course. Fortuitous events, not predictable from basic marketability, in part determine who in a given group will get the job offers made at a particular time. If recipients of job offers abandon earlier decisions to enroll and instead accept the jobs, they will have earnings during at least part of the time the training course is in progress. Since these job opportunities are not necessarily available to the trainees, the earnings of nontrainees during the period of the course will tend to overstate the earnings foregone by the trainees while they are in the course. If the jobs accepted by the nontrainees last into the period after the end of the course or if they enhance their subsequent job opportunities, their earnings in the period after the training course will also be higher than the trainees could have expected to obtain, had they not enrolled. The observed difference between trainee and nontrainee earnings in the period after the end of the course will similarly understate the earnings gain resulting from the training course. The net effect will be an understatement of the rate of return from retraining courses.

In our study, we found a number of persons who indicated that they turned down the opportunity to enroll in retraining courses because they had found jobs or had received promises of jobs. We could not determine, however, if the trainees had similar job opportunities available at the time of enrollment. Therefore, we had to rely on two compromise solutions. One was to distinguish analytically between three groups of respondents: trainees, nontrainees with jobs or job promises as reasons for nonenrollment, and nontrainees for other reasons. The second solution was to classify the respondents into four groups on the basis of whether they were trainees or nontrainees and whether they did or did not have jobs when the courses started. Both solutions might help in constructing upper and lower bounds to the estimates of earnings gain and opportunity cost.

A decision to use control groups of qualified and interested nontrainees naturally implies that courses for which comparable nontrainees cannot be found must be excluded from the study. If these courses differ in effectiveness from retraining courses for which comparable nontrainees can be identified, the estimates of gains in earnings will be biased. In our study, we found that in a few instances the local employment service had been unable to recruit enough qualified and interested persons even to provide the enrollment planned. More commonly, it chose to recruit just enough persons to have one or two alternates to fill course vacancies resulting from early dropouts. In some courses, no records had been set up for persons who had been reviewed as potential enrollees, or the records had been discarded. Since the lack of control groups resulted from a variety of circumstances, only some of which could be expected to relate to the social or private profitability of the course, it probably did not markedly bias the selection of courses in our study.

Quality of Earnings Data

In order to estimate the gain in national product one must utilize data on earned income for trainees and for nontrainees during at least two periods of time: a time period after the course to estimate the future earnings gain, and the period of the course to estimate the earnings foregone by trainees while they are in training. In addition, as discussed earlier, it is desirable to obtain data on earnings for a period before the course, in order to construct a rough index of the persons' marketability. Interviews may be used for collecting this information and may also be employed to collect detailed personal data and other important facts, but they pose several problems: the time and expense of locating and interviewing named individuals at scattered locations on the basis of old addresses, the bias in the results which accrues from the inability to locate highly mobile or otherwise distinctive persons, the uneven quality of interview data on earnings, and the possible correlation between reporting errors and the main experimental variables.

In our study we were faced with each of these problems. Obtaining the data on earnings proved to be a most difficult and time-consuming task. Of the slightly more than 1,000 persons whom we sought, we ultimately succeeded in interviewing about 830 persons. A number of persons who had migrated to neighboring areas of Wisconsin, Illinois, Indiana, and Ohio were interviewed, but interstate migrants with their potentially different changes in earnings were probably underrepresented. Very few persons refused to be interviewed, but the answers of some persons were too spotty or too inconsistent to be useful.

In the interviews we sought information on work history, earnings, unemployment experience, prior education, and various personal characteristics. The interviewing was scheduled to ensure that work histories would cover a period of at least 365 days after the end of the course. In addition, the interview covered the period the course was in progress and several years prior to the course. Wide variations in quality were apparent within the 800 interviews that we are currently using in our analysis, even though the interviews were made by trained interviewers according to a detailed and fixed schedule. Some respondents kept excellent records, offered pay check stubs and copies of income tax returns in evidence, and criticized us for not asking them beforehand to have these records available. Others lacked records or resented suggestions that records be consulted. It early became clear that fringe benefit data would be very poor. Our data consequently included earnings before taxes and transfer payments, not total compensation for work performed, as would be more appropriate to a social economic benefit-cost analysis.

Our experience with the interview data led us to ask whether the response errors were related systematically to the main variable in the analysis, that is, to the person's status as trainee or nontrainee, or perhaps to variables with which the main variable might interact in its effect on the dependent variable, that is, private earnings. We, therefore, made a comparison of earnings data collected in interviews and earnings data obtained from employer reports that were kept for purposes of unemployment insurance. These data were gathered in a study of retraining in Connecticut which resembled the present study in interview method, items covered, and types of respondents interviewed.

The two bodies of data appeared to agree very closely despite minor discrepancies in coverage and definition: the difference in mean weekly earnings between the two groups was $3.39 or about 5 per cent of the mean of earnings according to the interviews, and the simple product-moment correlation between the two sources was r=0.95. An analysis of the individual discrepancies showed, however, that these were significantly related to several personal data variables including age, sex, and education, all of which might be important in analyses of the economic benefits of retraining. Particularly relevant was the finding that the discrepancy between employer reports and survey data on weekly earnings was not the same for course dropouts as for graduates and nontrainees. This finding suggested that the true earnings advantage of the course graduates over the dropouts might have been as much as one-third greater than what was estimated from the survey data alone. Analysis of the economic benefits obtainable by reductions in the dropout rate would be seriously affected by errors of this magnitude, and even the estimates of the overall benefit-cost ratios would be weakened.

Measuring the Marginal Social Cost of Retraining

The direct social economic costs with which an analysis should compare the social economic benefits are composed essentially of two elements: the output foregone by society while the trainees are being retrained and the output foregone by society because of the diversion of resources to the retraining activity. The first element can be estimated from differences in private earnings during the period of the course, but with the weaknesses already discussed. The second element requires data of a different sort: the direct instructional costs incurred in the training facilities; the costs incurred in the activities of the public employment service for recruiting and selecting the trainees, administering the trainee allowance program, placing the trainees, and following up on their progress; the diffused costs of developing the training courses; and the costs of state and federal administration of the nation's manpower training program. The resources used in these activities will presumably be employed in other uses, if retraining does not exist. The value of the resources are then an estimate of the value of output foregone by their diversion to the retraining activity.

There are many problems in estimating the true value of the resources devoted to the retraining program. The available records for the instructional costs show as part of the costs of the training course the entire purchase price of of the equipment acquired or plant remodeling made for the sake of the course. This overstates the training costs, unless the capital assets have no further use and lack a significant salvage value. At the same time, facilities existing in the public vocational schools are used for some courses without any charge at all, which contributes to an understatement of training costs, while rented private facilities are used for some other courses with the rentals properly forming part of the current costs.

There are also problems with the accounts kept by the state employment service which is responsible for all of the local labor market aspects of the retraining program. The cost concept appropriate to the analysis of retraining is the additional cost incurred over and beyond what would have been incurred in the absence of retraining. At least some of the persons who are considered for retraining would receive services from the public employment service even if no retraining courses exist, and the cost for these services should properly be deducted from the ostensible costs of retraining. The proper size of the deduction is difficult to calculate, however, for it depends not only on the costs of normal services to an individual, but also on the impact that retraining has upon the subsequent demand from trainees and other persons for assistance from the public employment service, and this impact depends in turn on the extent to which trainees become securely employed in shortage occupations or merely displace others from surplus occupations. Sensitive benefit-cost analyses furthermore require an analysis of how the incremental costs of handling the retraining courses vary with such factors as course duration and the characteristics of prospective trainees, but no data adequate to this purpose are available.

The administrative costs of the overall program should also correspond to incremental costs, and the marginal costs of additional courses should be estimated. None of the detail required for this work can be obtained, and prorating the administrative costs in accordance with instructional costs seems the most acceptable compromise.

Conclusions

An understanding of how well the individual fares from retraining is an important element in judging the performance of the retraining effort. Many of the difficulties that we have described do not hamper the analysis of private benefits and costs, that is, the economic consequences for the individual person who contemplates being retrained. Except for the limited possibility that the post-training work experiences of trainees and control group members are interdependent, the effects of retraining upon other labor market participants are not relevant to the calculation of private benefits and costs and can be disregarded.

But what recommendations can be made about the analysis of social economic benefits and costs? Four main points may be made.

First, a benefit-cost analysis that infers social product gain from the differences in earnings between trainees and nontrainees tends to overstate the social economic benefits in periods of large, general unemployment, when retraining primarily redistributes jobs in favor of the trainees. It similarly tends to underestimate the benefits when there are pronounced labor shortages in particular occupations and the vacuum effect is likely to be important. Such situations may arise either in an economy of overfull employment without significant depressed areas or in a structurally maladjusted economy in which shortages and surpluses of manpower coexist. The calculations are most likely to be correct under conditions of balanced full employment if wages and prices are flexible. Since the passage of time tends to permit adjustment to new supply and demand conditions, it may be better to avoid making a benefit-cost analysis on the immediate effects of retraining, such as might be apparent in the first six to twelve months after the course, and to wait until the trainees have accumulated several years of labor market experience.

Second, if an analysis is to be made of the immediate effects, more weight should be given to results based on labor markets and time periods where unemployment is fairly low and evenly distributed than where a general recession or severe structural unemployment is evident. Efforts should also be made to refine the analysis by taking into account the shortages and surpluses existing in particular occupations in the period under observation. Short-run analyses might, perhaps, be corroborated in studies of employer adjustments to the filling and vacating of positions by trainees.

Third, the interview method of collecting data on private earnings is inaccurate and expensive, and it gives incomplete results. It should be replaced with a better method. This becomes increasingly urgent as the studies focus on longer-run effects or on programs in which geographic mobility may be a significant factor. The records of the Social Security Administration form a major useful body on the earnings of individuals. The limit on taxable reported earnings reduces the usefulness of these records when the analysis is to cover persons with medium or high earnings or when the limit has long remained unchanged despite a general growth in earnings. These difficulties are much less pronounced in analyses of human resource programs intended for low income groups. Another problem in using Social Security data results from the natural desire to prevent improper disclosure and use of information about individual persons. It should be possible, however, to find ways of enabling the analyst to use individual Social Security earnings data without actually disclosing the individual information to him. Having the statistical analysis performed in the agency is one possible solution. In another solution, which puts smaller burdens on the agency and which is probably also less costly and time-consuming to the analyst, the SSA would compute a sum of squares and cross products matrix based both on agency data and data supplied by the analyst. This would permit the analyst to use his own computational facilities and to explore the material in greater detail.

Fourth, the measurement of social economic cost, except for earnings lost while the person is in training, would benefit from a detailed in-house study of many aspects of costs. These should include the capital costs of instruction, the operating costs of the public employment service, and the dependence of overall administrative costs upon the number and nature of retraining courses undertaken.

VI

PREPARATION OF HOUSING PROGRAMS IN DEVELOPING COUNTRIES

Chairman, JAMES A. MOORE, Department of Housing and Urban Development

FINANCING PROBLEMS OF HOUSING PROGRAMS IN DEVELOPING COUNTRIES

Stanley E. Smigel, Department of Housing & Urban Development

I. Introduction

Before considering the financing problems of housing programs in developing countries, certain general points or conditions need to be enunciated. These points will either indicate the limits of, or add perspective to, the discussion which follows.

To begin with, references to housing will be to urban housing. In 95 out of 100 cases today the factors and situations in urban housing are so greatly different from those in rural housing that they are for all practical purposes two different subjects. It is conceivable that developments might alter this picture in the future. But for the present rural and urban housing are not only of different color, they are different animals. Except where noted, this paper deals only with urban housing.

The second condition relates to the content of the term "housing" as it should be viewed in the urban context and particularly in terms of the financial problems it poses. The strictest, most limiting definition of "housing" or "housing cost" includes three elements: (a) the structure itself, (b) the plot of ground on which the structure stands, and (c) the utility elements, i.e., water pipe, sewer pipe, etc., that are placed in or upon that plot. From the standpoint of the homeowner or the tenant this is, of course, the biggest part of his cost. But the street in front of his residence, the highway to which it leads and a number of other public or community services are frequently incorporated either into his housing cost or his taxes. Some of these charges or taxes are for repayment of capital investment, others are for maintenance and operation. Certain of these services or facilities are essential to urban housing, whether it is in an advanced or a developing country. The sewer, water and road networks, for example. Sometimes these investments or expenses, or a part of them, are not charged to the user or one who benefits but are met out of general tax revenues. In any event, these community or neighborhood facilities are essential to the individual housing units, and their capital costs are treated here as one part of the housing finance problem of developing countries. This is an increasingly common practice.

A wider perspective and appreciation of the content of the term "housing" and, consequently, the companion terms "housing cost" and "housing finance", can be gained by momentary reflection on the industry side of the housing process. It is easy to concentrate on the financing problems of the home-buyer or of the investor in rental housing, which, of course, are crucial. But the home-builder (the production side of the process) is also in need of financial assistance, particularly if he is to build housing units on a scale which will produce lower prices and provide the needed quantities. As contrasted to the financial needs of home-buyers, this is a short-term money need. Nevertheless, in the housing process it is an important requisite, and no long-term housing program can be successful if it does not meet this demand. In developed countries, it is not uncommon for the construction financing for the home-builder and mortgage financing for the home-buyer to be part of the same financing package.

Another activity that frequently poses housing finance problems is the preparation of land. This generally includes land purchase, grading, survey and plotting, placement of local roads, sewer and water system, etc. In some cases the financing of this activity is merged with construction financing, particularly in very large developments. And a final financial problem for housing is the capital cost of developing, expanding or improving building materials industries.

Although the treatment here will go beyond the financing problems of the home-buyer or the investor in rental housing, it cannot and does not pretend to cover all of these financing problems cited. Its principal preoccupation will be with long-term financing for home-buyers and, to a lesser extent, for investors in rental dwelling units. This is the supreme problem. Were it not for this aspect, the housing market would not be different from dozens of other consumer-good markets. Because of this aspect, the housing market is singular and incredibly more complex. The urbanite seeking to buy his own home has a market position on the demand side that is somewhat analogous to the supply position held by the small, independent farmer seeking to sell his crop in a free market. It is not at all strange, therefore, that governments all over the world have intervened in these two broad areas to provide a measure of balance and economic justice that pure market forces have not been able to effect. In the case of the housing market, it is noteworthy that in market economies government intervention has been very largely in the specific area of finance.

This paper does not have as one of its objectives the computation of an estimated dollar expenditure which over a period of ten, twenty or thirty years would put the world population into good or decent housing. In most respects this would be a fruitless exercise. Nevertheless, some idea of the dimension and the potential burden of housing finance to developing countries is helpful for perspective. There are considerable reliable data available from advanced countries that are useful for this purpose.

In the United States in 1963, Gross National Product (GNP) totalled $583.9 billion.(1) Residential non-farm construction - or urban housing construction - in that year was $32.6 billion or 5.6 percent of GNP. Of gross private fixed domestic investment in 1963, residential non-farm construction represented almost 31 percent of the total.(2)

*The views expressed by the author are his own and do not necessarily represent the views of the Department.

Based on annual investment, housing is the most important industry in the United States. According to national wealth data compiled by Raymond Goldsmith, housing represents more than 25 per cent of the national wealth and is the largest single component.(3)

In Europe, particularly Western Europe, housing's relationships and importance are closely analogous to that of the United States. In the period 1954-1956, for most West European countries, housing represented 4.5 to 5.8 percent of GNP. Investment in housing as a percentage of gross domestic fixed investment ranged generally between 19 and 25 percent.(4) In the USSR during this same period, housing investment as a percentage of gross fixed investment was reported in excess of 20 percent.(5) By any standards, these economic data for the United States and Europe are impressive. In light of the economic and financial significance of the housing sector, it is not surprising that the United States and other advanced countries have in the past used home-building as a counter-weight to recession-inducing forces.

As the foregoing indicates, the proportions of annual gross domestic investment that go into housing are large. The place of housing in the GNP is also important. For developing countries that are desperately short of capital, these proportions appear very high.(6) Yet a United Nations ad hoc group of housing experts representing every part of the globe has suggested that the underdeveloped areas of the world require rates of house-building substantially in excess of the rates maintained in Western Europe and shown in Table I.(7) An annual dwelling unit rate of more than 10 units per 1,000 inhabitants is considered an exceptionally high rate of housing construction. In Table I, covering selected Western European countries, only Western Germany, which is a special case, shows rates of over 10, their highest rate being 11.0. All other countries are substantially below this rate. The UN expert group, however, estimated that urban housing backlogs along with annual incremental needs would require rates of 13 and 13 to 16 dwelling units per 1,000 for Asia and Latin America respectively, and this for a period of thirty years. The efforts and the investments implied in these estimates for developing countries are staggering and under past and existing circumstances appear also to be impossible. The expert report itself concludes that "It is evident that, in order to sustain an adequate rate, housing construction would absorb practically the entire investment resources of developing countries."(8)

II. Domestic Considerations and Possibilities

It is customary to talk about "developing countries" as though they were a more or less homogeneous group. But in fact their only common characteristic is that they are all considered to have low levels of real income per capita relative to the standards of North America, Europe, Japan and Australasia. Within the group of developing countries differences are sometimes greater than between some advanced countries and some developing countries. Walter Rostow through his book, The Stages of Economic Growth, has helped put a proper focus on these differences. A 1957 UN publication which provides per capita net national product data for 55 countries adds additional concrete emphasis to this point. The developing countries included in this study had a per capita annual net national product (NNP) ranging from \$50 to \$540. Several of these developing countries had per capita NNP as high or higher than 5 European countries.(9) This circumstance, which in some albeit different measure persists today, tends to suggest that the definition of development level should include other factors, eg, levels of education, political and organizational stability and the like. For our purposes it is enough that those differences are appreciated and that it is understood that the considerations and possibilities being discussed will have uneven applicability.

The very large shortages of urban housing and related facilities that have been referred to above pose very important policy questions for governments of developing countries. Each government must decide first what role it should or must play. In many cases the social and political pressures have helped to formulate this decision and have put governments into a key formal role. In other cases, governments have resisted serious involvement in housing. The following, an official statement of the Government of India, is one example:

> "The reason why the provision is inadequate for housing and will continue to be so for quite some time is clear. Government will have to concentrate first on improving the standard of living of the people so that they can, amongst other things, pay if not the whole of the economic rent of a dwelling, at least a good portion of that rent. Raising the level of living means increasing agricultural production; exploring and exploiting the mineral wealth; promoting basic industries ... and developing the irrigation, power and transport potential of the country. These and other developmental activities thus come to get a higher priority in national planning."(10)

But though a government of a developing country can put off or delay serious involvement and a key role in housing finance and production, for most, eventual involvement and a key role seem inevitable.

The point has already been made that in the average market place housing does not fare well. Some of the factors responsible for this state of affairs have been mentioned. Additional information and discussion on this aspect is available in a number of good references.(11) It is appropriate, nevertheless, to mention briefly the most important single causal factor, namely rapid urbanization. The extraordinarily fast growth of cities in developing countries gives them problems of scale and urgency that did not face advanced countries in their take-off and maturity stages of development in the last century and the first couple decades of this century. For an excellent perspective in brief compass on the 20th Century phenomenon of rapid urbanization and its impact on developing countries, the reader is referred to the report of the Committee on Urban Development at a White House Conference on International Cooperation held late last year.[12]

Once a government of a developing country accepts the burden of a key role in housing it must then face additional important policy questions. It must consider the ingredients that are required - land, labor and capital - and must consider how much of each of these productive factors should be allocated in the short term and the long term to the task of supplying housing and essential related facilities. It must also consider the process or processes whereby housing is produced and take steps to strengthen and promote the efficiency of these processes, including the provision of needed institutions, legal framework, etc. In the consideration of housing policy decisions, and this applies strongly to the housing finance area, developing countries should relate housing and proposed policies and programs stemming therefrom to other policies and programs and to overall national objectives.

The first critical policy question concerns the allocation of resources. Of the total investment capital available for a given period, how much should be devoted to housing and related facilities? It soon becomes clear that no adequate answer is possible to this key question until a large number of complex aspects or sub-questions are examined, relationships determined and decisions on these are reached. Private as well as public resources must be considered. Existing programs and existing resources currently employed in this field must be examined from the standpoint of possible appropriate changes or of making resource use more efficient and productive. Are there resource possibilities or potential which are not being exploited? To what extent should these resource possibilities or potential be utilized for housing? What are the institutions or programs necessary to improve productivity or to exploit unrealized potential? Should resources be transferred to housing from other areas, or should increases in resources be earmarked for future housing use? These and many more knotty questions require answers. If they are not specifically examined and answered, they will nevertheless be answered implicitly or by default through the operation of the programs which implement adopted policy.

In consideration of an appropriate national government policy and role in housing, governments of developing countries quickly reach the question of national budget: How much of national revenues can or should be expended in programs for housing and related facilities? Uppermost in the minds of policy-makers, of course, are the shortages of available capital and the existing large demands for capital in other fields, some of which may not be wholly met. As the central agent in the national development program, the national government faces, consequently, weighty decisions and pressures. Governments of developing countries, moreover, face a disadvantage relative to advanced countries, in terms of their share of national income. Governments of richer countries generally secure a significantly larger proportion of annual national income than do the governments of less industrialized nations. Advanced nations not only have a larger per capita income, but their governments also capture a larger proportion of this income.[14]

Commitment of large amounts of public funds to housing by developing countries should not be done certainly without careful study of the experiences of other nations. Expenditures by governments for construction of new housing, and, where this housing is government operated, for the maintenance of an ever growing stock can be a large and serious burden on government budgets. How burdensome these expenditures can become may perhaps be best depicted by the experience of Eastern European countries. In the period after World War II, the governments in these states quickly took upon themselves the major burden of providing housing (confined initially almost wholly, and still very largely, to urban housing). The housing produced was rental housing. In accordance with what was described as Marxist principle the rents charged were not economic rents. In the USSR, for example, rents currently cover only about one-third of maintenance costs and provide nothing whatsoever for depreciation and thus recovery of the initial investment.[15] But the very heavy burden of new housing construction plus ever increasing maintenance costs has forced changes on several Eastern European countries and further changes are anticipated. Since 1959, the policy in Yugoslavia, represented by several increases in rental charges, has been to get closer and closer to an economic rent. Czechoslovakia and Poland followed this lead in subsequent years, raising rent payments. New policies have been instituted in virtually all Eastern European countries (USSR included), permitting greater private financing of housing, particularly through the promotion of cooperatives with the State providing land, loans and other assistance. Some idea of the importance to individual states of these measures to reduce the size of the housing finance burden may be gathered by reference to the situation in Poland (which, however, may be extreme). For new flats (apartments) the cooperative occupant in Poland pays six times as much as the renter of a new flat built by a peoples' council.[16]

Having touched upon certain considerations which weigh upon the possible decision of a developing country to use significant public funds for housing, the next problem that emerges is: "In what way should these funds be used?"

Direct government construction of houses, whether by private contractors or by the government's own work force, is a program used by countries in various parts of the world. This provides the greatest measure of control over the rate, timing and location of the units so constructed, over the selection of housing occupants and of the type of housing provided. Another technique that can be used is government lending. This also permits a large measure of control. A third technique for government funding of housing is one that has been developed largely by developing countries themselves and has been used with varying success by countries in Africa, Asia and Latin America. This technique has been called "land and utilities" or "site and services." Under this scheme, the government makes building plots available in surveyed and plotted, new or recently cleared areas. Utilities have also been made available, along with streets and pedestrian-ways. Generally the building plots are small and the services provided are minimal. Water taps, for example, may be at street-side, and one tap may serve four to eight dwelling units. Sometimes the service provision may be on a more ambitious scale.

This program requires the individual family to take responsibility for construction of the shelter. It is used most frequently as a means of providing better services and an organized neighborhood to squatter families. Where the government already owns the land, the costs entailed are survey, grading and provision of utilities and streets. When accompanied by some control of the structures built on the plots provided (and perhaps through the provision of some building materials at cost to the building families) this scheme has produced improved - frequently adequate -housing at low per unit costs to the government.

In each of the three programs briefly mentioned above, it is possible for the government to recover its expenditures. The housing directly constructed by the government could be rented at economic rents or be sold at cost. Similar conditions could be applied in the case of government lending. And the site and services plots could be transferred to families under conditions requiring payments to the government over a period of time which would offset the government expenditures incurred. As a matter of practice, however, this seldom occurs. Very frequently programs of the type described are launched on a recognized subsidy basis. Rents may be less than economic rents; homes may be sold to buyers at prices which do not include utility installation costs or land costs; loans may be made at interest rates significantly lower than interest rates paid by the government for its borrowing. In addition to these accepted subsidies, substantial implicit or concealed subsidies may be encountered. Rent collections may be poor and political pressures may make insistent collection or eviction unpalatable. Monthly payments of home-buyers may be forgiven for similar reasons. Maintenance and operating costs of rental units may soar due to unsophisticated tenants and poor administration. Developments as these may significantly increase the size of actual subsidies, and the burden on the government exchequer can balloon in magnitude.

For these and for other reasons, direct government funding of housing should have limited attractions to developing countries. Only the site and services program appears to have advantages which are not offered by alternatives. Government direct construction and direct lending for housing would appear to be best confined to special cases, or as pilot or demonstration efforts. The broadly fruitful area for government subsidy and financial assistance appears to be in another direction, in the stimulation and support of private efforts.

Reliance on private enterprise and private financing in developing countries may appear to be risky indeed, supporting one's self on a weak reed. And to be sure, as has been pointed out, the private market place has been unkind to housing. However, a partnership of government and private resources is a different matter. Herein lie many possibilities, some only fractionally exploited, others still to be discovered.

Private financing (as well as public financing) rests basically upon per capita income. Mention has already been made, in another context, that in the mid 1950's per capita annual income for a representative group of developing countries ranged from \$50 to \$540. Somewhat later estimates (1955-57) extend the upper limit of this range to \$700 per capita per annum.(17) But only 9 of the 100 countries and dependent areas listed in this latter source had estimated per capita annual income of over \$300. And a full 75 of the countries and territories had per capita annual income under \$200.

These data, in many cases admittedly crude estimates, are nevertheless derived by responsible people in as responsible a fashion as the limited and faulty data on many developing countries permit. At first blush the data appear to make the task of financing housing in these countries all but hopeless. But, as any statistician or economist worth his salt knows, aggregate data have their uses, these uses are quite limited, and there is an unfortunate tendency, particularly by the statistically untutored, to draw conclusions from aggregate data that are unwarranted. The particular point appropriate here is that these per capita national income data are deceptive and not necessarily applicable to housing.

How applicable are these per capita income data, for example, to the millions of rural families over the world who for centuries have constructed their own homes and still do today? The house in such case is the product entirely of local materials and family labor, and neither in whole nor in part does the dwelling unit enter the market place. Since the emphasis and concern of this paper is on urban housing, however, it is the considerations applying to urban housing that merit attention.

The first and obvious point is that national per capita income data include both rural and urban income. Productivity and income of the agricultural segments of developing countries are far behind that of urban areas. Income in urban areas is frequently double (or more) that in rural areas, and in the large metropolitan areas average per capita income is frequently three, four or five times average per capita rural income. In Peru, for example, overall annual per capita income for 1961 was estimated at $123. In the Sierra, however, in which area a large part of the population live, average per capita income was estimated at $48; the coastal region (including Lima) had an estimated per capita income of $244; and Lima itself had an estimated per capita income of $400 to $500.(18)

A second important factor to consider with respect to income is that the social customs and practices make family income a much more meaningful and logical statistic than per capita income. In many developing countries it is the rule in urban areas for the lower and middle income families to have more than one gainfully employed person.(19) The capacity to pay rents or to purchase homes is obviously much greater when viewed from the standpoint of family income than from the standpoint of per capita income.

A third pertinent factor bearing on income statistics in developing countries is the question of distribution. It is widely recognized that the distribution of national income in developing countries is quite sharply skewed relative to income distribution in advanced countries. Over time it has become popular to talk about three income groups: a very small and very affluent group at the top, a somewhat larger but still quite small middle income (middle class) group, and a third, huge group comprising low-income families. In the minds of many this bottom group, representing as many as 80 to 90 per cent of all families, is a mass of destitute, poverty-stricken families who possess little or nothing in the way of capital resources or of earning power in excess of subsistence needs. This view is just plain wrong and does not fit most developing countries. In most cases the lower income group shows a wide range of incomes and a surprisingly large proportion whose earnings must be considered significant and clearly above subsistence levels. The size of the error in this conventional wisdom is considerably greater when only the urban population and their incomes are considered. Annual per capita income for Ethiopia, for example, has been estimated at less than $50. A carefully derived distribution of family income for Addis Ababa, however, reveals a quite regular curve, albeit skewed. A full 35 per cent of the families were included in a group with incomes ranging from zero to $240 per annum. But fully 50 per cent of all families were in groups having annual incomes of $720 or more.(20) Family income data for squatter settlements in Peru provide additional evidence that the lower income group incorporates wide income differences. Squatters in Peru are very probably the lowest urban income group in the country. Yet according to 1962 data collected on selected squatter settlements containing over 35,000 people, the modal income group had average family income of about $440. Moreover, almost 50 per cent of the families enjoyed family incomes of $600 or more per annum.(21) Income data for Nicaragua provide a final example that, for meaningful analysis, it must be recognized that significant income differences exist within the lower income group. The survey data for Managua, the capital, are reported in Table II. About 70 per cent of Managua's population are covered. This very large "sample" covered occupants in low and middle class housing. It discloses that the modal group, 23 per cent of the total, had an annual family income of about $1100. At the top of the distribution, 7 per cent of the families enjoyed average annual income of over $4,000 per annum.(22)

Before leaving the subject of income distribution and what are meaningful income statistics, a conjecture or alert is in order. It is suggested that more often than not the income statistics reported by households in developing countries are apt to be understatements. When the opportunity presents itself, the researcher or scholar should attempt to make some check on their validity. Sometimes salary and wage data from employers can be useful for this purpose. Processed applications for public housing or for purchase of homes under favorable terms of a government supported program may be especially helpful.

The preceding paragraphs suggest that in developing countries, despite low national per capita data, there are indeed large numbers of families who have relatively substantial income. It is reasonably obvious that in a great many cases the income of these families is in excess of subsistence needs. The tropical or near tropical climates which almost all developing countries have helps make this possible.(23)

There is, of course, a great deal of independent evidence from economic history and from economic development experience which reports the existence of family savings or savings potential in developing countries in temperate as well as tropical climes. This savings is the key to economic development. Our concern and interest is in savings, realized and potential, that may be put to housing needs.

It is clear that under even the most adverse circumstances people will save for housing purposes. In a number of developing countries, for example, savings have taken the form of building materials, e.g., concrete blocks. When the household has accumulated a sufficient amount of building materials another portion of their home or an addition would be built.(24) This kind of saving is, however, wasteful and inefficient, both from the standpoint of society and the household. The hoarding of money, gold or silver is, of course, also uneconomic and undesirable. To counteract these wasteful practices and to put immobile assets to work, popular savings institutions can be created. Whether the savings thus collected should be put to housing purposes, in whole or in part, is a subject unto itself. It is understood that the paper by Professor Burns will deal with various aspects of this topic.

One of the most successful popular savings institutions is the savings and loan association. There are several variations in name, in practices and in ownership. All, however, are institutions geared primarily if not wholly to the financing of housing. A few short years ago the United States Government and the savings and loan industry of the United States began to give substantial technical assistance and encouragement to existing and proposed savings and loan associations in Latin America. The record is pleasing. In 1961 net savings on deposit in free savings and loan associations in Latin America amounted to $2 million. In 1965 net savings totalled $120 million.(25) And if a burst of optimism may be forgiven, this is still only a small fraction of the potential.

Most of the savings and loan associations now operating in Latin America are very new. Their ability to mobilize such a large amount of savings in such a short time indicates that they met a genuine need. Their establishment, requiring legislation for their chartering, supervision to protect the public interest, etc., depends generally upon friendly and positive action by national governments. At the same time, such savings institutions may pose an imaginary or real threat of competition to established financial institutions, which may resist their proposed establishment. The key role in such situation is played by the government.

The establishment and supervision of a new financial mechanism may, however, commit the conscientious, dedicated government to other policies and actions to ensure or to maximize success of the new instrument. In the case of savings and loans a deposit insurance plan, protecting depositors from loss, might be very desirable to create the psychological climate propitious to early success of a fledgling institution. Some government deposits in new institutions, particularly from Social Security or pension funds, might be another positive aid. Some means to provide liquidity to these institutions at times of stress is another possible and generally desirable action on the part of the government. The report of a UN Mission to Kenya, for example, recommended in light of Kenya's experience that some liquidity mechanism be provided for unusual withdrawals.(26)

Current experience in the United States illustrates the kinds of problems that can develop for savings and loan associations or other financial institutions mobilizing savings and investing these savings in long-term home mortgages.(27) Several economic factors, precipitated in recent months by Federal Reserve Board changes in regulations applicable to commercial banks, have led to greatly increased competition for savings. As a result interest rates for savings deposits have moved up sharply.(28) For savings and loan associations and savings banks which are so heavily invested in long-term home mortgages, this has resulted in the classic squeeze between relatively fixed income and rising costs. Higher interest rates are being paid on all deposits, most of which are represented by mortgages bearing relatively low interest rates that reflected the old market conditions. In developing countries when savings institutions run into this or similar problems, governments should be prepared to provide necessary assistance.

Although the role of governments with respect to the establishment and promotion of housing finance institutions is very important, there are many other areas wherein government action may eliminate roadblocks, redirect efforts to better purpose, or stimulate existing activities. One major problem in many countries which has been poorly handled by most governments is the high cost of urban land. Small improved building lots in urban areas of developing countries often sell for incredibly high prices. As a result, the cost of land represents in many cases an inordinately high proportion of the total housing cost or price, often ranging between 30 to 60 percent of the total price.(29) Various reasons account for this situation, but the principal ones are tradition, fear of inflation and speculation. Governments in almost every case have the means at their disposal to modify greatly the cost of urban land, thereby reducing the costs of housing and lightening housing finance problems. Improved urban property tax systems, penalizing vacant land and speculation, and recovering the costs of improvements from landowners, is a powerful weapon that has been too little utilized. Appropriate condemnation laws, perhaps relating prices paid for land to tax assessments, is another readily developed tool. Re-establishing an appropriate balance and relationship between land-costs and other factors in the housing picture in developing countries would in a great many instances upset long-standing practices and encounter powerful opposition. Yet no other course appears feasible.

The indirect benefits to housing finance problems from the institution and operation of an adequate property tax system are considerable. But they are certainly matched by the direct benefits that accrue. Property taxes could provide funds for the local provision and maintenance of community facilities and services. To the extent that these local services are financed out of national budgets - and this is quite common - the national exchequer is benefitted as well.(30)

Since the task of financing housing exceeds by far the resources being devoted to it, governments of developing countries should attempt to exploit all feasible possibilities, even though the possibilities individually are not in themselves impressive. In many developing countries, foreign companies are charged with responsibility for housing local employees. If this assigned responsibility is made under reasonable terms and without malice, the practice cannot be faulted. In many of these same countries, domestic employers - generally those exceeding a certain size - are also charged with some responsibility for housing their employees.(31) There are dozens of ways in which this can be done. Companies may build housing for their employees, which housing may then be occupied rent-free or at agreed rentals. Or alternatively, companies may pay payroll taxes to a municipal board which constructs housing and allocates this housing to the work forces of different employers according to formula. Loans may be made to employees under favorable terms, loans from other sources may be guaranteed, contributions to rental or amortization payments may be directly made, etc. Governments frequently permit employers tax benefits of various kinds to lighten burdens imposed by this housing assistance or to induce maximum participation.

In this country, company housing has a bad name and is frowned upon by labor unions and corporate bodies alike. Circumstances in developing countries are radically different and many of the programs have little of the disadvantages normally associated with company housing.

One of the most intriguing - and controversial - government actions supporting housing finance activities is the adoption of a link-index system to counteract the impact of inflation. Israel has been operating such a system for a number of years. In Israel it is not confined to housing but covers industrial and agricultural activities as well. Following the Israeli example a number of Latin American countries have incorporated the essential aspects of the Israeli system into their housing finance operations and others are considering it. The link-index system has also been recommended to developing countries in other continents.

The basic concept of the link-index system is quite simple. The objective is to overcome the capital erosion effects of inflation. Few people will save and continue to save when the purchasing power of their savings is sharply reduced through price inflation. Instead, the funds they have remaining after meeting their daily needs will tend to be converted into land, durable or consumer goods, precious metals or gems, etc. In brief, inflation induces investment in tangibles, in equities which generally rise in value along with inflation. Deposits or certificates of indebtedness do not, however, rise in value with inflation and thus saving and lending is hard hit by inflation. The link-index system attempts to correct this imbalance by periodic elevation of the explicit monetary value of savings and investments in accordance with the movement of some agreed price index.

Brief examination of a crude example will show the main elements and intended effects of this scheme: A savings institution has $100,000 in savings. This $100,000 is invested in home mortgages. At the close of the agreed period (quarterly, semi-annually or annually) the price index governing the link-index scheme shows a 20 per cent inflation. Accordingly, the $100,000 of deposits is raised by 20 percent, yielding a new total of deposits of $120,000. But the means of repaying the deposits and of paying the interest on savings are the outstanding mortgages. So it is necessary also to raise the $100,000 of outstanding mortgages by 20 percent. Both sides of the ledger have received the same treatment, and the books are in balance. Because of the link-index scheme, savers have been insulated from the effects of inflation. Thus incentive to save remains unimpaired. The homebuyer paying off the mortgage made possible by the savings has been able to purchase a home because savings have not dried up. The face amount of his mortgage and his resulting amortization and interest payments rise, but, of course, this is part of the mortgage contract to which he agreed. Moreover, at the same time the mortgage value and payments are escalated, wages are also raised by the same formula.

Link-index systems for specific purposes have a long history; most frequently they have been connected with the issuance of bonds, particularly by governments. Not too long ago, based presumably on Israeli experience, a proposal was made to produce an "inflation-proof economy", recommending adoption of link-index plans by inflation-prone countries which would apply to all elements of an economy. Receivers of wages, salaries, rents, interest, pensions and other transfer payments would all have such payments subject to escalation by the same index.(32)

The writer was in Israel earlier this year and spent some time discussing link-index operations in Israel.(33) Wages, housing loans, industrial and agricultural loans and bank deposits are covered. The operations, however, are by no means all-encompassing. Increases in wages and salaries for example, are limited and apply only on the first $165 earned per month. This, of course, penalizes the higher wage and salary earner. Not all bank deposits are covered, either. In many cases there is no linkage or only partial linkage. With respect to industrial and agricultural loans, a fixed annual charge of 4 percent substitutes for inflationary possibilities, and the government assumes the linkage risk. The arbitrary premium is paid by the bank to the government, and the government reimburses the bank should the annual inflation rate be high enough. The link-index scheme operates when annual inflation reaches 3 to 5 percent, the figure varying in different cases. Last year, 1965, the inflation was between 4 and 5 percent; over the last ten years it has averaged 6 percent per annum. The law governing link-index usage permits linkage for loans over 24 months. There is consequently a rather high percentage of bank loans for a 25 month period.

The link-index system in Israel for the housing sector began with linkage to a cost of building index. Subsequently, a linkage system to foreign exchange rates was utilized, a more or less natural result because of the large amount of foreign borrowing. Currently, two indexes are used, a cost of living index and a cost of housing index. The Government Housing Saving Scheme is linked to the cost of housing index which rises much more rapidly than the cost of living index, thus making the government scheme more attractive to savers. The housing saving plans promoted by banks utilize the cost of living index. Under both government and private bank savings schemes, the saver who saves for certain periods and meets the general terms of the plan becomes eligible for a loan. Among those made eligible for a loan, there is a lottery to select loan recipients. The number of lottery winners depends upon the number of flats available. The basic interest rate paid by the lottery winners is 8 percent plus linkage.

In recent years the Government decided that linkage under its loan program would be only 70 per cent. In special cases which the Government wished to promote, the Government also agreed to restrict linkage on loans to 50 percent. These conditions, obviously, involve elements of subvention. It is not clear how these terms would apply to housing because the Government also has special housing loans in connection with development schemes. Interest rates as low as 3½ percent are charged, grace periods before linkage begins are permitted and moratoria of 5 to 10 years before repayment may be granted. The private or semi-private banks do not participate in these schemes and for their housing loans they require 100 per cent linkage.

The writer was told in Israel, by banking experts that the link-index system had indeed succeeded in inducing people to save. Through it, a favorable attitude toward saving in the entire populace had been created. There were, for example, hundreds of thousands of savings accounts for school children. These accounts were growing although they were not covered by link-index adjustments. But the system is controversial and appears to have proponents and opponents in Israel as well as out. Inflation is feared by bankers throughout the world, yet the impression exists that not all Israeli finance men are happy living with the link-index system. The large (perhaps overly large) role of the government in savings schemes and the considerable amount of government regulation may be factors.

Some of the benefits of the link-index system have been made clear. What then are the disadvantages? Whether in Israel or elsewhere, the link-index system poses several disadvantages. To begin with, the plan almost automatically entails government subsidy. In the crude example described earlier of how the scheme works, 100 percent of savings was represented by a like amount of mortgages. This is an impossible situation. There must be reserves and cash in the till for example. Consequently, perhaps 15 to 25 per cent of savings will not be represented by mortgages. When savings and mortgages are readjusted, someone must accept the burden of escalating this 15 to 25 percent reserve or cash portion of deposits. In Israel this expense is assumed by the Government. But Israel's inflation has been relatively mild. Could this expense have been as readily assumed if inflation had been at a much higher rate? And, of course, if the total amount of mortgages represented a very high monetary figure? In Latin America, for example, some of the countries adopting this system or considering it have experienced cost of living rises of 20 to more than 100 percent in one or more years in the past decade or so, and annual exchange devaluation of as high as 25 to 600 percent.[34] Also to be considered is that Israel has special access to foreign loans and grants which other developing countries do not have. This subsidy cost, regardless of who bears it, is over and above the cost and effort of administering the system.

Another difficulty that faces the link-index system is the adjustment impact on the mortgagor. The principal value of the mortgage and the amortization payments are raised in accordance with the index. What happens when the mortgagor's income has not risen proportionately? In some countries there are covered industries which almost automatically receive periodic wage benefits related to cost of living indices. Other workers in different industries, however, may receive adjustments in haphazard fashion at best. Small tradesmen, self-employed, small establishments and some white-collar employees may or may not receive these adjustments. Obviously a 20 to 30 percent increase in mortgage payments might be an insurmountable escalation for some homeowners. If another escalation without equivalent income gain were to be imposed the following year, these homeowners would indeed be in a very tight, likely impossible squeeze.

One of the strongest objections made to the link-index scheme is that it treats a symptom and neglects the major cause. Inflation is a major economic malaise. Governments should take the necessary budgetary and tax measures to eliminate it. A continuation of inflation accentuates maladjustments and adds to national economic problems. For example, if housing is the only protected sector, capital may tend to flow into this sector to the detriment of other sectors. Almost everyone has heard about the very large amounts of capital that have taken flight from developing countries and are on deposit in Switzerland, London, and other western financial centers. Given political and economic stability, this capital would be at home making important contributions instead of adding to the affluence of advanced countries. This argument continues by pointing out that any success secured by link-index use simply postpones the day of reckoning by easing temporarily the pressures on the national government.

Another disadvantage of the link-index system is that government actions - and even the index itself - are subjected to much political pressure which is very difficult to resist. In 1965, for example, the Israel Government postponed an adjustment reportedly because of the imminence of an election.

The pros and cons of the link-index system have by no means been exhausted in the above discussion. Some day, perhaps, there will be a careful study of the scheme in its various aspects and in light of Israeli and Latin American experience. Such a study would have great interest.

Before moving from the subject of how governments of developing countries can profitably assist and encourage private enterprise to do a larger and more effective job in the housing field, one more possibility merits mention. This is assistance to non-profit or limited dividend rental housing. In urban areas, in all continents, there is a wide demand and need for rental housing for lower income families. The segment of the urban population that can afford rental facilities is much larger than the segment that can afford to purchase a home. The capital and land costs of rental units are lower than for home purchase. Yet in a number of countries, notably in Latin America, little effort is made to encourage or assist the rental sector. One reason may be that some developing countries have had poor experience with rental housing, such as poor rent collections and excessive neglect by tenants.

It would appear inescapable, however, that for many decades in the future large amounts of urban rental housing will be needed. Through government assistance, including some subsidy, it would appear that significant quantities of rental housing could be built. Such housing could be built under non-profit, limited dividend or cooperative aegis. Political difficulties could thus be minimized while at the same time the support and resources of the government could be utilized with important effect. In this connection it might be pointed out as an obiter dictum that the administrative and technical difficulties of a link-index system are probably much less with rental housing than with purchase housing.

III. External Financing.

External financial assistance to developing countries may come from two broad sources: Bilateral aid or country-to-country, and international aid that comes from multilateral organizations or arrangements. Dealing first with multilateral assistance, one may generalize safely and state that until the last five or six years, multilateral financial assistance for housing was very limited. Since that time the situation has changed significantly.

The preeminent international organization, of course, is the United Nations. Its headquarters organization and the United Nations Development Fund (which results from the merger of the Special Fund and the United Nations Technical Assistance Organization) provide technical assistance, aid in the establishment and financing of pilot projects and also finance very useful pre-investment or feasibility studies. These activities are performed not only for housing but for all economic and social development projects or programs. They are essentially teaching, "how-to-do" and pre-investment activities which are very important but provide little investment assistance. In the United Nations family, but essentially independent, is the International Bank for Reconstruction and Development, commonly known as the World Bank. The World Bank and a subsidiary, the International Development Association, have made a few loans for urban projects, principally for water and sanitation improvements.

In the main, the World Bank has made it quite clear that it is very reluctant to play an important role in the housing/urban field. Recently, however, it has shown greater awareness of the housing/urban field and has included housing or urban specialists on its country teams. It should be noted that the World Bank has always been willing to consider including funds in its loans for housing purposes in special cases. For example, if an oil field or a hydro-electric site is to be developed in a remote area, the World Bank may well regard housing facilities as part of the necessary facilities. Without satisfactory housing, supervisory and key personnel may not be attracted to or remain at the site.

A second subsidiary of the World Bank (its "third window") is the International Finance Corporation. Unlike the other two members of the World Bank family this body can engage in equity financing. Its financial assistance in the housing field to developing countries has been confined to a relatively small amount of funds loaned to building materials companies.

The important multilateral financial assistance given to developing countries in the past half-dozen years has come from a rather new regional organization, the Inter-American Development Bank. Founded in 1959, this Bank has rapidly become a very important institution for Latin America. Its authorized loans already exceed $1.5 billion, and these loans are for economic and social projects which have estimated total cost of $4.2 billion.(35) Fully 40 percent of the Bank's current portfolio, or loans amounting to $600 million are for housing.(36) An additional $288 million has been loaned for 59 separate projects to improve water supply and sewage disposal systems. This is a remarkable record of financial assistance for housing and community facilities, particularly when viewed in relation to World Bank Group commitments for 1963-65, wherein 3.6 percent of commitments were for housing and related community facilities.(37) It is no wonder that Latin American countries are the envy of developing nations in other continents.

A potentially important development in multilateral financing of housing is a proposed meeting of experts on finance, to be convened under UN auspices, and to consider carefully policies and programs of both domestic and international financing of housing and related urban facilities. The object of the meeting would be to bring potential lenders and borrowers, independent financial experts, and development economists together to explore exactly what real possibilities exist for providing additional funds to this sector and for increasing the effectiveness of funds currently being employed. The proposal for such an expert group was made by the UN Economic and Social Council's Committee on Housing, Building and Planning. The superior body of this Committee, the UN Economic and Social Council, had this proposal on its July 1966 agenda for consideration and appropriate action.

A number of advanced countries provide important bilateral assistance to developing countries. In few cases, however, has housing loomed importantly in these programs. The United States bilateral assistance program was begun in the late 1940's. Since its inception it has been the largest and most important among world bilateral programs. However, housing/urban programs have until recently played a relatively small role. In recent years, the U.S. Congress has taken keen interest in housing/urban conditions in developing countries and has taken legislative initiatives which have increased substantially the United States bilateral assistance in housing. A great deal of this assistance has been directed to Latin America and has been made part of the hemisphere Alliance for Progress. The encouraging success of savings and loan associations, referred to earlier, has been aided greatly by United States seed-capital loans to various Latin American countries, as well as technical assistance. Cooperative housing has also received important assistance, including some loans.

One of the major programs providing financial assistance for housing, whether multilateral or bilateral in nature, is the United States bilateral program of investment guarantees for housing. The executive agency for this program, as well as for all United States bilateral assistance programs, is the Agency for International Development (AID). The very large part of this program to date has been for the guaranty of investment in housing projects in Latin America. Only one or two projects in Asia and Africa have been initiated or are at the point of initiation. Other projects are under consideration, however. As of May 31, 1966, 16 projects were authorized and underway in Latin America. An additional 16 projects had been authorized but were not as yet under contract. Together these 32 projects for Latin America represent an investment in housing of $179 million and will provide 32,000 units. In 1965, the United States Congress expanded the authority of AID to guaranty investments in housing projects in Latin America, providing new authorization of $150 million. Revised criteria and regulations are being developed by AID which are intended to give the investment guaranty program a new emphasis.

The operations of the investment guaranty program are basically simple. An United States sponsor, perhaps in partnership with a Latin American firm, submits an application to AID proposing to build housing in a Latin American country. The financing must be part of the application, and to date financing has come from insurance companies, commercial banks and pension and retirement funds of AFofL/CIO unions. If all aspects of the proposal meet the criteria, the guaranty is issued for the full amount of the outstanding principal balance of the loan. An annual guaranty fee, on the unpaid balance and payable monthly is charged. This varies up to a maximum of 2 percent. For this fee the financial intermediary providing the funds secures an "all-risk" guaranty which covers almost every conceivable possibility of loss except fraud.

The currently tight capital market in the United States suggests that funds for investment in housing under the AID guaranty may be considerably harder to find in coming months than has been the case. The Housing Act of 1965 provided authority for the first time for Federal Savings and Loan institutions to participate in this program. These associations are now authorized to invest up to 1 percent of their assets in loans for housing projects in Latin America which have been guaranteed by AID. Given the squeeze that has affected savings and loan associations (mentioned earlier) the amount of funds from this source will probably be little for a time.

There is one small housing finance program of AID that has interest and potential far beyond the money involved. This is a $10 million loan to the Central American Bank for Economic Integration (CABEI). This loan, completed in accordance with a commitment made by the late President Kennedy, was granted to CABEI for the express purpose of establishing a secondary mortgage financing facility. In November 1963 the Home Loan Department of CABEI was formally established. Its successes to date, its problems and its overall experience are of great value to developing countries and to all who are interested in international finance. Here is a unique experience: an international secondary mortgage facility, regionally based, serving five developing countries. Readers interested in international finance or in housing finance may find its operations of interest.(38)

United States bilateral programs in housing assistance have made their marks and on some occasions have earned criticism. The writer sees little purpose in adding up possible pluses and minuses earned in the past. Certainly good as well as dubious results have emerged from the various programs. Currently, however, and for some time, the writer has felt that the US bilateral program in the housing/urban sector has been deficient by its almost complete disregard of rental housing. If the program is seriously directed at developing country needs, it cannot continue to ignore rental housing. It is not enough to promote housing for purchase. Laudable though this ojbective is, and much as the writer favors home ownership, the hard realities are that the greatest mass of low-income urban families in developing countries are logically renters. Yet programs locally initiated to meet some of this need are either lacking or grossly inadequate. Important programs of external origin that place the stress wholly on home-ownership would appear accordingly to be courting the dangers of imbalance. Some AID professional housing officers see the incongruity in their housing programs. It has been suggested, for example, that a portion of the investment guarantee program might be used for rental housing projects. The writer approves of this kind of thinking.

In the discussions immediately above relating to international financing of housing, nothing was said about the propriety or logic of international investment for housing. Housing, after all, is a local product, built by local labor, on local land, with largely local materials and for local use. Recognizing that the total international capital needs of developing countries far exceed the available supply of funds, how can one justify international loans for housing?

The answer to this question cannot be simply made. The fact that substantial international financing of housing is taking place and that serious people continue to consider additional assistance suggests that economic, social and political reasons for such action do exist.

In the brief examination to be given this question at this point, we may safely begin from the premise that external financial assistance is not intended to carry the bulk of the burden. Only a few zealots would disagree with this premise. "Housing is fundamentally a national problem requiring national efforts to solve it."(39) External financial assistance, therefore, is essentially of a "seed", psychological or generative nature. Viewed in this light, with appropriate purpose and sharply reduced proportions, even the purist becomes less outraged at the prospect of foreign lending for housing. One argument that is raised in opposition to foreign lending for housing points to the overwhelmingly large domestic component (domestic labor, materials, etc.) and continues that at most, foreign lending should cover only the foreign exchange component, which, depending upon the country and the housing standard varies between 5 to 15 percent of the total construction price. This argument possesses some validity but must be considered essentially non-germane if the lending is intended to generate or "seed" local effort.

Care and consideration are obviously in order, however, from the standpoint of the foreign payments burden that is shouldered when external loans are accepted for housing purposes. Dollar loans repayable in dollars place a future lien on export earnings. Unless export earnings increase, and by more than the indebtedness and service needs of other activities, the balance of payments can be expected to take an unfavorable turn. This in due course can jeopardize a country's international credit and the stability of its currency. Housing, of course, does not earn foreign exchange. As foreign loans become due, the borrowing institution must enter the foreign exchange market as another claimant for foreign funds. If the exchange situation should be tight, it will be made tighter and more serious. And many, if not most, developing countries have a more or less perennial shortage of foreign exchange.

Despite these serious considerations, a negative attitude relative to foreign loans for funding housing programs may be in error. The application of most economic principles, after all, is quite uneven and varies case by case. Realities, practical facts of circumstances, attitude and practice must be carefully weighed. Although it has already been conceded that domestic funds should be utilized in the very great part for domestic housing programs, these funds may simply not be available. In all too many cases, the institutions and habits that are prerequisites for mobilizing these resources may be lacking. The very first use of foreign funds for housing purposes may be to promote new institutions, new concepts and new methods which ordinarily have scant opportunity to command domestic money. And while these new institutions, concepts and techniques are in the formative period, it would appear that there may be good and sufficient reason to borrow abroad in order to hasten the initiation of a housing program. When there is gross inability to secure domestic funds for housing in competition with other domestic activities, foreign borrowing may play an especially strategic role.

Viewing the problem from the borrowing country one must recognize that certain sources of foreign funds (e.g., the Inter-American Development Bank) have been established with important objectives in the social field. Conditions and terms of financial assistance from these sources are very favorable, differing greatly from ordinary loan conditions. Loans and technical assistance available from such sources can promote housing programs with a minimum of balance of payment complications or other difficulties.

The impact of the balance of payments problem can be greatly minimized under proper safeguards or conditions attendant to the loan. The procedure adopted in the AID dollar loan to the Home Loan Department of the Central American Bank for Economic Integration (CABEI) is an example. The dollar funds from CABEI pass through the Central Banks of the countries concerned. The Central Banks, in turn, give local currency to the borrowing domestic institutions, which is what the local fiduciary needs. By previous agreement between CABEI and the Central Banks, the dollar funds retained by the Central Banks are utilized to support national development, particularly the support of export industries or industries which will lead to reduction in foreign exchange needs. Ensuring prudent use of foreign currency balances acquired from foreign loans for housing does much to remove what are otherwise important and valid objections.

IV. Conclusions

Implicit in this effort to cover such a broad subject as our title has been the necessity to pick and choose those aspects to be covered. The attention and space devoted to each aspect has also required arbitrary judgment. The views and interests of the writer were the sole criteria. Included among those interests was a hope that some of the provocative aspects touched may be picked up by other scholars and researchers as profitable areas for further research and study.

The writer cannot accept the deep pessimism that is expressed by some development planners when they view the potential financial ramifications of housing programs. This kind of attitude suggests inadequate understanding and an inexact focus. Improvement in the housing of developing countries is truly a long-term task of grand proportions. The housing finance aspect is one of the more difficult components. But there will be framed in the future better programs and answers, and these will be based on the lessons and successes of past and current programs.

TABLE I

Investment in Housing in Selected Western European Countries, 1953 - 1956

COUNTRY	ITEM	1953	1954	1955	1956
Austria	A	18	21	24	23
	B	4.4	4.6	4.4	4.2
	C	24	22	18	18
	D	5.5	5.8	6.0	6.0
Belgium	A	16	17	17	17
	B	3.8	4.1	3.7	3.4
	C	24	25	22	20
	D	4.5	5.1	5.0	4.9
France	A	16	16	17	17
	B	3.8	4.0	4.3	4.2
	C	24	25	25	24
	D	2.7	3.8	4.9	5.4
Greece	A	12	14	14	15
	B	4.3	4.5	4.9	4.9
	C	37	31	35	33
	D	6.6	5.9	6.7	6.9
Italy	A	19	20	20	21
	B	4.0	4.7	5.3	5.4
	C	21	24	26	26
	D	3.2	3.7	4.5	4.8
Sweden	A	20	20	20	20
	B	4.6	4.9	4.8	4.9
	C	23	24	25	25
	D	7.4	8.2	7.9	7.9
United Kingdom	A	14	14	15	15
	B	3.7	3.6	3.3	3.2
	C	27	26	23	21
	D	6.5	7.0	6.4	6.1
Western Germany	A	20	21	23	23
	B	5.4	5.8	5.8	5.7
	C	27	28	25	25
	D	10.5	10.9	10.7	11.0

Key: Item A - Gross Domestic Fixed Investment as percentage of GNP
Item B - Investment in housing as percentage of GNP
Item C - Investment in housing as percentage of Gross Domestic Fixed Investment
Item D - Number of dwellings completed per 1,000 inhabitants

Source: Financing of Housing in Europe, United Nations, Economic Commission For Europe, Geneva, 1958, pages 10-11.

TABLE II

SELECTED DATA ON FAMILY INCOME AND RENTS
FROM THE DECEMBER 1960 SAMPLE SURVEY
OF MIDDLE AND LOW-CLASS HOUSING IN MANAGUA, NICARAGUA

- In Cordobas* -

A. Average Monthly Income of Family	No. of Families	Per Cent
1. Less than C$ 150	1,316	3.9
2. C$ 150 to 250	3,873	11.4
3. C$ 150 to 250	6,201	18.2
4. C$ 350 to 500	7,302	21.4
5. C$ 500 to 750	7,897	23.2
6. C$ 750 to 1,000	2.949	8.6
7. C$ 1,000 to 1,250	2,050	6.0
8. C$ 1,250 to 2,000	1,417	4.2
9. C$ 2,000 to 3,500	911	2.7
10. C$ 3,500 and up	165	.5
	34,081	100.0

B. Families Broken Down by Monthly Rent Paid	No. of Families	Per Cent
1. No rent**	8,036	23.6
2. To C$ 25	721	2.1
3. C$ 25 to C$ 50	3,911	11.5
4. C$ 50 to C$ 75	7,252	21.3
5. C$ 75 to C$ 100	4,560	13.4
6. C$ 100 to C$ 150	3,569	10.5
7. C$ 150 to C$ 200	2,548	7.5
8. C$ 200 to C$ 250	1,327	3.9
9. C$ 250 to C$ 400	1,381	4.0
10. C$ 400 to C$ 700	716	2.1
11. C$ 700 and up	60	.2
	34,081	100.0

C. Families According to % of Income Paid for Rent	No. of Families	Per Cent
1. Less than 7.5%	1,110	4.3
2. 7.5% to 10.0%	1,264	4.9
3. 10.0% to 12.5%	1,940	7.4
4. 12.5% to 15.0%	2,555	9.8
5. 15.0% to 17.5%	3,364	12.9
6. 17.5% to 20.0%	3,560	13.7
7. 20.0% to 25.0%	5,277	20.2
8. 25.0% to 30.0%	2,502	9.6
9. 30.0% to 35.0%	1,512	5.8
10. 35.0% to 40.0%	1,272	4.9
11. More than 40%	1,689	6.5
	26,045	100.0

* C$7 equals U.S. $1.
**Includes Home Owners.
Source: NICARAGUA: Housing Finance and Related Aspects, Stanley E. Smigel, Rex G. Baker, Jr., and Lyndon R. Day, US Agency for International Development, December 1963, Table VI.

References and Footnotes

1. Economic Report of the President, 1965, Table B-12

2. 17th Annual Report, Housing & Home Finance Agency, 1963, Table B-38

3. Business in Brief, No. 22, Sept-Oct, 1958, Chase Manhattan Bank, New York

4. Financing of Housing in Europe, United Nations, Economic Commission for Europe, 1958, pp 10-11. (For specific figures for selected Western European countries, see Table I)

5. Ibid., Table II, p. 19

6. The Capital Development Needs of the Less Developed Countries, United Nations, 1962, esp. Chaps. I and II

7. Report of the Ad Hoc Group of Experts on Housing and Urban Development, United Nations, 1962, pp 9-10

8. Ibid, p. 13

9. Per Capita National Product of Fifty-Five Countries, 1952-54, United Nations, 1957, Table I

10. The Problem of Housing in India, Government of India, Ministry of Works, Housing & Supply, New Delhi, 1957, pp. 21-22

11. a) Finance for Housing & Community Facilities in Developing Countries, E/C.6/32/REV 1, Jun 1, 1966, UN ECOSOC Committee on Housing, Building & Planning, CH V, pp 46-60

    b) Financing of Housing in Europe, United Nations, 1958, pp 72-77

    c) Needleman, Lionel, The Economics of Housing, Staples Press, London, 1965 pp 119-141

12. Report of the Committee on Urban Development, 'Shaping Tomorrow's Cities: A Program to Meet Urgent International Urban Problems", National Citizens' Commission, White House Conference on International Cooperation, Nov 28-Dec 1,'65

13. Major Long-Term Problems of Government Housing and Related Policies, United Nations Economic Commission for Europe, 1966, 2 Vols, p 8 Vol I

14. a) Ibid, Vol I, p 45 & Vol II, Table B-1

    b) Kindleberger, Charles P., Economic Development, McGraw-Hill, 1958, Figure 8.1, p139

15. Major Long-Term Problems of Government Housing and Related Policies, op cit, Vol I, p 111

16. Ibid, p 123

17. Hoffman, Paul G, One Hundred Countries, One and One Quarter Billion People, Albert and Mary Lasker Foundation, Wash, DC, 1960, p 18

18. Housing in Peru, Pan American Union, Washington, D.C., 1963, pp 584,585,651,644

19. Housing Market Analysis in Latin America, Richard Metcalf, AID, Wash, D.C., 1965, p 24

20. Housing in Ethiopia, Country Report Series, Department of Housing & Urban Development, Wash, DC, 1965, pp 7, 31-32.

21. Housing in Peru, op cit, pp 588-591

22. NICARAGUA: Housing Finance & Related Aspects, Stanley E. Smigel, Rex G. Baker, Jr., & Lyndon R. Day, US Agency for International Development, Wash, DC, Dec 1963, Table VI

23. There are those who argue persuasively that the tropical climate itself may be a key reason for the underdeveloped state of many developing countries.

24. Financing of Housing & Community Improvement Programmes, United Nations, New York, '57,p8

25. Statement of the President, 7th Annual Meeting of the Board of Governors, Inter-American Development Bank, Apr 26, 1966, p 23

26. United Nations Mission to Kenya on Housing, Lawrence N. Bloomberg & Charles Abrams, United Nations, New York, Dec 1964, pp 31-32

27. The previous emphasis on savings & loan institutions is for illustrative purposes and does not imply that other savings institutions (e.g., savings banks and mortgage banks) may not be equally desirable, or in some instances even more suitable to local conditions of developing countries.

28. Although general references are to interest rates, it should be recognized that mutual savings institutions have "shareholders" rather than depositors and receive "Dividends instead of interest; but for all practical purposes, the differences are in terminology only.

29. a) Financing of Housing and Community Improvement Programmes, op cit, p 51

    b) NICARAGUA: Housing Finance & Related Aspects, op cit, p 14

    c) Housing/Urban Situation in Central America, Stanley E. Smigel, Survey for US AID, Housing & Home Finance Agency, September 1964

30. a) Housing/Urban Situation in Central America, op cit, pp 3-4

    b) Financing of Housing & Community Improvement Programs, op cit, pp 99-100

31. a) United Nations Mission to Kenya on Housing op cit, page 5

b) Financing of Housing & Community Improvement Programs, op cit, pp 38-39;49;22-23;9

32. Morag, Amotz, "For an Inflation Proof Economy" American Economic Review, March 1962, pp 175-185

33. The writer regrets greatly that the pressures of time prevented all but surface examination of this truly interesting subject. The writer's comments on the Israeli system and the specific facts given, therefore, should not be considered definitive.

34. a) Fourteen General Questions Concerning the Establishment of an International Home Loan Bank to Promote and Finance Savings and Loan Systems in Developing Countries, National League of Insured Savings & Loan Associations, Wash, DC, Jan 64, p 7

b) Capital Formation for Housing in Latin America, Editors: Walter D. Harris & James Gillies, Pan American Union, Wash, D.C. 1963, p 102.

c) In Chile, an adjustment of 35 percent for cost of living index increases was reportedly made on June 30, 1964

35. Scope and Problems of the Inter-American Development Bank, speech by T.G. Upton, Executive Vice President, March 16, 1966

36. Statement by Felipe Herrera, President of the Inter-American Development Bank, Jan 24, 1966 (Meeting on Municipal Financing.)

37. Finance for Housing & Community Facilities in Developing Countries, op cit, p 121

38. The background and operations of the Home Loan Department of CABEI are covered in The International Secondary Mortgage Financing System in Central America, Stanley E. Smigel. US Housing & Home Finance Agency, Wash, DC March 1965. An article based on the above report and bringing the subject up-to-date is to be published shortly in the magazine, Desarrollo Economico.

39. Financing of Housing & Community Improvement Programs in Latin America, UN, UN/TAA/LAT/7, New York, Oct 11, 1956, p 26.

HOUSING AND PRODUCTIVITY: CAUSALITY AND MEASUREMENT[1]

B. Khing Tjioe and Leland S. Burns, University of California, Los Angeles

Identifying the causes of changes in labor productivity, and measuring their magnitude, is a challenging research task. Even under conditions approaching those of the laboratory, the task is full of booby-traps. Observed changes may be the result of changes in personnel policy, the level or structure of wages, general or specific employee attitudes toward work, types of incentives, working conditions on the job or living conditions off the job, training or other educational programs, the demand for output or the availability of overtime work, the quantitative and qualitative supply of labor or of other factor inputs, any of which may be irrelevant to the postulated cause and exogenous to the measurement. Were the changes random? Or, could they be the result of other forces not identified in the problem?

The problems associated with accurate identification and measurement are compounded when important changes occur in the living environment rather than at the place of work. That changes in the employee's home environment, in contrast to changes at the work site, may bear on his ability to produce, constitute the subject of concern in this paper. Broadly, it is postulated that qualitative improvements in housing and community facilities affect productivity in the same fashion as improvements in working conditions. This paper narrates the difficulties of assessing this more subtle of the two broad classes of factors improving productivity.

The paper is divided along three lines: first, a description of the research design including a statement of facilitating assumptions; second, a test of the direction of causation; and finally, an estimate of the magnitudes of the changes in productivity induced by changes in the quality of housing. Evidence is provided from a case study of the relocation of Korean mine workers to a new community consisting of housing and related facilities.

The Research Design

A test site located in a remote section of South Korea was selected for this case study since the working and living conditions were close to ideal for research of this type. During the three-year period studied, new housing and a limited complement of community facilities were provided for 500 (44%) of the coal miners employed in Hambaek Village. The relative isolation of the new community to the old area permitted a neat division of the employees into a test group consisting of the rehoused for comparison with a control group of workers remaining in old quarters. The test and control groups each comprised 50 members selected at random from the rehoused and non-rehoused respectively.

The mine had been in continuous operation since mid-1955, eight years before relocation of the labor force. Plant records of the government-owned operation yielded productivity data covering a year prior to relocation and two years following.

Causality could only be determined with a four-cell "before-after/with-without" framework. For example, a cross-sectional analysis comparing the performance of a test group with that of a control group might demonstrate that the performance of rehoused workers was superior; yet, the conclusion fails on the grounds that the more productive may have enjoyed superior ability to pay for better residential accommodations. Hence, causation could run in either direction, from greater productivity to better housing, or the obverse. Moreover, a before-after comparison based solely on a group of rehoused workers could obscure any changes in non-housing factors, any of which might bear on productivity. For example, labor productivity could increase merely from familiarity gained by the repetition of work tasks, or any of the other factors listed earlier. Such influences, exogenous to this measurement problem, would be revealed in the records of a control group subject to all of the same influences except the one under consideration.

Our four-cell analytical framework permits comparisons of performance differences in three important dimensions:

(1) the test group compared with the control group during the period following the test group's relocation,

(2) the control group with the test group before rehousing to ascertain if the latter's ability to pay for qualitatively superior housing exceeded, or merely equaled, the former's, and

(3) the test group's performance after rehousing compared to before.

If output was machine-dictated, if employees were hired to work for a certain short time period without regard for output, or if there were no production incentives, then it would be unreasonable to assume that productivity is functionally related to the workers' energies or motivations, two qualities closely associated with environmental conditions. Although the mine is of comparatively recent vintage, the operation is still labor intensive, a necessary condition for our research design. Further, since transient laborers are not hired, temporary surges in demand are accommodated by overtime work. Finally, incentive wages are paid for output in excess of established norms.

Several other conditions also governed the

[1]The research reported in this paper is part of the International Housing Productivity Study, a project supported by the Agency for International Development. The authors are indebted to Leo Grebler, Marvin Hoffenberg, Leo H. Klaassen, and E. H. Mulder for their comments on an early version.

choice of this site. First, except for the notable change in housing quality, working conditions affecting performance remained unchanged over the three-year observation period. Second, because the vast majority of workers had long employment histories at the Hambaek mine, any productivity or skill increases resulting from work repetition were zero or negligible. This assumption was confirmed by the lack of statistical significance in the productivity trend for the year pre-dating relocation, a finding discussed later in more detail. Third, other conditions remaining constant during the observation period were plant and equipment (except for the replacement of equipment as it depreciated), and the wage and organization structure. Our data have been adjusted for across-the-board salary increases, however. Finally, since our information was obtained at the end of the three-year observation period, the knowledge that research was being done on performance could not have biased the data. In sum, all factors bearing on productivity except living conditions remained constant. The performance of the control group supports this general statement.

The new community, constructed with the financial assistance of the Agency for International Development, stands in sharp contrast to previously existing facilities. The 500 dwelling units are qualitatively superior particularly with respect to size, protection from the elements, and the availability of utilities and sanitary facilities. In contrast to the new dwellings, most of the old had neither running water, electricity, nor adequate ventilation. Occupancy intensities were higher by a substantial margin in the old units. On average, the intensities were 4.6 persons per room in the old dwellings against 2.5 persons per room in the new, and the living space per person was 61 square feet compared to 143 square feet in the new housing units. Limited community facilities, consisting of a public bathhouse, a kindergarten with recreation facilities for children, a barber shop and beauty parlor, and a grain distribution center, were also included as an integral part of the development. These structures accounted for about one-third of the total cost of 108.2 million won ($832,300) for the entire project.

Although limited, the community facilities can be considered as a substantial improvement over those available in the old area, where the only bathing arrangement, for example, is the river which also serves as a source of water and destination of sewage. Because the new community facilities are situated a mile distant from the old housing area, it is likely that they serve only the rehoused families and their effect on the productivity of those remaining unrehoused would be negligible. Yet, the journey-to-work distance from the new area was the same as from the old; hence, differences in commuting time or distance were not a factor influencing productivity changes.

The Hypothesis

The hypothesis to be tested is that better housing[2] bears significantly and favorably on labor productivity. The transmission of causation from housing to productivity is via dual channels. It may be reasoned that healthier, more sanitary living conditions affect physical well-being, reduce the incidence of illness, and, with it, the number of days on sick leave, and improve performance on the job. In addition to the physiological effects, the improvement in living standards may have psychological ramifications translating into improved motivation and, in turn, into greater output. The latter channel would be identified in data on output per hour. The two taken together should be indicated by an increase in production over the longer period, say a week. This is the information which serves as the basis for our test.

The most convenient common denominator for measuring heterogeneous output is income from wages and salaries. Income earned is a reasonable proxy for productivity since wages are paid on the basis of piecework and performance level expected. Performance-based rates paid for picking are illustrative. Because upward picking is easier and faster than horizontal picking which, in turn, is easier than downward picking, the least is paid per unit for coal mined by upward picking and the most for downward picking. Similarly, because mining rock formation is slower and more difficult than working in a sand formation, coal picked from rock commands a higher wage. As the wage system compensates for the difficulty of the task, it is a good indicator of labor productivity and a workable proxy for output in physical units. The wage proxy has the further advantage of stating output in commensurable units. The wage dollar earned for picking is comparable to the dollar earned for, say, repair work.

Before testing the hypothesis itself, a digression is in order to reinforce our heroic ceteris paribus assumption; that is, that nonhousing factors influencing productivity remained "constant" during the three-year observation period. If the assumption is indeed realistic, then any changes in productivity may be attributed to housing, the only factor not held constant.

The assumption would be acceptable for the year preceding rehousing if output per man remained unchanged throughout that period. Regressing each of the variables, hours worked per week, output per hour, and output per week against time, yields the estimates for the average worker listed in Table 1. Coefficients of determination are omitted as irrelevant since the purpose is to identify a trend rather than to "explain" fluctuations in productivity.

Comparing the regression coefficients with the corresponding standard errors (Table 1), reveals that none of the regression coefficients

[2]For convenience, the term "housing" will henceforth refer simultaneously to the dwelling units themselves plus the inseparable, but less important, community facilities component.

TABLE 1. TIME TRENDS IN LABOR PRODUCTIVITY PER MAN, YEAR PRIOR TO REHOUSING

Variable	Test Group: Regression Coefficient of Time	Test Group: Standard Error of Regression Coefficient	Test Group: Intercept	Control Group: Regression Coefficient of Time	Control Group: Standard Error of Regression Coefficient	Control Group: Intercept
	During Normal Working Hours					
Hours worked per week	4.24	2.68	2,157.58	5.07	2.56	2,094.33
Output per hour (in won)	-0.06	0.05	54.99	0.06	0.07	55.81
Output per week (in won)	96.49	118.48	118,183.28	139.75	189.06	117,032.82
	During Overtime Working Hours					
Hours worked per week	-1.81	1.66	357.92	-2.39	1.87	368.75
Output per hour (in won)	-0.09	0.05	51.68	-0.06	0.05	50.52
Output per week (in won)	-126.10	88.66	18,802.65	-110.83	97.91	18,432.85

are significantly different from zero at the .05 level. Consequently, the lack of a significant trend points to the absence of factors affecting productivity at least during the year before rehousing. This applies equally to both groups of workers. Heroic as it may be, the *ceteris paribus* assumption would seem to hold for this period.

TABLE 2. DIFFERENCES IN AVERAGE WEEKLY OUTPUT DURING NORMAL WORKING HOURS, SECOND YEAR FOLLOWING REHOUSING COMPARED TO YEAR PRECEDING REHOUSING

Group	Mean Output (in won earned)	Standard Error
Test Group	34,100	11,000
Control Group	29,000	14,500

If the same holds for the period after rehousing, there should be no significant difference in the control group's average productivity levels after rehousing compared to before. For the test group, however, a significant difference in the before and after levels would support the hypothesis; that is, that the productivity increase could be traced to the housing improvement. Comparing the second year following rehousing[3] with the year before rehousing yields the differences in weekly output during normal working hours listed in Table 2.

Quite clearly as this evidence indicates, the same conditions are no longer applicable, for the productivities of the control group, as well as the test group, each taken in the aggregate, increased significantly. Either of two explanations is plausible. The *ceteris paribus* assumption fails to apply to the interval following housing; that is, changes of the type noted in the first paragraph of this paper "caused" an increase in the productivity of all workers whether rehoused or not. A second alternative explanation is that better housing has a "demonstration-effect" on the performance of those not rehoused.

Consider the second possibility first. If better housing is popularly accepted as a goal for achievement,[4] and further, if productivity governs progress toward this goal, it follows that workers will be motivated to boost their

[3]The second year following rehousing (or the third year of the observation period) was chosen for this test since, as noted later, the impact of the change in living conditions is not fully realized during the first year.

[4]The rehoused family received gratis a dwelling conservatively "shadow-priced" at a monthly rent value of 500 won for which it forsook a monthly rent subsidy of 300 won paid while they were resident in old company housing. Hence, aside from the prestige of being chosen for relocation in the new project, the rehoused family enjoyed an effective annual standard of living increase netting 2,400 won. As further evidence supporting acceptance of the new units, it should be noted that since the housing project was opened, the management of the mine has been flooded with requests from miners in other districts for transfer to the Hambaek operation.

output levels in order to raise their housing standards. Occupancy policy followed by the housing corporation is relevant here. Although the comparative statistics for test versus control group fail to reveal the difference, productivity was held out as the principal criterion for assigning occupancy priorities to the 1,100 workers seeking a supply of less than half as many dwellings. During the one year prior to rehousing, average weekly output for test and control groups -- 84,513 won and 84,508 won respectively -- was nearly identical. This "hard evidence" indicates that, in fact, those ultimately chosen for occupancy produced no more efficiently than those who were left behind. Aside from the reality, the relevant consideration is that performance was perceived as governing the probability for initial selection for relocation, for maintenance of occupancy once rehoused, or for the opportunity for subsequent rehousing among those left behind.

The alternative explanation for the behavior observed for the control group is that the *ceteris paribus* assumption is unwarranted during the post-rehousing period. Probing our data further indicates the reverse; that is, non-housing factors in fact remained constant during this period as well. Our support is gained by examining the behavior of earnings paid for overtime work in 1963. As the negative signs of the regression coefficients in Table 1 indicate, average earnings were declining slightly as a result of a decrease in the demand for coal during that year, and a consequent decline in the availability of overtime work. Hence, to account for these conditions, exogenous to the firm and its employees, overtime output is measured in terms of hourly production in order to hold constant the availability of overtime work. Again, comparing the differences between the second year following rehousing with the year preceding rehousing, yields the averages reported in Table 3.

TABLE 3. AVERAGE HOURLY OUTPUT DURING OVERTIME HOURS, SECOND YEAR FOLLOWING REHOUSING COMPARED TO YEAR PRECEDING REHOUSING

Group	Mean Hourly Output		Mean Difference	Standard Error of Mean Difference
	Before Rehousing	After Rehousing		
Test Group	49.321	72.233	22.912	10.391
Control Group	48.932	60.466	11.534	13.765

The differences are readily apparent. For the test group, output per hour of overtime work increased after rehousing by a substantial margin (about 46%). This is not the case for the control group whose increase in hourly output during overtime cannot be considered significant due to the relative size of the standard error, and is thus merely the result of incidental, hence irrelevant, factors.

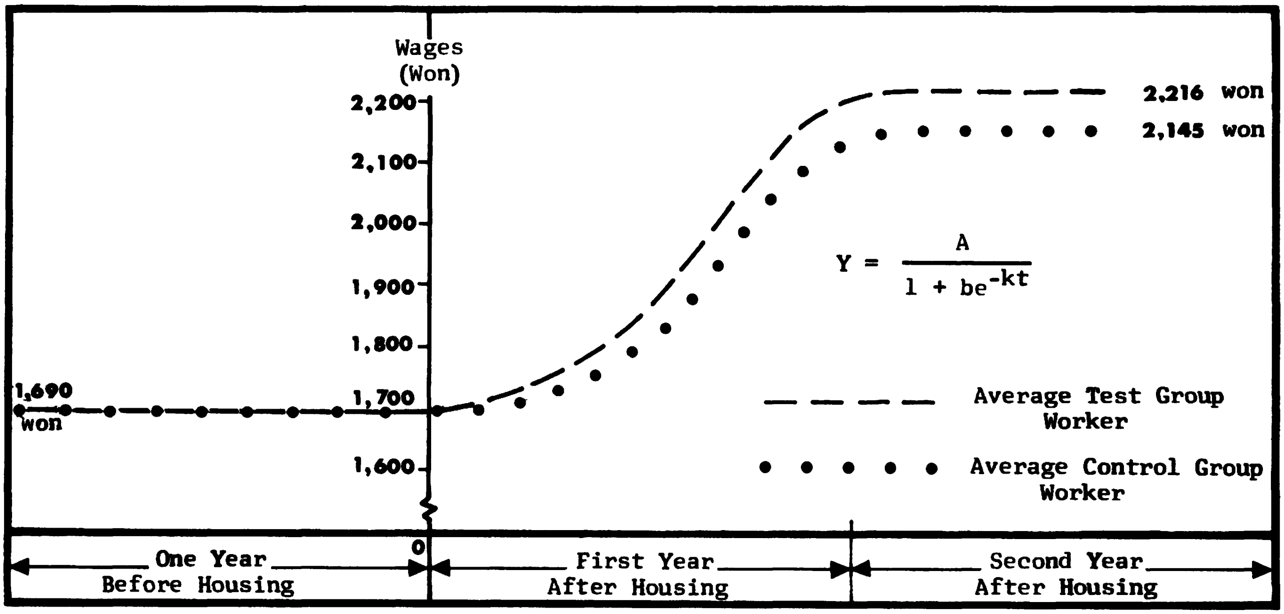
In sum, the *ceteris paribus* assumption would appear to persist during the post-rehousing period as well as before. If a change exogenous to our problem had been operative -- for example, if the quality of equipment had been improved substantially[5] -- the effects would have shown up in the productivity data for the overtime work of control and test group alike since both would have been affected similarly. That it did not, and that the assumption holds, lends credence to the "demonstration-effect" as the true explanation of the control group's productivity increase during normal working hours. Finally, the analysis of this section demonstrates that qualitative improvements in housing translate favorably and significantly into increases in labor productivity, and that causation runs from the former to the latter.

The Estimates

The best fit for the trend development in earnings, or output per week during normal working hours, for the average worker approximates a logistic curve (Figure 1). As noted, the weekly output of the workers in both groups showed no significant trend during the period before rehousing; that is, output fluctuated around a horizontal line. Moving into the post-rehousing period, output increased very slightly until roughly six months after relocation when output began climbing rapidly followed by a decline in the rate of increase until a maximum or capacity level was reached at roughly the end of the first year following rehousing. Beyond this point, further changes were statistically zero with the level of output maintaining itself on a new plateau. The parallelism between control and test group behavior in the timing of changes gives more credence to the "demonstration-effect" for it suggests the

[5]Examination of company records extending from the start of operations in 1955 to the present indicates that the increase in annual capital investment was no greater during the observation period than in earlier years. For the data, see B. Khing Tjioe and Leland S. Burns, *Report on Productivity in Relation to Housing Conditions and Community Facilities in Hambaek, Korea*, (Los Angeles, California: International Housing Productivity Study, Real Estate Research Program, Graduate School of Business Administration, University of California, 1966), Table 3.

FIGURE 1. TRENDS OF WEEKLY EARNINGS, AVERAGE TEST GROUP AND CONTROL GROUP WORKERS, ONE YEAR BEFORE AND TWO YEARS AFTER REHOUSING



extent to which the control group members competed with their rehoused counterparts.[6]

For purposes of estimating the quantitative impact of the housing improvements, we shall ignore the one year post-rehousing adjustment period, and focus on experience of the second year. The post-rehousing development of output is best described by a logistic curve of the form,

$$Y = \frac{A}{1 + be^{-kt}} \tag{1}$$

where Y is earnings, A is the earnings maximum to be estimated, and t is time. The maximum level in weekly output can be described by rewriting (1) and differentiating:

$$be^{-kt} = \frac{A - Y}{Y} \tag{2}$$

$$\frac{dY}{dt} = \frac{kAbe^{-kt}}{(1 + be^{-kt})^2} \,. \tag{3}$$

Substituting (2) in (3) gives

$$\frac{dY}{dt} = \frac{Ak\frac{A - Y}{Y}}{\frac{A^2}{Y^2}} \text{, or } \frac{dY}{dt} = \frac{k}{A}(A - Y)Y \tag{4}$$

which leads to

$$\frac{1}{Y} \cdot \frac{dY}{dt} = k - \frac{k}{A} Y \,. \tag{5}$$

This means that the relative or percent increase in Y, the left-hand side of (5), is linearly dependent on the absolute level of Y. Both the relative increase in output and the absolute output for the two-year period following rehousing can be determined in order to estimate the parameters of (5). Regression of these variables yields the following for the test and control groups respectively:[7]

$$\frac{1}{Y} \cdot \frac{dY}{dt} = \underset{(0.00624)}{-0.02649}\, Y + 41.9258 \tag{6}$$

---

[6]Interestingly, this same pattern has been observed in numerous other cases. See, for example, J. P. Davison, P. S. Florence, B. Gray, and N. Ross, Productivity and Economic Incentives, (London: Allen and Unwin, Ltd., 1958), who conclude from an investigation of the effects on productivity of changes in incentives that "such increases in output, most of them large,... were found not to be just a 'flash in the pan' but were sustained over the whole period of study"(p.203).

[7]The output used in these regressions applies for ten-day periods and is measured in hundreds of wons for each of the whole group of 50 workers.

$$\frac{1}{Y} \cdot \frac{dY}{dt} = \underset{(0.00594)}{-0.02252}\, Y + 34.5051 \qquad (7)$$

(Standard errors appear in brackets below the corresponding regression coefficients.)

From equation (5), the coefficient of Y is $-\frac{k}{A}$. Since A is the output maximum, this level can be calculated as the intercept values, 41.9258 and 34.5051 (=k) for the equations describing test and control group samples, respectively, divided by the corresponding regression coefficients. Adjusting the above to weekly production per man yields estimates of output maxima of 2,216 and 2,145 won per week for the average test group (rehoused) worker and average control group worker respectively. Compared with the pre-rehousing average performance level of 1,690[8] won per week for the average worker in each group, the productivity of the test group increased about 31 percent while the level of the control group increased 27 percent.[9]

Conclusions

The Korea test case has demonstrated that improved living conditions resulted in a significant and sizable positive impact on productivity levels among rehoused workers.[10] In addition, although the benefits of these improvements were not directly shared by the non-rehoused, in their competition for available space these workers also bettered their production records and, with increased incomes, raised their standards of living. Although our data limit us to short-run estimates of both this "demonstration-effect" and the direct productivity increases of rehoused laborers, the trends of the output curves strongly suggest that the jump in production is sustained at a new capacity level over the longer period. This speculation would seem justified, for once families accustom themselves to new, higher standards of living, they generally seek to maintain, if not increase, them further.

A few caveats are in order, however. The extent of broad generalization is always limited when conclusions are based on the evidence of a single case. Unfortunately, this seems particularly true in the context of economic development, an area of inquiry desperately searching for generalized principles but one where they are difficult to attain when behavior is so strongly conditioned by cultural traits indigenous to particular areas rather than common to many. Along the same line, unique factors of this case may account for the phenomena we have observed. For example, were it not for the performance-based occupancy criterion, it seems doubtful whether the "demonstration-effect" would have materialized. The effect could also depend on the number of units built in the project. In other cultures where housing may rank low in desirability, the psychological reactions in the form of motivation may never occur.

The results of other studies of the relationship between housing and productivity, currently in progress, will confirm or modify the conclusions and estimates of magnitude reported here. The purpose of all of these studies is to develop operational framework for evaluating the returns to investment in residential construction in order that housing can compete for scarce development resources on the same basis as alternative investments.

---

[8] Obtained by dividing the mid-point of the regression equations for weekly output (Table 1) by 50, the number of sample members per group.

[9] Most likely, the changes are conservative since the maximum productivity estimates of each of the two groups of workers were unfavorably affected by a temporary decline in productivity toward the end of the first year following rehousing.

[10] As we have noted elsewhere, when the benefits are fully counted and priced (most particularly, when the benefits to health are added), the rate of return on this housing investment was attractive and comparable with rates of return earned on capital invested in alternatives. See Tjioe and Burns, *op. cit.*, Section 7.

HOUSING DEMAND ANALYSIS IN DEVELOPING COUNTRIES

Lawrence N. Bloomberg, U. S. Bureau of the Budget

It is my belief that a successful housing program cannot be developed without intimate insight into the social, economic and political structure of a country. This is even more critical in the case of developing countries where there are few empirical data and the stresses of emergence cause conflict in the attainment of social, economic and political objectives.

The stolid application of dogma is futile. A pat uniform "do it our way" approach is the essence of sterility and can only produce a thorny straight jacket from which Houdini-like escape is both inevitable and painful.

For the purpose of this paper, housing is defined as shelter, with adequate protection against the greatest severity of the elements expected to be normally encountered, with space for its occupants sufficient to permit the performance of household functions and to meet the privacy requirements of the culture and with such amenities as are needed for the preparation of food and for the occupants' personal health and the public health of the community.

It will be noted at once that this is a very restrictive definition and is so intended. There is no reference to aesthetics nor to environmental or planning features, nor to the infrastructure required to support housing development. These, certainly, are matters of great concern in the preparation of housing programs but my intention is to trim fine and concentrate on basic housing demand.

We all know that demand for a house of a particular type, size and price can vary widely, depending upon its location within a community and its relationship to transportation and major sources of employment. This type of demand analysis is highly specialized in reference to a particular housing project and requires refined techniques and consideration of factors not covered here. I shall not, therefore, attempt to treat housing demand below the level of the community.

Even though I do not attempt to cover demand in sub-areas of communities I am convinced that national housing demand estimates must be built up by local areas within countries.

Generally, a housing demand estimate, based entirely upon national aggregates is virtually meaningless and can lead to grave program miscalculations. I only say "generally" because of two possible but not probable exceptions: that of a country whose population remains relatively fixed in its geographic location with little, if any, internal migration; or a country in which there is internal migration but with little or no effect upon the population of any local area. The latter is even more improbable than the former because the very reasons that cause population to move from an area are generally the same reasons that prevent others from moving in.

Use of national figures on population, or household, growth does not account for the effect of internal migration upon housing demand. Since housing is relatively fixed in its location, it is possible to have a substantial unfilled demand as a result of migration to one part of a country while numerous vacancies exist in another. This situation could occur in a country whose total population is static or even declining.

Ideally a national demand estimate should be the summation of estimates for all local market housing areas. This, however, is rarely, if ever, practicable. An absolute minimum would be a division between urban and rural. A second level, in ascending order of significance, would be to break total urban at some population size to obtain a distinction between smaller and larger urban places. Next, is to take all individual cities, towns or villages above some population and to group the balance of urban. Actually, in many cases it will be found that not much more can be done than to select one, or maybe two or three principal cities, for intensive analysis. Availability of data and ingenuity in producing them from diverse sources will largely dictate what can be done.

It is necessary in the first instance to draw a sharp distinction between housing "need" and housing "demand". I have defined these terms, with special reference to housing analysis, several times in the past [1] and will not cover them in detail here.

In summary, housing <u>need</u> is expressed in terms of the number of physical units required to meet an acceptable minimum standard, for example, a number of standard housing units of a given size distribution or a number of roofs of an acceptable utility and quality.

Elements of <u>need</u> consist of: changes in the number of households or occupant groups through net new household formation and net migration; undoubling of families in two or more family households; replacement of dwellings lost through demolition, fire and other disasters; replacement or rehabilitation of dwellings which fall below a minimum standard; amount of additional space required to relieve overcrowding; provision of components required for health or safety.

The basic concept of need is that price, or ability to pay, does not enter into its computation. To illustrate: if in order to maintain health, all children between the ages of five and twelve need a quart of milk a day, it makes no difference in an estimate of the total volume of milk needed for this purpose whether milk costs twenty-five cents or one dollar per quart or indeed, whether some children can pay nothing and others as much as five dollars.

Unlike housing need, the most important elements of an estimate of housing demand are the prices at which housing is, or may become available, the ability of families to pay and the proportion of income they are willing to devote to shelter.

Computation of demand is a particularly elusive exercise because of the many subjective elements involved. As difficult as it is in many cases to estimate family, or household, income distributions, a much more troublesome problem is to determine willingness to pay.

A principal virtue of the need-demand approach to housing analysis, particularly in developing countries, is that it shows clearly the gap between minimum housing requirements and the ability and willingness of the people to pay.

This gap can be closed by one, or more probably, by a combination of the following: reducing the cost of housing through operations in the private sector; increasing the incomes of families by economic expansion; or, intervention of the government through subsidies or loans.

Except for the basic analytical structure for housing demand estimates, which it is not my purpost to deal with here, the analyst must be prepared, in the case of developing countries, to modify substantially his concepts based on experience with advanced economies and take a fresh and uncluttered view. He must resist the enticement, however alluring and comfortable, of attempting to rationalize unfamiliar facts according to preconceived notions.

Too often, the analyst tries doggedly to fit what little statistical data he is able to accumulate into a tested empirical framework. He makes what I like to call Procrustean adjustments to his data. In his efforts to rationalize unfamiliar behavior of a population according to his preconceptions he is apt to be of greater disservice than benefit to a country in assisting in the formulation of a housing program. It is to some of these unique differences that I will primarily address myself.

Boulding, in his 1965 address to the American Economic Association [2] made the point that economists must be humble about the field of economic development of the poor countries. He said:

> "In the rich countries we have done fairly well; in the poor countries our record is distinctly spotty. This is almost certainly because we are dealing in this case with a total social process, and the economic abstractions are simply not sufficient to deal with the problem. Here, what we need is clearly economic anthropology, and this science, unfortunately, hardly exists."

While I agree entirely with Boulding as far as he goes, I think it is necessary to go even further. The social, or behavioral, scientist today must, in large degree, be an interdisciplinary entity. Actually, it is primarily his emphasis that places him in one discipline rather than another. In dealing with developing countries, the economist must not only be part anthropologist, but also sociologist, political scientist and even psychologist.

Malinowski, in a Scientific Theory of Culture, [3], says eloquently:

> "Again, economics as an inquiry into wealth and welfare, as means of exchange and production, may find it useful in the future not to consider economic man completely detached from other pursuits and considerations, but to base its principles and arguments on the study of man as he really is, moving in the complex, many-dimensional medium of cultural interests. Indeed, most of the modern tendencies in economics, whether labelled 'institutional,' 'psychological,' or 'historical,' are supplementing the old, purely economic theories by placing economic man within the context of his multiple drives, interests and habits, that is, man as he is molded by his complex, partly rational, partly emotional cultural setting."

Neither can the statistical equipment of the social scientist remain untouched. While the laws of mathematics and statistical analysis remain almost immutable, the data that the social scientist deals with in developing countries may have different meaning and he will secure unaccustomed results. For example, in our western culture [4] we are used to a high negative correlation between income of the male parent and the number of children he produces. In a polygamous society, the statistician may find exactly the reverse. He will discover, on analysis, that there is nothing wrong with his data but the higher the income, or wealth, the greater the ability to secure wives and these, under ideal conditions, inevitably lead to more offspring.

Before discussing some specific differences that might be encountered in developing countries, I want to mention a few considerations that affect housing need and demand estimates generally.

The analyst from temperate or cold climates, for example, may look askance at the flimsy small houses he encounters in tropical areas but the mud and wattle houses he finds in Africa may be entirely acceptable there. The materials are mostly free in nature; little interior space is required because most living is done outside; thatch is superior insulation against heat and moreover, it is not noisy in the rainy season. While the house may require almost constant repairs these are both easy and inexpensive to make.

To the contrary, if the analyst were accustomed to bamboo, palm and straw construction he might find it very difficult to consider that substantial masonry structures in a large northern U. S. city could constitute a slum.

But even climate and physical environment is not a completely predictable variable in its effect upon estimates of housing demand. As Forde points out in Habitat, Economy and Society [5]: man's activities, type of society and even religious beliefs have often been explained on the basis of climate and vegetation but, despite the intimate relationship, there are clear limits to this explanation, for in closely similar regions there are sharply contrasted societies.

Writing about shelter in primitive societies Herskovits [6] says:

> "Though a food supply is indispensable, it is quite possible to get along without shelter. Not many people, it is true, omit this item from their cultural equipment, but it always surprises the novice in the study of culture to what a minimum it can be successfully reduced.... In actuality, we find that the minimum of shelter necessary to preserve life is lower than casual consideration would indicate. Conversely, the maximum degree of shelter attained by most societies is so much greater than necessary for survival that we must seek other reasons to account for it."

It is obvious, I think, that housing is not generally a matter of survival. Although adequate protection against the elements is a first order, as I have indicated, in the provision of satisfactory housing, there are others of almost parallel importance. One of these is space. Part of the definition of housing I have given is that it must have space for its occupants sufficient to permit the performance of household functions and to meet the privacy requirements of the culture. Space is an important consideration in housing demand estimates because of its direct effect upon quantities of labor and materials and thus, the cost of dwellings.

Inter-cultural differences in attitudes toward privacy are important determinants affecting the amount and distribution of housing space in the present day urban environment. Of course, in many societies, customs with respect to entertainment of friends has a marked effect upon the sizes of dwellings. In others, home industry often requires space over and above that strictly needed for living. But it is to privacy differences that I want to direct particular attention.

Hall, in his excellent new book, The Hidden Dimension [7] discusses the need for privacy in various cultures. The American way is for each child to have "a room of his own." It is difficult for the English to understand this because the middle and upper-class Englishman is brought up in a nursery shared with brothers and sisters. "The difference between a room of one's own and early conditioning to shared space, while seemingly inconsequential, has an important effect on the Englishman's attitude toward his own space. [8]

An interesting footnote is the bearing that this space conditioning has upon inter-personal relations. When the American wants to be alone he goes off and shuts the door. He would never refuse to talk to someone else present in a room except as a sign of displeasure or rejection. The Englishman, on the other hand, not accustomed to architectural privacy, remains physically, but shuts himself off from conversation by withdrawing and bitterly resents attempts to talk to him while he is in this state [9].

According to Hall, there is no such word as "privacy" in Arabic [10]. The form of the Arab home is such as to hold the family together. "They avoid partitions because Arabs do not like to be alone." Like the English, their way of invoking privacy is to stop talking, indicating that they want to be alone with their own thoughts and do not wish to be intruded upon.

Attitude towards the home as a possession, or as a mark of prestige, is another general consideration affecting the determination of housing demand. This factor largely governs the extent of sacrifice people are willing to make in order to have a satisfactory dwelling.

It has been found in some developing countries, particularly those in which economic stress is greatest, that many people look upon the ownership of good housing as an investment for profit rather than the purchase of a home for themselves.

The Foundation for Cooperative Housing, in its 1962 study of housing cooperatives for the Agency for International Development, noted, in Kenya for example, that a number of private individually owned dwellings constructed with direct government financing and a land subsidy were being operated as investment properties [11].

Two years later we investigated this matter in detail in Kenya, and the above observation was fully substantiated [12] by an examination of all tenant-purchase plans being operated by local authorities. An experience in Nakuru is illustrative. We obtained the addresses (actually a description of the location) of a sample of "owners" who were not living in the housing they were purchasing and inspected their dwellings (from the outside only). Almost without exception these were found to be living in the lowest quality housing in the town, actually as close to a slum as one finds in Kenya. Further investigation showed that rents received by the owners were from 30% to 90% more than the payments to the authorities. Since down payment on the houses were only 10% of cost, the return on the original investment is fantastic since, aside from taxes which are not high, there are few other expenditures.

We observed the same practices in Uganda and Tanganyika (now Tanzania). From discussions with others I have learned that this kind of operation is by no means limited to East Africa.

To avoid being misunderstood, I want to say clearly that what I have described is by no means universal, even in the countries I have cited. It is, however, of enough significance to warrant attention of the housing analyst. Furthermore, in the formulation of housing programs there must be adequate safeguards, because the government of developing countries cannot afford to waste its scarce resources /13/.

Determination of housing demand, as I have indicated earlier, begins with an estimate of housing need. Need has basically three elements: growth in occupant groups, replacement of losses, and bringing the housing supply to an acceptable standard. So far as I know, replacement of losses in developing countries present no unique considerations. The other two factors do present some unusual problems and I should like to start with the matter of substandardness.

Although I want to deal with housing standards, I would be less than honest if I didn't say that, as a matter of fact, the analyst will probably give up in many cases before he is through. Either he will be unable to determine the criteria for substandardness in developing countries or if he does succeed, even with the most minimal measures, their application will result in such magnitude that any reasonable amelioration is almost beyond possibility. However, there is merit in attempting an estimate, even though crude, if it is possible to do so. If properly qualified, such a figure does give the government a rough measure of total housing need and more clearly indicates the gap between need and demand.

Measures of substandardness have been applied as, for example, by Badgely, Logtens and Schechter in their housing analysis of Surinam /14/. A housing survey was conducted on a sample basis in Paramaribo by sanitary inspectors. The housing stock was divided into adequate, repairable and beyond repair. On the basis of this survey the mission concluded that about 4,000 of 21,000 units had to be replaced - 3,000 already beyond repair and 1,000 of 5,000 repairable units which the mission estimated would not in fact be repaired.

Dodge, Young and Wilson say in their report on housing in Ethiopia /15/ "Setting minimum standards of acceptable livability in an economy such as that of Addis Ababa is an extremely difficult thing to do. No such minimum standards have been pronounced, or suggested, either officially or unofficially, unless the census classification by type of construction may be taken as a suggestion." The mission used the census figure of "non-weatherproof construction" which amounted to 57% of the housing supply in Addis, or 71,000 out of 124,000 units. Thoughtfully, the mission said that this is only to be considered a benchmark for present purposes of analysis; "This properly leaves to the future, and to local determination, the more specific definitions of what should be considered to be minimum standards of construction and occupancy in the interest of safety, health and livability /16/."

The Statistical Commission of the Economic and Social Council of the United Nations has been trying for many years to formulate "statistical indicators of housing levels of living." A report was prepared for the eleventh session of the Commission in 1960. Some of the many problems involved in formulating criteria for substandardness, so that countries can be compared, are illustrated by this report.

The report /17/ says "... the general aims of housing policy may be to increase the housing stock so that crowding beyond certain standards may be eliminated, the average density of occupation may be brought down to certain levels. All married couples may occupy separate dwellings or so that families may be able to move from improvised or other types of inadequate housing to conventional and suitable dwellings."

One of the statistical indicators proposed is the percent of the population living in housing units classified as "rustic," "improvised" and "not intended for habitation." Rustic housing units are defined /18/ as those "rudely constructed or erected (e.g. having mud walls, thatched roof, etc.) with locally available rustic materials such as stones, sun-dried bricks, bamboo, palm, straw or any similar vegetable materials for the purpose of habitation by a private household...."

In East Africa, for example, some 90-95 percent of the housing would be considered "rustic" under the above definition. Yet, if there were no other deficiencies, a substantial proportion of the housing, particularly in the rural areas and small towns, should not be regarded as substandard, under the conditions that exist in that area.

Two other criteria, applicable only to urban areas, are proposed, piped water and flush toilets /19/. The case for these, especially in urban areas, is a good one and probably can be applied universally.

They both illustrate, however, the analyst's problem, for the need may not necessarily be for new complete units but only to remedy deficiencies in existing housing. It is possible, for example, to bring piped water to, and even into, dwellings that are substandard solely because of this deficiency. Installation of flush toilets, or water-borne sanitation is more difficult, but nonetheless possible.

A related case arises in connection with overcrowding. A proposed indicator is the percent of occupied dwellings with three or more persons per room /20/. This standard generally represents over intensive use of living space but a culturally based analysis of identical data for different countries may yield significantly different conclusions as to housing need from this source.

It will be recalled that one of the stated aims of housing policy according to the United Nations is to provide all married couples with separate dwellings. Undoubtedly this aim derives from a concept of privacy and the desirability of developing a family life without outside influences.

Under this assumption it is necessary to analyze overcrowding data to determine whether it is due to families that are too large for the space they occupy, or a combination of families that crowd together in order to save rent, or the attachment of individuals to what is commonly regarded as a family in our western culture, that is, the extended family.

To the extent that overcrowding is due to large families, the need in aggregate terms may be for more rooms, not necessarily more dwelling units. It is certainly possible, theoretically, that overcrowding can be eliminated under certain circumstances by a better distribution of space. It could also be eliminated by the addition of rooms. Another possibility is the consolidation of some small dwellings for use of larger families and building new units of small size. Finally, the new dwellings could be of larger size and the newly formed smaller families moved into the units vacated by the overcrowded.

However, to the extent that overcrowding is the result of two or more families living together through economic necessity, not all of the above options are possible. In this instance the need is for more dwelling units of appropriate sizes.

Finally, interpretation of aggregate overcrowding data depends upon accurate knowledge of the country involved. As already indicated there are wide inter-cultural differences in the concept of space requirements.

I should like now to turn to that part of the need estimate resulting from growth. Since the house, or dwelling, is the primary unit for need and demand estimates from this source, any projection of requirements for additional housing is a function of growth in the number of occupant groups. In a standard housing analysis, population figures must be adjusted for any expected change in the number of persons per family or household. There are a number of ways in which this may occur: changes in the birth, death, marriage and divorce rates, to mention a few. These kinds of changes, however, are generally not cataclysmic and will not make a great deal of difference in a relatively short term projection.

I do wish to mention two cases in point that have substantial impact. While these situations could occur anywhere, they are more apt to be encountered in developing countries and in developing and declining parts of developed countries.

The first of these cases involves a population increase that is significantly greater than the increase in occupant groups.

In Kenya, according to its 1962 census, the urban African population consisted of 163 males per 100 females although for the country as a whole, the ratio was 98 per 100. For Nairobi, the capital city, there were 187 males per 100 females among the African population which represented about two-thirds of the total. For the adult population only, the ratio was 250 per 100.

Only in Kuwait, among the fifty-one countries for which the U. N. publishes this information [21] is there a city of over 100,000 population (Nairobi has about 300,000) with a sex ratio that even approaches that for Nairobi; only 11 of the 51 countries have cities of this size in which the number of males is greater than the number of females and but 4 exceed 130 per 100, with Kuwait (Kuwait City) having a ratio of 178 per 100.

An examination of the situation in Nairobi revealed, on a large scale, a case not unlike that of our own familiar mining camps of the early west. In Kenya, men were leaving their shambas and coming to the cities, particularly Nairobi, in search of employment. They leave their wives and children in the tribes and send for them only when settled employment is found.

In the 1948 census of Kenya, the sex ratio for urban Africans was 295 males per 100 females; this was reduced to 163 per 100 in 1962. Thus, a large part of the in-migration to urban areas between 1948 and 1962 consisted of women and children.

With the high ratio still existing in 1962, and with somewhat improved economic conditions in prospect, it is reasonable to expect a modest acceleration of the movement of women and children to the cities.

All of this has an important bearing on housing analysis. A large number of the adult males are lodgers, but a substantial proportion live by themselves in separate dwellings. Evidence of this is found in the fact that 22 percent of the dwellings in urban areas in Kenya are occupied by one person. In the U. S., one person households constitute about 15 percent of urban housing; a large part of these, however, are occupied by the elderly.

Based on an extrapolation of past trends, it was estimated that the population increase for Nairobi would be at a compound rate of 4.9 percent per annum for the period 1962-1970. However, a study of the recent trends in the sex-age distribution led us to the conclusion that a substantial part of the increase would consist of women and children joining men already in the city. Thus, the growth in households would be at a rate of 3.9 percent per annum, or about one-fifth less than the population increase.

The effect of this conclusion upon housing need is at once obvious. To the extent that the males were already occupying housing units, there would be no increase in the <u>number</u> of units needed. There would, however, be a need for units of larger size. It was therefore necessary to fit the expected population increase to the existing situation. The assumption was that newcomer adult males would occupy the units relinquished by the males who brought in their families and that new housing units produced had to be substantially larger. Moreover, adjustment was made for males who were living as lodgers rather than occupying separate dwellings.

In terms of net effect, the situation in urban Kenya could be expressed as follows. To take care of the increase in household, or family, size alone, the mean number of rooms per unit had to increase from 1.376 in 1962 to 1.761 in 1970. This computation assumed the same degree of overcrowding (half of the urban units had three or more persons per room) in 1970 as in 1962. Thus, the increase in number of rooms per unit would only keep the occupancy situation from deteriorating but would do nothing to improve it.

The other case I want to treat is the converse of the one already discussed. It was encountered a number of years ago in Puerto Rico which at the time I suppose it is fair to say, was a developing country.

In many areas, there was substantial outmigration but few, if any, vacant dwellings could be found. Study of the situation revealed that the migration consisted largely of younger persons, either unmarried and previously attached to family groups or young married couples who had been living with their families.

There was accordingly, a reduction in household size but a very small change in the number of households. A situation such as this can materially affect estimates of housing need. As in the case of an increase in household size, already discussed, it is important to analyze the causes of a decrease in size in order to assess its effect.

This is a complex type of analysis and the resulting effects upon housing need and demand depend upon the relationship of changes in household size to a number of other factors. It is probably best to illustrate the problem through several examples.

The first of these is the case in which there is no in-migration to the area and the loss in population is directly proportional to the decrease in household size. Under this circumstance the same number of dwellings would be required at the end as at the beginning of the period of estimate. The size distribution of the units would, of course, be different.

Need, and ultimate demand, in this instance would depend on the volume of housing to be replaced during the period and the ability of families to pay for replacements.

Another case is one in which there is a population loss as a result of net change over the period and the decrease in household size is relatively greater than that in the number of persons.

Suppose, for example, there are 1,000 persons in an area with five persons per household, or 200 occupied dwellings. Over a period, all of these households remain in the area, but on the average, one person leaves. This would reduce the population of these households to 800 but they would still occupy 200 dwellings.

Over the same period, suppose 25 families move in with an average of four persons per household. This would be a population increase of 100. The net change in population would be a reduction of 100, with 100 persons moving out and 100 moving in. The number of occupied dwellings required, however, would be an increase from 200 to 225.

These are very simple examples, I know, but the essential point is that a projected loss in population does not necessarily mean a corresponding diminution in housing need and demand, provided that there are reasons to believe that a change in household size is likely to occur.

Probably the most critical, but more frequently glossed over element in housing demand analysis, is the willingness and ability of people to pay for housing. While this can be said of housing analysis generally, it is even more cogent in the case of developing countries.

It is in this critical phase that the analyst, if his experience has been restricted to more advanced economies, is most likely to go awry. He must learn many new things. He will discover that the priority he may place on adequate housing is vastly different from that of persons in poorer countries, particularly those with tropical or mild climates. He has to see the problem through the eyes of the native population. He must not be judgmental as to what people should spend or save for housing.

Demand is the result of both price and the ability and willingness of people to pay for housing. I shall not, in this paper treat the supply side beyond saying, that in developing countries a substantial proportion of dwellings is constructed by the occupants; whatever home construction industry there is, often operates either for the government, or the limited number of higher income persons; there are a number of individual craftsmen who may be hired to assist a family to build its house or to build the entire house; in some cases, an individual will construct extremely shoddy houses as an investment, which generally means exploitation of the low-income population. There are many exceptions to this general statement but I believe that it is a fair picture.

Ability and willingness to pay for housing represent one of the most complex problems an analyst can face. The diversity among families of motivations, satisfactions and selection from among multiple choices creates a complicated pattern of housing expenditures.

At one extreme is the family that can barely exercise any choice in the allocation of its income.

Below this marginal family are those who cannot purchase sufficient food and clothing, much less housing. At the margin it is necessary to determine the amounts required for other primary necessities such as food and clothing (assuming that these have priority over shelter) and the balance of income represents the amount that could be spent on housing and other minimum requirements. Basically, at this level, demand would, therefore, depend almost entirely upon the minimum price at which housing could be supplied.

As income increases from the margin and families can begin to make choices in its apportionment, the principal factor is the degree of priority families place upon housing. Some families will give education of their children a precedence over housing. Others would rather have consumer goods such as transistor radios, better clothing, bicycles or even automobiles. In some families, the man, who often controls the purse strings, may place a higher priority on drink and other pleasures.

But none of these problems is limited to developing countries. Yet in these countries they are exacerbated by a number of additional factors, a few of which I want to cover.

In many developing countries a substantial part of the population is not used to paying for housing at all. Even in urban areas there is a vast amount of "squatting." Typically the use of a piece of land is acquired either under a law which makes it easy, or because authorities will not, or cannot interfere. The squatter generally builds a flimsy shelter of odds and ends. Authorities do not often take action, whether the squatting is legal or illegal, unless they can provide alternative housing. To do so, would in some instances, be politically indefensible.

Another factor that complicates analysis of expenditures for housing in developing countries is the often encountered custom, particularly in former colonies, of providing housing as part of the wage structure. Housing may be provided without cost or at a modest amount.

In Kenya, one-third of all urban housing for Africans is supplied by the employer. In Nairobi, nearly 45 percent of the housing is of this character. A senior officer in the Kenya Government occupies a government-owned house at a rental of £42 per annum although a fair market rental would appear to be about £375. If an officer should own his home he receives an annual allowance of 10 percent of its capital cost.

It appears easy to say that the entire wage structure should be reivsed to cut out payments in kind by both government and private employers. This, however, presents many difficulties.

Not all employees have been treated alike, either in the amount of payment or in the quality and amenities of the housing provided. To work out a uniform and equitable revision of the wage scale is a gigantic task (the Kenya Government, however, has announced its intention to do so.) Moreover, in talking with government officers in East Africa I discovered that, in some cases, they were balancing the savings on income taxes through payments in kind, which were not taxable, against increased salaries which form the bases for their pensions.

Both of these factors, squatting and employer furnished housing, makes it very difficult to place expenditures for housing in perspective. Even if data are available for the non-affected part of the population, the inter-effects of these two segments on the housing market as a whole makes analysis a nightmare, but this is a factor that cannot be ignored when it is encountered.

I have already said that determination of ability and willingness to pay for housing is the most critical phase of demand analysis and have given several examples of complicating factors. Despite complexities the analyst must try, because very small miscalculation on this point has enormous consequences. Given the estimated income distribution of Africans in Nairobi, Kenya we find that if families can or will pay no more than one-tenth of their incomes for housing, 22% can afford 30 shillings or more per month. However, if the proportion of income is increased to one-seventh, the percentage of families affording this amount doubles. At one-fourth of income, the percentage triples. [22]

Complicating the expenditure pattern still further, at least in East Africa, are charges against income not generally encountered, particularly among lower income families. One of these is school cost, because free public education is virtually non-existent at present.

Another is the custom of sharing income with the family that remains on the land. It was explained to me by a government housing manager in Jinga, Uganda - an African who had been well educated at Makerere University. He said the payment did not spring entirely from altruism but that it should be regarded more in the nature of insurance. So long as contributions are made by the urban family to the part of the family remaining in tribal areas, the urban portion may return and be cared for in the event of illness or unemployment. Although this custom is declining, it is still strong.

Of course, adequate budget studies would reveal this situation. But these are scarce and at best, there will generally be only data on total income. In this case, the analyst might well overestimate the proportion of income that might be devoted to housing if he is unaware of the fact that part of the income is not available to the immediate family for consumption purposes.

I do not say this in criticism - because I have been guilty myself, although I have tried - but most of the studies I have seen pay little attention to this significant factor. It is true that adequate data are almost entirely lacking, not only for developing countries but for advanced economies as well.

An often used device, when the data are available, is either to relate housing expenditures to total income or total expenditures. Many times this is done without regard to the level of income and yet we know that generally, expenditures for housing as a percent of income, decline as income increases, but not so rapidly. That is to say: expenditures for housing increase as income rises but not in proportion.

Even here, I can give an exception. A rather carefully done study of Mombasa, Kenya showed conclusively that until a relatively high income was reached, absolute expenditures for housing remained almost constant. This is a fair indication, I believe, of a low priority placed on housing, relative to other goods and services.

Even within given income groups there is wide dispersion of amounts families spend on housing, reflecting the tastes, preferences, ability to obtain housing and a multitude of other factors that go into individual decision making. I have not even mentioned the effect of family size upon housing expenditures. This, of course, introduces even more complications.

The simple fact is, and I say this advisedly, that we have devoted little research and effort at data collection that would throw light upon the significant subject of what people can afford and are willing to pay for housing.

Finally, in this respect, as in others I have mentioned, the housing analyst should not, especially in developing countries, approach his subject with a fixed set of ethical or moral principles. He must recognize that the high value we place on hearth and home may not be so high elsewhere in the world and that the sacrifice of today's consumption in order to save for the future may not be regarded elsewhere as the virtue we think it to be.

I have not attempted in this paper to describe the serious housing conditions in many developing countries, particularly those with extremely rapid urbanization. Nothing I have said should be interpreted as downgrading the significance of the housing problem. My intent has been solely to point out that housing problems are relative to the characteristics of countries and that no single standard for measuring them can apply.

Despite the fact that differences among nations cause conflict, it is true that cross fertilization often yields new and vigorous species while pure strains frequently run a course of sterility. We have much to give from our own background and experience but it cannot be applied blindly. We can assist but must not impose. We need to understand and interpret - not in our terms but in those of the countries with which we are dealing.

References

1. Bloomberg, Lawrence N. Housing Market Analysis for Puerto Rico Urban Renewal and Housing Corporation, 1959 (English), U. S. Agency for International Development, Housing and Planning Division, Santiago, Chile, 1963 (Spanish).

Bloomberg, Lawrence N. and Charles Abrams. United Nations Mission to Kenya on Housing, Department of Economic and Social Affairs, United Nations, 1964.

2. Boulding, Kenneth E. Richard T. Ely Lecture, American Economic Review, Vol. LVI, No. 2, 1966, page 11.

3. Malinowski, Bronislaw. A Scientific Theory of Culture and Other Essays, Oxford University Press, 1960, pages 5-6.

4. Women by Number of Children Ever Born, PC (2) 3A, Table 37, U. S. Bureau of the Census, 1960. Note: In 1960, the negative correlation did not hold for the age groups under 25 years for white women and was not high for the age group 25-29 years.

5. Forde, C. Daryll. Habitat, Economy and Society, E. P. Dutton & Co., 1956 ed. page 3.

6. Herskovits, Melville J. Cultural Anthropology, Alfred A. Knopf, Inc., 1955, pages 128-129.

7. Hall, Edward T. The Hidden Dimension, Doubleday and Co., 1966.

8. Ibid., page 130

9. Ibid., page 131

10. Ibid., page 148

11. Schechter, George, Alfred Mercado, William S. Tennant. A Report on Cooperative Housing and Related Activities, FCH Company, Inc. and the Foundation for Cooperative Housing, Inc. for the U. S. Agency for International Development, 1963, Kenya, page 10.

12. /1, Kenya, page 10/

13. Some possible safeguards are described in the Kenya Report /1, pages 40-41/. The purposes of the proposals is to make it clear that Government is not subsidizing the property but it is willing to assist the purchaser so long as he uses the property for purposes for which it is intended.

14. Badgely, L. Durward, Peter C. Logtens, Henry B. Schechter. Housing Mission to Surinam, Report of September, 1962, U. S. Agency for International Development, pages 43-46.

15. Dodge, J. Robert, Burton O. Young, L. Albert Wilson. A Survey of Housing in Ethiopia with Special Emphasis on the Capital City of Addis Ababa, U. S. Agency for International Development, 1965, page 56.

16. Ibid., page 58.

17. Statistical Indicators of Housing Levels of Living, prepared for Eleventh Session of the Statistical Commission of the United Nations, E/CN.3/R.2, 1959, par. 8.

18. Ibid., par. 11.

19. Ibid., pars. 20-26.

20. Ibid., par. 15.

21. Compendium of Social Statistics, United Nations, 1963, Table 6, page 84.

22. See [1, Kenya,] page 21.

DISCUSSION

Byron L. Johnson, University of Colorado

It is clear from these papers that American statisticians have to add a bit, to use henceforth a measure of variance about per capita income. We need a "Sigma" for p.c.i.

It is also clear that statisticians must help us economists by taking into their toolbag the divergent concepts of "flow" and of "stock" as Prof. Boulding has suggested. Housing is clearly not just a "flow" but also a most important "stock".

But most of all, statisticians and economists should join the public in a discussion of the purposes of development. Two lines of approach are visible today:

One view of development is oriented to "productivity", to things. It sees development in terms of building steel mills, power plants, fertilizer factories, assembly lines, bottling works, paper mills, highways, etc. But it leaves out housing, because it is confused as to whether housing is "productive".

The other line approaches development in terms of people--it is "consumer oriented" and wants better and more adequate food in human diets, better health and facilities to assure it, and it wants more adequate, varied, and attractive clothing and SHELTER. It wants to raise standards of living, and enlarge human opportunities. For this group, housing is part of the basic purpose of development. It is <u>the</u> major consumer good, providing status, privacy, liberation from chores, as well as shelter from the elements.

In passing, I regret the use of the term, in Stan's remarks, of the "burden" of housing. One doesn't speak of the burden of providing food or clothing, or education. Let us speak of the benefits.

Of course, housing is actually productive. What I take as a first premise has now been proven by Tjioe and Burns. But what a tragedy that we should require proof of what ought to be obvious. I'd like to suggest that as their studies continue, they extend the concept of productivity of housing beyond the coal mines and other material productivity of the worker-tenant, to the productivity of housing in serving as a long term consumer good--yielding continuing satisfactions to all consumers of it, the whole family and their successors. Housing produces a demand for public services, for amenities, for the conveniences of life, for the water, sewer, gas, electricity, streets, telephones, sidewalks, parks, household appliances, as well as the building materials and labor used in the house itself.

Development planners trying to increase the employment of the labor force ought to put housing high on their list for its direct and indirect employment--and for the training which it provides that carries over into skills needed in other production, and other construction.

Housing underlines the need for economists and statisticians to be more cautious and prudent in their sharp distinctions between consumption and investment. Housing is both, and so are outlays for health, for education, and much more. We allow our labels to get in the way of our good judgment. Certainly the quality of the composition of the gross national product is quite as essential to welfare as its purely numerical quantity.

Let me turn to Smigel's paper on the Financing of Housing.

One must applaud vigorously his call to adopt a concept of economic rent. It is essential to seek an end of housing subsidies, as fully as possible. They are almost inescapably going to lead to favoritism, to corrupt politics, and to be only partial successes. For budgetary logic runs against them. Other demands have higher priorities. I've helped prepare budgets, and helped vote them. Look at the realities: Housing takes a lot of money to help a relatively small number of families. A smaller amount of money, spent on programs that immediately reach the entire community, win more votes, win more public acclaim, and thus carry the day.

Subsidized housing programs come to an end whenever this logic prevails, or whenever money runs out. Housing that pays an economic rent can provide most of the funds needed for more housing to be built tomorrow and the day after.

Housing is a major source of savings, or can be. In the United States, it constitutes now a debt commitment, in mortgages, around $300 billion. The repayment on these contracts is providing almost $40 billion a year. In addition, the hope to be able to buy a house brings prospective buyers into savings institutions, to accumulate their down payments. Their purchases of appliances, furniture, etc., on time payment plans is another form of contract savings.

By contrast, subsidized housing is a drain upon other savings, a drain upon taxes and operating budgets. It is bad economics. It encourages a "something-for-nothing" psychology. It abandons the utility of the savings idea, the return-flow of funds idea, and in the process, subsidized housing discourages economic housing in all sectors of the market near the subsidized housing.

Subsidized housing is also bad politics and bad social policy. If there are two prices for the same good in one market, someone will try to profit by trading on the difference--and he will succeed. The favoritism will hamper development. Even the indirect subsidy of employer and government housing is open to criticism. A developing society needs to plan for mobility. Subsidized housing means a scarcity of housing, and it means that people tend to be frozen in their localities and in their jobs. Such employer housing provides a powerful control over people. It is much better to plan to facilitate easy shifts from job to job, and from place to place. To plan for a future in which there is more freedom, and more development, will require more of the mobility of persons that is common in a developed society. This calls for a freer market in housing, and the use of economic rent.

And to ask, as Smigel does, for more rental housing, is all to the good--but little rental housing can be expected to be forthcoming commercially so long as subsidized housing is common, except at the other end of the income spectrum.

One could wish that papers on financing housing had room to explore the difficulties of housing finance with the legal systems in some developing countries. The high cost and delays in processing acquisition or transfer of property, in completing contracts, in verifying title, in paying taxes, and in securing essential services so that development may occur--these are hidden, but very real costs of housing.

And because housing depends so much on local government for many services to the land and structure, it would be well if there were broader understanding of the differences between the highly centralized governmental structures we associate with French and Spanish influence, with the decentralized structures that the U.S. takes for granted and finds in Germanic or Anglo-Saxon influenced areas. When a local government can levy its own general property tax to provide water, sewer, roads, schools, and the like, housing can be developed more easily. The talk of land reform often overlooks the general property tax, in local hands, as the best single tool of land reform.

Finally, it is necessary to say another word as footnote on Housing and Foreign Loans. Of course housing is built with local currency, spent on local land, local labor, and almost entirely on local building materials. Foreign currency is needed to pay only for those components that cannot be produced locally. But that is not the end of the argument. Long-term low-interest foreign loans are generally hard to get. Depending on the purpose of the loan, it has political problems in the lending or the borrowing country, or both. Housing happens to be free of this hostility. It is a purpose that appeals to both lender and borrower--both are willing to have it be a fairly long-term loan. The lender has the assurance that the income from the housing provides a reasonable assurance of funds out of which repayment can be expected--though currency conversion questions remain. And under investment guarantee programs, private sources will lend. Hence if a country shows maturity and good judgment in allocating to other development requirements the foreign currency fruits of such housing loans, these loans deserve consideration. The converse also holds.

There are good reasons for making pilot loans, for demonstration projects, and for using small foreign loans to provide technical assistance to the whole housing industry.

In this connection, the use of the cooperative as a way of accumulating funds, planning large housing development, cutting overhead costs, and training persons in sound management methods, deserves special attention. The growing success of cooperative methods overseas attests to the wisdom of the Humphrey amendment to the AID act which brought the Savings and Loan organizations, the Credit Unions, and the Foundation for Cooperative Housing into the whole issue of financing, organizing, planning, and building housing--and then managing it, as a part of development.

Let us now turn to Larry Bloomberg's paper on housing demand. This is a provocative piece, listing many complications that arise in making a housing demand analysis. I have no quarrel with what he said. But since this is the occasion to look at the conceptual framework of such analysis, let's add that had he wanted to make the picture even more complex, he would properly have added several sections not found here: In the interest of time, I use only an outline approach -

1--What is the purpose of a housing demand analysis?
   a. Government programs: loans, grants, insurance, cooperative promotion, savings programs,
      Or economic budgeting for health, education, or other public services,
      Or other socio-political analysis
   b. Private development--a straight market sales analysis
   c. City Planning--where to build how much of what? The market requirements sector by sector
   d. Design analysis

2--Income Class and Social Class analysis
   a. Housing as an incentive to marriage and family--its lack a deterrent. Its importance to the emerging classes. Policy questions as to priorities to be encouraged. The risks of being too hard-nosed about housing.
   b. Problems of up-grading amenities and community services as a contribution to housing analysis (water, sewer, health, roads, community facilities)

3--The Time Dimension
   What development itself will do to demand over time as to requirements for numbers, for size, for amenities, for privacy--and the adjustments to urbanization, industrialization, rising incomes, breakup of the extended family into nuclear families.

These were all good papers, contributions to the literature.

DISCUSSION

AGENCY FOR INTERNATIONAL DEVELOPMENT

BUREAU FOR AFRICA

Richard Metcalf

I will discuss this morning's three papers in a different order than presented. I believe this will set forth more clearly the relationship of each of the topics discussed by panel members to the preparation of a housing program for a developing country.

I have been indirectly involved in the International Housing Productivity Study from its inception. In effect what UCLA is trying to do is to "better the position of housing as a form of social overhead investment in the competition for development capital." There have been over 40 studies in the U.S. and aboard which have analyzed the relationship between improved housing and health and various social characteristics. But to my knowledge this is the first study which successfully relates improvements in housing to productivity.

Because of lack of time imposed on the speakers, Dr. Tjioe and Dr. Burns were unable to give you a complete explanation of the background and thinking that led A.I.D. into supporting this study. I believe it may be of value if I summarize some of the steps preceding the Hambaeck research.

One of the problems of international development is to pinpoint those areas where dollar investments in the form of loans or grants will make the most impact toward economic development. A.I.D. over the years has made loans and grants, particularly in Latin America, amounting to many millions of dollars. Some of this money has gone into housing and urban development.

Up until now there has been an unfortunate lack of knowledge of the effect of investment in housing on economic development. The Hambaeck study supports the hypothesis that investment in housing is a proper tool of economic development.

It is hoped that once UCLA has completed all of its studies in various parts of the world we will have a better picture of how housing affects productivity under different social and cultural conditions.

As the next step it would be interesting to see which, if any factors other than housing, would have an equal or larger impact on productivity. We may find that the same or greater increases in productivity could be obtained with a lower capital investment. For instance, maybe clean water or air or electricity would have a similar effect on productivity at a lower cost. I for one would like us to try to pinpoint particular improvements in living conditions which would provide the largest increment in productivity with the smallest investment. This may turn out to be an improvement other than housing in certain cultures although undoubtedly housing will continue to be close to the top of the list.

Whenever programs are initiated to improve housing the question of housing for whom, how much, and at what price is immediately raised. Dr. Bloomberg's paper directs itself at this problem.

Housing market analysis, even where current and reliable statistical data are available is a tremendously difficult field because of the many variables which influence the analytical process and the unpredictability of human attitudes. Dr. Bloomberg is most sensitive to this problem. I would like to add an additional complicating factor to those referred to by Dr. Bloomberg which adds further to the difficulties of the practicing market analyst. This is the virtual absence in the developing countries of any attempts to estimate housing needs in terms of the ability of families to pay for housing. For example in a developing country with which I am familiar, government agencies and private builders continued to build high priced housing even though vacancy rates in some of the completed units amounted to 90%. They were surprised that the units stood vacant in a city confronted with a terrible housing shortage. In this instance they simply did not understand the economics of housing demand. Even the most rudimentary market analysis would have indicated a very limited demand for high priced units.

A housing market analyst stepping into this kind of situation is confronted, in addition to the usual complications of his profession, with the task of convincing local officials of the value of housing market analysis and to overcome the skepticism that valid estimates of absorptive capacity for housing can be made.

Sometimes this skepticism is founded on a suspicion of the accuracy of local statistics. I would add a word of caution, already implied in Dr. Bloomberg's paper, that the analyst should always be aware of the fact that in some countries statistics become a political tool. Population statistics in particular are frequently juggled for political reasons.

In one particular case I recall that population count of a city had been inflated by a healthy percentage. Everybody knew that the population count was unrealistic but no one knew by how much.

I would also like to amend Dr. Bloomberg's statement that "space is an important consideration in housing demand estimates because of its direct effect upon quantities of labor and materials and thus the cost of dwellings." To this I would add the thought that cultural traditions may have an equal effect on housing costs.

Let me give you an example. In some cultures, tradition requires that the wife be out of sight of guests while she is preparing food. This tradition called for the relocation of the kitchen and addition of a wall in a housing project being built with U.S. private capital. The extra cost involved was considerable and to this extend affected the market.

The most serious problem in international housing market analysis, as Dr. Bloomberg points out, is to obtain reliable family income data and estimates of the proportion of income families are willing to allocate for housing. This is an area that requires a lot more study than it has received because a very small error can have a significant impact on the ultimate schedule of effective demand.

Dr. Bloomberg gives several examples in his paper of the sliding relationship of income to housing expenses to which I would like to add my experience. Usually expenditures for housing as a percent of income will decline as income increases. However, I found that in some Latin American countries the percent of income allocated to housing increased as family income increased, although not as rapidly. The trend reversed itself as income increased beyond the middle income level. The percent of income allocated to housing declined as may be expected as incomes increased into the upper income level.

Closely related to the estimates of ability and willingness of families to pay for housing is the ability to make a down payment on sales housing. This is an area of particular interest because of the large scale housing guaranty program administered by A.I.D. It is another area requiring more study.

Experience shows that the market for sales housing may be cut by as much as 50% because of the inability of families to save sufficient cash for a down payment. This is particularly true in countries with a substantial inflation.

Assuming that the market analyst has estimated with reasonable precision the demand for housing and is ready to recommend construction of a specified number of dwelling units at a given price the next question is how to finance these units.

Mr. Stanley Smigel's paper directs itself to this problem.

The principal problem of building housing in developing countries is lack of capital. This holds equally true for the private developer and the public agency. As Mr. Smigel points out the other components of housing are available. For instance land, while it may be high priced, it nonetheless will be available, to the builder. (I may point out that urban land costs in developing countries may comprise 50% or more of total development costs.)

Building materials are also usually available. In some countries certain materials such as glass and plumbing fixtures may have to be imported and may be expensive.

Labor is also available. It may be relatively unskilled and productivity may be low. While wages may not amount to more than one or two dollars a day the total cost will be relatively high because the builder might have to hire 10 workers in a developing country to do a job one or two men would do more efficiently in the U.S. The reason for the low productivity is lack of skills, lack of power tools, lack of organization, and lack of planning. For instance, and I've seen this actually happen, masons will erect a wall, the plasterers will come and finish the surface, to be followed by the plumbers who tear great holes into the newly completed wall to place water and sewer pipes. The masons come back and patch up the scars followed by the plasterers. As soon as everything is back in order, in come the electricians tearing down the plaster, knocking holes into walls, again to be followed by masons and plasterers, etc. I remember estimating that destruction costs probably amounted to close to 10 or 15 percent of construction costs in one construction job I inspected.

But these are problems that can be overcome. However, the critical lack of capital for housing is a problem that will remain with the developing countries for a long time.

A few comments appear in order on the link index system which Mr. Smigel discussed at some length. My impression from what Stan has said is that the system when properly applied is extremely complex. In the African area, to my knowledge no country has so far adopted the system. One reason may well be that the African countries for the most part have not experienced the tremendous inflation of some of the Latin American countries. Another reason may be that it takes a certain amount of sophistication and administrative "know how" to properly manage a link index system. The needed skills may still be lacking in many parts of Africa. And I for one would like to see a lot more study of the system before recommending adoption of the link index system by African nations.

Mr. Smigel in his paper has touched upon the reasons why housing receives such as small share of the limited available capital resources -- both domestic and international. Without going into this problem at any great length I do believe the following comments are in order to bring this morning's session into sharper focus.

The relatively limited investment in housing in the developing countries may be attributed to what I consider two misconceptions: that housing investment only serves a social purpose with no or little economic significance and (2) that the people of the developing countries are unable to accumulate savings and that housing therefore must be financed by the government.

Today's discussion pointed up the fact that housing is an integral part of economic development, not only in the industrialized countries, but also in the developing countries. Certainly the experience of AID in assisting savings and loan associations in the developing countries fully supports the conclusion that the people in developing countries have the capacity to save.

A frequent argument heard in the developing countries is that agricultural and industrial development must be placed ahead of improvements in standard of living and expansion of consumer goods. This conclusion is based on the need to develop foreign exchange producing industries and income producing investments ahead of gratifying the consumer. However, this argument overlooks the fact that housing is required in developing countries to assure continuance of economic development.

In the United States we are fully aware of the vast contribution the home construction industries make toward full employment and economic development. There is to my knowledge no economic evidence which would support the conclusion that the home construction industries do not make a similar substantial contribution to the economies of the developing countries.

VII

STANDARD OCCUPATIONAL CLASSIFICATION

Chairman, ANN MILLER, University of Pennsylvania

THE CLASSIFICATION OF OCCUPATIONS: SOME PROBLEMS OF SOCIOLOGICAL INTERPRETATION*

Robert W. Hodge and Paul M. Siegel
National Opinion Research Center, University of Chicago

According to Webster's an occupation is, among other things, "the principal business of one's life: a craft, trade, profession or other means of earning a living."[1] Albert J. Reiss, Jr., takes us a little further toward a viable definition of the objects considered in this paper:

> The social valuations attached to work in a society may be thought of as referring to both the kind of work a person does and the situation in which one works. The specific kind of work a person does in a socially evaluated work situation generally is thought of as a job, while an occupation refers to job characteristics that are transferable among employers.[2]

Reiss continues by setting forth the major components of work and work situations:

> The major characteristics related to the kind of work a person does are (a) the task, whether one manipulates symbols, physical and/or social objects (this is what is usually given in the "job description" of the Dictionary of Occupational Titles); (b) the prerequisites for entry into the work, such as educational, training, certification, and experience requirements; (c) the kind of social organization for the task, whether individual or group, for example; (d) the structure of interpersonal relations in the task, as the nature of work supervision; and (e) the structural demands made of the worker, such as responsibility for physical or social objects. The work situation includes a variety of conditions in addition to these, such as (a) the institutionalized setting, whether a factory, office, hospital, etc.; (b) the community setting of the work situation, as for example, its place in a labor-market area or the structure of the community in which it is located; (c) the perquisites of the job, as tenure, retirement benefits, etc.; (d) the rewards, such as income, recognition, etc.; (e) the type of industry; and (f) the class-of-worker, whether private, public, or self-employed. The particular configuration of these elements defines a particular job The classification of these jobs into larger categories of work requires the selection of characteristics in respect to which the jobs remain homogeneous. Work categories called occupations generally are defined by characteristics designating the kind of work and include, as a minimum, a definition of the task[3]

So conceived, the particular jobs which must be grouped to form occupational classifications are themselves defined by the logical intersections of a wide range of mutually exclusive and exhaustive classifications of jobs according to more refined aspects of their formal definitions. Heterogeneity in the taxonomic principles utilized to create occupational classes and diversity in the dimensions needed to write detailed job descriptions create the major difficulties encountered in the sociological use and interpretation of occupational data. In this paper, we develop the view that research on the division of labor is hindered by the multidimensional character of occupational typologies currently employed and would profit from efforts (1) to reduce the dimensions employed in constructing occupational typologies and (2) to work out more systematically the logical possibilities admitted by the variables used in developing an occupational classification The great need for continuity from decade to decade in the form of published statistics requires, however, that any energies expended in the creation of new occupational groups should ultimately supplement rather than replace existing modes of classification.

*The research reported in this paper was undertaken as background to a larger project dealing with occupational stratification in the United States. The study is being conducted under a grant to the National Opinion Research Center from the National Science Foundation (NSF #GS725, "Occupations and Social Stratification"). We are grateful to the National Science Foundation for its support.

[1] Philip Babcock Gove, ed., Webster's 3rd New International Dictionary (Springfield, Mass.: G. and C. Merriam Company, 1961), p. 1560.

[2] Albert J. Reiss, Jr., and others, Occupations and Social Status (New York: The Free Press of Glencoe, 1961), p. 10.

[3] Ibid., pp. 10-11.

PRINCIPLES OF OCCUPATIONAL CLASSIFICATION

On the whole, sociologists and other social scientists have shown but little interest in the problem of allocating individual reports upon jobs and occupations into relatively detailed occupational categories, such as those identified in the 3-digit census occupational code. Instead, they have been more concerned with the problem of aggregating relatively detailed occupational codes into a few broad classes of occupations. For the moment, however, we postpone commentary upon questions of aggregation and address ourselves to some disturbing particulars of the Classified Index of Occupations and Industries and the detailed census occupational code.

Most of the criteria of occupational classification enumerated by Reiss are manifest in the 3-digit or detailed occupational code of the U. S. Bureau of the Census. Many of the occupations identified in the detailed classification are, of course, groupings of individual jobs largely upon the basis of similarity in the tasks performed by their incumbents. For example, only individuals who report their occupation on the census schedule as "optologist" or "optometrist" are classified into the detailed census occupational title identified as optometrists. These two jobs engender the performance of nearly identical tasks and grouping them into a single occupation seems to be effected primarily upon the basis of their task similarity. Detailed occupational codes created in this manner are the ones best adapted for sociological use. They do not contaminate the occupational code with the rewards, work contexts, and other variables associated with the duties of particular jobs and, consequently, do not undermine the investigation of the associations between job tasks and these variables.

While detailed occupations such as optometrists, dentists, and bus drivers which adhere primarily to the principle of task similarity in the individual jobs comprising them are frequently found in the census code, other detailed occupations are clearly formed by invoking other criteria such as those mentioned by Reiss. With the notable exceptions of nurses, professional; nurses, student professional; and apprentices in various trades, the training, experience, and certification requirements of jobs do not appear to have been extensively used in forming detailed occupational codes. These criteria are, however, used to form the major census occupational groups from the detailed classification, a point to which we return below.

The social organization of tasks is likewise used rather infrequently in creation of the detailed occupational codes. Such detailed occupations as social and welfare workers, except group; and deliverymen and routemen (as distinct from truck and tractor drivers, hucksters and peddlers, and messengers and office boys) appear, however, to have been formed primarily with reference to the social organization of the task. Relationships of subordination and superordination--a feature of the structure of interpersonal contact in work tasks--are utilized to define some managerial titles in the detailed census code and are explicitly introduced to identify foremen (not elsewhere classified); carpenters' helpers, except logging and mining; and truck drivers' helpers. The structural demands made of workers are also employed in the detailed census code, usually by making explicit reference to the specific physical or social objects for which employees are responsible. Some examples of detailed census occupations employing this criterion are personnel and labor relations workers; public relations men and publicity writers; buyers and shippers, farm products; inspectors, scalers, and graders, log and lumber; and fruit, nut, and vegetable graders and packers, except factory.

In addition to various aspects of the kind of work a person does, the detailed census occupational classification also employs various aspects of work situations to define occupational groups. The institutionalized setting of jobs is, for example, frequently employed to identify the detailed codes in the census classification. Various types of attendants are separated almost wholly according to this criterion: attendants and assistants, library; attendants, physician's and dentist's office; attendants, auto service and parking; attendants, hospital and other institutions; attendants, professional and personal service (not elsewhere classified); and attendants, recreation and amusement. Not all of these attendants, of course, perform similar tasks, but most of them do perform tasks essentially the same as those performed by individuals whose jobs are classified in other detailed census occupations, as, for example, appears to be the case with attendants, physician's and dentist's office and receptionists. Other examples of the use of institutional settings to define occupational groupings are scattered throughout the detailed census code. Some particular instances are buyers and department heads, store; dressmakers and seamstresses, except factory; and the distinctions between many kinds of service workers according to whether or not they are employed in private households.

Other aspects of work situations utilized to define detailed occupations in the census code are industry and class of worker. Examples of the use of industry include the distinction of conductors, railroad from conductors, bus and street railway; the distinction of painters, construction and maintenance from painters, except construction and maintenance; the distinction of graders and sorters, manufacturing from inspectors, scalers and graders, log and lumber, and fruit, nut, and vegetable graders and packers, except factory; and of sewers and stitchers, manufacturing from dressmakers and seamstresses, except factory. The use of class of worker is evident in such detailed occupational titles as inspectors, public administration (public administration itself being an industrial code defined by class of worker); farm laborers, wage workers; and farm service laborers, self-employed. In addition the Bureau of the Census employs both industry and class of worker distinctions to subdivide managers, officials, and proprietors (not elsewhere classified) in many published tabulations. Industry alone is used similarly to subdivide salesmen and sales clerks (not elsewhere classified); operatives and kindred workers (not elsewhere classified); and laborers (not elsewhere classified) in publishing tabulations.

The foregoing discussion makes abundantly clear that the occupational classification currently employed by the U. S. Bureau of the Census is at best a multidimensional typology. Occupational codes are not formed solely by reference to the similarity in tasks performed by individual incumbents of specific jobs. Instead, work settings, industrial affiliation, and other factors are used to define detailed occupational groups. To the extent that this is the case, the operational meaning of occupation begins to depart from the notion of a classification of jobs requiring skills transferable among employers. At this point, difficulties are posed in the sociological use of occupational statistics.

The student of the division of labor more than likely will carefully distinguish between several forms of labor specialization and regard the relationships between these forms as empirically problematical. Common distinctions are those between the territorial, the task, and the functional division of labor. The division of labor according to tasks refers to the grouping of particular work activities--such as typing, painting, reading, operating a floor polisher, etc.--into specific jobs performed by a single individual. Further grouping of these specific jobs according to the similarity in their tasks or work activities produces an occupational classification. The division of labor according to function refers to the larger purpose or goal toward which work activities are ultimately directed such as the production of knowledge in specific fields, the administration of justice, the extraction of raw materials, or the manufacture of specific products. When specific jobs are grouped according to the goals or purposes which vindicate their performance, then one creates an industrial classification of jobs. Everyone knows, of course, that the occupational and industrial division of labor are interlarded; some purposes require the performance of tasks which the pursuit of other goals either does not require at all or requires to a lesser extent. Precisely how the industrial and occupational division of labor are intertwined and the manner in which they are intertwined in different sectors of the economy should be an empirical question. But one can hardly conduct any definitive empirical study of their interrelationships if the answer to the question has already been partially built into the data by the use of industrial and class of worker distinctions to define occupational groups. Much the same can, of course, be said for the use of any criterion other than task similarity to define occupational groups. The sociologist has much interest in the relationships between job tasks and such distinct variables as the social organizations of these tasks, the nature of interpersonal relations on the job, their institutional setting, the prerequisites (education, experience, on-the-job training, etc.) for performing them, and the rewards derived from their performance. However, when these variables are utilized to construct occupational typologies it becomes circular to inquire subsequently about their relationship to the occupational division of labor.

THE HOMOGENEITY OF DETAILED OCCUPATIONAL GROUPS

Sociological use and interpretation of occupational statistics is impeded not only by the employment of diverse criteria to create an occupational typology, but also by the consistency with which the criteria are employed and the homogeneity of the occupational groups so formed. The detailed occupational groups of the U. S. Bureau of the Census are organized according to such major occupational levels as professional, technical, and kindred workers; sales workers; operatives and kindred workers; and laborers, except farm and mine. Each of these major occupational levels contains one detailed occupational group which incorporates all jobs assigned to the major occupational level that are not elsewhere classified (n. e. c.) into a specific detailed occupation assigned to the major occupational group. Thus, 313, 858 of the estimated 7,335,699 professional, technical, and kindred workers in the experienced civilian labor force in 1960 were assigned to the detailed occupation, professional, technical, and kindred workers (n. e. c.).[4] At the professional level, the detailed occupation incorporating workers not elsewhere classified is relatively small, comprising less than 5 per cent of the total number of professional, technical, and kindred workers in the experienced civilian labor force. However, as one can see in Table 1, the not elsewhere classified groups comprise substantial fractions of other major occupational levels. Needless to say, at every major occupational level the specific job titles falling in the not elsewhere classified category are quite heterogeneous. For example, merely to list the job titles falling in operatives and kindred workers (n. e. c.) requires sixty-six, double-columned, single-spaced, typed pages in the Classified Index of Occupations and Industries.[5] For the sociologist who wishes to use detailed occupation as an indicator of socio-economic status, the lack of homogeneity in the not elsewhere classified codes can only comprise a major source of inaccuracy which afflicts 32.8 per cent of the total experienced civilian labor force.

There is no reason to discuss the problem of homogeneity created by the not elsewhere classified groups in great detail, since the problem is widely recognized by most users of occupational statistics. Table 1 shows that, on the whole, changes in the system of occupational classification between 1950 and 1960 tended to reduce the fraction of the experienced civilian labor force falling into the residual, "n. e. c." categories at each major occupational level. This conclusion may be drawn by comparing the percentages of the 1950 experienced civilian labor force in "n. e. c." categories when the 1950 labor force is alternatively classified by the 1950 and 1960 systems of occupational classification. Changes in the labor force distribution tended to have little overall effect upon the proportion of the experienced civilian labor force falling in residual, "n. e. c." categories in 1950 and 1960. This conclusion may be reached by comparing the proportion of the 1950 and 1960 experienced civilian labor force in "n. e. c." categories when both are distributed according to the 1960 system of occupational classification.[6]

Apart from the problem of the large "n. e. c." categories, there are other curiosities in the way

[4] The figures are taken from U. S. Bureau of the Census, 1960 Census of Population, Detailed Characteristics, United States Summary, Final Report PC(1)--1D (Washington, D. C.: U. S. Government Printing Office, 1963), Table 201, p. 522.

[5] U. S. Bureau of the Census, 1960 Census of Population, Classified Index of Occupations and Industries (Washington, D. C.: U. S. Government Printing Office, 1960), p. 133-199.

[6] At least one commentator, though rightly concerned about the size of the "n. e. c." categories, appears to confuse growth in the labor force size with growth in the relative importance of these residual categories. See J. G. Scoville, "Making Occupational Statistics More Relevant," Proceedings of the Business and Economics Statistics Section of the American Statistical Association: 1965, p. 318.

Table 1.--Per Cent of Persons in Each Major Occupation Group Falling in Generic "Not Elsewhere Classified" Codes, Experienced Civilian Labor Force, 1950 and 1960

Major Occupation Group	Reported In 1950[a]	1960 System of Occupational Classification Applied In: 1950[b]	1960[b]
	Total Experienced Civilian Labor Force		
Total, all groups	38.8%	33.6%	32.8%
Professional, technical, and kindred	2.3	1.8	4.3
Managers, officials, and proprietors, except farm	85.7	85.7	83.5
Clerical and kindred	42.6	33.0	31.4
Sales	84.3	84.9	81.0
Craftsmen, foremen, and kindred	0.9	0.9	1.2
Operatives and kindred	54.9	40.4	38.9
Private household	85.3	80.3	70.2
Service, except private household	12.0	4.4	3.3
Laborers, except farm and mine	84.4	79.4	78.3
All farm workers	....	....	....

Table 1.--Continued

Major Occupation Group	Reported In 1950[a]	1960 System of Occupational Classification Applied In: 1950[b]	1960[b]
	Male Experienced Civilian Labor Force		
Total, all groups	35.7%	32.4%	32.5%
Professional, technical, and kindred	3.1	2.2	5.5
Managers, officials, and proprietors, except farm	86.4	86.4	84.2
Clerical and kindred	57.3	40.4	39.3
Sales	78.8	79.7	75.2
Craftsmen, foremen, and kindred	0.9	0.9	1.2
Operatives and kindred	46.0	38.9	38.0
Private household	89.4	86.9	82.2
Service, except private household	8.6	5.1	4.0
Laborers, except farm and mine	84.1	78.9	77.8
All farm workers	....	....	....
	Female Experienced Civilian Labor Force		
Total, all groups	47.1%	36.7%	33.3%
Professional, technical, and kindred	1.1	1.2	2.3
Managers, officials, and proprietors, except farm	81.2	81.2	79.5
Clerical and kindred	33.6	28.4	27.5
Sales	94.8	94.8	91.1
Craftsmen, foremen, and kindred	0.5	0.5	0.7
Operatives and kindred	78.8	44.4	41.1
Private household	85.1	79.9	69.8
Service, except private household	16.3	3.4	2.7
Laborers, except farm and mine	92.8	92.8	91.4
All farm workers	....	....	....

[a]Source: U. S. Bureau of the Census, 1950 Census of Population, Vol. II, Characteristics of the Population, Part 1, U. S. Summary, Chapter C (Washington, D. C.: U. S. Government Printing Office, 1953), Table 124, pp. 261-266; U. S. Bureau of the Census, 1950 Census of Population, Vol. II, Characteristics of the Population, Part 51, Alaska, Chapter C (Washington, D. C.: U. S. Government Printing Office, 1952), Table 46, pp. 59-62; and U. S. Bureau of the Census, 1950 Census of Population, Vol. II, Characteristics of the Population, Part 52, Hawaii, Chapter C (Washington, D. C.: U. S. Government Printing Office, 1952), Table 52, pp. 99-105.

[b]Source: U. S. Bureau of the Census, 1960 Census of Population, Detailed Characteristics, United States Summary, Final Report PC (1)--1D (Washington, D. C.: U. S. Government Printing Office, 1963), Table 201, pp. 522-527.

specific job titles are assigned to detailed occupational groups which are less well known. Here we discuss a few examples which suggest that the Classified Index of Occupations and Industries might be profitably reviewed with the explicit goal of attempting to form detailed occupations characterized primarily by the task similarity in the specific job titles allocated to them.

For the researcher who wants to identify the detailed occupational classification of any particular job, the best starting place is the Alphabetical Index of Occupations and Industries.[7] Should, for example, a person look under the heading "engineer, specified type: flight" he would find that in 1960 persons

[7]U. S. Bureau of the Census, 1960 Census of Population, Alphabetical Index of Occupations and Industries, revised edition (Washington, D.C.: U. S. Government Printing Office, 1960).

reporting their occupation as "flight engineer" were classified into two detailed occupations depending upon their industrial affiliation. Those employed in industries manufacturing "transportation equipment: aircraft and parts" are assigned by the census code to engineers, aeronautical--a detailed title which is subsequently associated with the major occupation group identified as professional, technical, and kindred workers. Alternatively, "flight engineers" employed in the "air transportation" industry are assigned by the census occupational code to mechanics and repairmen, airplane--a detailed title which is subsequently allocated to the major occupation group identified as craftsmen, foremen, and kindred workers. The use of industry to assign persons reporting themselves as "flight engineers" is probably wholly justified and is doubtless an example of the situation in which industry must serve as a guide to occupational classification "because the same occupation title is sometimes applied to entirely different kinds of work in different industries."[8]

While the Alphabetical Index of Occupations and Industries informs us how the individual returns should be classified, it still leaves us uncertain about the job duties of "flight engineers." We know, however, that we need to be careful to distinguish between persons so designated in the aircraft manufacturing and air transportation industries. The "flight engineer" in the air transportation industry is probably better known to the public as the second officer on commercial airliners--a person who aids the first officer or captain in flying the aircraft and who is hardly distinguishable from the pilot in dress and outward appearance. A more careful job description of "flight engineers" in the air transportation industry is given by the Dictionary of Occupational Titles:

> Makes preflight, inflight, and postflight inspections, adjustments, and minor repairs to insure safe and efficient operation of aircraft: Inspects aircraft prior to takeoff for defects, such as fuel or oil leaks and malfunctions in electrical, hydraulic, or pressurization systems according to preflight check-list. Verifies passenger and cargo distribution and amount of fuel to insure that weight and balance specifications are met. Monitors control panel to verify aircraft performance, and regulates engine speed according to instructions of AIRPLANE PILOT, COMMERCIAL. Makes inflight repairs, such as replacing fuses, adjusting instruments, and freeing jammed flight control cables, using handtools, or takes emergency measures to compensate for failure of equipment, such as autopilot, wing heaters, and electrical and hydraulic systems. Monitors fuel gauges and computes rate of fuel consumption. Keeps log of fuel consumption and engine performance. Records malfunctions which were not corrected during flight, and reports needed repairs to ground maintenance personnel. May perform repairs upon completion of flight. Must be licensed by Federal Aviation Agency. May be required to be licensed AIRCRAFT-AND-ENGINE MECHANIC or AIRPLANE PILOT, COMMERCIAL.[9]

Apart from the fact that the "flight engineer" may be required to be a licensed airplane pilot, there are certain other similarities in the tasks of the two jobs. The Dictionary of Occupational Titles informs us that an "airplane pilot, commercial" is engaged by such activities as reviewing "ship's papers to ascertain factors, such as load weight, fuel supply, weather conditions, and flight route and schedule" and reading " gauges to verify that oil, hydraulic fluid, fuel quantities, and cabin pressure are at prescribed levels prior to starting engines."[10] These duties seem very similar to those of the "flight engineer," yet the census code assigns "airline pilots" to the detailed occupation designated as airplane pilots and navigators while, as noted above, the "flight engineer" in the air transportation industry is included in the detailed occupation designated as mechanics and repairmen, airplane.

For the occupational analyst it is perhaps not too disturbing that "flight engineers" and "airline pilots" be classified into different detailed census occupations. The distinction between the two occupations is also given formal expression in the social organization of the two jobs into different unions--the International Air Line Pilots Association and the Flight Engineer's

---

[8] U. S. Bureau of the Census, 1960 Census of Population, Classified Index of Occupations and Industries, op. cit., pp. v-vi.

[9] U. S. Department of Labor, Dictionary of Occupational Titles, Vol. I, Definitions of Titles, third edition (Washington, D. C.: U. S. Government Printing Office, 1965), pp. 288-289.

[10] Ibid., p. 11.

International Association.[11] The subsequent assignment of the detailed occupations into which "flight engineers" and "airline pilots" fall to widely different major occupational levels is, however, less than satisfactory, especially in view of the overlapping tasks and cooperation between the two jobs. As it turns out, mechanics and repairmen, airplane fall into the major occupation group described as craftsmen, foremen, and kindred workers, while airplane pilots and navigators are part of the major occupation group described as professional, technical, and kindred workers. Thus, the "flight engineer" in the air transportation industry is ultimately considered a blue collar worker, while the airline pilot is regarded as a professional. Such a wide distinction in these two jobs hardly seems justified, especially when one ponders the fact that the detailed occupation described as baggagemen, transportation--a group which includes "cargo handlers" in the air transportation industry—falls into the major occupation group described as clerical and kindred workers and would be regarded as falling together with "airline pilots," but not "flight engineers," in the broader group of white collar workers.

We have explored the case of the "flight engineer" at some length in order to indicate just how difficult it is to evaluate the homogeneity of the detailed census occupational code. Examination of the Classified Index of Occupations and Industries reveals many other instances of heterogeneity in detailed occupations and curiosities in the way particular jobs and pairs of jobs are classified. A few particulars may help convince the reader that, whether or not he agrees individual jobs should be classified according to their task similarity, the detailed census occupational code falls short of fully satisfying such a criterion.

The student of the Classified Index of Occupations and Industries may at first be stunned by the enormous task it accomplishes. Closer scrutiny may reveal some job titles whose classification demands explanation. Why is a "professional golfer" assigned to the detailed occupation designated as athletes, while a "golf pro" is assigned to detailed occupation described as sports instructors and officials? Does the census really believe that this minor difference in the way a person reports his occupation enables one to distinguish between Arnold Palmer, Jack Nicklaus, and the hundreds of professional golfers who also serve as the professionals-in-residence at country clubs and the like? Or, for that matter, is the distinction even meaningful, particularly when one recognizes that even the top professional golfers perform the functions of instructors in a different fashion by authoring columns in newspapers and writing manuals of best golfing practices? (In 1970, Arnold Palmer and Jack Nicklaus will doubtless turn up as editors and reporters or authors.)

Other instances of dubious classification of specific jobs are scattered throughout the Classified Index of Occupations and Industries. For example, a "seeing-eye dog teacher" is placed in the detailed occupation described as entertainers (n. e. c.). A "student-activities director" in a college is placed in the detailed occupation identified as social and welfare workers, except group, rather than in the detailed occupation--recreation and group workers--which contains such similar jobs as "recreation directors" and "Y. M. C. A. program directors." "Caterpillar operators" are classified as truck and tractor drivers, a detailed occupation subsequently placed in the major occupation group identified as operatives and kindred workers. To the best of our knowledge, "Caterpillar" is a brand name which is colloquially used to identify certain kinds of heavy machinery, notably earth moving equipment and bulldozers. But a "bulldozer driver" and a "heavy-equipment operator--except manufacturing" are specific occupational titles which are allocated to excavating, grading, and road machinery operators, a detailed occupation subsequently placed at the major occupation level described as craftsmen, foremen, and kindred workers. "Dance hall supervisors" are classified as policemen and detectives, but "bouncers" (in the miscellaneous entertainment and recreation service industry) are classified as guards, watchmen, and door-keepers. "Floorwalkers" and "ushers" in the retail trade industry are classified as floormen and floor managers, store, a detailed occupation subsequently grouped at the major level with managers, officials, and proprietors, except farm. However, the man who shows you to your seat at the symphony and "floorwalkers" in hotels and lodging places are grouped, respectively, with ushers, recreation and amusement and policemen and detectives, two detailed occupations ultimately assigned at the major level to service workers, except private household.

Other cases in which the detailed occupational code departs from a criterion of task similarity

---

[11] For information concerning these unions, see Frederick G. Ruffner, Jr., and others, eds., Encyclopedia of Associations, Vol. I, National Organizations of the United States, 4th edition (Detroit, Michigan.: Gale Research Company, 1964) pp. 859ff.

in job classification could be supplied by scanning the Classified Index of Occupations and Industries. But the point is apparent enough: to the extent that the detailed occupations identified by the Bureau of the Census are themselves heterogeneous, one is frustrated in attempting to form relatively broad occupational groups, in assigning prestige or socio-economic status scores to detailed occupations and in using the distribution of the labor force over detailed occupations to form anything but a vague impression of the job skills currently being exercised by employees. Certainly, the detailed occupational titles provide no sure guide to the actual content of the detailed occupational codes; the user of census occupational statistics must be prepared to examine the Classified Index of Occupations and Industries before drawing any definitive conclusions from inter-occupational comparisons.

THE FORMATION OF OCCUPATIONAL CLASSES

Sociologists have evidenced little interest in detailed occupational statistics; like the political economists in the late 19th and early 20th centuries they have been primarily concerned with the aggregation of individual reports upon jobs into relatively broad occupational classes designed to reveal the social class configuration or divide the population according to the principal axes of structural differentiation in society. Many traditional, sociological approaches to the aggregation of occupational data doubtless had and, to a large extent, still have their theoretical roots in Karl Marx's distinctions between the three great social classes of wage laborers, capitalists, and landlords. Remnants of the distinction still exist in the three-fold classification of white collar, blue collar, and farm workers frequently employed in sociological analyses. But Marx, unlike many subsequent writers, was at least conscious that questions bearing upon social class should not be confused with--even though they are related to--questions concerning the occupational division of labor. He observed that it is "well known that modern class differences are not in any way based upon handicraft differences, and that, on the contrary, the division of labour produces very diverse occupations within the same class."[12]

[12] Karl Marx, Selected Writings in Sociology and Social Philosophy, edited by T. B. Bottomore and Maximilien Rubel and translated by T. B. Bottomore (London: C. A. Watts and Company, 1956), pp. 201-202.

However one evaluates the final product, the uneasy wedding between political economy and econometric theory seems to have transformed the earlier concern of economists with the class bases of society into more specific questions about the supply and demand of various job skills and the effects of technological change upon them. During the decades when the economists were changing the focus of their interests in occupational data, sociologists began to question the utility of a unitary concept of "socio-economic status." Rather than regarding income, education, and occupation as resources which individuals transformed into a single, unique position in the social hierarchies of communities, theorists and researchers alike suggested that attitudes, values, and behavior were contingent not only upon the "sum" of these resources, but also upon the particular way in which they were combined. Thus, while the "social value" attached to the resources of a highly educated person with a low income might be roughly equivalent to that associated with the resources of a high income person with little education, there is some question whether or not they belong to the same "socio-economic group." To assign two such respondents to the same class ignores differences in the nature of their resources--differences which might well be related to their consumption habits, child-bearing patterns, and tastes. However, despite the apparent abandonment of interest in social class configurations by economists and the growing conviction of sociologists that a unitary measure of "socio-economic status" or class is not sufficient to describe a person's location in the social stratification system, the aggregation of occupational data into broad groups continues to be dominated by a conception which equates "occupational classes" with "socio-economic groups."[13]

The earliest census reports on the occupations of the people did not admit of aggregation according to "socio-economic status." For example, the British Censuses of 1801, 1811, and 1821 identified only three broad occupational pursuits: (1) those chiefly employed in agriculture, (2) those chiefly employed in trade, manufacture, or handicraft, and (3) those not

[13] This point and many others raised in this paper are also discussed by Scoville, op. cit., pp. 317-323. Indeed, in writing this paper it became increasingly clear as we moved from one point to another that we were treading ground already covered by Scoville's short and generally excellent review of problems of occupational classification.

employed in the preceding two classes.[14] The Sixth Census of the United States (1840) called only for the enumeration of the number of persons in each family employed in mining, agriculture, commerce, manufactures and trades, navigation of the ocean, navigation of canals, lakes and rivers, and learned professions and engineers.[15] These early efforts to describe the occupational distribution of the population over a few broad groups were, however, soon replaced by the collection and publication of more detailed statistics which admit of summary by aggregation in various ways.[16] In many instances, researchers chose to invoke criteria of aggregation which make no reference at all to "socio-economic status." For example, in his many discussions of British occupational statistics in the late 19th and early 20th centuries, Welton employed a distinction which amounts to whether occupational pursuits are required ubiquitously in order to service populations or are necessitated primarily by the economic bases of particular communities. In his own words,

> The secondary occupations are those connected immediately with the consumption of articles of necessity, and with the supply of the daily wants of the population. The primary occupations are those which are connected with the production and manufacture and traffic in articles, afterwards to fall into the hands of the secondary class, and in general all occupations which do not subserve merely the supply or benefit of the neighboring population, but also that of distant places, or which are necessary for the fulfillment of national requirements.[17]

Attempts to create broad occupational classes according to the way particular jobs or detailed occupations are articulated with the needs of local or national populations tend, however, to confuse the distinction between industries and occupations. More importantly, such a basis of occupational classification was hardly suitable for the investigation of the poverty of wage laborers and the effects of industrial change upon labor--questions perennially raised, but only periodically consuming economic and social investigators as they did in the late 19th century and show evidence of doing today.[18]

By 1911, Isaac A. Hourwich could still observe, "Our population statistics, however, have heretofore taken no cognizance of economic classes."[19] He then proceeded to undertake

---

[14] For a discussion of 19th century, British occupational statistics, see Charles Booth, "Occupations of the People of the United Kingdom, 1801-81," Journal of the Statistical Society of London, 49 (June, 1886), pp. 314-435.

[15] Carroll D. Wright, The History and Growth of the United States Census (Washington, D. C.: U. S. Government Printing Office, 1900), pp. 142ff.

[16] The 1855 New York State Census adopted, for example, a classification involving several hundred detailed occupational groups. In many ways this early classification is superior to that employed by the U. S. Bureau of the Census prior to 1930, since it avoids much of the confounding between industry and occupation which was especially pronounced in early Federal censuses. See New York (State), Secretary of State, Census of the State of New York for 1855 (Albany, N. Y.: Charles Van Venthuysen, 1857), pp. 178-195.

[17] T. A. Welton, "On the Classification of the People by Occupations and On Other Subjects Connected With Population Statistics of England," Journal of the Statistical Society of London, 32 (September, 1869), p. 274. Among Welton's many analyses and discussions of British occupational statistics are the following papers: "On the English Census of Occupations, 1871," Transactions of the Manchester Statistical Society, Session of 1875-76, pp. 51-110; "On Forty Years' Industrial Changes in England and Wales," Transactions of the Manchester Statistical Society, Session of 1897-98, pp. 153-243; "On the 1891 Census of Occupations of Males in England and Wales, so far as Relates to the Large Towns and to the Counties after the Exclusion of Such Towns," Transactions of the Manchester Statistical Society, Session of 1897-98, pp. 244-266; "The Occupations of the People of England and Wales in 1911, from the Point of View of Industrial Developments," Transactions of the Manchester Statistical Society, Session of 1914-15, pp. 47-170.

[18] On statistical activities in the late 19th century, see T. S. Ashton, Economic and Social Investigations in Manchester, 1833-1933 (London: P. S. King and Son, 1934), especially pp. 99-107.

[19] Isaac A. Hourwich, "The Social-Economic Classes of the Population of the United States, I," Journal of Political Economy, 19 (March, 1911), p. 188.

a socio-economic grouping of the detailed occupations identified in the reports of the decennial censuses. Hourwich did not choose to elaborate at length upon the problems encountered in construction of his grouping, remarking that

> . . . just as we find no difficulty in assigning a vertebrate and a tree to distinct realms of nature, although a bacteriologist might hesitate where to place certain microbes, so in a complex society like ours the existence of a few leading classes should not be obscured by the recognition of the parallel existence of intermediate or transitional social groups. To be sure, it is impracticable to apply to the occupation groups of the census the familiar classical division of society into landowners, entrepreneurs, and laborers. We may accept as satisfactory a scheme of classification which divides the population into indecomposable social groups or strata. . . [20]

Without further elaboration, Hourwich proceeded to present his classification, which, within the limitations of the then current detailed occupational classification, resembled rather closely the major occupation groups still employed in occupational tabulations of the U. S. Bureau of the Census.

Hourwich clearly hit upon a popular note, for Alba Edwards was soon at work on precisely the same problem.[21] By 1933, Edwards had culminated his investigations, publishing a socio-economic grouping of occupations which in many essential details remains equivalent to the major occupation groups currently identified by the U. S. Bureau of the Census.[22] Dividing manual workers into three levels on the basis of skill, Edwards formed six broad divisions--(1) professional persons, (2) proprietors, managers, and officials, (3) clerks and kindred workers, (4) skilled workers and foremen, (5) semiskilled workers, and (6) unskilled workers--which he regarded as ordered according to their "socio-economic status." The resemblance of this classification to the major occupational groups currently in use is apparent and requires no further comment.

We have already provided above some indications that the major occupation groups currently employed by the U. S. Bureau of the Census do not adequately fulfill a criterion of classification based on task similarity. It is also the case that, despite the obvious use of socio-economic criteria to form them, they do not comprise a grouping of detailed occupations effected on the basis of status alone. From recent studies of occupational prestige, we know, for example, that the ratings assigned by respondents to specific jobs or occupations falling in different major occupational groups overlap appreciably.[23] When the average income and average years of school completed by the incumbents of detailed occupations are used to assign them a socio-economic status score, one also finds an appreciable overlap in the scores achieved by the detailed occupations falling in different major occupation groups.[24] Thus, many sociologists would doubtless prefer to see detailed occupational statistics aggregated into broad classes explicitly on the basis of social status, without any reference to the task similarities of the specific occupations classified into identical major levels. In a study of graduate student finances, James A. Davis, for example, rejects the major occupation groups employed by the U. S. Bureau of the Census because they "present some difficulty when used as a measure of prestige."[25] One can, of course, appreciate the sociologist's need for a classification of occupations on the basis of status alone. Where occupation is the only measure of economic well-being or social evaluation available to a researcher--as is the case with death registration and marriage license application statistics and is often the case in sample surveys of intergenerational mobility--then the investigator may choose to analyze his materials by effecting a grouping of occupational returns on the basis of their

---

[20] Ibid., pp. 189-190.

[21] Alba M. Edwards, "Social-Economic Groups of the United States," Journal of the American Statistical Association, 12 (June, 1917).

[22] Alba M. Edwards, "A Social-Economic Grouping of the Gainful Workers of the United States," Journal of the American Statistical Association, 28 (December, 1933), pp. 377-387.

[23] See, for example, Robert W. Hodge, Paul M. Siegel, and Peter H. Rossi, "Occupational Prestige in the United States, 1925-1963," American Journal of Sociology, 70 (November, 1964), pp. 286-302.

[24] See Otis Dudley Duncan, "A Socioeconomic Index for All Occupations," in Albert J. Reiss, Jr., and others, Occupations and Social Status, op. cit., pp. 109-138.

[25] James A. Davis, Stipends and Spouses (Chicago, Ill.: University of Chicago Press, 1962), p. 25.

prestige or socio-economic level. While this particular need might best be met by revising the current major occupation groups into occupational classes homogeneous with respect to status and/or economic position, we do not feel that this need is urgent enough to offset the general disadvantages of the resultant heterogeneity of occupational groups with respect to task similarity of the specific jobs and detailed occupations aggregated to form its broad divisions.

If the occupational codes currently in use were to be extensively revised, we would prefer to see an attempt made to create more homogeneous groupings of specific jobs at the detailed level and an attempt to use generic features of job tasks--such as their complexity--to form revised divisions at the major level. There are several reasons why we believe such a revision would be superior to one in which the Bureau of the Census either produced groupings of detailed occupations solely on the basis of some estimate of their socio-economic position or abandoned current tabulations by major occupation groups in favor of tabulations by some composite status index. First, if detailed occupations were formed according to the task similarity of specific jobs, one might expect relatively more detailed occupations to be identified in published statistics than is currently the case. These occupations, though perhaps numerically small in many instances, would at least be relatively more homogeneous. Under such a situation, individual researchers would violate empirical reality much less by assigning the specific jobs grouped into a detailed occupation the same status or prestige score and, of course, published tabulations at the detailed level would still leave the researcher free to form his own classification of occupations according to status as he best saw fit. Second, resort to a major grouping of occupations based solely on status distinctions between them ignores the fact that many interoccupational differences are not attributable to socio-economic differences, but to other features of occupations such as their work settings, the kinds of objects and symbols manipulated by their incumbents, and the responsibilities incurred by their performance.

BROAD OCCUPATIONAL CLASSIFICATIONS

Two different standardized procedures for grouping detailed occupations into relatively gross occupational categories are currently available to research workers. One, the major occupation groups of the U. S. Bureau of the Census, employs socio-economic considerations among other criteria to effect an occupational typology. The other, the classification presented in the revised Dictionary of Occupational Titles, employs the criteria of work tasks and work settings to a much greater extent, especially among manual workers. A comparison of these two classifications and their relationship to socio-economic correlates is of some interest, since a decision to adopt either one or both of them in a particular research situation should be informed by knowledge of the relationship between them and of their associations with other standard indicators of social status and related phenomena.

Distributions of the labor force are not published according to the categories recognized in the revised Dictionary classification. Consequently, in order to compare census major occupation groups to those recognized in the Dictionary of Occupational Titles, it is necessary to aggregate published, detailed occupational statistics into the Dictionary codes. Nine broad occupational groups are recognized in the Dictionary classification: (1) professional, technical, and managerial occupations, (2) clerical and sales occupations, (3) service occupations, (4) farming, fishery, forestry, and related occupations, (5) processing occupations, (6) machine trades occupations, (7) bench work occupations, (8) structural work occupations, and (9) miscellaneous occupations. In order to tabulate the labor force according to these divisions, we used the Dictionary of Occupational Titles and the Classified Index of Occupations and Industries to assign each title identified in the detailed occupational classification of the U. S. Bureau of the Census to one of the nine major divisions recognized in the Dictionary classification. The assignment of the detailed occupational categories to the nine broad groups recognized in the DOT (Dictionary of Occupational Titles) code was not wholly satisfactory, for some of the specific occupations assigned to the detailed census codes actually fall into different major categories of the DOT classification. However, the allocation of detailed census occupations to the major DOT divisions is sufficiently precise to permit gross comparisons between the census major occupation groups and the broad DOT categories.

For the total experienced civilian labor in 1960, exclusive of those not reporting occupation and "former members of the armed forces," Table 2 shows the relationship between the major occupation groups of the U. S. Bureau of the Census and the broad categories of the DOT classification. As one can

Table 2.--Relationship of Census Major Occupation Groups to Revised Dictionary Classification, Total Experienced Civilian Labor Force, 1960

Major Occupation Groups	Classifications Used in Revised Dictionary of Occupation Titles									Total Per Cent	Number In Total ECLF (In Thousands)
	Professional, Technical, and Managerial Occupations	Clerical and Sales Occupations	Service Occupations	Farming, Fishery, Forestry, and Related Occupations	Processing Occupations	Machine Trades Occupations	Bench Work Occupations	Structural Work Occupations	Miscellaneous Occupations		
	Percentage Distribution										
Total ECLF	20.0	22.5	13.7	6.8	10.2	5.9	3.2	7.1	10.6	100.0	64,518
Professional, technical, and kindred workers.	98.6	...	0.5	0.5	...	...	...	...	0.4	100.0	7,336
Managers, officials, and proprietors, except farm	98.2	0.2	...	...	...	...	...	...	1.6	100.0	5,490
Clerical and kindred workers.	3.0	95.5	...	...	...	...	...	...	1.4	99.9	9,617
Sales workers . .	...	100.0	...	...	...	...	...	...	...	100.0	4,801
Craftsmen, foremen, and kindred workers	0.0+	0.6	0.4	...	2.1	37.2	3.2	35.0	21.6	100.1	9,223
Operatives and kindred workers.	0.0+	3.4	4.7	...	43.8	2.8	13.5	3.4	28.3	99.9	12,846
Service and private household workers	...	...	100.0	...	...	...	...	...	...	100.0	7,590
Laborers, except farm and mine .	...	...	15.6	7.3	20.8	1.5	0.9	26.6	27.4	100.1	3,530
All farm workers .	...	...	...	100.0	...	...	...	...	...	100.0	4,085

Compiled from U.S. Bureau of the Census, 1960 Census of Population, Detailed Characteristics, United States Summary, Final Report PC(1)--1D (Washington, D.C.: U.S. Government Printing Office, 1963), Table 201, pp. 522-527.

see from the tabulation in Table 2, two of the DOT categories are essentially combinations of divisions recognized in the major occupation code of the U. S. Bureau of the Census. Although there are minor exceptions, the DOT category identified as professional, technical and managerial occupations is basically just a combination of the major occupation groups identified as professional, technical, and kindred workers and managers, officials, and proprietors, except farm. Similarly, the DOT category of clerical and sales occupations is basically just a combination of the major occupation groups specified as clerical and kindred workers and sales workers. However, among manual workers the fit between the census major occupation groups and the DOT categories is much less perfect. The major occupation group identified as craftsmen, foremen, and kindred workers is basically split into three of the DOT groups: machine trades occupations, structural work occupations, and miscellaneous occupations. The major group operatives and kindred workers is likewise dispersed between the DOT categories; two-fifths fall in processing occupations, three-tenths in miscellaneous occupations, and a little more than one-tenth are assigned to bench work occupations. The census category laborers, except farm and mine is also divided among the broad DOT divisions--roughly one-fourth to one-fifth fall in the DOT groupings of processing occupations, structural work occupations, and miscellaneous occupations, while another 15 per cent of laborers are assigned by the DOT to service occupations. The DOT category of service occupations is comprised of all persons in the major census group identified as service and private household workers, together with a sizeable number of persons falling in the census categories operatives and kindred workers and laborers, except farm and mine. Finally, the DOT category referred to as farming, fishery, forestry, and related occupations is basically comprised of the major occupation groups identified as farmers and farm managers and farm laborers and foremen (collapsed in Table 2 as all farm workers) along with 7 per cent of the census category laborers, except farm and mine.

In sum, the broad DOT groups retain for most practical purposes the distinction between white collar and blue collar occupations as found in the major occupation groups of the U. S. Bureau of the Census. However, among white collar occupations, the DOT employs a more coarse grouping than the census major occupation classification, combining the four census codes among white collar occupations to two groups. Among blue collar occupations, the DOT categories completely revise the census groups, substituting a division based on work tasks for the skill gradient employed in the census code. In addition, the DOT classification does not identify farm workers as such, placing them together with occupations in the related extractive industries of fishing and forestry.

SOCIO-ECONOMIC CHARACTERISTICS OF OCCUPATIONAL GROUPS

In addition to the direct comparison of the broad DOT categories and the major census occupation groups, we have also aggregated the income and educational distributions of the detailed census occupations falling into each of the DOT categories. This enables us to contrast the relationship of the census and DOT occupational groupings to two standard indicators of social status. Relevant summary statistics are shown in Table 3, which gives the mean income and mean years of school completed for the male experienced civilian labor force in each DOT group as well as for the major occupation groups of the U. S. Bureau of the Census.

As the reader can see from Table 3, the major occupation groups of the U. S. Bureau of the Census tend to fall along a socio-economic gradient defined by education and income. Both the manual and non-manual groups are differentiated by income and education, a point which is not surprising since skill levels were used as explicit criteria for defining the manual categories and such factors as income and education taken as guidelines for devising the distinctions between white collar occupations effected by the major occupation groups. The socio-economic grading of the DOT categories is not so readily apparent as that observed among the major occupation groups of the U. S. Bureau of the Census. This is especially true for the manual categories: nearly identical levels of average income and education are observed for processing occupations, machine trades occupations, bench work occupations, structural work occupations, and miscellaneous occupations. The DOT classification makes clear that the socio-economic differentiation of manual workers does not depend in any important way upon a gross differentiation of the kinds of tasks performed and the situations in which they are carried out. Instead, the differentiation of manual workers appears to rest upon entry requirements and skill levels which cut across gross divisions of task characteristics and functional settings.

In addition to the mean levels of education and income within each of the broad occupational groups identified in the census and DOT classifications, Table 3 also shows the correlation ratios of individual income and years of school completed with both the major occupation groups of the U. S. Bureau of the Census and the DOT classification. Despite the differences in the two classification schemes at the manual level, both explain roughly the same proportion of the total variation in the income and education of individuals. As one would expect, the correlation ratios of income and education on the DOT classification are less than the corresponding correlation ratios on the major occupation groups of the U. S. Bureau of the Census. The two sets of correlations are not, however, substantially different, indicating that the use of skill levels to distinguish grades of manual workers has little overall effect in the proportion of variation in individual income and education explained by the census' major occupation groups. That the correlation ratios are not identical, though roughly of the same magnitude, is testimony to the point that there is no single set of correlation coefficients which adequately summarize the relationships between occupation and other indices of social status. The correlation between occupation and any variable is in part a function of one's

Table 3.--Income and Years of School Completed by Census Major Occupation Groups and the Revised Dictionary Classification, Male Experienced Civilian Labor Force, 1960

Occupation Groups	1959 Income (In Thousands of Dollars	Years of School Completed
	Means	
Major Occupation Groups		
Professional, technical, and kindred	7.7	15.3
Managers, officials, and proprietors, except farm.	8.1	12.3
and kindred	4.9	12.0
Sales workers	5.8	12.0
Craftsmen, foremen, and kindred	5.4	10.2
Operatives and kindred	4.4	9.5
Service and Private household	3.6	9.6
Laborers, except farm and mine	3.2	8.5
All farm workers	2.7	8.6
D. O. T. Classification		
Professional, technical, and managerial . . .	7.9	13.8
Clerical and sales	5.2	11.8
Service	3.5	9.5
Farming, fishery, forestry, and related . . .	2.7	8.6
Processing occupations	4.4	9.3
Machine trades occupations	5.1	10.0
Bench work occupations	4.5	9.9
Structural work occupations	4.7	9.6
Miscellaneous occupations	4.9	9.8
Total, All Occupations	5.2	10.8
Association with	Squares of Correlation Ratios	
Census Major Occupation Groups	.2052	.2948
D. O. T. Classification	.1860	.2457

Source: Compiled from U. S. Bureau of the Census, 1960 Census of Population, Subject Reports, Occupational Characteristics, Final Report PC(2)--7A (Washington, D. C.: U. S. Government Printing Office, 1963), Table 9, pp. 116-129, and Table 25, pp. 296-315. Persons not reporting on occupation and "former members of the armed forces" are excluded from the figures shown in this table; income and education means were derived by scoring the intervals identified in the census tabulations according to their midpoints and setting then open-ended, upper intervals equal to the arbitrary values of $18,000 for income and eighteen years for education.

choice of occupational classification and the criteria upon which it is based. The indeterminancy of the association between occupation and other variables is also augmented by the problem of intellectual, if not wholly logical, circularity when the criteria of occupational classification are also the variables whose relation to occupation is in dispute.[26]

From the perspective of the sociologist, there is no comprehensive answer to questions

[26]This point is discussed by Pascal K. Whelpton and Edward Hollander, "A Standard Occupational and Industrial Classification of Workers," Social Forces, 18 (May, 1940), pp. 488-494.

involving the relative merits of the census and DOT classifications. In studies of differential mortality, fertility, life style, and social values where occupation is used as a surrogate for social status, then the major occupation groups of the U. S. Bureau of the Census are preferable. However, in these situations there are alternative strategies for coding occupations according to their social status which are superior to using the major occupation groups.[27] In other research contexts, as for example in studies of work satisfaction, the broad DOT groups may be more suitable than the census classification. Ultimately, however, there is little point in evaluating the relative merits of the two broad occupational groups under discussion. What is needed is discussion of revisions of the entire occupational code, at the detailed level, so as to effect occupational categories more homogeneous with respect to the actual tasks performed in the specific jobs assigned to each occupational code.

PROBLEMS OF COMPARABILITY

Although we have argued in this paper that several advantages might be expected to accrue from attempts at revising the detailed census code in order to achieve greater homogeneity within detailed occupations, we do not believe that those revisions and the advantages which might flow from them warrant sacrificing intercensal comparability in any degree. Elaborations of the existing occupational classification which permit of aggregation to the groups identified in previous censuses are, of course, entirely acceptable. Changes which destroy the validity of intercensal comparisons--such as the shifting between 1940 and 1950 of accountants and auditors from clerical and kindred workers to professional, technical, and kindred workers--are not acceptable. However plausible such changes might be for the cross-sectional uses of the census, the price of intercensal changes in definition seems too high to pay. One's ability to undertake studies of social change in the labor force and to effect age cohort comparisons in occupational attainment is restricted by such revisions, yet the study of the trends obscured by changes in definition forms, at least for the sociologist, a major focus of interest in census statistics.[28] Our view is that modifications of the currently employed occupational code should be effected only if one of the following three conditions are met: (1) the changes are only elaborations of the existing code and permit, therefore, aggregation to classifications employed in previous censuses, (2) extensive revisions are made, but large subsamples of previous censuses are drawn, recoded with the new scheme, and tabulated so as to effect comparability with current reports of occupation specific to age, sex, race, education, income, etc., or (3) a revised occupational scheme is used only to supplement rather than replace tabulations according to the classification employed in previous censuses. Doubtless this makes us, at ages under thirty, stodgy old men with fat cigars. But it is our life expectancy which makes us recalcitrant to change: we can look forward to the prospect of examining occupational changes from 1940 to 2000, but we shall never be able to accomplish it if in 2000 we look backward over a tangle of major and minor intercensal revisions in the occupational code effected without adequate regard for problems of comparability.

A fairly large number of changes in the detailed occupational classification of the U. S. Bureau of the Census were effected between 1950 and 1960. Many of these revisions involve the identification of new titles and were designed to ease the numbers of the labor force falling into the "not elsewhere classified" categories observed at each major occupation level. Among the new lines appearing in the detailed occupational classification for 1960 are receptionists, postal clerks, payroll and timekeeping clerks, and file clerks in the major occupation group clerical and kindred workers; assemblers, checkers, examiners, and inspectors, manufacturing, graders and sorters, manufacturing, knitters, loopers, and toppers, textile, packers and wrappers, n. e. c., and sewers and stitchers, manufacturing, in the major occupation group operatives and kindred workers; and carpenters' helpers, truck drivers' helpers, and warehousemen, n. e. c. in the major occupation group laborers, except farm and mine. On the whole, these changes seem like modest and useful elaborations of the 1950 detailed occupational code. The only trouble with them and with other changes effected between 1950 and 1960 is that it is impossible to reconcile the changes so as to effect exact comparability

[27] See Duncan, op. cit.

[28] Recent studies dependent upon one's ability to effect age cohort comparisons of occupational attainment between different censuses are Paul M. Siegel, "On the Cost of Being Negro," Sociological Inquiry, 35 (Winter, 1965), pp. 41-57, and Otis Dudley Duncan, "Occupation Trends and Patterns of Net Mobility in the United States," Demography, 3 (May, 1966) pp. 1-18.

between the 1950 and 1960 classifications. At least one of the changes which we can pinpoint crosses the lines of major occupation groups--the very kind of change which tends to invalidate even gross occupational comparisons. In the 1950 occupational classification, the detailed occupation insurance agents and brokers was placed in the major occupation group sales workers. Adding Alaska and Hawaii returns to the 1950 census reports indicates that there were 304, 633 employed persons in this group. In the 1960 census code, we no longer have a line for insurance agents and brokers; instead we find a line identified as insurance agents, brokers, and underwriters. The reports of the 1960 census inform us that as of 1950 there were exactly 272, 663 insurance agents, brokers, and underwriters, a figure which falls 32, 000 short of the number of insurance agents and brokers we can find in the returns of the 1950 census. Further examination of the 1960 occupational classification suggests that these 32, 000 insurance agents and brokers did a disappearing act from the major occupation group sales workers to the major occupation group clerical and kindred workers. The 1960 occupational classification provides for a line among clerical and kindred workers identified as insurance adjusters, examiners, and investigators. This line did not appear in the 1950 occupational classification but the 1960 census reports inform us that as of 1950 there were exactly 32, 000 persons employed as insurance adjusters, examiners, and investigators. This is just the number of insurance agents and brokers we are lacking and one presumes that in creating the new line--insurance adjusters, examiners, and investigators--among clerical and kindred workers the census officials decided to fill it up with some sales workers, in particular with insurance agents and brokers.

The reports of the 1960 census are themselves of little help in pursuing the kind of detective work, illustrated above, which is necessary to reconcile the 1950 and 1960 occupational classifications. To the best of our knowledge there is no published document which details the changes in classification and their effects on intercensal comparisons. In the 1960 census reports, we can find only the comforting assurance that,

> The occupational classification system used in 1940 and 1950 is basically the same as that of 1960. There are a number of differences, however, in title and content for certain items, as well as in the degree of detail shown for the various major groups.[29]

The precise nature of these differences is not apparently to be revealed in published documents, a deficiency which can only discourage one's faith in intertemporal comparisons. As Charles Booth complained of the British occupational statistics some eighty years ago,

> . . . there is such a want of fixity of principle or method, that even competent authorities have been seriously misled regarding the apparent results. Possibly these changes were to a large extent necessary or unavoidable, but surely attention might have been drawn to them and some explanation given, instead of which there is not even so much as a footnote. The seeker after information is left to grope his way in the dark; if by chance he stumbles on the truth, well and good, if not he but adds his quota to the enormous total of false information before the public.[30]

Changes in classification without proper attention to comparability are abhorrent enough; failure to detail these changes and provisions of guidelines for effecting comparability with earlier returns is inexcusable.

CONCLUSIONS

The purpose of this paper was to review problems of occupational classification from the perspective of the use of occupational statistics in sociological research. The most important deficiency in the occupational classification currently employed by the U. S. Bureau of the Census is heterogeneity of the specific jobs assigned to many detailed occupational codes. Lack of homogeneity in detailed occupations makes it risky for the sociologist to assign a unique socio-economic status or prestige score to all individuals falling into each detailed occupational category. A further problem with the current classification is the method of its construction: implicitly and explicitly many characteristics, ranging from industrial affiliation and class of worker to skill level and entry requirements, are used to define the occupational aggregates identified in the detailed census code. However, for

---

[29] U. S. Bureau of the Census, 1960 Census of Population, Subject Reports, Occupational Characteristics, Final Report PC(2)--7A (Washington:D. C. : U. S. Government Printing Office, 1963), p. xiv.

[30] Booth, op. cit., p. 318.

sociologists and economists alike, some of the most signal questions which can be posed of occupational statistics concern the relationship between occupation, skill levels and training requirements and the interlarding of the occupational and industrial divisions of labor. It is difficult to devise cogent answers to these queries when the variables of central interest have already been employed, either in whole or in part, to help define occupational groups.

Some of the difficulties posed with current classification schemes could be eased, in our opinion, by a complete revision of the system of occupational classification aimed at devising new categories of specific jobs which are very homogeneous in the work tasks performed by their incumbents. Such a classification should give a much better overall impression of the job skills actually utilized by the labor force; it would also involve a more elaborate coding system in which the total number of detailed occupations identified would be several times the number of detailed occupational codes recognized in the present system of classification.

Comparability with previous censuses poses the most severe barrier to effecting an extensively revised system of occupational classification. Unless government agencies are willing to publish statistics by both the old and a revised system of classification or recode and retabulate older censuses with a new system, changes in occupational classification must be limited to modest elaborations of the existing system. This is perhaps a high price to pay for comparability, but we believe that most sociologists, economists, and even many local users of census statistics would prefer to cope with the inadequacies of the present, detailed occupational codes than lose the reasonable, if not precise, correspondence of the present system with the classifications employed in 1940 and 1950.

SOME RECENT DECENNIAL CENSUS OCCUPATIONAL EXPERIMENTAL WORK

Stanley Greene, Bureau of Census

The Decennial Census data on occupation are receiving a great deal of attention of late. This very meeting on occupational classification is but one manifestation of this concern. Others are demand for special tabulations, special studies, articles in learned journals, and Bureau of the Budget research subcommittees. Although it is nice to be recognized, not all this attention has been flattering. In fact, some of it has been downright critical. "Why do you lump professional athletes with chemists?" "One-third of the labor force is ill-defined." "Response variation makes the data little better than nothing."

Some of these criticisms are valid. Valid in the sense that there is a satisfactory and at least theoretical alternative. Response variation can theoretically be reduced. If we ask the proper question(s) of the proper respondent we should get the correct answer.

Underlying the other critical comments is basically a dissatisfaction with the classification scheme. This is so because for many operational and administrative programs, projection work, training needs, educational facilities, the data just are not sufficiently refined. This is a broader concern and one for which no working alternative is posed and simply reflects a legitimate dissatisfaction with the keenness of the only available comprehensive tool. At any rate, any classification scheme is no easy task. For example, in a purer science than that of job classification, that of biology, various characteristics of an organism must be taken collectively into account in attempting even to distinguish between animal or vegetable. In some cases, especially of lowly organized forms, the distinction is difficult or uncertain.

So we see that implementing a classification scheme is difficult. Some schemes are impossible, for example classifying jobs by skill level, though desirable, is not directly possible. What common denominator is there, besides dollar earnings, to measure the comparable skills between the television repairman and the economist? The economist is completely baffled by the maze of wires and parts constituting the innards of the simplest electronic circuits and I daresay the repairman would not find Samuelson comprehensible.

Despite the difficulties the Census Bureau is in the job of classifying occupations and has done so since 1820. During that time a great many changes have been introduced into the system, varying from an industrial frame of reference to an extensive listing comprising 600 occupational groups. I mention this to indicate the flexibility of the Census approach to these problems.

This brings me to the subject of the paper, the concrete steps we are now engaged in, or seriously considering, to solve some of our occupational classification and reporting problems.

Within our existing classification system, that is, our job grouping system, there are three basic areas of concern: The not reported group, those cases for which we get no reports or the job is described in words that cannot be intelligently interpreted; n.e.c. or not elsewhere classified groups, such as "clerks n.e.c." where the category may be so broad as to provide little information; and the reliability of the data--is the distribution accurate?

Our experimental work in the not reported area has taken the form of a field test to determine if one of the elements comprising our not reported cases could be eliminated. The not reported category is one of our major concerns and this problem has been aggravated by the increases in the rates in the 1960 Census. In the hope of minimizing, to some extent, this problem, we examined a sample of occupation returns from the 1960 Census. These returns represented written entries that could not readily be coded and had to be reviewed by expert classifiers. These are called "referral cases." We noted that about 15 percent of such referral cases represented written entries falling into a single category. They are the "department" or "area of work" type of returns, which are actually subdivisions of industry returns rather than occupations. Some examples of this type of entry are "shipping department," "office work," "stock room." Since these responses provide no clue regarding any particular type of work activity, and, theoretically, they may reflect a broad spectrum of job activities, we had no alternative but to classify them as "not reported." A rough estimate indicated that they may account for around 6 percent of the total not reported figure.

A case could be made in support of the hypothesis that most of these "department" type entries do not cover a broad spectrum of activities within the area, but do, in fact, refer to specific types of work. That is, a "department" type entry may be strongly correlated with one occupation, the person doing work of a general nature in that area. If such is the case, and the groups of "department" type occupational entries are strongly correlated with a given occupation or fixed occupational proportions, then a system of assigning an occupational classification would be feasible. Such a system, in addition to reducing the number of occupation not reported cases, would also result in savings on operational costs by reducing the number of referral cases for the experts to resolve.

To test this hypothesis, a sample of the Cleveland Special Census (basically a methodological test of the mail out - mail back technique) was drawn from among such "department" entries. To this group we mailed a form noting the job entry provided in the Special Census and asking for further details. The additional probing questions related primarily to "job title" and priority order of activities and duties.

The returns from this test support the original hypothesis. As may be noted in the handout of Table A, 71 percent reporting "shipping department" were determined through the additional probes to be shipping and receiving clerks; 42

percent of the "stock room" entries could be classified to "stock clerks." Although these relationships are promising and do provide evidence to support the original premise that a sound allocation system might be devised, the most fruitful feature is that the responses to the additional probing items permitted the classification of occupation. Almost all (approximately 95 percent) of these cases formerly assigned to the not reported category, could, after examination of the entries to the additional questions, be assigned to an occupational group. However, there may be an adverse effect on the other component of the not reported, i.e., complete blanks. The possibility exists that by adding to the reporting burden we discourage response completely. I shall return to that point later.

The Census Bureau's Occupational Classification System's "Not Elsewhere Classified" categories are also a source of attention and concern by both users and the Bureau.

The Bureau of the Budget has noted in Working Paper No. 66-2: "These nonspecific categories included close to one-third of total employment in 1960 and their size has been the occasion of considerable complaint on the part of users of Census data." A similar observation was noted by Dr. Scoville in his paper on the relevance of occupational data where he states: "Further evidence of the present loss of analytical value of the existing classification scheme appears in its treatment of the 'not elsewhere classified' categories. It is probable that many of the key jobs for analysis of technological change are incorporated in these groups...Not only do the different groups grow at markedly different rates, but it is impossible to assert that the placing of one-third of the labor force into 'not elsewhere classified' groups does not affect its analytical value."

To put these views in proper perspective, it may be advisable to discuss briefly what these n.e.c. groups are and what they are not.

There are over 30,000 different job titles to codify, or classify. Because of limits of page space and tabulation, any statistical display of such detail is patently impossible. Thus the problem of combining and grouping. What criteria should be used in the decisions to determine the titles to be combined into a group? It is not done arbitrarily, nor by simple arithmetic division. Within the limits of our resources we provide the maximum detail of job families. The criterion used to determine the detailed categories is significance--significance in regard to analytic usefulness, policy need, and number of workers represented. What remains and constitutes one part of our residual n.e.c. categories are those job groups in which relatively few persons are employed and fail to have basic analytical or policy significance. Of course, combining many such groups will and does result in sizable employment figures.

One other element enters into the residual groups; they comprise that set of job entries provided by the respondents not sufficiently detailed to enter more than a generalized code. This may be necessitated by the fact that some jobs require a variety of duties to be performed. Thus our n.e.c.'s comprise two elements--a very detailed job entry--but not significant, and a generalized respondent entry. We may conclude from this that the n.e.c. categories are not simply "catch-all groups." Moreover, some of the distinct 31 occupation n.e.c. groups are quite limited in scope. As an example, our "Natural scientists (n.e.c.)" line is only n.e.c. by virtue of the fact that data for "Chemists" are shown separately. Had chemists been subsumed in the "Natural scientists," leading ironically to a broader category, there would be no n.e.c. designation required. Furthermore, the n.e.c. groups are in and of themselves of intrinsic value, certainly the "Natural scientists n.e.c." group is clearly distinguished from "Service workers, n.e.c."

One other matter concerning the n.e.c. groups--they are not fixed and static. In past decades the n.e.c. groups have been examined for specific job titles which occur frequently enough to warrant setting up new occupation categories. For example, after an examination of the "Clerical and kindred workers (n.e.c.)" group in 1950, five new specified clerical occupation groups were set up for 1960 (File clerks; Payroll and timekeeping clerks; Postal Clerks; Receptionists; and Stock clerks and storekeepers.)

Nonetheless, despite the logic of the n.e.c. groups and periodic reviews resulting in their streamlining, their composition suggested further approaches to the problem. You recall that the groups theoretically comprise two parts--a general part and a specialized part. If each of these parts contributes a sizable proportion to the whole, then an approach of splitting the n.e.c.'s into their two basic components would be advisable since it would result in a much better analytical understanding of their makeup, for we could then define and present separate data for the two components in our Census tabulations. For as Dr. Ann Miller has noted: "Even a simple separation of the n.e.c. categories into two component parts, 'miscellaneous' and 'not specified,' for example (analogous to the way such categories are handled in the industry tables of the Census of Population) would make some contribution to a reduction of the problem. For instance, analysts working with the cross tabulation of occupation by industry would probably be aware of the particular 'miscellaneous' occupations, as listed in the Classified Index, that are important in specific industries and would be able to interpret the significance of the 'miscellaneous' category in this context."

To determine whether the two components were numerically significant, it was necessary to review schedules. This was so, since a single 3-digit code applies to both components. Therefore, we analyzed each job title comprising our major n.e.c. groups--distinguishing between the general titles and the specialized titles. We used both a sample of Census returns for this analysis and a sample from our monthly CPS returns. As you may note from Table B, each of the two components contribute a sizable proportion in each particular n.e.c. group. The lowest proportion of general titles amounts to 19 percent for the operatives n.e.c. group. The Table also shows that these significant levels are true for both Census and CPS data and have held up for a period of more than a decade.

Barring any unforeseen operational difficulties, we propose a sizable reduction in our n.e.c. groups for 1970 by the device of splitting them into two distinct homogeneous categories - specific and general - resulting in much better analytical usefulness. Of course, in addition we shall perform our normal procedure of investigating each job title comprising our n.e.c. groups to determine if further subdivision or allocation to other occupation categories is warranted.

In regard to the still sizable general component, it would be very helpful if we could reduce this portion, resulting from vague, broad spectrum descriptions. It is certainly conceivable that these broad descriptions result from vaguely defined job tasks such as are associated with new workers where there are a multitude of secondary tasks. If this is the case, then little more can be done to more sharply depict this real job situation. There is no doubt, however, that some of this component is contributed to by communications failure. It is this phase of the problem that played a role in our experimental questionnaire design, adopted for our First Content Pretest for the 1970 Census. I shall return to this point shortly.

The third area of major analytical concern that I mentioned earlier had to do with the reliability of the data--is the distribution accurate? Much solid and valuable analytical work has been accomplished over the years with the existing body of occupation statistics. Nonetheless some of our evaluation work indicates that there is much room for improvement in the reduction of net and gross error when related to independent distributions.

The main thrust of our thinking on this matter is in terms of additional questions on occupation. This is a departure from our usual approach on these matters of question wording. For usually when an item is reported poorly we can trace the cause to a communications misunderstanding and a revision of the wording of the question will usually remedy the problem. But for occupation, we felt our basic question on "What kind of work were you doing?", with a series of examples--of any single approach was doing the best possible communications job. But what to do about the too vague idefinite responses? the upgrading problem? Perhaps if we elicited more information, other facets, more pieces in the jig-saw puzzle--the job picture would come into clearer focus. With this idea in mind, we formulated a series of supplementary questions to follow the basic item. These relate to job activities in priority order and to the employer's title. This new formulation of questions, among other items, was tested in the Bureau's First Content Pretest which went into the field in two test sites around the middle of May of this year. Field work was completed near the end of June and though it is too early for a definitive report on the results, we do have enough evidence to support some tentative conclusions.

My general impressions from field interview observations and a scanning of schedule returns are favorable. I was encouraged by the responses. The additional parts of the item on "Most important activity," Other important activities," and "Employer's title for this job" seemed to help clarify the concepts. It also may break down the difficult concept of occupation into more easily understandable parts, in addition to providing more information for our coders to use. In any case, many of our old general titles that were of necessity coded not reported or n.e.c. could now be more clearly defined. For example, occupation entries of office work had supplementary entries of "bookkeeper," "cashier," "clerk" was followed by cashier.

More quantitative measures of the success of this item are revealed in the handout Table C, which compares the NA and n.e.c. rates for the pretest against like measures from the 1960 Census which used only the basic questions.

As explained earlier, the n.e.c.'s contain a large component of generalized responses. If the level of this component is reduced, it would be a clear indication of more precision in our returns from the additional probes. As you can note in the table prepared especially for this paper from a sample of the St. Louis Park portion of the pretest, all but one of the major n.e.c. categories for the pretest showed lower proportions than for the 1960 Census. Some of the reductions were fairly substantial and overall the n.e.c. category was reduced close to 15 percent. The improvement may be even greater than this, for some of the generalized responses though clarified may have shifted to the specific component of the n.e.c. category, by the very nature of the n.e.c. makeup.

A problem noted with the wording is that the employer's title for the job is sometimes misconstrued to mean the title of the employer's job. So we get curious inconsistencies, such as housekeeper reporting the employer's title for her job as "owner of house," or "stock clerk" and "sweeper" followed by "manager." Therefore, we are going to modify the wording in our next Pretest and ask for the formal job title.

Another problem we face in implementing this item expansion is one faced by other Census items. The difficulty has to do with the methodological changes in census taking. Since we now operate on a self-enumeration, with follow-up of failures-basis, we require compatible item wording and format, that is, the schedule must be both respondent readable and easily adaptable for direct interview, if we are to avoid the problem of dual forms. However, the present wording, though appearing to satisfy the self-enumeration phase, is somewhat awkward in many situations of direct interviewing. How can an interviewer matter-of-factly ask someone who reports to the first part of the item "plumber" or "registered nurse" the additional probes on "activities" and "employer's title?" Problems of this nature can usually be resolved, most readily, by special training and instructions to interviewers.

There are two other questions about these supplementary probes before we can give them our unqualified endorsement. The first one has to do with the fact that we are getting more information. This introduces the problem of abstracting the relevant information from the additional entries which may be beyond the capability of the temporary relatively inexperienced Census coders. We will check on this feature by having such coders (Census style) attempt to do the job, and then comparing it to our expert coders' results shown in the Table C.

The other question has to do with the earlier fear mentioned that increasing the occupational reporting burden might harm the overall return rate. This fear may be groundless, for a comparison of the not reported rates as shown in the last line of Table C clearly shows an improvement in the not reported rates for the Pretest relative to the Census. Moreover, the Cleveland follow-up test of "department" type entries had a higher return rate than many mail surveys conducted in the same manner.

You may observe that all three parts of our problem - the "department type entries" portion of our nonresponses; the "general" component of our "not elsewhere classified;" and the reliability phase - all have a common thread intertwining them. This linkage has to do with the acquiring of more detailed responses. As noted in our First Content Pretest, the additional probing items used to solicit these detailed responses reduced our n.e.c.'s. Similar style questions in our Cleveland follow-up test clarified many of our department type entries. Such results surely point to improved reliability.

This completes the subject topic on census experimentation in occupation classification. However, there are some other basic considerations I would like to present to this forum.

The first concerns the allocation of not reported occupation cases. In the planning of the 1960 Census, the basic change in enumeration procedures and the extensive use of computers led us to adopt a system of allocation of nonreports for many items. This allocation is basically a system of assigning an entry for an unknown characteristic on the basis of other reported characteristics. Presumably this is a help to the analyst, and superior to simply distributing the unknowns in accordance with reported distributions. This was not proposed for occupation because of the variety and complexity of the distribution. We are investigating techniques for such an occupation allocation and attempting to determine if a system of allocation according to important socioeconomic characteristics can be developed within a reasonable budget.

Concerning the reliability of the data, one view holds that respondents cannot report adequately because they are too limited in technical knowledge. This view concludes that we may have progressed as far as possible with the present approach. What is needed is an approach that secures the occupation information from the employee, through a modified W2 form or some other direct contact with the employer. This technique could be extended to other items, presumably more reliably reported by employers - wage and salary income would immediately come to mind. The basic problem here would be employer and Internal Revenue cooperation and the feasibility of linking these returns with the basic demographic data gathered from the households. At any rate, discussions are being held on these matters.

In regard to the basic consideration of the adequacy of the classification system itself. Do the groupings provide salient data for today's problems? Would another system be more appropriate? Is the worker's relationship to the machine a significant characteristic of job determination?

The only point I would want to make in this regard is to scotch the underlying assumption of such proposals that there is one ideal basis of classification. There is no such thing. No one system can supply the statistics necessary to meet the multitude of needs of users and analysts. Although a statement of this kind, at one time, would mean the end of discussion after agreement on the one best compromise system, in this day of the computer it is not impractical to think of a variety of systems, each tailored to a specific set of problems, and our thinking is exactly along these lines.

TABLE A.--OCCUPATION DISTRIBUTION OF "DEPARTMENT" TYPE RETURNS BASED ON CLEVELAND FOLLOW-UP PROBES

Occupation by Original Department Entry	Percent of Employed
Shipping Department	100.0
Shipping and Receiving Clerks	71.4
Laborers (n.e.c.)	14.3
Operatives (n.e.c.)	14.3
Stock Room	100.0
Stock Clerks	41.7
Purchasing Agents and Buyers (n.e.c.)	8.3
Operatives (n.e.c.)	8.3
Laborers (n.e.c.)	25.0
Clerical and Kindred Workers (n.e.c.)	16.7
Kitchen Help	100.0
Kitchen Workers (n.e.c.)	73.3
Cooks exc. pr. household	10.0
Housekeepers and Stewards	3.3
Managers (n.e.c.)	6.7
Prof. Tech. and Kind. (n.e.c.)	3.3
Service Workers exc. Pvt. Household	3.3
Office Workers 1/	100.0
Specified clerical occupations	60.9
Clerical and kindred workers (n.e.c.)	39.1
Specific duties	30.4
Broad duties	8.7
Miscellaneous Dept. Entries	100.0
Operatives and Kindred Workers (n.e.c.)	15.6
Clerical and Kindred Workers (n.e.c.)	9.4
Tech., Medical and Dental	9.4
Laborers (n.e.c.)	12.5
Furnacemen, Smelters and Pourers	6.3
Office Machine Operators	6.3
All Others	40.6

1/ Entries of "Office Workers" were classified as "Clerical and Kindred Workers (n.e.c.)" in the 1960 Census.

TABLE B.--MAJOR NOT ELSEWHERE CLASSIFIED (n.e.c.) OCCUPATION CATEGORIES BY GENERAL AND SPECIFIC COMPONENT FROM THE 1950 CENSUS AND 1965 CPS

(Percent)

Major n.e.c. category	1950 Census		1965 CPS	
	General Title	Specific Title	General Title	Specific Title
Both Sexes				
Managers, officials and Proprietors (n.e.c.)	61	39	69	31
Clerical and Kindred workers (n.e.c.)	68	32	49	51
Salesmen and sales clerks (n.e.c.)	70	30	60	40
Mechanics and repairmen (n.e.c.)	82	18	56	44
Operatives and kindred workers (n.e.c.)	26	74	19	81
Laborers (n.e.c.)	69	31	45	55
Males				
Managers, officials and Proprietors (n.e.c.)	61	39	69	31
Clerical and kindred workers (n.e.c.)	55	45	39	61
Salesmen and sales clerks (n.e.c.)	62	38	31	68
Mechanics and repairmen (n.e.c.)	NA	NA	56	44
Operatives and kindred workers (n.e.c.)	NA	NA	19	81
Laborers (n.e.c.)	NA	NA	45	55
Females				
Managers, officials and Proprietors (n.e.c.)	59	41	75	25
Clerical and kindred workers (n.e.c.)	75	25	55	45
Salesmen and sales clerks (n.e.c.)	82	18	95	5
Mechanics and repairmen (n.e.c.)	NA	NA	NA	NA
Operatives and kindred workers (n.e.c.)	NA	NA	19	81
Laborers (n.e.c.)	NA	NA	27	73

TABLE C.--EMPLOYED IN MAJOR N.E.C. OCCUPATION CATEGORIES FOR ST. LOUIS PARK, MINN., 1966, AND HENNEPIN COUNTY, MINN., 1960

(Percent)

Major n.e.c. category and not reported	St. Louis Park 1966 1/	Hennepin County 1960	Percent change
Major n.e.c. group total	26.6	31.1	-14.5
Managers, officials and proprietors (n.e.c.)	6.5	7.6	-14.5
Clerical and kindred workers (n.e.c.)	5.3	6.4	-17.2
Salesmen and sales clerks (n.e.c.)	6.4	7.7	-16.9
Mechanics and repairmen (n.e.c.)	0.3	1.6	-81.2
Operatives and kindred workers (n.e.c.)	5.3	5.0	+ 6.0
Laborers (n.e.c.)	2.8	2.8	-
Occupation not reported	1.8	4.6	-60.9

1/ Adjusted for differences in basic distribution of major groups.

CLASSIFICATION OF OCCUPATIONS: SOME PROBLEMS OF ECONOMIC INTERPRETATION*

G. G. Cain, W. L. Hansen, B. A. Weisbrod, University of Wisconsin

Introduction

A well-conceived system of occupational classification can be a valuable tool for facilitating empirical work in economics on topics concerned with labor markets, manpower supplies and requirements, wage structures, and the like. This paper considers the nature of such a "well-conceived" system.

Despite the potential importance of occupational data and the interest of government in them, little effort has been given by academic economists to examining the conceptual basis of the occupational classification system or to suggesting modification in it in the light of their own objectives.[1] Now that we are confronted with a new occupational classification system (set forth in the third edition of the Dictionary of Occupational Titles), and now that the Bureau of the Budget has established a committee to re-examine occupational classifications, it is past time for economists, along with others, to indicate their own needs and to assess the usefulness of the existing and alternative classification systems.

The fundamental position of this paper is that occupational classifications can serve economists as part of a larger information system, a system designed to reveal more about the current and prospective labor-resource flexibility of the economy. This concern about labor-resource flexibility is the essential theme of this paper. And we suggest that occupational classifications need to be defined and developed with this objective in mind to be useful to economists. We are aware that other disciplines may view the objectives differently.

Labor-resource flexibility is, of course, only part of a larger class of resource substitution issues with which economists are concerned--substitutions between capital and labor, in particular, receive and deserve considerable attention. For this reason, the emphasis on flexibility among types of labor that is implicit in the emphasis on occupational data may be too narrow for certain problems, as is noted later in the paper.

The following remarks are organized in two parts. The first focuses on the meaning of the term, "occupational data classification," and also on the uses to which such a classification system can be put. The second discusses the attributes of an "ideal" system of occupational data classification. We concern ourselves for the most part with conceptual issues rather than with the empirical implementation of any particular occupational classification and data system.

The Meaning and Uses of Occupational Classifications

By a "system of occupational data classification" we mean two things: First, one or more sets of categories that provide (1) job-skill descriptions and (2) worker-skill descriptions; and second, sets of data relevant to those categories--that is, for a given period of time, data on (a) number of jobs available at specified levels of skill and at specified levels of wages and (b) numbers of workers possessing specified levels of skills and willing to supply those skills, at specified levels of wages.[2] In short, a system of occupational data consists of useful sets of boxes filled with corresponding series of quantity measures.

As economists our interest in such a system falls into two major categories: (1) We are interested in the process by which employers choose among the alternative types of labor that are capable of producing given goods and services. (2) We are interested in the process by which individuals choose amongst alternative job and career opportunities. To understand and predict the outcomes of these processes, economists need three types of information: (a) information about current and expected factor prices, (b) information about the factor substitution possibilities that are technologically feasible, and (c) information about the preference patterns that determine the willingness of people to take one job or another. Without such information, we can say little or nothing about the choices employers and workers will make in response to changes in technology or relative factor prices. The lack of this information, moreover, is the principal reason why manpower projections and occupational forecasts are so often empty of economic content.[3]

---

*This is a revised version of a paper presented at the American Statistical Association meetings, Los Angeles, California, August 1966.

[1] The following two articles have recently examined the problem from the standpoint of economic analysis: John Dunlop, "Job Vacancy Measures and Economic Analysis," The Measurement and Interpretation of Job Vacancies, A Conference Report of the National Bureau of Economic Research, Columbia University Press, New York, 1966, pp. 27-47; and James Scoville, "The Development and Relevance of U.S. Occupational Data," Industrial and Labor Relations Review, October, 1965, pp. 70-79. An earlier article that discussed the concepts of occupational classifications in the course of analyzing the distribution of income is, Jan Tinbergen, "Some Remarks on the Distribution of Labor Incomes," International Economic Papers, Vol. 1, 1951, pp. 195-207.

[2] The fixing of the point in time controls for the state of technology, price structure, and other variables that need to be given to permit an economically meaningful count of jobs and workers.

[3] For further discussion of this point see W. Lee Hansen, "Labor and Force and Occupational Projections," Processing of the Industrial Relations Research Association, 1965, pp. 10-20.

Occupational data, as defined above, can be helpful in providing information about (b) and (c)--that is, about substitutions that employers and workers are willing and able to make. But in any case that information must be supplemented with data on current and anticipated factor prices.

Existing data on occupations do provide some information relevant to substitution possibilities, since workers within occupational classes--as those classes are currently defined--tent to be better substitutes for each other than are workers in different occupational classes. But the extent to which this is the case is worthy of further study.

As we have noted, occupational data can be part of an information system which facilitates decision-making regarding (a) production planning (short-run and long-run), and (b) job and career selection (short-run and long-run). We now consider each of these.

(a) Production Planning. For given conditions of product demand, production plans will be made by cost-minimizing decision-makers on the basis of information concerning the availability of various skills or occupational groups --information which is given by an occupational data system--and on the basis of schedules of prices at which these skills are offered--which is not normally given by such a system. This knowledge is then translated into "lowest-cost" production techniques by employers. These adaptations of production plans to resource availability and prices are a principal cause of substitutions in the mix of labor inputs demanded.[4] The substitution possibilities are, of course, a principal determinant of the elasticities of demand for various types of labor.

Actually, the process of determining "lowest cost" production techniques is more complex than just suggested. Decision-makers in fact are not confronted solely by choices among various types of labor, but also by choices among various types of capital and, probably even more important, substitution between various types of labor and capital, particularly in the long-run. For this reason, any occupational data system ideally should be meshed with a broader information system which takes into account these other substitution possibilities--the full range of which is encompassed by the concept of the production function.

In this context we might note that discussion of "needs" and "requirements" for workers with various skills are likely to be seriously deficient because they imply that substitutions among types of labor and between labor and capital cannot be made. The empirical estimates that have been made of production functions have concentrated on substitutability between capital and a single undifferentiated labor input.[5] There is reason to expect and to hope, however, that future work will disaggregate labor into various occupational subtypes, which will provide estimates of elasticities of substitution among such subtypes of labor. A significant degree of detail involving many occupational classes, however, will probably not be incorporated in the statistical models of production functions for some time.

(b) Job and Career Choice. These choices are made by individuals largely on the basis of information about the relative attractiveness and remuneration associated with various kinds of work, the costs of securing the education and training necessary for entry into various occupations, and the ease of shifting from one type of work to another. The third type of information might well be provided by an occupational data system, while the first two are unlikely to be provided. Again, we see that an occupational data system must be regarded as part of a larger information system, in this instance one producing indicators of long-run elasticities of supply for various types of labor.

The two uses of occupational data just noted included the possibility of adopting new production techniques and of acquiring new skills and both of these require time. But even in the very short run when these possibilities may be limited, occupation data can improve the functioning of the labor market through reducing the costs of search to both employers and workers. If jobs and workers' skills were identified and described more precisely, the process of employers finding workers and workers finding jobs would be facilitated. And it is interesting to note that the resulting reduction in the costs of search could come about because of a standardized classification alone, i.e., even in the absence of data on the number of jobs and workers available at various prices.

Attributes of an Effective Occupational Classification and Data System

To serve effectively the purposes of economists, an occupational classification and data system should possess a variety of attributes that would contribute to the likelihood that the system would be useful for the purposes indicated above.

1. The first attribute we suggest stems directly from our insistence that the basic purpose of occupational classifications for the economist is the provision of information about factor substitution possibilities. On this basis occupational classes should be relatively homogeneous, in the sense that a high degree of

[4]These demand patterns will sometimes produce their own supply responses as employers adjust their level and composition of training activities.

[5]For a useful review of the extensive literature on this subject, see Marc Nerlove, "Statistical Production Functions: A Selective Review," to be published in the volume in the series, Conferences on Income and Wealth, the National Bureau of Economic Research.

substitutability should exist within each class. Specifically, for any given level of aggregation of occupational classifications:

(a) Each "class" of jobs should be such that the elasticity of substitution among jobs in that class (or, rather, among various workers who can perform those jobs) will on average be higher than the elasticity of substitution between jobs in different classes. We use the term, "elasticity of substitution," in its conventional sense--as a measure of the technical ease with which one input may be substituted for another to obtain a given output. The higher the elasticity, the greater are the substitution possibilities.

(b) Each "class" of workers should be such that cross-elasticities of supply among workers will on average be high than the cross-elasticities between workers in that class and those in other classes. (Here, too, the higher the cross-elasticities, the easier it is to substitute one worker for another).

The first condition views the classification from the employers' standpoint and depends on the technical production function, which specifies the extent to which factor substitutions among types of labor are possible. The second condition views the classification from the workers' standpoint. On the workers' side, the substitutability of one job for another depends on workers' preferences along with their abilities to perform various tasks. In short, the elasticity of labor supply for any given type of work depends on both the ability and willingness to perform those tasks. The greater the change in remuneration required to cause workers to switch types of work (which may involve the acquisition of more education or training), the lower is the cross-elasticity of supply, and the more disparate are the two types of work.[6]

It may be noted that our emphasis on cross-elasticities as a criterion for defining classes of jobs and classes of workers follows the usual definition in economics of an "industry" as compromising those firms that produce goods for which the cross-elasticity of demand is high.[7] Similarly, a "commodity" is often thought of as a group of (not-necessarily identical) items for which the cross-elasticity of demand is high--for example, automobiles. Our criterion for determining "job" classes also uses this cross-elasticity-of-demand concept, while our criterion for determing "worker" classes is that cross-elasticities of supply of workers should be high.

In principle, the elasticities between any and all pairs of occupation can be measured. Although it may be objected that this would be prohibitively expensive, given the paucity of current information on this subject and the potentially large number of occupational classes, the same objection could be made concerning economists' definitions of "commodities" and "industries." But this has not prevented us from devising useful--if less than ideal--groupings of commodities and industries that have widespread acceptance. For example, although all automobiles are not the same, it is nonetheless useful to discuss and to forecast the demand for a group of diverse vehicles that are defined, at least implicity, in terms of substitutability criteria, and which we label "automobiles."

Our approach to the problem of classifications may be illustrated in the context of the policy-oriented debate over "shortages" of engineers. In preliminary work on another research project[8] we have found that occupational mobility (by several measures) appears greater between the occupational groups of surveyors and civil engineers than between the[9] groups civil engineers and aeronautical engineers Assuming these findings hold up in the completed analysis, we would draw the following policy implications:

(1) any projected shortage of civil engineers would be significantly lessened if large numbers of surveyors existed;

(2) any projected shortage of aeronautical engineers would be lessened only slightly by the presence of large numbers of civil engineers;

[6]John Dunlop has suggested an approach to categorizing occupations involving two dimensions: (1) "job families," that include the characteristic of a "common mobility pattern," and (2) "job content," that is related to the tasks performed. This approach is consistent with the emphasis we have given to workers' substitution decisions--which ties into Dunlop's first point--and to employers' substitution decisions--which relates to point (2).

[7]For example, "a commodity group with high cross-elasticities (of demand) within the group but with low cross-elasticities with respect to other commodities is often said to constitute an industry." Richard H. Leftwich, The Price System and Resouce Allocation (New York: Holt, Rinehard and Winston, 1966), p. 43.

[8]The research, by Cain and Hansen, uses the 1962 Post-Censal Survey of scientists, engineers, and technicians, and is sponsored by the National Science Foundation. A first report will be available in early 1967.

[9]This finding actually refers to observed inter-"occupational" mobility. Strictly speaking, for the illustration to be valid evidence for our point it should be true that an equivalent percentage change in wage rates of surveyors and civil engineers, and of aeronautical and civil engineers (or better, "net remuneration") should bring about greater occupational mobility in the former case. The available data on occupational mobility, unfortunately, do not disclose the magnitude of changes in relative wage rates that led to the actual occupational shifts--even assuming that the only reason for the shifts was the change in wages.

(3) the occupational category of "engineers" is either too broad or, if a gross level of aggregation is desired, the category should include surveyors.[10]

Furthermore, the implications of these findings for individuals making career choices in the field of engineering are apparent and significant. As one example, a "large" pool of surveyors would tend to moderate wage increases among civil engineers while a large pool of aeronautical engineers would not.

The new edition of the Dictionary of Occupational Titles provides another illustration to which our approach applies. We note that there are separate classifications of "salesman" and "salespersons," and within these classes there are 150 subtypes. By contrast, the occupation, "faculty-member college or university," contains no sub-types. By the cross-elasticity criterion we propose, the sales classifications have excessive detail while the faculty-member class has too little. The cross-elasticities for employers and workers between many of the sales sub-types appear to exceed greatly the cross-elasticities among, say, professors of physics, English, Latin, etc.--not to mention between labor economists and mathematical economists. Incidentally, this example shows that the numbers of individuals in an occupational class may be a poor indicator of the amount of within-class homogeneity. The number of people in a subtype of sales workers would surely be larger than the number in a subtype of professors.

2. As a second attribute, closely related to the cross-elasticity notion, the occupation classification system should be applicable to both types of substitutions--by employers and by workers. A set of categories useful for analyzing one type of substitution may not be useful for the other. For example, an employer may regard two technicians as equivalent to that of one engineer for the performance of certain tasks. Thus, from the employer's point of view, the system should be such that these two occupations can be combined, perhaps with some differential weighting scheme. From the workers' point of view, however, the different training required, the varied work activities performed, and the disparate salary levels paid may make the two occupations relatively poor substitutes. Thus, with respect to choices made by workers, the two should not be put into the same class.[11]

3. The informational function of occupational categories clearly requires that job descriptions be codified, in terms of skill requirements, and that worker-skill descriptions be codified, in terms of skill capabilities. In this way the process of staffing by employers, and job-finding by workers would be facilitated. Just as in commodity markets where standarization of size, quality, and so on has been used effectively, similar gains could be achieved by the standarization of skill descriptions.

The desirability of some degree of standardization can hardly be disputed, but the difficult question is exactly what features of jobs and characteristics of workers should be standardized. This issue is beyond our area of expertise, but we are intrigued by the possibility of using a set of quite basic or elemental skill attributes as the building blocks for classification. Thus, it would be desirable to know the level of attainment that a person has in such skills as manual dexterity of various types, the ability to get along with people, or the ability to do abstract reasoning, as well as various combinations of these skills. Occupational titles or categories would then correspond to various combinations and levels of such "elemental attributes."[12] The particular strength of this feature is the possibility of developing a continuum of gradations of job requirements and skill attainments, and of being able to recombine different groupings to suit the needs of various users. We are dubious that the present state of knowledge permits the speedy development and application of such a system, but we applaud the attempts along this line which appear in the new 1965 edition of the Dictionary of Occupational Titles.

4. The occupational classification system should be adaptable to changes that occur over time. Changes in technology and educational policy may bring about new types of jobs and different skill levels which, in turn, alter the range of substitution possibilities. An important advantage of classifications based on rather basic and elemental skills, is that they could be restructured without great loss in continuity.

5. Occupational categories should be presented at a level of detail that pays heed to the costs of obtaining it. In short, there is a level that balances the benefits of additional detail--information that would enable workers to pick jobs more suitable to their skills and preferences, and that would permit employers to select more easily workers with the desired skills--against the costs of obtaining the additional detail. In view of the heterogeneity of uses for labor market data, the optimal level of detail will vary among uses and among occupa-

---

[10] We would not advocate obliterating the distinction between the classes, "civil engineers" and "surveyors"; rather we suggest that it will be useful to combine and rank these occupations in different ways than have been customary.

[11] This example may be compared with the case of certain types of engineers and physical scientists. They may be good substitutes for employers in production, and at the same time the two occupations could also provide alternative career paths--lateral occupational mobility--for people trained as one or the other, and thus they would be considered good substitutes to workers as well.

[12] Such a system was proposed by Tinbergen, who wanted to analyze the supply and demand of these types of skills to determine the distribution of wages and salaries that would result. Tinbergen, op. cit.

tional groups. We could point, again to the new Dictionary of Occupational Titles and sales people and professors as specific examples of excessive and insufficient detail, respectively, in occupational classifications.

Conclusion

In attempting to sketch out the kind of occupational classification and data system that would be of interest to economists, we have tried to make clear the main uses to which such a system could be put. These uses--involving substitution possibilities among different types of labor--are not likely to be the same as those of sociologists or others. Moreover, a system useful to academic or government researchers, be they economists or sociologists, may not be of greatest value to employers, workers, and government officials responsible for action programs. In short, many goals and objectives must be taken into consideration before we can determine what is an optimal system of classification, and, in fact, we have suggested that no single system is likely to be optimal for all purposes.[13]

In developing and implementing a system it should be borne in mind that more-detailed standardization, added flexibility, and, for that matter, added precision in measurement, all come at increased costs. Thus, users of the data system--researchers as well as labor and employer groups--have a serious obligation to justify in terms of real benefits to be produced, the demands they make on governmental agencies responsible for developing an occupational classification system. At the same time, government agencies responsible for these programs have an obligation to be responsive to the needs of users of the system.

Irrespective of the particular type of occupational classification system chosen, too much should not be expected of it. An occupational classification system and the data it generates serve at best as a proxy for one class of variables--labor and job substitutions--with which we as economists are concerned. Thus, we reiterate our hope that any system of occupational classification will be recognized as only one part, albeit a potentially important one, of a larger system of information for decision-makers.

---

[13] James Scoville offers a useful discussion of the various purposes for which occupational classifications and data were collected as these purposes evolved over the years. He mentions briefly the economic, analytical purpose concerning the "characteristics of manpower and technological change, such as training and education requirement," and suggests that the present system has serious shortcomings for this purpose. James Scoville, op. cit. especially p. 78.

DEVELOPMENT OF A CONVERTIBILITY LIST BETWEEN THE DOT AND CENSUS CLASSIFICATION SYSTEMS

Leon Lewis, United States Employment Service

The Convertibility List of Occupations published by the Bureau of the Budget in 1940 related the occupational classification structure used in the 1940 Census of Population with that in the first (1939) edition of the Department of Labor's Dictionary of Occupational Titles. The third edition of the Dictionary of Occupational Titles, published December 1965, represents a basic change in the structure of occupational classification. It is not merely an updating of the original classification system in the Dictionary. The old Convertibility List, therefore, is no longer usable in connection with it. Since a number of Federal agencies and a number of non-government organizations use both the Census and DOT classifications in developing and presenting manpower data, it is essential that a new Convertibility List be developed.

The third edition of the Dictionary of Occupational Titles contains the most comprehensive information about occupations and their worker traits requirements developed to date. For the first time we have identified, defined and classified jobs not only in terms of the traditional tasks performed, but also in terms of requirements made on the workers. For every job in the Dictionary we have obtained not only information about the nature of the work performed; the materials, products, subject matter, and services involved; and the machines, tools, and equipment used; but also identification of the specific functions performed by the worker; the significant aptitudes, interests, and temperaments involved; the critical physical demands and working conditions; and the training time required. We are now in the process of developing a convertibility table between the DOT and the Census classification systems to permit relating this information about occupations to numbers of people in the work force.

More specifically, we have the following new items of information for each defined job in the DOT:

Classification of the job in terms of 23 worker functions related to data, to people, and to things. Examples of these functions are "analyzing," "computing," "supervising," and machine "tending."

The degree required, on a 5-point scale, of each of 11 aptitudes. These include the 9 aptitudes of the U.S. Employment Service's General Aptitude Test Battery, plus two additional aptitudes, eye-hand-foot coordination and color discrimination.

Selection of those interests significant in the job from a list of 10 factors.

An indication on a 5-point scale of the degree of strength required, and a selection of other critical physical demands from a list of 5 factors.

An indication of whether the job involves working inside or outside, and a selection of other critical working conditions from a list of 6 factors.

An indication of the degree of general educational development required from a 6-point scale of reasoning, mathematical, and language development.

An indication of the specific vocational preparation required on a 9-point scale of months and years.

Although this information is available about occupations, it has not yet been related to numbers of workers. The Bureau of the Census provides information about the occupational distribution of the work force. By relating these two classification systems, these two kinds of information could be brought together. This would be an important step in providing information on the supply side of the labor force equation.

For example, we know the general educational development level required for a worker to perform in each occupation. If we can relate this to the number of workers employed in each occupation, we could develop a distribution of the minimum general educational development levels of the American work force as demonstrated by the requirements of the occupations they hold. Similarily, by determining the distribution of all of the worker traits factors, a worker traits profile of the working population could be developed. This could be a milestone in such manpower planning activities as facilitating utilization of available skills in the work force.

Two approaches are being undertaken to develop this convertibility table: (1) Assignment of DOT classifications to all job titles in the Census Classified Index of Occupations and Industries, and (2) Assignment of DOT classifications to several thousand Current Population Survey household returns.

The first approach involves dividing the 296 Census 3-digit occupational categories among seven occupational analysts. Their assignment is to identify for each Census job title, in a specific industry, the DOT title and code to which it probably converts. The term "probably" is used because Census titles are not defined and conversion depends on the analysts' judgments based on available data. A high degree of similarity exists between Census categories and DOT 3-digit groups with respect to professional, clerical, sales, service, agricultural, and craft occupations. We do not expect conversion in these categories to be difficult. However, there is little similarity between Census categories and DOT groups with regard to the bulk of industrial occupations and most of the conversion problems occur among the industrial occupations. Of the 296 Census categories, preliminary conversion to DOT titles and codes has been completed for 111 categories containing almost 5,000 of the 24,000 Census job titles.

About 27% of these Census conversions presented problems. Most of these problems stem from ambiguity resulting from the lack of definition of the Census title in terms of specific job duties. In about 10% of the cases conversions were attempted but it was indicated that some doubt remained about the adequacy of the conversion. An example is the conversion of the Census title "Heavy-Equipment Operator" to the DOT title "Operating Engineer." About 9% of the Census titles could be converted to several DOT titles and codes so that no single conversion could be made. For example, the Census title "Graduate Nurse" converts to most titles in the DOT 3-digit group "Registered Nurses." However, there is no problem in converting from a title in one system to a classification group in the other system. For about 8% of the Census titles, no equivalent DOT title and code could be located. Examples are "Acid Painter" in the Glass and Glass Products industry; and "City Employee."

It is estimated that about 15% of the conversions will present problems requiring resolution by a joint Census-USES team of experts at the conclusion of the preliminary conversions by the seven occupational analysts.

The second approach is that of classifying several thousand Current Population Survey household returns. About 50,000 returns from one month's survey are available for this study. 4,919 entries have been coded so far. Only 6% of this number have presented classification problems. The most common problems involve either insufficient information to relate the job to any DOT classification, or a job that could convert to too many DOT classifications. Examples of this latter problem would include Truck Driver, where the DOT distinguishes between drivers of light and heavy trucks; Dishwasher, where the DOT distinguishes between hand and machine dishwashing; and Farmer, where the DOT distinguishes among different types of farming at the 2-digit classification level.

The distribution of the 4,919 returns classified to date indicates that we probably already have a sufficient sample for this study. These returns are distributed throughout 80 of the 83 2-digit divisions in the DOT and represents about two-thirds of the 603 3-digit groups. The next step will be an analysis of the data classified in order to evaluate the meaningfulness of the data resulting from using the type of occupational structure embodied in the DOT for Census and Current Population Survey returns. There is also under consideration the possibility of classifying the 1970 Census of Population returns in terms of the DOT system so that a complete convertibility list between the DOT and the 1970 Census could be produced by machine.

A proposal has been made to ask a few large-scale establishments to attempt to code their work forces by the new DOT classification system. Interest has been expressed by several employer associations in the establishment of a general occupational classification scheme that could be utilized for their own statistical purposes, and it is hoped that perhaps from among these some volunteers may be obtained. Such an undertaking would be useful from a number of viewpoints: establishment evaluation of the new DOT classification structure would be very valuable; experiments with having establishments code their employees by a standard occupational scheme, rather than merely return occupational descriptions to be coded by a central agency, would have implications for the development of any large-scale system of collecting occupational data in this way; the success with which establishments can convert their own occupational designations to a standard classification would be an indicator of the feasibility of developing a U.S. Standard Occupational Classification System.

We are also concerned with the relationship between the International Standard Classification of Occupations as developed by the International Labour Office and the classification systems used in the United States. A review of the tentative revision of the ISCO leads us to believe that its comparability with the new DOT classification system is very high. Similarities appearing for the first time include: (1) The expansion of the professional and technical area, (2) The more functional classification of clerical occupations with a distinction between machine and non-machine related jobs, (3) The classification of industrial occupations in such categories as "processing" and "structural work," and (4) The classification of first-line supervisors in the same group as those supervised. At the 3-digit level even the language of the group titles is much closer to that of the third edition DOT than was the case with the 1949 DOT and the 1958 ISCO.

The development of a convertibility list between the ISCO and the DOT will be a less complex task as compared with that between the DOT and the Census system. For the most part, it will be possible to relate the two systems on the basis of equivalent groups, with relatively little necessity to go to the individual occupation before a conversion can be made.

The Interagency Committee on Occupational Classification of the U.S. Bureau of the Budget plans as one of its long range objectives to investigate the possibility of establishing a Standard Occupational Classification, analogous to the Standard Industrial Classification, and to establish such a system if it appears possible. It is interesting to note that a Standard Occupational Classification was one of the original objectives of the committee when it was first instituted over 25 years ago, and that it was never able to establish this although the development of the original Convertibility List was a first step in that direction. Today, the development of such a system appears more feasible than at any time in the past. The existing comparability between the DOT and ISCO systems, the development of a Convertibility List between the DOT and Census systems, and the interest of employer associations are all leading in the direction of such a Standard Occupational Classification System.

Discussion

Stanley Lebergott, Wesleyan University

A.

Messrs. Hodge and Siegel present a mass of thoughtful provocative analysis. I shall regretfully have to ignore this and concentrate on their three key conclusions.

1. "The most important deficiency in the (Census) occupational classification" they say, "is heterogeneity of the specific jobs assigned to many detailed codes." P. 31. "For the sociologist who wishes to use detailed occupation as an indicator of social economic status, the lack of homogeneity of the n.e.c. code can only comprise a major source of inaccuracy which afflicts 32.8% of the...labor force." P. 9. Two comments on this conclusion.

a) That the Census labels occupations as "not elsewhere classified" hardly creates any "inaccuracy." Just the opposite: it gives a precise and accurate definition.

b) The central problem here, however, is lack of information on the part of the respondent. Indeed although I applaud the enterprise and responsiveness of the Census in investigating the point, I do not imagine that the reduction of the n.e.c. category from 36% to 30% of the labor force -- I utilize Mr. Greene's table of ratios as a guide -- will cheer Messrs. Hodge and Siegel very much either. Manipulating the occupational classification is going to do little about the respondent's lack of information of the kind we seek, and the lack of time in a national decennial inquiry for extensive inquiry.

2. They urge "a complete revision of the system of occupational classification" to rely on a single criterion -- "similarity of tasks performed." P. 3. Doing so would replace the Census use of "many characteristics" p. 32, the Census code being "at best a multitudinous typology" at its best. p. 6. Given a single dimension for classification sociologists could more clearly test the relationship between occupation on the one hand and industrial affiliation (p. 32), subordination and superordination (p. 4) etc. on the other.

a) I can well understand the desire for a single criterion. But the Census must serve many uses, as the Cain, Hansen, Weisbrod paper notes. Its first responsibility must be to use as many relevant dimensions for classification as it has available. It is up to the analyst in a specific area then to try to redeploy the data for his uses.

b) Judging from the examples the paper gives, however, this is an unnecessary battle. For example, the title "optometrist" is given high marks because it relates to tasks performed, while "foremen" "messengers" and "office boys" and "buyers of farm products" are given low marks -- because they are defined on the basis of "the relationship of subordination and superordination," place of work, etc. I really don't believe the Census is that fancy. The title of foreman is usually a handy proxy for a set of tasks, involving supervision. Similarly that of messenger also indicates a set of tasks, although also redolent of place of work.

c) At the same time, they contend, that the Census unwisely distinguished a variety of occupations -- e. g., "receptionists" from "attendants in physicians' offices" although "many perform tasks essentially the same." Here I am puzzled. Why should we not accept as full detail as is feasible? For a particular problem we always have the option of pooling Census data to a grosser level.

3. They do not want to sacrifice "intercensal comparability <u>in any degree</u>" - this in order to permit cohort analysis, etc. How the other changes they want are to be achieved without a break in comparability is difficult to see. However, I should like to differ with this proposition, and for two reasons. As probably the only person present who has utilized the occupation estimates back to 1820 to make comparable series, I am intellectually aware of the delights of comparability. And having spent an agonized six months trying to adjust Census of Mfrs. data for comparability I am emotionally on their side. Nevertheless, I differ flatly with this proposition for two reasons.

a) It is wrong. The Census data are primarily used for clear and present problems. As these problems change, as our insight grows, we must make changes. Since the structure of any government agency, and its scholarly advisory committees, normally slows down the rate of change, I also think it superfluous to be worried about the rate of change in the classification.

b) It is unnecessary. Given modern sampling and modern computers, samples of the micro observations can be maintained and recombined in any scheme that a scholar may desire.

I should like to turn to two general observations that bear on both papers. One gets the feeling that we have been discussing Hamlet without referring to the Dane, or the royal family, and concentrated instead on Ophelia. We have had no reference to one of the most important occupational classifications in use in the nation, namely that used in the BLS wage rate surveys. The wage rate helps define "classified" and "not elsewhere classified" occupations as few other single bits of evidence do. For economic analysis we should be relying much more on the surveys that can give us wage rates. The report that one "mechanic" receives $1 an hour and another $5 tells us a good deal about what kind of a mechanic each one is.

4. Amidst all this pother about classifications we have had not even a passing regret about the quality of the data. In a real world if we persuade the Census to spend a lot of time on jiggling the classifications, the staff concerned will have little time, and no energy, to pursue ways of improving the data. I find it hard to know why that is a wise choice. The 1960 evaluation results have been provided us by the Census, with a frankness in testing its own work not matched by any public or private agency with which I have any acquaintance. And we are ignoring its findings. They tell us, for example, that on the average, 28% of the employees classified by Census in any major occupation group belong in <u>another</u> major occupation group if one relies on their employers' reports. The percentages reallocated from individual occupations must be at least as staggering.

Some years ago I proposed that we utilize employer reports for occupation industry and wage rate. Given wage rate data we could classify more accurately. Given employer reports we could come closer to tabulating data that are worth tabulating.

B.

I assume that we are all in hearty agreement with Messrs. Cain, Hansen and Weisbrod that an effective system should have standardized and homogeneous categories, should report an optimal level of detail. Certainly Messrs. Greene and Lewis ought to agree. The rub comes in the rules of application.

To classify jobs so that the elasticity of substitution is greater within jobs in a class than among classes is triply difficult. (1) Elasticities change as the structure of relative prices do: Hence a 1960 classification might no longer suit by 1970. (2) We still need a principle for aggregating 75 million workers into less than 75 million occupations. Perhaps there is less substitution between GM foremen and Ford foremen than between operatives and foremen within either firm? Do we show an occupation title for "GM foremen" or do we put up with the combined foremen category despite the substitution ratio? I can imagine no congregation willing to substitute the Reverend Billy Hargis for the Reverend Martin Luther King, or more generally, Baptist ministers for Greek Orthodox ones -- although within denominations substitutions of ministers for social welfare personnel occurs. Do we therefore cease to tabulate data for ministers? (3) I would much like to see the authors compute elasticities for even a few pairs of occupational classes so that we could also join in evaluating this novel, potentially very helpful criterion.

VIII

MIGRATION AND ECONOMIC OPPORTUNITY

Chairman, HENRY S. SHRYOCK, U. S. Bureau of the Census

Page

SCHOOLING, EXPERIENCE, AND GAINS AND LOSSES IN HUMAN CAPITAL THROUGH MIGRATION

Mary Jean Bowman and Robert G. Myers, University of Chicago

Since the revival of the concept of human capital in the 1950's, many areas of economic research, including the economics of migration, have undergone a rethinking. Although neither the idea of human capital nor its application to migration can be termed "new," current estimating procedures and their economic applications are much more sophisticated than earlier ones.[1] In considerable part new approaches and greater sophistication reflect improvements in the quality and availability of data. However, new appreciation of critical problems is always some jumps ahead of the data -- and also, we might add, behind them.

We will first sketch briefly several recent treatments of human capital in migration. Some of their methodological implications will become evident in a reexamination of one-way migration that takes into account place and timing of schooling and work experience. We turn then to the more complex analysis required when we include remigration possibilities. Along the way suggestions will be made for useful new tabulations of 1960 census data together with some pleas or hopes for the 1970 census.

In discussing the relation between "Migration and Economic Opportunity," we might emphasize migration as a means of responding to economic opportunity or as a means for creating economic opportunity. Heretofore more attention has been given to the passive view of migration as an allocative mechanism than to the view of migration as playing a dynamic training role. But migrants often move out and return with new skills, and in-migrants bring and transfer know-how. We emphasize these linkages among investments in schooling, experience, and migration. We present decision models incorporating various migration and re-migration sequences and argue their value as tools for analysis of migration behavior, human resource policies, the transfer of know-how, and the diffusion of development.

I. RECENT APPLICATIONS OF HUMAN INVESTMENT MODELS TO MIGRATION

A. Migration as a private investment

In a 1962 article, Larry Sjaastad[2] looked at migration primarily as a form of private, rational decision-making -- as a private investment that entails costs and engenders increments to lifetime earnings streams. To oversimplify, people discount expected earnings streams to set present values on themselves for alternative courses of action: migration or remaining put. In theory, people will move if they can increase their present value by an amount greater than the cost of moving. This added value might be attained if, by moving, a person finds a better market for his existing skills, upgrading himself within his occupation, or if migration provides an opportunity for him to change occupations, thereby acquiring a new skill and increasing his remuneration. Sjaastad emphasizes occupational change in attempting to analyze rural-urban migration and to account for observed age patterns in such migration.

Costs, according to Sjaastad, include direct costs of moving, earnings foregone while moving, earnings foregone while searching for employment and training for a new position. He gave special emphasis to the training or retraining costs of urban newcomers, which are identified with the initial excess of foregone rural earnings over realized urban wages. Returns (benefits) are the expected income stream at the destination.[3]

Sjaastad's primary concern was with the efficiency of migration as a process of resource allocation. He argued that much of the seemingly non-rational response or lack of response to economic incentives to move is really a matter of measurement methods which look at net migration rather than gross figures. He also criticized failures to disaggregate populations sufficiently before associating them with differential opportunities and earnings. In order to identify statistically the potential earnings streams of migrants at places of origin and destination, Sjaastad suggested that comparisons must be made among more homogeneous sub-groupings, specifying particular use of age-occupational classifications. (Curiously, he said nothing about classifications by schooling.)

Statistical averages of costs and earnings for each population category are used as proxy measures in a decision model that is fundamentally individualistic. Despite his work with the Upper Midwest Project, Sjaastad's interest in spatial resource allocation remains spatially neutral; hence we find neither aggregated human capital measures nor "regionalism." The important thang is the micro-economic allocative process.

B. Migration and a rationale of community investments in education

In his treatment of migration, Burton Weisbrod[4] takes a very different tack -- even though he starts from the same base in human investment theory as Sjaastad and, like Sjaastad, computes "human capital" in present value terms.

Weisbrod sets up a model in which the community is treated not only as an aggregative entity that receives benefits and incurs costs, but also as a decision-making unit analogous to the individual decision-maker of micro-economic theory. The decision on which he focuses is local community investment in schooling. He is concerned with how migration will affect benefits accruing to the community from investments in education, and hence with how it affects the cost-benefit balances that would determine rational investment decisions. He argues that rational community behavior in this context will lead to under-investment in education on a national scale because of "spill-over" effects -- "external benefits" of a community's investments that accrue to other communities. This treatment of the community as a decision unit has been attacked from several sides. However, the empirical part of Weisbrod's work is relevant to aggregative analysis of human capital gains and losses, flights of the imagination quite aside.

A careful study of Weisbrod's method for computing "spill-overs" points up the problems involved in assigning values to human capital flows. In working from incremental income[5] streams associated with varying levels of educational attainment through high school graduation, he assumes, for instance, that the relevant income streams for valuing both in and out migrants of given age, race, and sex are the same. All those who migrate to and from the Clayton, Missouri community are valued at non-South rates, regardless of their origin or the location of their previous education. This is an empirical compromise dictated by data limitations, and is discussed by the author at some length.

C. Cost valuations of internal migration

Grubel and Scott[6] focus on international migration of "human capital," and on arguments concerning "brain drains." Although they discuss individual decision-making in a framework similar to Sjaastad's, in their published work they emphasize the effects of out-migration on social welfare. There is no community decision unit, such as Weisbrod's; in fact, they discard aggregates and GNP in their theoretical presentations. Social welfare they define as the per capita incomes of all initial residents, whatever their place of residence after migration.

However, appealing as the welfare measure they suggest may be, it is no accident that in the end Grubel and Scott make no attempt at direct empirical assessments in such a framework; the practical difficulties are manifold. Instead, they go to quite the other expreme of empirical pragmatism. Not only do they forsake their welfare concept for an aggregative type of social assessment, but they also by-pass human capital measurement in present value terms, choosing to use cost assessments instead.[7]

National gains are measured by the cost-savings realized in acquiring human capital formed elsewhere without paying for its formation. National losses are incurred when a nation pays for the formation of human capital that others then acquire free.

Using figures from an annual census of foreign students in the United States, interesting estimates are derived for United States gains and losses. The value of the gain (saving in costs) to the United States of acquiring "non-returning foreign students" is taken as what it would have cost to "produce" an American equivalent. Against these cost-savings are set the costs incurred by the United States in providing education for all foreign students. Similar human capital estimates are made for scientists and engineers who have migrated to the United States.[8]

Measurement of human capital in cost terms as a way of assessing the resources that have gone into the making of a man is one thing. Using cost valuations in assessing gains and losses through migration is quite another. The most serious distortions are apt to occur when reproduction cost estimates exceed present value. Thus, a fallacious cost view of human capital can explain many of the complaints that have been so common concerning Southern or Appalachian losses through migration.[9]

D. Present value estimates of inter-regional migration

Rashi Fein,[10] like Weisbrod, was concerned with the measurement of aggregate human capital gains and losses of spatially defined units -- in Fein's case, regions within the United States. Like Weisbrod, Fein used a present-value measure of human capital (discounting future income streams at 5%). There is no decision theory in Fein's analysis, however, and he draws no inferences for public policy.

Fein's first task, a considerable one, was to lay out the census data in a form that would permit examination of gross 1955-1960 migration streams for males. Male migrant flows are disaggregated by age, race, and educational attainment. Focusing on the South and its sub-regions, he then computes net in or out flows for each age-race-education category. To get estimates of social net capital gains or losses associated with migration he first multiplies net flows by average discounted values of Southern income streams associated with each age-race-education category. These sub-aggregates are then summed to give the overall net gain or loss to the region. Present values have, in effect, become weightings for the various statistical categories. Although this procedure preserves the disaggregation by sex, age, race, and schooling, the distinction between characteristics of inflows and outflows is unfortunately lost in the procedure for valuing human capital. (We will come back to this later.)

II. ALTERNATIVE EARNINGS STREAMS AND ONE-WAY MIGRATION

From the above discussions of recent work, it is evident that analysis of migration in human investment terms has opened up over a wide front. There are at once common elements and sharp distinctions in problem foci, in theoretical frameworks applied, in the kinds of compromises made in using incomplete and sometimes inappropriate sets of data. Back of all this are some fundamental theoretical problems in decision theory on the one hand, some critical issues for public policy on the other. Theory and policy are in fact quite closely related. It is with such basic questions in mind that we ask: which earnings streams are the relevant ones in either private or social assessments of migration? Whichever the viewpoint, identification of the conceptually appropriate future earnings streams is a necessary first step. We start in this section with the simplest case -- one-way migration. This in fact is the only case for which earnings streams have been examined in anything we have seen.

A. The net flow fallacy and estimates of human-capital gains and losses

In estimating human-capital gains and losses in the South, Fein took a short cut. As noted earlier, instead of valuing the gross inflows and outflows he took the net flows for each age-race-schooling category of males, computing the aggregate present values of these net flows on the basis of Southern life-income patterns. To the extent that he was dealing with flows among regions within the South he had no alternative; the needed income data are not available for sub-regions. Ideally we would want such figures, and we would want to maintain clear distinctions among parts of the South, for it is a very heterogeneous region. However, when it comes to flows between Southern and other regions the situation is changed; we have data on life-income streams by race, sex, age, and schooling that can differentiate at least on this gross regional dichotomy.

Weisbrod also used a single set of values in his study of gains and losses through migration into and out of the Clayton school district, in Missouri. Our chief criticism in his case is that he did not check for the sensitivity of his results to this decision. Since he was only incidentally interested in his quantitative results, and was not posing the more substantive and analytical questions concerning migration in which we are most interested, this omission is understandable, even if regretable.

We assume, with Sjaastad, that an individual will not normally migrate unless his potential discounted earnings stream in the new area is going to be at least as high as that at the area of origin. When the average stream for the area of destination is obviously lower than that in the area of origin, but people move anyway, this would suggest that the rational in-migrant is not typical of the area into which he is moving. When this happens even within particular race-sex-age-schooling categories, the fallacy of using average incomes at the lower income destination for migrants to that area is underlined. Until data that distinguish migrants are available there can be no firm answer as to what is happening, and how. At present we can only hypothesize, meanwhile making some crude estimates as to what alternative hypotheses may imply.

The potential importance of taking market differences into account can be illustrated by making some simple calculations. The distribution of schooling among 1955-60 out-migrants from the South as a whole is very close to the distribution of schooling among 1955-60 migrants into the South. However, except at college levels, incomes of white males with the same years of schooling are higher in the North than in the South. After a bit of rough pencil-and-paper work we picked $40,000 as an average present-value figure for income streams of white male Southerners leaving the South and $50,000 for white male migrants from other regions into the South.

Using Fein's method of valuing all migrants at Southern values, we have:

	Number Migrating[11]	Present Value (per migrant)	Total Value (millions)
In	579,100	$40,000	$23,164
Out	554,900	$40,000	$22,196
Net			$968

This, we argue, is a minimum estimate of the South's gain. Using the $50,000 figure for in-migrants -- that is, valuing them at their origin rather than their destination, we have:

	Number Migrating	Present Value (per migrant)	Total Value (millions)
In	579,100	$50,000	$28,955
Out	554,900	$40,000	$22,196
Net			$6,759

The net gain by this computation is almost seven times as great as that above and represents a substantial absolute increase, about 5.8 billion dollars. Actually the $40,000 and $50,000 with which we started are low estimates; raising them would raise the absolute discrepancy.

Using a more conservative ratio between non-South and South values will obviously decrease the differential, but even if we cut it drastically to 1.10 rather than the 1.25 ratio above, still applying it to a Southern $40,000 figure, we have:

	Number Migrating	Present Value (per migrant)	Total Value (millions)
In	579,100	$44,000	$25,480
Out	554,900	$40,000	$22,196
Net			$3,284

On this conservative basis, the estimate still increases by a 3.4 multiple and the absolute increase is well over 2.2 billion dollars. However, we have not considered what proportion of the in-migrants may have been unsuccessful returning Southerners, or what proportion of the out-migrants were originally from the North. Our estimates are still gross oversimplifications. Until data that distinguish migrants and their associated income streams are available, there can be no firm answer.

Typically applications of human investment models to migration have suffered from insufficient disaggregation of the population of migrants, though industrious scholars in and outside of the Census Bureau are beginning to remedy the most serious gaps.[12] It is obvious by now that within-each-region disaggregation by age, sex, race, educational attainment, and income (preferably earnings) is needed. Both Weisbrod and Fein recognize this and adjust their human capital estimates accordingly. Actually for the 5% sample of the 1960 census we have within South and within North cross tabulations of all these variables plus 1960 occupations. However, there is no breakdown that would separate out migrants within each of these cells.

A possibility we would very much like to see followed up in 1970 would be refinement of the regional tables on incomes by sex, race, age, and schooling of 1960 residents in order to distinguish the new migrants (1955-60) from others, and each of these groups in turn by place of birth and occupation - ideally both at origin and destination. Obviously this will be possible only if all these items are obtained on the 25% sample.

Suppose we had all of this information. What else should we know to distinguish migrants and their earnings in a meaningful way? It is obvious that many other factors affect productivity and earnings streams of any given population of migrants. Ability, attitudes toward work, quality of schooling, environmental experiences as a youth, and job experience differentials will all be reflected in differential earnings streams. Important also are labor market imperfections that may partially segregate migrant from native populations of the same race, sex, age, and reported years of schooling.

Concerning ability and attitudes we will be brief. We can reasonably assume that the ability distribution for a given age, race, sex, education category of the population will not differ much among regions; hence we need not be concerned on this score with places of origin of migrants. However, in these pages we go further, to make the more vulnerable assumption that the average ability of migrants is no different from that of non-migrants once disaggregation has established homogeneity of population sub-groups in other critical respects. The saving feature of this assumption is the critical disaggregation presupposed. Differences in attitudes toward work are a problem that we will simply have to ignore for the present. It should be noted, however, that both ability and attitudes can be objectively assessed, and for special migrant groups this has been done.

Experience and quality of schooling differentials are at the heart of our analysis, and command more careful attention.

B. Schooling and experience

School-quality differences can be critical in the economics of migration from many points of view, one, but only one, of which is estimation of regional gains or losses in human capital through migration. If the distributions of quality of schooling among men of the same race and age with the same reported schooling were the same in one region as in another (or in urban as in rural areas,) we could forget about this problem. However, despite large within-area differences in quality of schools there are also large and significant differences between rural and urban schools and between schools in one region as against another.

To get at effects of differences in average quality of schooling we need to know where the migrant was schooled (ideally for secondary school by sub-region or state and by type of community). Present census data do not permit disaggregation by locus of schooling even on a broad regional basis. However, one could make some assumptions about the location of both schooling and previous job experience by looking at place of birth, place of 1955 residence, and 1960 residence in relation to age at the time of migration.

The first component of experience, a very important one, is the learning that goes on outside of school during school years. There is ample evidence to show that youth from rural communities start at a disadvantage when they come to the city; the extent of the disadvantage varies with the nature of the rural area. This is one among many reasons why data concerning location at the time when a young man attended secondary school (or if he stopped short of that, where he last attended school.[13] In fact we strongly urge inclusion in the 1970 census of information concerning the state and type of community in which a man resided at the time when he last attended school below the college level (or alternatively, at, say, age 16).

The importance of what men learn on the job has been reappraised recently and given new theoretical respectability in the economics journals. If we accept Mincer's estimates, over a life-time investments in learning on the job typically exceed investments in schooling in the United States today.[14] In any case, whatever the relative magnitudes, learning through on-the-job training and experience is a major component in the formation of human competencies. Given this fact, we must ask what and how much of such learning has been built into which groups of migrants.

Taking it for granted that we will not be working with full life histories -- such data would swamp us in any case -- what might we hypothesize as the best statistical clues to sort out categories of migrants (and non-migrants) by experience? Obviously age is part of the picture. So is schooling if we look at the national scene, since over-all there is a positive correlation between schooling and on-the-job training and learning. However, this is much too crude an approximation. How do people of the same age and years of schooling differ in competencies acquired at work? As a first step we would look for two kinds of information: what a man was doing (his occupation) prior to migration, and where he was doing it.

The 1960 census does not provide information on prior occupations, and collection of data on past occupations is both difficult and expensive; however, this possibility has been considered for the 1970 census. If collection of such data proves feasible for a large enough sample, it will open up a range of possibilities for hypothesis testing that must challenge many researches on both migration and the economics of education.

There can be no doubt that work opportunities and with them opportunities for on-the-job training and learning vary substantially from one place to another. This is glaringly obvious if we look across nations on a world scale. It is sufficiently evident within the United States, and even if we control for age, sex, and prior schooling. There is a strong presumption that knowledge of the location in which men have acquired their work experience will improve statistical predictions of their competencies. How far a migrant's previous experiential learning may be transferrable to his new setting is another matter. Undoubtedly there is selectivity in such transferrability; he can move into the new environment carrying his experience with him only to the extent that the new environment gives scope for its use. This may contribute to differentiation of labor markets between natives and in-migrants, especially in the middle and higher occupational brackets. It is a reasonable generalization that unless a man can take enough of his acquired competencies with him to ensure earnings at least as high as those he would receive at home, our "average" migrant will not move -- at least not if he already has a substantial investment in such competencies and there is a demand for them at home. For youth who have little or no such investment the problem does not arise. Rather, the question may be, where can I go to get the best learning and long-term income opportunities?

We may systematize these comments concerning effects on potential earnings associated with location of schooling and experience by setting up a rough typology of schooling-work combinations as between two regions, A and B. Within any age, sex, race, educational, and, ideally, prior occupational grouping of migrants who moved from place A to place B, we would distinguish the following (still deferring consideration of temporary migration and return):

1. Those born in A, but who migrated to B for schooling and work.

2. Those born in A and schooled in A, but who migrated to B as soon as their formal schooling was completed and who have worked only in B.

3. Those born in A and schooled in A who remained to work for some time in A, but moved to B before they were 40, continuing to work in B thereafter.

4. Those born in A and schooled in A who worked in A to at least the age of 40, migrating to B after that age.

1. The first category of individuals might be expected to have an age-earnings profile very close to or even the same as comparable life-time B residents of the same race, sex, age, years of schooling -- and perhaps prior occupation -- because all schooling and work experience is in B.

2. Those in the second category were schooled in A but their entire work experience is in B. We would expect their average earnings profiles to fall somewhere between those of life-time residents of A and of B, but closer to the latter. Deviations from average B earnings streams would be greater the greater the regional differences in average quality of schooling.[15] It will be greater the greater the differences between out-of-school environment of the migrant's adolescent years (rural or urban, and where) and the environment into which he is moving. Related to this, though partially independent of it, is also the extent to which migrants from A enter labor markets in B that are distinct from those in which lifetime residents of B sell their services.

3. Migrants in category 3, should also, on the average, have future income streams falling somewhere between those of lifetime residents of A and of B. The older they are at migration and the more experience they bring with them, the closer we should expect them to be to the lifetime residents of their area of origin. Here (as also in case 2), if migrants to B work in branch organizations which have head offices in and are operated by individuals from A, their earnings streams should approximate A-type earnings. This is merely an extreme of differentiated labor markets -- islands of A located within the geographic boundaries of B. On the other hand, the younger the migrant the larger the part of his experience he will accumulate in B (segregation of labor markets aside), and the more clearly will his future income stream approach the B pattern.

4. If the movement to region B comes after age 40, it is probable that earnings in B will be similar to those in A where schooling and work experience were obtained. Non-economic locational preferences for B aside, the A income potential should in fact give us a minimum estimate of earning streams in B. We assume that a rational person established in a career in A will not move after age 40 unless he can earn at least as much in his new location virtually from the start. Although there are exceptions, there is sufficient empirical evidence to show that after age 40 most men are recouping investments in themselves rather than making new ones, intuitive reasoning quite aside.

A fifth category which is really a special case of the fourth has quite different implications, however, Suppose an individual born in A, schooled and with his work experience in A, whose skill has become obsolete after age 40 (or somewhat earlier). In this case the migrating individual has suffered a severe "human capital loss." He has little or no learned competence to apply in either A or B, and his expected future earnings stream cannot be measured by an average in either area. Though a special case, this is an important one, which merits the special study that is beginning to be given to it.

C. Package migration and the transfer of experience

In discussing the locus of schooling and experience as related to likely future income streams of migrants we commented upon the possibilities -- in some cases the strong likelihood -- that there would be some differentiation of markets in which migrants and non-migrants resident in a given area sell their services. We specified further that this differentiation could and does occur even among men of the same race, age, and years of schooling. Even if we went all the way from regions down to state economic areas in an attempt to assure greater homogeneity we would not necessarily get rid of such labor market differentiation.

Taking the more dramatic in-migrant labor market segments to illustrate the extreme case, there are "Northern firms" in the deep South, firms from the United States in Sao Paulo, French firms in the Ivory coast. Traditionally such "foreign" firms have preferred to import talent from their area of origin rather than to employ locals who lack not only the technical know-how acquired on a job but also, equally important, know-how with respect to how a modern Northern organization operates.[16] (In addition the Northern highschool graduate is likely to be better schooled than the Southern.) Many of the men who have been brought into the South when a Northern firm established a branch there would not otherwise have migrated; there would have been no opportunities for them to work in the South at jobs in which they could use their Northern skills and experience. Transfer of those skills would not have been possible. Conversely, the firm would not have moved into the South had it not been possible to bring Northerners with know-how along. This is what we mean by "package migration."[17]

To the extent that migration is associated with these segmentalized labor market structures, differences in earnings stream between migrants and permanent residents should appear. However, this assumes further that there are not parallel opportunities of a very different kind to which the local population has the readier access. This is more likely to happen at the top than in the middle of the income and status scale. Thus a traditional elite, experienced in its own culture and perhaps protected by other barriers to entry[18] may well exist along with a "foreign" class of high-salary professionals and technicians. How far this pattern characterizes the South is not entirely clear, though the statistics on education, occupation, and income by region do suggest that it exists in some degree and is not uniform. Such a pattern is unstable, however. It fades away with development. Ultimately, we suspect, it is eroded more by the progress of native populations at the middle level than by any direct impact at the top.

As the number of "Northern" firms in the "South" multiplies, and as a few and then a few more Southerners filter into and through these enterprises at the middle levels, the transfer of skill and experience from North to South becomes easier. Migrants now come in greater numbers without any special prior association with the enterprises in which they find jobs. The process of diffusion of development is well under way.

How much validity there may be in this theorizing concerning relations between migration and the process of diffusion of development remains to be seen. Existing evidence is spotty. But there is clearly a challenge to test hypotheses of this kind. To what extent are Southern labor markets differentiated or even sharply segmentalized? For any given race, age, and schooling, how do earnings of Southerners working in Southern-type firms compare with those in Northern-type firms in the South? How do earnings of similar individuals in Northern firms in the North compare with those in Northern firms in the South? Looking at the shapes of income streams and applying Mincer's method of analyzing on-the-job learning, do we find that Northern profiles transfer to the South intact? If there are differences between Northern firms in the South and Southern firms with respect to training and learning opportunities, who is reaping the benefits -- migrants from the North (temporary or permanent) or locals? Over time are profiles in Southern firms changing -- and can we observe this in cross-section by comparing the border South with the deep South? Turning to data on occupations, types of firms quite aside, are occupation and occupation-earnings patterns quite different among migrants into the South and native Southerners of comparable race-sex-age-schooling categories? Are these differences persistent, and repeated over large parts of the South, or are they changing?

Many of these questions concerning segmentalization of labor markets, transferrability of experience, the relative role of experience and schooling in shaping productivity, and the processes of diffusing skills might be tested by using human capital models and U.S. census data.[19]

III. MIGRATION AND RE-MIGRATION

The common tendency to treat migration as though it were a once-and-for-all affair, often combined with a definition of migration that in itself suggests permanency, has many unfortunate results.

One of the more obvious of these may be faulty assessments of "brain drains" and gains (the latter rarely noted). It is evident that once remigration is considered new difficulties arise in such assessments. Are the migrants college students returning home, prodigal sons, disappointed job-seekers, retrained workers, retiring elders, or former political "outs" who are now "in"?

Further, how does the human capital value of these re-migrants differ from what it was when they first migrated? This question points to a more basic weakness of one-way migration simplifications; they disregard the duration of migration. There is a continuum from brief periods away at school or on a short-term job assignment to permanent residence at destination. To disregard temporary migration is to disregard the important linkages among private investments, on-the-job learning, and migration which, when coupled with regional differentials in quality and availability of schooling and experiential learning, can provide extremely important channels for the transfer and local acquisition of know-how. A developing country seeking to build up its cohorts of qualified people is faced with important choices as to how this can best be done. How many and which sorts of individuals should be sent for study and/or for work experience "abroad"? How many will return? What about the importing of outside experts on a temporary basis? How can these choices be evaluated from the points of view of the receiving and of the sending nations? What about problems of recruiting such men, of the strategies of inducing higher rates of return among those subsidized for study or training abroad, or in another region of the United States? Human investment decision models that incorporate re-migration sequences can aid in such evaluations. We will develop this theme in section III. First, however, we look briefly into the quantitative importance of re-migration as return to the sub-regions (Divisions) of birth within the United States and a few distinctive patterns in these respects.

A. The quantitative importance of re-migration in the United States

Although U. S. census publications do not permit breakdowns by earnings, sex, age, schooling, occupations for migrants -- let alone re-migrants -- they do permit a partial assessment of the extent of re-migration. This is made possible by the tabulations of 1960 division of residence against residence in 1955 and according to whether division of birth was or was not the same as that of residence in 1960. We selected three divisions as illustrations and computed for selected age and education groups the proportions of the white male 1955-60 in-migrants who were coming "back home" and of out-migrants who were returning to the division in which they were born. The results are shown in Tables 1 and 2.

The most striking feature of Table 1 is the generally high rates of return to division of origin. The only exception is the migrants into the Pacific Division, only a small proportion of whom were returning to the part of the country in which they were born. This is, of course, to be expected. Almost equally striking in the Pacific ratios is the extremely high proportion of the younger out-migrants who were returning to their home divisions and virtually regardless of educational attainment levels--though the proportion is less extreme for college graduates. As Table 2 shows, the ratio of in to out-migrants among young college men is exceptionally high for the Pacific Division, though in all age and schooling groups that division had net inflows.

At the opposite extreme is the East South Central region, which had net outflows in every case. The differences in the gross in and out movements were small, however. The most interesting thing to us in the East South Central figures is the rising proportion of in-migrants who were returning home as we move down from college graduates to the highschool drop-outs. This pattern is repeated for every age group. We strongly suspect that had we computed ratios below highschool 1-3 those ratios would have been even higher than for men with one to three years in highschool. Here a very distinctive migration and re-migration phenomenon is evident. It should come as no surprise, for this is a manifestation of "the poverty problem" in one of its most serious and difficult forms -- and of the failures of migration to solve problems of Galbraith's "insular poverty" where men have become obsolete or have never been sufficiently equipped to enter into jobs that offer opportunities for new training or learning.

The least predictable of the results, at least to us, were those for youth returning to the East North Central Division. Even though this has not been a growing area, we had expected that lower proportions of in-migrants would be returnees. Even the lowest ratios are around 30 per cent. It is interesting that the highest ratios are for young men, and are associated with absolute net in-flows, while in the older age categories the East North Central Division was a net loser in each of the age-schooling groups we examined. Such findings invite further speculation, but we resist that temptation in favor of a more generalized (and safer) procedure -- the presentation of migration and re-migration decision models with some of their analytical and policy potentials.

Table 1

RETURN MIGRATION RATES AMONG 1955-60 MIGRANTS INTO AND OUT OF THE EAST SOUTH CENTRAL, EAST NORTH CENTRAL, AND PACIFIC DIVISIONS: WHITE MALES IN SELECTED AGE AND SCHOOLING CATEGORIES

Schooling	Proportion of 1955-60 in-migrants who were returning to division of birth, 1960 Age			Proportion of 1955-60 out-migrants who were returning to division of birth, 1960 Age		
	25-29	30-34	35-39	25-29	30-34	35-39
EAST SOUTH CENTRAL						
Highschool 1-3	.59	.56	.53	.24	.18	.17
Highschool 4	.52	.42	.36	.32	.23	.23
College 1-3	.46	.35	.26	.37	.26	.23
College 4	.32	.32	.28	.30	.30	.27
EAST NORTH CENTRAL						
Highschool 1-3	.43	.30	.27	.29	.29	.26
Highschool 4	.56	.36	.30	.25	.21	.19
College 1-3	.53	.34	.30	.22	.18	.18
College 4	.40	.34	.30	.18	.21	.20
PACIFIC						
Highschool 1-3	.09	.06	.05	.60	.46	.47
Highschool 4	.10	.07	.07	.61	.42	.36
College 1-3	.15	.09	.09	.60	.38	.30
College 4	.11	.12	.09	.46	.37	.30

Source: Computed from data in the 1960 U.S. Census of Population, Lifetime and Recent Migration, PC (2) 2D, Table 8. Ratios are computed excluding persons for whom residence was not reported.

Table 2

NUMBER OF 1955-60 WHITE MALE IN-MIGRANTS AND OUT-MIGRANTS FOR WHOM RESIDENCE WAS REPORTED: THE EAST SOUTH CENTRAL, EAST NORTH CENTRAL, AND PACIFIC DIVISIONS

	1955-1960 In-migrants, 1960 Age			1955-1960 Out-migrants, 1960 Age		
	25-29	30-34	35-39	25-29	30-34	35-39
EAST SOUTH CENTRAL						
Highschool 1-3	7300	5750	4280	9585	6955	5145
Highschool 4	13367	7960	6716	16279	9296	7690
College 1-3	6781	4067	3441	8064	4573	3679
College 4	9752	7485	5372	13743	9577	6303
EAST NORTH CENTRAL						
Highschool 1-3	18305	10884	7455	15488	15558	13185
Highschool 4	36831	14659	12030	25922	21715	20856
College 1-3	17920	8615	6752	13918	11793	10456
College 4	33906	24707	15326	31472	28233	20950
PACIFIC						
Highschool 1-3	22146	20800	16556	14764	10011	7453
Highschool 4	36646	27640	26411	31305	14452	13301
College 1-3	22645	14578	12942	16439	7592	7432
College 4	35438	27629	19899	16825	13764	9626

Source: Same as Table 1

B. Decision models for migration and re-migration

The interpretation of migration behavior in economic terms and the analysis of potential strategies in social policy require something more than aggregative estimates of gains and losses.

With this in mind, we have developed models that begin with the individual viewpoint but are transformed into social decision models as parameter values are readjusted to allow for cost and income transfers, as individually expected earnings are replaced by socially expected or realized productive contributions, and as probability values are applied to allow for rates of return and non-return of migrants. The models allow choices with respect to locus, duration, and sequences of schooling and experiential learning. Relevant earnings streams are those of an "average individual" within a sex, race, age, and initial educational attainment and occupation category. Our presentation is necessarily summary.[20]

The following notation is used:

- a- age at the first decision point relating to migration
- b- date of actual or intended out-migration
- m- age at return from residence abroad
- n- retirement age
- R_t- expected earnings in the year t at the place of origin prior to (or in the absence of) any migration
- D_t- expected earnings abroad in the year t
- Y_t- expected earnings in the year t at the place of origin for migrant returnees ($t > m - a$)
- C_t- direct cost in the year t of schooling or training in the area of origin
- K_t- direct cost in the year t of schooling or training abroad
- Z_t^o- direct cost of out-migration incurred in year t
- Z_t^h- direct cost of return migration incurred in year t
- r- discount rate
- V- present value of future income streams
- j- superscript denoting a particular race, age, sex, school attainment, and occupation (if any) at time a

We limit ourselves to the following possible migration sequences:

1. Remain in the area of origin permanently (no migration)

$$V_1^j = \sum_{t=a}^{n} \frac{R_t - C_t}{(1+r)^{t-a}}$$

2. Migrate immediately and remain permanently in the area of destination

$$V_2^j = \sum_{t=a}^{n} \frac{D_t - K_t - Z_t^o}{(1+r)^{t-a}}$$

3. Migrate immediately, remain temporarily in the area of destination, then return to the area of origin

$$V_3^j = \sum_{t=a}^{m-1} \frac{D_t - K_t - Z_t^o}{(1+r)^{t-a}} + \sum_{t=m}^{n} \frac{Y_t - Z_t^h}{(1+r)^{t-a}}$$

4. Remain temporarily in the area of origin, then migrate permanently

$$V_4^j = \sum_{t=a}^{b-1} \frac{R_t - C_t}{(1+r)^{t-a}} + \sum_{t=b}^{n} \frac{D_t - K_t - Z_t^o}{(1+r)^{t-a}}$$

5. Remain temporarily in the area of origin, migrate and remain temporarily in the area of destination, then return to the area of origin

$$V_5^j = \sum_{t=a}^{b-1} \frac{R_t - C_t}{(1+r)^{t-a}} + \sum_{t=b}^{m-1} \frac{D_t - K_t - Z_t^o}{(1+r)^{t-a}} + \sum_{t=m}^{n} \frac{Y_t - Z_t^h}{(1+r)^{t-a}}$$

1. Individual Choices

An individual has many alternatives within each of the five sequences. To illustrate the complexity of choice, we begin with sequence one, the simplest case. Think of a high school graduate deciding about his future. He must weigh going to work immediately against spending one, two, five, or more years in school during which his earnings are low or zero. He must weigh schooling now against schooling later If he is in the United States, he is faced with a bewildering array of institutional settings. Each of the above carries implications for cost streams (C_t) and for future earnings (R_t).

Each of the schooling choices can be associated with a set of alternative occupational and job choices with varied amounts of experiential learning. The potential occupation and/or job choice implies visualizing a series of future yearly earnings (R values). These separate streams may differ in their present value and in the amount of associated experiential learning.

A person may choose between positions with higher learning potential which pay little at first but will bring higher pay later and positions with less or negligible potential which pay well at first but promise little or no earnings increase in the future.[21] The relation of experiential learning to earnings stream differences and the place of experiential learning in our models is less observable than that of schooling. Schooling obviously involves foregone earnings: most students do not work or they have earnings far below what they might have if employed full-time. And schooling also carries direct costs -- tuition, books, fees -- that are only too evident to all who have paid them at one time or another. However, costs to an individual of experiential learning, embedded as they are in the earnings streams and usually involving less contrast in earnings, are less conspicuous even when they are cumulatively substantial.[22]

From the above it is apparent that even within sequence one, there are potentially almost an infinity of schooling and experience mixes from which to choose, each with its earnings stream and associated costs. Within sequence two, the possibilities for mixint schooling and experience are increased as a geographical dimension is introduced; within sequence three, the number of possibilities is further multiplied as the geographical progression becomes more complicated; and so forth.

In theory, one could identify the costs and earnings streams for every conceivable combination within and between each of the five migration sequences and select that one giving the highest present value. However, many combinations can be immediately discarded as unpromising ones. Furthermore, in practice any one individual will be constrained by his native ability, his interests, his financial situation, behavior of acquaintances, etc. The possibilities actually considered are reduced to the most promising and feasible ones, even to choices among two or three specific combinations for which costs and benefits are weighed and compared.

The most interesting aspect of this exercise for analysis of migration concerns how the locale of schooling and/or experiential learning up to any given age affects subsequent income streams for each work location which might be chosen after that age.

How great are the differences in particular cases? How stable are they? What explains them? How fully do they, in turn, explain migrant behavior? By what processes and how rapidly are they changed in the wake of migration? These questions are key ones for a positive economics that would go beyond traditional resource allocation to merge decision theory into a theory of development dynamics.

2. Social Choices and Migration Probabilities

Now, suppose we shift our viewpoint to a social one, for instance that of a regional or national body which is deciding whether or not to subsidize study outside the area. Social gains and losses may be evaluated by using the same basic models used for individual decisions. However, the costs and returns are now those to the society. For example, all costs of educational services (teacher time, physical facilities, etc.) are real social costs, even though subsidies may reduce or eliminate such costs to the individual.[23]

Permanent migration, as in sequences 2 and 4, is commonly regarded as unambiguous loss from a social viewpoint. This position, which assumes that to add to the productivity of an area an individual must be physically present, has been challenged.[24] However, in most cases policy-makers are concerned with the problem of non-return and regard physical presence as a crucial consideration from the national or regional point of view.

Temporary migration, as in sequences 1, 3, and 5, when viewed socially as a training alternative, requires a major adjustment.[25] Unless a government coerces in some way, it cannot insure that those trained at home will stay at home or that those trained abroad will necessarily return. Therefore it is necessary to allow for the possibility that students trained at home will emigrate and that students trained outside will not return. This is done by including probabilities in the models.

As an example, we focus on comparisons between longer and shorter periods of training abroad. Let α_t^m be the probability that students trained outside the area for (m - a) years will return before age $\underline{t}$. To illustrate, sequence three then becomes:

$$V_{3m}^{j} = \sum_{t=m}^{n} \frac{\alpha_t^m Y_t^m - (1-\alpha_t^m) Z_t^h}{(1+r)^{t-a}} - \sum_{t=a}^{m-1} \frac{K_t + Z_t^o}{(1+r)^{t-a}}$$

For each value of (m-a) there will be a different set of income variables, Y_t^m, and of probabilities of return, α_t^m. The probability set α_t^m is likely to be a declining function of (m-a); the rate at which α_t^m declines will be one of the critical elements in the comparison of social net returns from longer or shorter periods of training abroad. In this form the equation refers to one individual (or a fraction thereof) but with appropriate identification of Y, K, and Z it is also a marginal social benefit-cost summation.

Another type of comparison involves locus of training for any given training period as, for instance, in a comparison of sequences 1 and 3. A concrete example is the argument over establishing local medical schools (versus training doctors abroad or out-of-state). The answer will depend on rates of migration and retention as well as on the relative cost and quality of the training in each locality. The argument usually given is that probabilities of retention for those trained at home so exceed the probability of return for those trained elsewhere as to outweigh all other considerations. From a "nationalistic" or "localistic" point of view, the economic validity of this argument will depend, among other things, upon who bears the costs of training in the alternative locales.

Another practically important issue is whether and how long students sent overseas should be allowed or encouraged to stay beyond formal schooling in order to acquire experience. The solution will depend, in part, on what "wastage" from non-return is associated with the experiential learning. This "wastage" is incorporated into our social decision model by introducing the migration probabilities.

The dependence of social policy upon understanding individual behavior is particularly evident in attempts to decrease wastage by influencing rates of out-migration and remigration. Also, knowledge of how migration probabilities vary with individual characteristics within each schooling-work alternative would provide guidelines for efficient selection of individuals to the training programs, further reducing wastage.

3. Long Versus Short-term Importations of Highly Qualified Manpower

As a final example of human capital models applied to migration, we turn to decisions in developing nations concerning imported skills. We by-pass comparison of more training of local people against more importing of outside talent to focus on models weighing three alternatives in the purchase of foreign experts' services: a one-man, long-term contract, a sequence of short-term contracts to several people, and a system of two-year rotations between two individuals.

When decisions involve a continuous series of replacements, the succession of short income streams can be regarded as one long one. The equivalence is not complete, however. The sharpest contrasts and associated problems can be most clearly illustrated by looking at human capital migration from the point of view of the developing nations and their needs for highly qualified manpower from the more advanced economies.

There is much discussion today of alternatives and combinations in the flow of high-level manpower for shorter or longer stays in the developing nations. In the terminology of the previous section, should a man be kept abroad for a period $(m-a) = 2$ or for, say, $(m-a) = 5$, or 10, or even 20 years? For that matter, what about double appointments in which two individuals alternate with each other at home and abroad?

Let us designate the present value to the importing country, X, of a single individual for 10 years as V_x^1, that of two individuals who replace each other every two years over a period of 10 years as V_x^2, and that of a sequence of five individuals staying two years each as V_x^5. In the third case, on our assumption that each successive individual is the twin of his predecessor D_t for $t = 1$ will have the same current year value as D_t where $t = 3, 5, 7,$ or 9 and D_t where $t = 2$ will have the same value as where $t = 4, 6, 8,$ or 10. We make the simplifying assumption that the importing country pays for all travel expenses. The maximum present values and hence amounts that the importing country would be justified in paying for the production streams generated by each of these alternatives can be represented as follows:

$$V_x^1 = \sum_{t=a}^{a+10} \frac{D_t^1}{(1+r)^{t-a}} + \sum_{t=a}^{a+10} \frac{Z_t^0}{(1+r)^{t-a}} + \sum_{t=a}^{a+10} \frac{Z_t^h}{(1+r)^{t-a}}$$

$$V_x^2 = \sum_{t=a}^{a+10} \frac{D_t^2}{(1+r)^{t-a}} + \sum_{t=a}^{a+10} \frac{Z_t^0}{(1+r)^{t-a}} + \sum_{t=a}^{a+10} \frac{Z_t^h}{(1+r)^{t-a}}$$

$$V_x^5 = \left(\frac{D_1^5}{(1+r)} + \frac{D_2^5}{(1+r)^2}\right) + \left(\frac{D_1^5}{(1+r)^3} + \frac{D_2^5}{(1+r)^4}\right) \ldots + \left(\frac{D_1^5}{(1+r)^9} + \frac{D_2^5}{(1+r)^{10}}\right) + \sum_{t=a}^{a+10} \frac{Z_t^0}{(1+r)^{t-a}} + \sum_{t=a}^{a+10} \frac{Z_t^h}{(1+r)^{t-a}}$$

We can simplify in comparing these alternatives by assuming that the amounts and timing of travel costs paid by the receiving country are the same in all three cases. This is evidently the case as between V_x^2 and V_x^5, and it is consistent enough with common practice respecting travel allowances for vacation at home in the case of expatriates on long-term appointments. Which alternative will yield the highest present value then depends upon the summations or sets of terms incorporating the D's.

If there is any learning on the job at all, the V_x^5 stream will clearly have the lowest present value. In fact, if we had made any allowance for direct outlays on training, this disadvantage of the V_x^5 stream would be still more apparent. There is good reason for the widespread concern over the cutting off of so many technical assistance activities at two years per man, and even stronger reason for the increasingly firm attempts to adhere to a two-year minimum, to permit men to learn about the situation in a strange environment and to attain a reasonably high level of effectiveness in it. Two year appointments may serve very well when there is a special job that needs to be done by a man with unusual qualifications, after which the requirements of the task are less demanding--in other words, when the calibre of the first man has to be higher than that of his successors. But that is a different sort of situation, and not the "successive twin" case with which we started and by which we defined V_x^5.

Comparison between V_x^1 and V_x^2 must be a bit more subtle. In case V_x^1 there is the advantage of continuity on the job in country X, but offsetting this is loss of contact with dynamic centers of activity in the expert's home country; he tends to fall progressively further behind his colleagues there. Case V_x^2 has the advantage that experiences in the home country and in the importing country may well feed into each other, to enhance a man's effectiveness in both. This is one of the arguments in favor of developing career opportunities in technical assistance by establishing supernumerary university posts (double staffing) in selected fields and locations.

If V_x^5 is so likely to be the inferior choice for value to the importing country, why is this alternative so common in practice? Evidently there are two reasons. First is the political reaction to colonial experience and the desire to avoid entrenchment of foreigners who might build up too much power in the country. In many cases ex-colonial countries have quite deliberately made a trade-off between economic and political ends by their policies with respect to expatriates. However, this phase of the transition is fading, and with this change the economic decision models may have greater potential impact on policies. In strictly economic terms, what can be said for V_x^5? It seems clear that set against the lower present values of a V_x^5 stream are lower recruitment costs; it is easier to get good men for a short than for a long time, and primarily for two reasons: First, the pace of learning in the first year is likely to be especially high, and what is learned over a two-year period may have more transferrability to the job market in the expert's home country than the learning that cumulates with longer time abroad. He loses little, if anything, in carrying this learning home with him. Second, he has suffered less loss in getting out of touch and losing contacts with colleagues at home when his stay is not too prolonged. (In addition, a reasonably short term abroad may be enticing for quite non-economic considerations that would pall if the stay were extended.) Thus set over against the lower value of the V_x^5 stream is the greater ease and lower cost of maintaining it.

This brings us back to V_x^1 against V_x^2 once again. Let us take another look at V_x^1. Unless a man becomes a permanent migrant, he is likely to suffer a disadvantage when he returns home after a long stay abroad. He has foregone learning opportunities suited to his home country, and the experience he has accumulated in his years abroad often has limited transferability back into an advanced industrial nation. (In lesser degree this may happen with migration from North to South in the United States, unless the migrant is associated with a Northern or a Northern-type firm.) In order to attract a man for 10 years, it would be necessary to pay him a very high salary to compensate for this accumulation of obsolescence. There is evidently a point at which long-term stays abroad must become permanent ones if they are to prove beneficial both to the receiving country and to the individual involved.

V_x^2 is quite another matter. In this case the learning process continues and contacts at home are maintained. No special bribe need be paid to attract a man into such a career. On balance it looks very much as though V_x^2 might come out as the best alternative in a large proportion of cases. Systematic application of human investment decision models to particular cases will help sort these alternatives out. They just might lead to some important innovations in technical assistance and in relations between Universities and the Department of State.

IN CONCLUSION

1. <u>Current methods of calculating human capital gains and losses from migration take a too simplistic view of migration.</u>

a. Even adhering for the moment to the prevalent treatment of migration as if it were a one-way affair, it is of the greatest importance that gross flows be analyzed: the critical problems and evidence concerning the effects of migration are concealed when net flows only are assessed -- even when the latter are broken down into finely disaggregated population categories.

b. Re-migration is important quantitatively. Furthermore, disregard of re-migration leads to serious misinterpretations of even the gross flows and even when the latter are disaggregated on a number of key variables. The importance of analyzing re-migration is underlined where regions differ in quality of schooling and experiential opportunities and where there is rotating migration of obsolescent and undereducated men.

2. Human investment decision models provide useful conceptual and empirical tools when applied to migration, from both individual and social perspectives.

a. The models add insight into motivations of migrants. Understanding migrant behavior provides points of leverage for channeling migration to social purposes, and provides a means of determining to what extent the socially rational may coincide or conflict with effects of individual behavior that are rational.

b. Social decisions involving the locus of training may be put in a cost-benefit framework and evaluated. Major decisions such as whether to train elsewhere or at home can be weighed. Losses from non-return and their probabilities can be valued and included in assessing investment alternatives. The potential gains from policies to reduce rates of non-return could also be estimated.

c. Migration, coupled with regional differentials in quality of schooling and experience, can be examined as it relates to the diffusion of know-how among regions or nations. Of particular interest would be application of human capital concepts to understanding "package migration" as an agent of change.

d. Effects of social decisions to import manpower on a short-term continuous replacement basis, an alternating basis, or a long-term basis can be sorted out using human capital investment models. For each situation, individual decision models provide estimates of outlays necessary to attract the talent desired. Costs and benefits can be compared to determine the best alternative.

3. Availability of census tapes with data for samples large enough to permit refined breakdowns would permit new kinds of research on critical aspects of migration as a human investment. Potential contributions to both theoretical developments in the social sciences and to public policy formation are substantial.

a. Even if no data other than those collected in 1960 were obtained, larger samples would permit multiple breakdowns that distinguish migrant status by origin, destination, and place of birth for Divisions, within existing categories on age-sex-race-income-education-occupation tables.

b. All too sizeable a list of other items might be suggested, but we will use restraint. A high priority item would be occupation just prior to migration, but we recognize that occupation data are costly. Home residence of college students is presumably being included. We would be interested in dates of migration, together with state and type of community in which a migrant last resided, but we would not want to give up the identification of residence and other traits at a fixed time interval (5 years) before the census. Obviously not all these things can be done.

c. Because of the importance we attach to it, we list separately information concerning residence when last attending high school (or, for those who never entered high school, elementary school). Alternatively, the question could be asked for age 16, though this might be more difficult for some to answer. State of residence and type of community would both be desirable; together they should provide valuable indexes of the combined effects of quality of schooling and experiential learning from adolescent environment. Such information could be extremely useful not only in analysis of migration but in many other aspects of the economics of human resource development and utilization.

FOOTNOTES

[1] In the 19th Century, estimates of human capital gains and losses were made for the United States, England, and Germany, using both present values and cost replacement methods. Discussions of the estimates and methods may be found in: Richmond Mayo-Smith, Emigration and Immigration, New York: Scribner's Sons, 1892, Chapter VI, and in Grace Abbott, ed., Historical Aspects of the Immigration Problem, Select Documents, Chicago: The University of Chicago Press, 1926, pp. 370-381.

[2] Larry Sjaastad, "The Costs and Returns of Human Migration," The Journal of Political Economy, LXX, No. 5, Part 2, October, 1962, 80-93.

[3] In his theoretical formulation, Sjaastad also considers "psychic" costs and benefits.

[4] Burton Weisbrod, External Benefits of Public Education, Princeton, N.J.: Industrial Relations Section, Princeton University, 1964.

[5] Census cross-classifications are by income, not earnings. There are substantial difficulties

involved in establishing a "correct" figure for such groups as proprietors. Neither income nor earnings is quite accurate.

[6] See the following three articles of Herbert Grubel and Anthony Scott: (1) "The International Flow of Human Capital, the Brain Drain," American Economic Review, Papers and Proceedings, LVI, No. 2, May, 1966, 268-274, (2) "The Immigration of Scientists and Engineers to the United States," The Journal of Political Economy, LXXIV, No. 4, August, 1966, and (3) "The Characteristics of Foreigners in the U.S. Economics Profession," American Economic Review, forthcoming.

[7] Herbert Grubel, "Nonreturning Foreign Students and the Cost of Student Exchange," International Educational and Cultural Exchange, (a publication of the U.S. Advisory Commission on International Educational and Cultural Affairs), Spring, 1966.

[8] Herbert Grubel and Anthony Scott, "The Immigration of Scientists and Engineers to the United States," loc. cit.

[9] This, by the way, is where Grubel and Scott's social welfare view of effects on per capita incomes could prove especially fruitful--and empirically operational up to useful approximations.

[10] Rashi Fein, "Educational Patterns in Southern Migration," The Southern Economic Journal, XXXII, No. 1, Part 2, July, 1965, 106-124.

[11] The figures for numbers migrating in and out of the South, exclusive of inter-South migration, are taken from William N. Parker's comment on the Rashi Fein article (see footnote 10), in The Southern Economic Journal, XXXII, No. 1, Part 2, July, 1965, p. 126, Table I.

[12] A number of outstanding scholars have approached this topic from demography -- among them our chairman, Henry Shryock.

[13] The critical importance of community characteristics in determining differences among schools in distributions of achievement has been well documented. See, for example, H. T. James, J. Alan Thomas and Harold J. Dyck, Wealth, Expenditure and Decision-Making for Education, in U.S. Department of Health, Education, and Welfare, Cooperative Research Project No. 1241, Stanford (School of Education), 1963, and Charles Benson, et al., "State and Local Fiscal Relationships in Public Education in California," State of California Senate Fact Finding Committee on Revenue and Taxation, Report, Sacramento, March, 1965.

[14] Jacob Mincer, "On-the-Job Training: Costs, Returns, and Some Implications," The Journal of Political Economy, LXX, No. 5, Part 2, October, 1962, 50-79.

At any given time, however, the aggregate on-the-job training embodied in the labor force will be the lower figure, since young people who have completed school have yet to accumulate the learning on-the-job that Mincer measures by opportunity costs.

[15] This makes the further assumption that migrants average the same schooling quality as comparable nonmigrants in their area of origin.

[16] Roy L. Lassiter reports Southern business preference for Northerners with high school education or more in "The Experience of Selected Manufacturing Firms with the Availability, Skills and Training of Manufacturing Workers in Florida," Occasional Paper No. 1, Bureau of Economic and Business Research, College of Business Administration, University of Florida, 1961. The reluctance of U.S. firms abroad to employ local management is illustrated by John Shearer in his High-Level Manpower in Overseas Subsidiaries: Experience in Brazil and Mexico, Princeton, New Jersey: Industrial Relations Section, Princeton University, 1960, especially Chapter VI.

[17] For some related discussions of transfer of know-how, see Mary Jean Bowman, "From Guilds to Infant Training Industries," in C. A. Anderson and Mary Jean Bowman, eds., Education and Economic Development, Chicago: Aldine Publishing Company, 1965, pp. 98-129.

[18] However, the high incomes of the "traditional" or native elite groups may be largely property income (for instance, Southern farmers with college education) or their salaries may be bureaucratic sinecures (as in many developing areas).

[19] In the 1970 census and thereafter, social security numbers may be included, making possible special studies that can be linked into census information for the same individuals. This should allow many kinds of research on migration that have not been feasible previously without incurring prohibitive costs. For example, a special survey might obtain rosters of individuals working in Southern firms and in Northern firms in the South. By matching, these individuals could be located in the census and their age, schooling, and migrant status identified. Such data would permit analysis of the degree of labor market differentiation and of native Southerners and immigrants of comparable age and education employed in establishments with Southern and Northern managers. Many other possibilities, including migration sequences as revealed by social security data, come to mind.

[20] The models used here are part of a family of such models for analysis of investment in human resources (and before that in business economics). Their relation to Gary Becker's approach is evident. (See his "Investment in Human Capital," The Journal of Political Economy, LXX, No. 5, Part 2, October, 1962, and his Human Capital, Columbia University Press, 1964). However, for our purposes, comparison of present values taking an assumed external discount rate proved more flexible and is generally more appropriate than the "internal rate of return" comparison of Becker's.

[21] Note that this description of choices, and our use of present value comparisons, requires no assumption with respect to year-to-year choices of alternatives or constancy of internal rates of return to successive self-investments such as Mincer used (op. cit.). For some remarks on Mincer's as one of a broader set of models that values the length of a contract term, see M. J. Bowman, "The Costing of Human Resource Development" in E. A. G. Robinson and J. E. Vaizey, eds., The Economics of Education (Proceedings of the 1963 Conference of the International Economics Association), London and New York: Macmillan and Co., 1964.

[22] Stress here is on cost to the individual because we are speaking of an individual decision. There is no need to assume year-to-year matching of incomes and productivity even in social assessments, however, since only the present value enters into our social decision models; more than one income sequence can yield the same present value. Becker's "general" learning will cause no trouble. Some part of what he terms "specific" on-the-job training will escape measurement in the estimations of present values of income streams accruing to individuals.

[23] Evidently the use of shadow pricing for public cost and benefit assessments will sometimes be required -- where there are substantial discrepancies between present values of what men will be paid and what they will produce over ensuing years; this is why, earlier, we spoke of transformations from "individual earnings" to productive contributions. The discussion in this section identifies "social" returns with returns to the entire society, as measured by national product. Often foundations may take similar ultimate goals as a basis for decisions to allocate funds to one or another type of educational project in one or another location. Narrower definitions of "social" that balance flows into and out of one versus another public exchequer are a very different matter.

[24] For a listing and discussion of contributions to the area of origin which might be made from outside the area, see Harry Johnson, "Economics of the 'Brain Drain': The Canadian Case," Minerva, III, No. 3, Spring 1965, 299-311, and Herbert Grubel and Anthony Scott, "The International Flow of Human Capital, the Brain Drain," op. cit.

[25] One of the authors -- Myers -- has been developing implications of this approach more fully. Our discussion here is limited by space considerations.

INTERNAL MIGRATION AND ECONOMIC OPPORTUNITY: FRANCE 1872-1911*

Gordon A. Marker, Regional Economic Development Institute

I

In the period between the Franco-Prussian and the 1914-18 wars, population redistribution in France was characterized by the rural-to-urban shift widely observed in modern economies. Urban growth due to net migration ranged from a high of over 1 million to a low of 620 thousand during the eight intercensal quinquennia between 1872 and 1911. The net contribution of rural out-migration varied from a high of 820 thousand to a low of 417 thousand. The balance between net in-migration to urban areas and out-migration from rural areas is explained by foreign sources which in only two periods were negative, indicating net out-migration from France. Since the international contribution was relatively minor, much of the urban growth through net migration may be regarded as internal. [1]

While such internal shifts have become typical of economic development in the past and are presently occurring in the non-industrialized world, they take on additional significance for France, a country of slow population growth. France was exceptional for her low reproduction rate, which turned down before and remained less than those of other industrial economies. Thus, the internal shifts of population may be of relatively greater economic significance in supplying labor than would be experienced in countries of high natural increase. An additional feature of French migration of interest to economists and statisticians is the presence of long swings or Kuznets cycles of fifteen or twenty years duration. These are particularly evident in the series of urban net migration which exhibits peaks in 1876-81 and 1896-1901, troughs in 1886-91 and 1901-06.

Rather than a comprehensive examination of this migration pattern, the objective of this paper is much more limited. We seek to examine the hypothesis that area differences in net migration within an intercensal period reflect to a large measure differences in economic opportunity; i.e., areas of high net in-migration will show higher indicators of economic opportunity than areas of relatively low net in-migration, and vice versa.

The economic variables used to measure levels of economic opportunity are income per worker and a proxy for labor supply. The 87 départements and their rural and urban subdivisions constitute the geographic units for the observations on the dependent and independent variables. The hypothesis is examined primarily in terms of Kendall's rank correlation for three intervals -- 1876-81, 1886-91, and 1901-06. Quinquennial census data provide the basis for the migration estimates and for some of the economic variables examined.

II

Net migration levels are published or can be derived as residuals for each intercensal period. That is, for any given area the balance of births and deaths between censuses would lead you to expect a particular population at the next census if we were to consider natural increase alone. Should the actual population at the next census show a higher population than would be expected from natural increase,then net in-migration is inferred as the difference or residual between expected and actual population. Of course, for net out-migration, the expected population coming solely from natural increase factors would be more than the population actually enumerated. These net migration residuals are the net balances of opposing gross flows over an interval. By themselves, such residuals indicate nothing of either source or destination of migration.

To facilitate comparison, the net migrations have been converted to rates per thousand of population at the beginning of the five-year intercensal period. The rates cover not only the 87 départements but also their rural and urban sectors. In each département these sectors are defined by the character of the principle place (chef lieu) of each of the more than 35,000 communes. When the chef lieu has 2,000 or more inhabitants the commune is "urban", with fewer inhabitants, "rural". Accordingly we can utilize three different sets with 87 items for the migration data within each département. Each set of migration rates have been ranked with the first being assigned as follows: to the highest plus-value (in-migration) for urban sectors and for entire départements; to the highest minus-value (out-migration) for the rural sectors.

Since the departement net migration is available as total, urban and rural rates, it is desirable to have parallel département income categories -- total, non-agricultural and agricultural income per worker. The non-agricultural incomes do not match exactly urban and rural sectors of départements. For example, income from fishing, which is a part of agricultural income, may result from activities of urban population along the coast. Nevertheless, this income delineation is the closest approximation to the incomes of urban and rural populations it is possible to make under the circumstances.

Income data provided by Delefortrie and Morice serve as the basis for estimation of the relative income positions of the départements and their rural and urban compenents.[2] The detailed inquiries of agriculture and manufacturing (Enquêtes of 1861-65) were used for the 1864 estimates; the economic census, for 1954 data. Generally judicious use of these sources by Delefortrie and Morice provides incomes estimates by département for agriculture from the product side and for non-agricultural activity from the income side.

Since our study lies within this period, the Delefortrie-Morice volume provides no cross-sections for direct comparison. However the relationships between the data for 1864 and 1954 and the statistical methodology do provide a basis for approximating income positions for intervening year. First, the analysis runs in term of rank orders of département incomes and not the magnitudes. Hence emphasis is placed on relative income positions, on ordinal not cardinal considerations.

Second, the analysis of ranks is within income types. The income discussion does not compare different types of income -- e.g., agricultural vs. non-agricultural -- where the approach of Delefortrie and Morice is subject to particular criticism. And in the discussion of total income which combines agricultural and non-agricultural incomes, the département income structure tends to be dominated by either agricultural or non-agricultural activity.

Third, the rank orders within département income categories exhibit remarkable stability between 1864 and 1954. Between the two dates, the rank orders of total, agricultural and non-agricultural incomes per worker were positively correlated at the .001 level of significance.

Given this stability in département income positions of 1864 and 1954, we have made crude estimates of income positions within the period, assuming (1) that the income positions of the départements in 1864 and 1954 are a reasonable approximation to their secular positions, and (2) that the changes were distributed evenly over time. This allows a first approximation of the intervening relative income positions of the départements.

The per capita income data have been converted into income relatives, indicating the position of one area with respect to the national average. [3] Relatives were computed for 1864 and 1954 and interpolations made for 1879, 1889 and 1904. For each département this was done for total income per member of the département labor force, non-agricultural income per member of the non-agricultural labor force, and agricultural income per member of the agricultural labor force. At each interval, the départements were ranked for each income category; the highest income relative being assigned the first rank. [4]

With labor supply conditions as an independent variable, our hypothesis suggests that a more rapid growth from natural increase in one area compared to others might in time be expected to make employment conditions in that area less favorable, resulting in a higher net out-migration rate. Other areas experiencing relatively slow natural growth of population might be expected to be net in-migration areas, other things remaining unchanged.[5]

Changes in labor supply conditions do not exert equal pressures toward migration on all elements of a population. Empirical studies of other countries and scattered observations for France suggest the 20-29 year age group to be the most vulnerable to migratory influences, and hence we seek a measure which suggests pressure on this group. [6] Direct measures of labor supply by age are not available. Consequently the rate of change due to natural increase of the population aged 20-29 years for each of the départements serves as a proxy variable. [7]

The expected rate of increase of a particular age group tells by how much the population in the age group would be expected to change over a given period, in the absence of subsequent migration, as the cohort[8] initially in that group moves out and is replaced by its successor cohort. For the period under study the censuses are quinquennial; a more generalized statement of this variable is as follows:

$$\text{Exp. } \Delta \text{ pop. } (20\text{-}29)_{t_1 - t_2} = \frac{(15\text{-}19)_{t_1} - (25\text{-}29)_{t_1}}{(20\text{-}29)_{t_1}},$$

where t_1 and t_2 are consecutive censuses. In the discussion to follow, this has been expressed in terms of a rate per 1000 inhabitants aged 20-29 years at t_1 in each département. This rate has been calculated for each département and ranked in descending order of expected rate of increase.

III

The associations between net migration and income are given in Table 1 which shows the rank correlation of total income per member of the département labor force with total, urban, and rural net migration rates. Likewise, agricultural and non-agricultural income relatives have each been correlated with each of the three net migration rates.[9]

Table 1 can be variously interpreted, as suggested by the following questions: (1) For any particular interval, what was the "strength" of the link between the three net migration types and a given income variable? (2) Which of the three income variables, if any, provides the "best" explanation of the ranking of the various migration rates? (3) Is there any trend in the correlations between the incomes and the migration rates? Consider first the panels of Table 1 horizontally; secondly the columns of all the panels; finally changes over time within each panel.

Total income per worker is positively and significantly correlated (.01 level) with the net migration rates of the total and rural populations. The correlation is substantially higher with the net migration rates of total population -- all above the .001 level. There is no significant correlation between département levels of total income per worker and the rates of net migration in urban areas, except for one interval.

Agricultural income per worker as related to migration rates follows the same pattern for total income. The relationship is positive and statistically significant, generally at the .01 level, with the migration rates of total and rural populations. As one would expect, no significant correlation is apparent between agricultural income and the behavior of urban net migration rates.

Non-agricultural income per worker is positively and significantly correlated to the .001 level with the net migration rates of département total populations at each of the three intercensal periods. No clear relationship is apparent between non-agricultural incomes and either rural or urban net migration rates. Non-agricultural income correlates positively and significantly (.05 level) at two intervals with rural net migration and at one interval with urban net migration.

TABLE 1

RANK CORRELATION OF DEPARTEMENT INCOME RELATIVES WITH RATES OF NET MIGRATION: 1876-81, 1886-91, 1901-06

[Kendall's Tau]

Income Type:		Period	Rates of Net Migration of: Total Population	Urban Population	Rural Population
Total income per member of the total labor force	1879	1876-81	+.304[a]	-.038	+.193[b]
	1889	1886-91	+.367[a]	+.192[b]	+.193[b]
	1904	1901-06	+.405[a]	+.027	+.389[a]
Agricultural income per member of agricultural labor force	1879	1876-81	+.309[a]	-.037	+.155[c]
	1889	1886-91	+.333[a]	+.121	+.187[b]
	1904	1901-06	+.334[a]	-.041	+.343[a]
Non-Agricultural income per member of non-agricultural labor force	1879	1876-81	+.321[a]	+.064	+.004
	1889	1886-91	+.406[a]	+.199[b]	+.150[c]
	1904	1901-06	+.327[a]	-.002	+.222b

Sources for basic data: Statistique Général de la France, Résultats statistiques du dénombrement de 1881, Table 9, pp. 82-85. Résultats statistiques du dénombrement de 1891, Part 1, Table 6, pp. 394-99. Résultats statistique recensement général de la population de 1906, Vol. 1, Part 1, Table 4, pp. 96-99.

Nicole Delefortrie and Janine Morice, Les revenus départementaux en 1864 et en 1954, pp. 217-19, 270-72.

Note: Highest rank according to highest algebraic order.
Significant to: [a].001 level; [b].01 level; [c].05 level.

Considering the columns of Table 1, total population rates of net migration correlate somewhat higher with total and non-agricultural incomes, although all are significant at the .001 level for each of the three income variables. For the three intervals, the highest coefficients appear twice in the non-agricultural income panel (1876-81 and 1886-91) and once in the total income panel.

Urban net migration rate behavior is illuminated very little by any of the income variables examined. At only two of the nine correlations is there a significant association (.01 level), and these are for the same intercensal period (1886-91). For this interval urban net migration correlates significantly with both non-agricultural and total income per worker.

Rural net migration rates are somewhat more closely correlated to total income than to agricultural incomes; in either case only loosely associated with non-agricultural incomes.

A trend in the correlations suggests an increasing degree of positive association between some migration rates and income categories -- if three observations in each case can be regarded as meaningful. The correlations of net migration rates of total population rise over time with both total income and agricultural income, but not with non-agricultural income. The rural net migration rate correlations exhibit a tendency to increase over time with all three income variables. No similar tendency is evident between urban net migration rates and the income relatives.

A more general interpretive statement requires reference to the extent of correlation within each of the types of variables. For the dependent variables, the net migration rates of both rural and urban areas correlate significantly and positively with the rank order of the rates of the département total populations. The highest correlations were between the net migration rates of the total and rural populations; and lower although still significant correlations between the total and urban net migration rates.[10] These correlations largely reflect the extent to which rural net migration dominated the migration pattern of the total population of most of the départements.

With regard to the designated independent variables, total income is more closely related to agricultural than to non-agricultural income. Agricultural income as an influence on the total income pattern becomes less pronounced, and the role of non-agricultural income becomes more important as France becomes more industrialized.[11]

This interdependence among the migration variables and among the income variables increases the difficulty of assessing the relative strengths of the income factors as explanatory variables. With this reservation in mind, however, the simple correlations between the income and migration variables suggest the following generalizations:

1. Départements with relatively high net in-migration rates were also areas with relatively high incomes per worker (total, non-agricultural and/or agricultural). The reverse would tend to be the case for net out-migration areas. For the rural sector, high rural net out-migration was linked to relatively low total and agricultural incomes.
2. Urban net migration rate differentials do not generally appear to be associated with the income relatives. While higher correlation is evidenced with non-agricultural income relatives, the coefficients are not generally significant. If an income factor is operating in explaining urban migration differentials, a more appropriate variable may be found in a disaggregate of non-agricultural income.
3. The upward tendency noted in some of the coefficients, as the period under study progressed, suggests the relationship between the migration differentials and the economic variables became stronger. This might imply that net migration became more responsive to direct market forces or, perhaps, a reduction in the real costs of transfer.

Turning now to Table 2 where net migration is associated with the labor supply variable, we compare the actual net migration structure during selected quinquennia with the likely "pressure" on the supply side for the same period which would be felt in the absence of any net migration of 20-29 year olds. It will be recalled from the discussion of methodology that the net migration rate at the end of the intercensal period is compared to the expected rate of increase of this group at the beginning of the intercensal period.

TABLE 2

RANK CORRELATION OF DÉPARTEMENT EXPECTED RATES OF INCREASE OF POPULATION 20-29 YEARS WITH RATES OF NET MIGRATION, BY TYPE OF POPULATION 1876-81, 1886-91, 1901-06

Intercensal Period	Kendall's Tau when Correlation is with Net Migration Rate for:		
	Total Population	Urban Population	Rural Population
1876-81	-.315[a]	+.006	-.206[b]
1886-91	-.383[a]	-.219[b]	-.217[b]
1901-06	-.555[a]	-.182[c]	-.468[a]

Sources for basic data:

Statistique Général de la France, Résultats généraux du dénombrement de 1876, Table 12, pp. 112-18, 136-42. Résultats statistiques du dénombrement de 1881, Table 9, pp. 82-85. Résultats statistiques du dénombrement de 1886, Tables 23, and 27, pp. 156-59, 168-69. Résultats statistiques du dénombrement de 1891, Part I, Table 6, pp. 394-99. Résultats statistiques de recensement général de la population de 1901, Vol. IV, Table 7, pp. 382-85, 402-05. Résultats statistiques de récensement général de la population de 1906, Vol. I, Part 1, Table 4, pp. 96-99.

Note: Significant to: [a].001 level; [b].01 level; [c].05 level. Ranking in each case beginning with highest plus-rate in first rank.
For a given intercensal period, t_1-t_2, the migration estimates are from the census at t_2; the labor proxy estimates from the census at t_1.
With quinquennial censuses, the expected rate of population increase aged 20-29 years is stated as follows:

$$\text{Exp. } \Delta \text{ pop. } (20\text{-}29)_{t_1-t_2} = \frac{(15\text{-}19)_{t_1} - (25\text{-}29)_{t_1}}{(20\text{-}29)_{t_1}},$$

where t_1 and t_2 are consecutive censuses.

These correlations suggest the following relationships:

1. The correlation is higher with total than with either the rural or net migration ranks. However, all correlations except one (1876-81 for urban rates of net migration) are significant to at least the .05 level.
2. With regard to the migration rates of total and rural population, the rising coefficient suggests an increasingly strong link between département labor supply conditions in the most migratory ages and migration rates.

If one were to hazard a more general interpretation at this point, the table would suggest that the relatively high expected rate of increase of the 20-29 age group is associated with relatively high rates of net out-migration from those départements. A similar, though less marked relationship obtains for rates of net out-migration from rural areas.

Of course, these correlations pertain only to three of the eight quinquennia in the 1872-1911 period. Also, the expected change in the population is not obtainable for rural and urban components. The analysis would profit from more detailed information on the nature of the labor force changes by sector and by age; but the available data do not make this possible. Nevertheless, the evidence presented here suggests a strong association between differentials in net migration rates and expected rates of change of the more migratory ages.

Having shown some significant relationships between migration rates and income on the one hand and the expected rates of change of the more migratory ages on the other, it remains to be seen how migration rate differentials are related to both of these variables considered simultaneously. First, at each income level are the net out-migration rates typically higher for départements with relatively high rates of expected increase of the 20-29 year age group, and vice versa? Second, for each age-specific labor supply interval do the départements with lower incomes generally tend to be those with higher net out-migration rates, and vice-versa?

In considering these questions, however, there is an issue of whether the income and labor supply variables exert independent influences because the rates of expected increase of population ages 20-29 years and annual income per worker are themselves negatively and significantly correlated.[12] Further, one might argue that there is really no need to consider labor supply, since it is already reflected in income. This may be the case, but the income estimate is for the whole labor force not just the younger population on which our labor supply variable centers. Income is a general rather than a specific index of economic opportunity.

TABLE 3
MULTIPLE RANK CORRELATION OF MIGRATION, INCOME, AND LABOR SUPPLY VARIABLES: 1876-81, 1886-91, 1901-06

Intercensal Period	Partial Tau BC-A	Tau BC	Partial Tau AC-B	Tau AC
	(1)	(2)	(3)	(4)
1876-81	+.228	+.304*	-.448	-.315*
1886-91	+.258	+.367*	-.600	-.383*
1901-06	+.140	+.405*	-.860	-.555*

Source: Computed from basic data cited in Tables 1 and 2.

Notes: A = Rate of expected increase of population aged 20-29 years.
B = Annual income per worker.
C = Rate of net migration.
* = Significant to .001 level. According to Kendall, no test of significance is possible for Partial Tau.
Also of interest are the remaining simple correlations (AB) for 1876-81 to 1901-06 which are respectively: -.278*, -.246* and -.368*.

The interrelations between these three variables are shown in Table 3, where the rank order of the net migration rates is correlated (1) with income per worker when the labor supply variable is held constant (col. 1), and (2) with labor supply when income is constant (col. 3).[13]

Comparisons of columns 1 and 2, 3 and 4, 1 and 3 suggests the following conclusions:

1. When labor supply is held constant, the partial tau between income and net migration is less than the simple tau between the two.
2. When income is held constant, the partial tau between labor supply and net migration is greater than the simple tau between the two.

Hence, of the two independent variables, the labor supply variable appears to exert relatively more influence on the rank order of net migration rates. This is shown too by the co-efficients of column 3 being higher than those of column 1 at each interval. However, this indication may reflect only the inadequacy of the income estimates and the fact that they do not necessarily reflect the incomes in the more migratory age groups.

At this level of generality, urban net migration does not appear to be associated with either income and only slightly with the labor supply variable. This may follow from the nature of the income estimates. Hence, further examination of this relation would involve some disaggregation of the income and labor supply variables and should include examination of economic activity particularly attractive to potential migrants -- e.g., net capital formation, construction activity and the like. Preliminary investigation indicates the importance of the rate of change of the non-agricultural labor force.[14]

In a limited exploration of this type, the additional avenues for investigation far outnumber the few variables examined. Adjustment for the foreign component is necessary, if the net migration pattern is to reflect forces and responses internal to France. More reliable income and product accounts should be constructed. This paper has not explained why the link between net migration and several economic variables exhibits a closer association later in the period under consideration. Was the market mechanism communicating relative advantage (and disadvantage) more efficiently? Were there substantial improvements in the cost of transfer, thus lowering the real cost of migration? Nor has this study considered either the extent to which the parcelling of land holdings or the differences in agricultural mechanization may have influenced migration. All these issues can be investigated at the <u>département</u> level with the information presently available mainly from official sources and will have to be considered in any extension of this exploratory study. Fortunately for the analytical economic historian, France provides an excellent laboratory for historical-statistical investigation of the development process.

FOOTNOTES

*This study was supported financially by the Population Council and intellectually by Hope T. Eldridge, Everett Lee, and, especially, Richard A. Easterlin among others. The author gratefully acknowledges this assistance, but he alone is responsible for remaining errors.

[1]Some areas, Paris and Marseilles for example, were focal points of net migration from abroad. But since this study concerns the general rural-urban pattern of economic differentials influencing migration, these exceptions are ignored in the present paper. For a detailed discussion see this writer's "Internal Migration and Economic Opportunity: France, 1872-1911"(unpublished doctoral dissertation), Department of Economics, University of Pennsylvania, 1964.

[2]Nicole Delefortrie and Janine Morice, Les Revenus départementaux en 1864 et en 1954 (Paris: Colin, 1959).

[3]Suppose, for example, that in 1864 annual income per member of the labor force is 750 thousand francs in département X, while it is 600 thousand for all France. An index for département X would be 125, indicating income 25% above the French average which is assigned a value of 100.

[4]Corsica was removed from all of the ranking because of its distance from the French economy. The départements in Alsace-Loraine ceded to Germany in this period were also excluded.

[5]Louis Chevalier has asserted that between 1880 and 1910 excessive levels of population brought about rural-urban migration. See "Localisation industrielle et peuplement", Population, 1 (1946), p. 28. But nowhere has this view been rigorously examined.

[6]The one generalization about migration differentials which can be considered definitely established, although even this one cannot be stated precisely, is the following: "there is an excess of adolescents and young adults among migrants, particularly migrants from rural areas to towns, compared with the non-migrating or the general population." See Dorothy S. Thomas, Research Memorandum on Migration Differentials (New York: Social Science Research Council, Bulletin 43, 1938), p. 11. Dudley Kirk, Europe's Population in the Interwar Years (League of Nations 1946), pp. 157-60. G.H. Daniel, "Labour Migration and Age-Composition," The Sociological Review, XXXI, 3 (July, 1939), pp. 287-304. A.K. Cairncross, Home and Foreign Investment, 1870-1913, (London: Cambridge, 1953), p. 219. J. Saville, Rural De-Population in England and Wales, 1851-1951 (London: Routledge & Kegan Paul, 1957), pp. 108-25. Dorothy S. Thomas, "Economic and Social Aspects of Internal Migrations: An Exploratory Study of Selected Communities," in Economic Essays in Honor of Wesley Clair Mitchell, (New York: Columbia, 1935), p.462. Hope T. Eldridge,Population Redistribution and Economic Growth: United States, 1870-1950, Vol. III Demographic Analyses, (Philadelphia: American Philosophical Society, 1965), Part I, Chap. 6. Louis Chevalier, La Formation de la population parisienne au XIXe siècle (Paris: Presses Universitaires de France 1950), pp. 263-68. Georges Mauco, Les Migration ouvrières en France au debut du XIXe siècle (Paris: Lesot, 1932), pp. 47, 51. Marcel Croze, "Un Instrument d'étude des migrations interieures: les migrations d'electeurs," Population, 11 (1956), pp. 242-43.

[7]Other possible variables -- the rate of natural increase of total population of the current quinquennium and of twenty years earlier -- were examined and rejected in preference to the expected rate of increase of the population aged 20-29 years, which proved better both on theoretical and empirical grounds.

[8]Deaths among the new cohort during the period are ignored. Considering survival ratios, this does not seem unreasonable for the 20-29 year age group. The forward census survival ratios of the 20-29 male and female cohorts combined were .9274 for 1876-81 and .9668 for 1901-06. Using the 1901-06 interval as an example, this ratio is defined as:

$$\frac{\text{Male and female residents aged 20-29, France, 1906}}{\text{Male and female residents aged 15-24, France, 1901}}$$

[9]It must be emphasized that non-agricultural income is not necessarily urban income. Mining, for example, may occur largely in rural communes of some départements.

[10]For 87 départements, Kendall's tau for the migration variables are as follows:

Rank orders correlated	1876-81	Intercensal period 1886-91	1901-06
Total v. urban net migration rates	+.308	+.419	+.344
Total v. rural net migration rates	+.366	+.457	+.639

All coefficients are significant to the .001 level. The rankings of urban and rural net migration rates are not sufficiently correlated.

[11]The values for Kendall's tau for the rank order of the départements by (1) total income per worker and non-agricultural income per worker and (2) total income per worker and agricultural income per worker are as follows:

Income per worker	1864	1879	1889	1904	1954
Total v. non-agrucultural	+.278	+.323	+.362	+.463	+.599
Total v. agricultural	+.778	+.764	+.741	+.703	+.561

All coefficents are significant at the .001 level.

[12]The values for Kendall's tau are -.278, -.246, and -.368 for 1876-81, 1886-91, and 1901-06 respectively.

[13]See the method discussed in Maurice G. Kendall, Rank Correlation Methods, (3rd ed. rev.; London, Griffin, 1962), Chapter 8.

[14]Rank correlation of urban net migration rates and the rates of change of non-agricultural labor force for 24 départements important in in-migration areas give values for tau of +.239 and +.362 for 1896-1901 and 1901-06 respectively. The latter is significant at the .05 level.

DISCUSSION

R. Marvin McInnis, Queen's University

Miss Bowman and Mr. Myers have provided us with a stimulating and particularly helpful contribution to research on human migration. Let me say at the outset that I am most sympathetic to the approach that they have followed. Besides sorting out the crucial distinctions between several of the existing contributions to this area of study, they correctly emphasize that future research into the economics of migration must deal with migrations rather than migrants. Both existing studies and the presently available statistical data concern migrant status and, conveying even less information, that fictitious being the "net migrant." I think that I can summarize Bowman and Myers' discussion of net and gross, one-way and re-migration by saying that economic models of decisions to migrate must be concerned with specific migrations. Partly because of the nature of the data that have been available but partly because much of the research has not been formulated in behavioural terms, the literature of migration exhibits an almost exclusive concern with the characteristics of migrants, either present or past. As Bowman and Myers so nicely show, unless migration is given a specific time reference and unless particular migrations are sorted out, misleading results can be expected.

I should like to add two qualifications. Firstly, the emphasis on behavioural models should not be construed as a complete denial of the usefulness of more aggregative and impersonal analyses. In many situations our problem is the historical one of assessing the role of migration in a particular case of economic development. Then the main outlines of the development under study may be perceivable without the detailed information and sophisticated model proposed by the authors of the paper we have heard today. A second quibble concerns the statistics on return migration presented by Bowman and Myers. They look at the proportions of persons reported in the census of 1960 as having migrated in the preceding five years that were migrating back to their region of birth. I am sure that the authors would agree that this is far from a satisfactory measure of re-migration. One would be most curious, for example, about the numbers of 1955-60 migrants who moved more than once during the quinquennium. But I appreciate the severe paucity of statistical data on migrations.

Bowman and Myers issue a plea for census or other data on migration in sufficient detail of cross-tabulation that the kind of disaggregated analysis which they propose might become statistically feasible. In this I heartily join them. The problems of the researcher in this area in Canada, where I am working, are even greater since census migration data, whatever the extent of cross-tabulation, have been available only for ten provincial units - far too small a cross-section sample for sophisticated statistical analysis. Moreover, classification by both origin and destination, a prime requirement of the kind of methodology the authors propose, has not been undertaken in Canada. One very useful body of data on migration exists in the microfilmed enumerations of the 1941 Census of Canada. On that census questions were asked about both previous residence and previous occupation but the two were never cross-tabulated. The form in which the data have been preserved would make a fairly large sample with substantial cross-classification feasible and not outrageously expensive. I can only hope that on a future occasion I shall be able to report on this material to this association and at that time I might be able to put Bowman and Myers' methodology to a more substantial test.

My comments have concerned the analysis of migration and I have been generally in agreement with the authors' position that migration must be considered in a more complex way than it has in most past analyses. Their paper emphasizes the interrelationships between migration and education and training as elements in the spatial transfer of human capital. Just as the complexities of migration have to be sorted out in more detail, education and training should be considered in more dimensions than Bowman and Myers give them. It is widely recognized that education and training involve both the acquisition of greater skill and increased specialization of the worker, although not always to the same extent or in the same degree. It is useful in this context to think of skill and specialization of human capital counterposed in much the same manner that Ames and Rosenberg have applied these concepts to physical capital.[1] Where greater skill implies less specialization, or less strict attachment to particular occupations, the implications for migration decisions may differ considerably from highly specialized or occupationally specific training.[2] The importance of the distinction will vary from one country to another since it will depend also on the spatial distribution of demands for particular skills and specializations. Recognition of further dimensions of education and training will assuredly complicate further Bowman and Myers already intricate treatment of different migration sequences but it is a complication that may be every bit as important as the recognition of differences in time-juxtaposition of migration and education decisions.

I have already joined the authors in their plea for improved census data on migration, with adequate cross-classification for the kinds of analysis both they and I would like to undertake. In concluding my comments, however, I should like to say that I am not optimistic about the adequacy for this type of research of even greatly expanded census tabulations. It is extremely difficult to conceive of census questions that would collect the needed information on migration as opposed to migrant status. This kind of information is best acquired in other ways. Here, as in a number of other areas of economic and demographic research, a continuous cross-section sample would be preferable. Perhaps Bowman and Myers will join me in a plea for sample statistics of that sort.

---

[1]Edward Ames and Nathan Rosenberg, "The Progressive Division and Specialization of Industries", Journal of Development Studies, Vol. 1, No. 4 (July 1965), p. 370.

[2]For some purposes human capital might be considered in undifferentiated "liquid" form but it must be kept in mind that at any time the stock of human capital like the stock of physical capital is largely committed to specific forms.

Discussion

William Paul McGreevey, University of California, Berkeley

Gordon Marker's paper on French migration in the interval between the Franco-Prussian War and the Great War is bound to occasion some thoughts about the relevance of his work to the new economic history. I propose in this brief comment to contrast Marker's work with the new economic history, to show that his paper is largely a-historical, and on the basis of this critique to make some suggestions for further investigations into French migratory history.

I

There seem to me to be three principal features of the new economic history of interest to us here: (1) It is history, i.e., there is a clear and demonstrable interest in historical data and information per se, not just for that information's relevance to contemporary policy problems or as a source of empirical support for some theory or other. (2) It uses statistical and econometric techniques in the analysis of historical data, not as ends in themselves but only as an adjunct or aid to the central task of understanding historical phenomena. (3) So far the new economic history has been largely confined to studies of American economic history, and that principally in the nineteenth century. Based largely on the work of an earlier generation of quantitatively-oriented economists (many of them connected with the National Bureau of Economic Research), the present stage of U. S. studies finds one with the possibility of aggregating the primary data into macro-economic categories which have become the touchstone of economic analysis since Keynes. In no other country (Gregory King notwithstanding) has so much preliminary work been done prior to the 1950's and 60's to permit the kinds of advances made of late in U. S. economic history.

Marker's paper, in contrast to the best work of the new economic historians, is not history. His primary interest is in using French migration data to support the hypothesis that "area differences in net migration within an intercensal period reflect to a large measure differences in economic opportunity. . ." And, I might add, he is not above insinuating a closing remark that "France provides an excellent laboratory for historical-statistical investigation of the development process," which is a highly elastic substitute for: "My work, though apparently related to the dark past, is really relevant to the problems of the developing nations. . ." In this, of course, Marker is no worse than many of our contemporaries: Conrad and Meyer's classic study of the ante-bellum slave economy mentions 'the near slavery existing today' and 'key policy questions in former colonial countries' as justification for undertaking their study.* Perhaps the improving status of economic history as a fit subject for study by economists will in the end justify abandonment of these unnecessary rationalizations of historical research. The historical sterility of Marker's investigation here reported resulted directly from the incidence of some feeling that historical study of French migratory movements is not in itself interesting. In this respect his work is less effective than that issuing forth from the new economic history of the United States.

If this study is not history, what is it? --Hypothesis testing pure and simple, for which the 'real data' are of only secondary importance. In operational form the hypothesis comes down to $m = F(Y_L/L, \Delta L)$, and the outcome is that ΔL is somewhat more important as an explainor of net migration than Y_L/L, though both are significant correlates of net migration. Is anyone surprised? Does anyone know very much more about French migration, given that many of us would have accepted the proposition as axiomatic?** The level of sophistication reached in recent years in comtemporary studies of migratory movements is probably not attainable for historical studies simply because of the limitations of data. The economic historian must perforce make his contribution in the understanding of history rather than the development of methodology. But my complaint as I read Marker's paper is not that the investigation was carried out, for surely one man's axiom is another man's hypothesis; rather it is that so much of the quality of French migration had to be drained out of the data in order to move us inexorably to the conclusion. Without asking for twinkly-eyed Parisians, stolid French peasants, or the rural-urban sins of a migratory Madame Bovary, I still would hold that something of France must be a part of French migration. Let me offer some suggestions which might breathe some of France into this discussion.

II

The observation of city size distributions within political units reaches back at least to the

---

* Alfred Conrad and John Meyer, The Economics of Slavery (Chicago, 1964), p. 47.

** The paper by Mary Jean Bowman and Robert G. Myers, "Schooling, Experience, and Gains and Losses in Human Capital Through Migration," this volume, pp. 0-000, carries forward its argument with these assumptions.

neoclassical economists Marshall and Pareto, the latter having developed from his Italian experience the so-called rank-size rule. Others, including the geographer Mark Jefferson, were intrigued by the development in a number of countries (England and France included) of the primate city, one which developed such scale economies that all secondary cities were much less developed than the Paretian rank-size rule would suggest. Paris is of course a prime example of primacy being several times larger than Marseilles, France's second city. Given this particular city size distribution today, one would expect some important discontinuities between Paris and all other urban places so that even the ordinal ranking employed by Marker would fail to distinguish important differences in the causes of migration to Paris itself and to all other urban places. We would like to know, moreover, whether during the period under study the agglomerative powers of the Parisian area waxed or waned: Given that some large share of total urban growth due to migration can be directly attributed to the growth of Paris (I hazard the guess of one-third), one might gain in the analysis by giving that city separate treatment. In any case it seems hard to justify the rank correlation techniques when the absolute population growth of the various départements vary so greatly. I am not satisfied that use of migratory rates per 1000 solves the problem of essential discontinuities inherent in the primate city size distribution of France.

In my own investigations of internal migration in Colombia, South America's third largest nation but with a populated area roughly the size of France, I have become intrigued with patterns of migratory movement. Perhaps surprisingly, about three-quarters of the movement to the largest cities is from smaller towns rather than from strictly rural areas. Though a pattern of step migration may seem to be the dominant theme of physical mobility, the available data are better explained by an hypothesis of fill-in migration: As small-town residents leave for the big city, their places are filled in by short-distance rural-urban migrants. I believe these two movements respond to essentially different economic variables -- the local movement to land pressure in the rural hinterland of each small town, the inter-urban movement to differential opportunities in the fraccionated, heterogeneous urban labor markets. For that reason one might suggest separate hypotheses for the two kinds of movement. In particular, a land/labor ratio in rural areas should predict rates of rural out-migration. Inter-urban movements might respond to the variables Marker lists, but I would suggest that absolute urban size may be (during certain periods of rapid urban growth) a good predictor of migration rates -- this view emanating from a theory of scale economies associated with urban size.

III

Finally, some questions suggested by this study:

(1) Marker notes that per capita income rankings of départements are remarkably similar at the two dates for which information is available, 1864 and 1954. How did the variance of this per capita income distribution behave over the ninety-year interval? Did interdepartmental income inequality continue in spite of (or because of) the significant internal migration which Marker has described? If workers are truly responsive to economic opportunity in their migratory plans, one would expect reductions in inequality. This subject deserves more exploration than Marker was able to devote to it in this brief paper.

(2) Did factor markets become more efficient in allocating the supply of labor? Marker's evidence which shows an increasing Tau for all migrant subsets over the interval 1872-1914 suggests that the answer is yes. His data does not, however, permit consideration of the hypothesis advanced by W. H. Nicholls* that nearness to an efficient urban labor market will make the rural labor market more efficient. During the period under study did rural-urban wage differentials vary with the absolute size of urban places or with local migration rates? These difficult problems are compounded by the lack of usable bases for comparing rural and urban incomes and levels of living. It is perhaps too much to ask Marker to solve this problem.

(3) Was land tenure an unimportant element in the determination of rates of out-migration? Certainly there are significant regional differences in tenure arrangements in France which might be correlated with migratory flows. In approaching answers to this question which would require intimate knowledge of agrarian conditions in France, the author may well arrive at a 'real' new economic history in a French setting. This paper does take us several steps in the direction of combining historical research and hypothesis testing.

---

* "Industrialization, Factor Markets and Agricultural Development," Journal of Political Economy, Vol. 69, No. 4 (August 1961), pp. 319-340.

IX

POPULATION GROWTH ESTIMATION STUDIES

Chairman, CONRAD TAEUBER, U. S. Bureau of the Census

OBJECTIVES AND REQUIREMENTS FOR MEASUREMENT OF VITAL RATES

Forrest E. Linder and Walt R. Simmons
National Center for Health Statistics

I. Introduction

Perhaps the most traditional and firmly established cornerstone of any national statistical system is the periodic decennial census and the continuous registration of births and deaths. In every advanced country of the world these statistics have long been well organized and operate to a high degree of effectiveness. In the lesser developed countries, efforts are being made to initiate and perfect these systems. Numerous international recommendations underline the basic character of these systems as elements of the national statistical program. Many of the activities of the United Nations and the World Health Organization are directed toward developing standards, definitions, and classifications for these systems and assisting countries to create the statistical organizations to compile census and vital event data.

Traditionally the periodic census has been recognized as the vehicle for tracing demographic and economic history of a nation; for describing the growth, migration, and composition of its people; and for providing a wide spectrum of information for administrative and research purposes.

The vital statistics system, an essential companion to the census for many analytical purposes, has provided death rate information which is the crucial index for appraisal of a nation's health problems and a measure of success in reaching national health goals. The birth rate and the current rate of natural increase, the other major products of the vital statistics system, have in the past been of interest primarily to the theoretical demographers. These fertility measures had, except in a few instances, no significant relationship to explicit national objectives or governmental action programs.

Recently, however, the statistical output of the vital statistics system has been raised to an entirely new level of importance. More and more countries of the world are now developing specific plans and programs to accelerate their economic and social development. There is sharp realization that if this rate of social and economic development, however fast it may be, is less than the rate of growth of the population, the resultant effect is not likely to be improvement in the level of living for that nation's people. The birth rate, the death rate, and the difference between these two measures have now become crucial indicators as to whether any developmental plan, however successful technically, will achieve its real objective.

Viewed in this light, many countries are now accepting the fact that their present rate of population growth presents a too difficult pacemaker for their economic development capacity. Such countries are having to develop policies, programs, and organizations which accompany plans for economic development with programs to reduce the rate of population expansion.

These family planning programs if they are to be effective are necessarily large, complicated, and expensive. The current birth rate, its components, and related factors now emerge as essential criteria for developing, guiding, and appraising the effectiveness of family planning activities.

Unfortunately, in most countries where the race between economic development and population expansion is most acute, the traditional vital statistics system and even census tabulations are inadequate to provide data of sufficient currency and accuracy to satisfy requirements for national planning and current evaluation of family planning programs. There is thus an urgent need for new statistical techniques to provide the necessary indicators. The statistical profession has not yet provided an adequate answer to this challenge.

In considering more precisely the problems that the statisticians must solve in meeting these needs for vital rate measurement, the objectives and requirements for several distinctly different situations must be distinguished. These different situations correspond somewhat to the geographic scope of the area for which vital rate measurements are to be made, but are more fundamental than geography alone implies. We shall consider the objectives for measurement at (1) the national level, and (2) the level of subnational intensive family planning program development areas. A third level is recognized but not here discussed: namely, the smaller and more intensive studies in which, under laboratory or experimental conditions, still more incisive study is made of the relationship between fertility and one or more of its components.

II. Measurement of Vital Rates for the Nation as a Whole

At the national level the prime objective is to obtain indexes of the trend of the total birth and death rate or related total measures of fertility and mortality. These are the prime indexes which will reflect total population growth and its relation to economic development. It is this purpose that an accurate periodic census and a current vital statistics system can serve. Eventually these systems will be created, since over the long run they are indispensable.

Family planning programs are not applied with uniform effectiveness throughout a country, and programs of either birth or death control represent large and ponderous national activities. National levels of birth or death rates will ordinarily change slowly and any change in the rate of total population growth will show great inertia.

For these reasons the measurement methods for the nation as a whole can be relatively slow in producing results. The week-to-week or month-to-month change is not likely to be of great significance, but observations at, say, decennial intervals are hardly sufficiently current. Although a certain amount of deliberateness can be accepted, the final measurements for longer periods must be obtained with some care since a small change in rates could have important meaning and could influence major policy decisions relating to national development plans or the degree of emphasis placed on family planning programs.

While at the national level the crucial items of statistical information are the vital rates that permit an assessment of overall rates of population growth, it also would be highly advantageous to have additional variables which permit a somewhat more penetrating analysis of what factors are related to any change. Age-specific rates would be essential as well as data relevant to broad effects of national changes in such components as age at marriage, urbanization, or employment of women. Since socioeconomic changes would not be occurring uniformly throughout a nation, data for geographic subdivisions become highly desirable.

Some attempts have been made to find statistical mechanisms that will provide these important vital measurements in the absence of a routine vital registration system. The Indian National Sample Survey has made serious efforts to measure vital rates through nationwide household interview surveys, and the Indian Statistical Institute has published a number of papers on some of the methodological problems involved. The Pakistan Population Growth Estimation Project and the Thai Survey of Population Change have attempted to obtain national estimates through an interrelated combination of intensive registration efforts in sample areas and independent interview surveys covering the same population. In Egypt, there are experiments with a so-called "busybody" technique, in which a vital-event inventory is attempted through the offices of a careful observer stationed in each village of a sample of several communities. In several other areas a solution is sought through periodic cross-section population surveys.

A very great deal needs be done, however, before adequately proven survey methods are a part of our statistical arsenal. For this national or global level, it is agreed that minimum requirements are modest in scope, at least in the early stages of study: perhaps age-specific vital rates for each of a few sectors of the economy. The difficulties are legion.

This is not the occasion and indeed we are not able to present a blueprint for an ideal survey procedure. We do offer three principles as guidelines for research and for action.

The first of these is that many methods must be tried—the more unsuccessful being discarded, and the better being retained, improved, and adapted to the specific land. This process of research and gradual development should characterize almost every phase of measurement of vital rates, and it is scarcely revolutionary to take note of it. Even so, it is well to remember that this is the state of the art.

The best approach will vary from one place to another. For example, the National Center for Health Statistics, Agency for International Development, and Research Triangle Institute are engaged in a cooperative project in North Carolina, attempting to develop under nearly ideal conditions a one-time retrospective interview which will satisfactorily reflect birth and death rates in North Carolina. If the effort is successful, it will be one step along the road we're trying to go, although it obviously will need modification and further testing in lands of differing cultural background. If it fails, under even the favorable circumstances, additional questions are raised as to whether it is feasible or possible to devise an acceptable interview procedure of this type.

Somewhat parenthetically, and illustrative of the search for new techniques, mention is made of a specific interviewing procedure which is being tested in the North Carolina project. It is a modification of an idea advanced by Stanley L. Warner, known as a randomized response technique. The procedure is one in which the individual respondent is able to reply "yes" or "no" to one of two questions, without the interviewer knowing which question is being answered. Yet at the close of the survey, the statistician can estimate the proportion of the population replying "yes" to each question. This machinery facilitates handling a delicate question which the respondent might otherwise refuse to answer. It is one instance of a whole family of related approaches which deserve investigation, and which may augment substantially the informational power of social surveys.

A second principle is that the *mechanics* of survey design and execution are critical to success. Here also alternatives need be tested and compared. Shall the principal sampling units be geographic areas, or tribal or family entities, or formed on some other basis? The length of reference period is a significant factor, as are the related questions of whether a single sampling unit should be measured more than once, or perhaps observed continuously; and if more than once whether the reference periods of the different measurement occasions should be overlapping. A long list of relevant sampling factors could be displayed. But the matter can be covered, without exaggeration, by saying simply that all the tenets of good sample design may profitably be taken into account. And no matter how superior the design, it is likely to fail its purpose unless the conduct of the survey is faithful to that design. This precept is true of any survey, but it is especially true of vital rate population surveys because they are subject to so many operational hazards under good conditions, and so often are carried out under unfavorable conditions. Accordingly, the total survey plan must include a surveillance system which provides a review of validity. The best system will vary with the particular design, but commonly it will be based on a second or alternative measurement for a probability subsample of the main sample.

Our third principle is that an adequate technical plan must be supported by effective management, administration, and public relations. It is necessary that the survey operate in a climate of adequate budget, political and public acceptance, sufficient competent personnel, and continuing management control. Officials and managers cannot expect or promise miracles and must be prepared to face and carry out extensive and difficult tasks. Large samples will be required, even when the design has maximum possible efficiency. For example, with a 1-year reference period, a sample of the order of 12,000 persons is needed to estimate with 95 percent confidence the birth rate within 10 percent and the death rate within 20 percent. If one asks for age-sex specific birth and death rates for each of several domains, the required sample size moves quickly into the hundreds of thousands of persons. The administration must recognize, too, that the survey data still will fail their purpose unless after collection, they are carefully processed, converted into estimates, displayed in such useful forms as life tables or traditional population projections, and appropriately interpreted by demographers.

III. Measurement of Vital Rates in Family Planning Program Developmental Areas

The planning and conduct of a definite action program to affect the level of the birth rate of a large and populous country is a major undertaking and in most situations cannot be carried on initially on a full nationwide basis. The required organization may take a long time to build, and effective operational procedures must be tested and perfected.

For these reasons it is probable that most national family planning programs will concentrate initially on what might be called "Family Planning Program Developmental Areas" which will constitute social laboratories for the development of the organizational and operational programs that eventually can be applied to the whole country.

Even an assumption that biological or clinical methods have been perfected does not imply that a control device is infallible or that the public health problem of applying such methods to a large population has been solved. The problem of application embraces the total range of questions related to levels and interrelations of governmental organizations, questions of effective communication with the population, the logistics of training, supply, treatment, clinical follow-up, acceptance, and continued utilization by the population. The ultimate effect on the population growth rate is the consequence of the net effectiveness of all the elements of the total program.

The design and appraisal of a family planning program comprising these elements can be undertaken only in an area with a population large enough to present at least a principal part of the total range of problems to be considered. For this reason, Family Planning Program Developmental Areas must be visualized as including populations of the order of magnitude of 1,000,000 persons.

In this situation the requirements for vital rate measurement are rigorous and extensive. The stakes are large because the investment of substantial public funds is involved. But, more important, the stakes are large because decisions based on the results obtained in such social laboratories will presumably be applied on a national basis and will determine the character of the total national family planning effort.

Vital rate measurement must consequently be rapid, sensitive to change, and precise. For these objectives it will not be adequate to obtain a measure of the crude birth rate nor merely of age-specific rates. The variables obtained here must permit a more profound analysis of the fertility changes that are transpiring. Factors such as child spacing, number of previous children, and changes in the level of pregnancy wastage may be critical variables for analysis. Of particular note is the fact that analysis of family planning programs usually will concentrate on the impact over 1, 2, 3, 5, or 10 years, and yet is a part of the broader study of likely changes over 1, 2, or 3 generations.

These are severe specifications—particularly when it is added that the timetable is an urgent one. Under some circumstances, service statistics, such as numbers of condoms sold, or numbers of other conventional devices accepted by families, or vasectomies performed, can be helpful. But this type of information only begins to be useful when it is accompanied by knowledge of a considerable variety of other specific factors which permit translation of service statistics to an estimate of impact on population change.

We should like to argue that the most promising attack on this measurement and analysis problem is a coordinated two-pronged endeavor. One of the two lines of attack is intensified effort to develop and utilize more incisive survey procedures. It was earlier noted that the application of survey techniques to determine relatively simple birth and death rates requires much care and effort. Yet the survey process has been and will be used further as a main instrument for measuring the even more elusive components of population growth: such factors as marriage, separation, and remarriage rates; number of children wanted; desired spacing of children; effectiveness of specific methods of birth control in the applied biological sense; and the relationship to birth rate of education, urban-rural residence, religion, race, economic status, parity, migration, health status. The difficulties are not quite so severe as they may seem at first, for at least two reasons. One is that the first stage of sampling can be of geographic areas, and thus operations need be carried out only in limited and concentrated regions, with consequent fewer demands on personnel and resources, and easier control. The other reason is that for many purposes the required estimate is a ratio or relative number, which may make much less demand on survey capabilities than would the estimate of an aggregate. Further, the universe of study will be the Family Planning Developmental Area rather than an entire nation.

We speak of a two-pronged attack. The companion to the tine of survey measurement is theoretical analysis accomplished through microsimulation and Monte Carlo techniques. This methodology, being developed by Orcutt, Hyrenius, Ridley, Sheps, Lachenbruch, Horvitz, and others is a flexible and powerful tool. It enables one to build a miniature universe in which people are identified only by those characteristics that are relevant to vital rates, and by modifying appropriate parameters condense into hours or even minutes of running time on electronic computers the course of affairs which could be observed in the real world only in years or perhaps generations, if at all. We must forego extended discussion of the process, including its weaknesses and difficulties. In building simulation models, the strategy is to take advantage of all the knowledge one possesses, including results from conventional analysis and the evidence from surveys, to discover what would happen to vital rates and population change given specified values of the key parameters.

This recommended route toward solution can be summarized in the following terms. The critical process is use of a micro-simulation model to determine the sensitivity of vital rates to changes in their component factors, and then either (1) to be able to *measure* those component factors in the real world, mostly through surveys, or alternatively (2) to be able to identify program actions which would change specific components by identifiable amounts, and thus to quantify their effect on population change. It follows that we seek to identify also those components to which birth rates are *not* sensitive, and so to avoid wasteful expenditure attempting careful measurement of such components. And further, analysis and model-building need be tempered by including only elements which are *capable* of measurement in the technological sense and which are *feasible* in the political or social sense.

We urge others to consider this type of two-pronged attack. The problems are large enough for all who wish to tackle them. The rewards of success may be substantial. The price of failure to solve the problems will be high.

REFERENCES

Ahmed, Nazir, and Krotki, Karol J. "Simultaneous Estimations of Population Growth—The Pakistan Experiment." *Quarterly Journal of the Institute of Development Economics,* Vol. III, No. 1 (Spring 1963), pp. 37-65. Karachi, Pakistan.

Chandrasekar, C., and Deming, W. Edwards. "On a Method of Estimating Birth and Death Rates and the Extent of Registration." *Journal of the American Statistical Association,* Vol. 44, No. 245 (March 1949), pp. 101-115.

Chintakananda, Anuri, and Lauriat, Patience. *Technique to Measure Population Growth: Survey of Population Change in Thailand,* mimeographed paper prepared for the World Population Conference, Belgrade, 1965. Also: "Survey of Population Change." Unpublished manuscripts of the Office of National Statistics, Bangkok, Thailand, July 1964, November 1964, January 1965, and March 1965.

Hyrenius, H., and Adolfsson, I. *A Fertility Simulation Model.* Demographic Institute, University of Goteborg, Goteborg, Sweden. Report No. 2 (1964).

Hyrenius, H. *New Technique for Studying Demographic-Economic-Social Interrelations.* Demographic Institute, University of Goteborg, Goteborg, Sweden. Report No. 3 (1965).

Lachenbruch, Peter. *An Introduction to Monte Carlo Methods With Selected Applications.* System Development Corporation, SP-293, Santa Monica, Calif. (1961).

Mauldin, W. Parker. *Evaluation of National Family Planning Programs.* Population Association of America, New York (1966).

Mauldin, W. Parker. *Estimating Rates of Population Growth.* International Conference on Family Planning Programs, Geneva, Switzerland (1965).

National Center for Health Statistics. *Computer Simulation of Hospital Discharges.* Series 2, No. 13. Washington, D.C. (1966).

Orcutt, Guy H. *Simulation of Economic Systems: Model Description and Solution.* American Statistical Association, Chicago, Ill. (mimeo draft report) (1964)

Orcutt, Guy H., Greenberger, Martin, Korbel, John, and Rivlin, Alice M. *Microanalysis of Socioeconomic Systems.* Harper and Brothers, New York (1961).

Potter, Robert G., and Sakoda, James M. *A Simulation of Human Reproduction.* Population Association of America, New York (1965).

Ridley, Jeanne Clare, and Sheps, Mindel C. *An Analytic Simulation Model of Human Reproduction With Demographic and Biological Components.* Population Association of America, New York (1966).

Ridley, Jeanne Clare, and Sheps, Mindel C. *A Simulation Model for Analysis of Human Reproduction.* American Sociological Association, Montreal, Canada (1964).

Sheps, Mindel C., and Ridley, Jeanne Clare. *Studying Determinants of Natality: Quantative Estimation Through a Simulation Model.* World Population Conference, Belgrade (1965).

Som, Ranjan Kumar. *On Adjustments for Non-Sampling Errors and Biases in the Estimation of Vital Rates.* (Personal contribution to the United Nations, undated)

United Nations. *Guanabara Demographic Pilot Survey: A Joint Project of the United Nations and the Government of Brazil.* United Nations, New York (undated).

Vukovich, G. *The U.A.R. Project for Measuring Vital Rates in Rural Areas.* World Population Conference, Belgrade (1965).

Warner, Stanley L. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association,* Vol. 60, No. 309 (March 1965), pp. 63-69.

PROBLEMS IN DESIGNING INTERVIEW SURVEYS TO MEASURE POPULATION GROWTH*

D. G. Horvitz, Research Triangle Institute

Introduction

There is an urgent need for reasonably accurate measures of the rate of population growth in many areas of the world today. The demand is particularly great in countries where programs for economic development are hampered by rapidly growing populations. Satisfactory measures of the components of population change are easily obtained in countries which have adequate registration of births and deaths and which also conduct accurate periodic censuses. Unfortunately, the efficiency of the vital registration systems in many countries leaves much to be desired. It will take some years before the statistics generated from registration data in these latter countries can be considered sufficiently accurate for demographic analysis and planning purposes. In the interim, other procedures are necessary.

A number of alternative procedures for obtaining basic demographic measures in situations where census and registration data are lacking or defective have been reviewed by Brass (World Population Conference held in Belgrade September 1965). Of particular interest here is the measurement of birth and death rates by means of personal interviews in a sample of households.

In recent years, periodic personal interview surveys to estimate vital rates have been conducted in a number of countries throughout the world. At last year's World Population Conference, papers reporting on experimental projects which utilize data on vital events collected by survey methods included such countries as Pakistan (Krotki), United Arab Republic (Vukovich), Morocco (Sabagh and Scott), Senegal (Cantrelle), and Thailand (Lauriat and Chintakananda). Since it is generally recognized that the accuracy of household interview data suffers from reporting errors of unknown magnitude, each of these experiments, Morocco excepted, also makes use of vital registration records. It is of more than passing interest to note here that both the Pakistan Population Growth Estimation Experiment and the Thailand Population Change project are using the method developed by Chandra Sekar and Deming in 1949[1] for estimating birth and death rates more accurately by individual matching of registration data with household interview data.

It is apparent, however, that there is no generally accepted procedure for accurate, yet inexpensive, measurement of vital rates in countries where there is a critical lack of this information. The Research Triangle Institute (RTI) has undertaken a research project to study and develop appropriate basic procedures for the measurement of population growth by means of interview surveys. It is our purpose in this paper to discuss various problems which arise when one attempts to design an interview survey to measure birth rates, death rates and net migration for a designated population. Particular attention will be given to the nature and magnitude of non-sampling or measurement errors. Some preliminary results obtained from interviews in a sample of households located in seven counties in North Carolina, and in which births and deaths were known to have occurred, will be reported.

There are several reasons for conducting the research in the United States, not the least of which is that the means for evaluating alternative survey methods are readily at hand. In particular, there is access to birth and death registration records in a system which is highly efficient and almost complete. It is not expected that techniques found to work satisfactorily in the United States would be directly transferrable to other areas of the world. Some local studies will be necessary for adaptation of recommended methods.

Population Survey Design Problems

There are numerous problems to be faced in the design of a sample survey to measure population growth. They may be classified broadly as problems of coverage and problems of response.

A. Problems of Coverage

In ideal terms, one would like to estimate accurately (1) the size and age distribution of the population of interest at the beginning of a designated reference period, (2) the number of live births that occur in the resident population during the reference period, (3) the number of deaths by age that occur in the resident population during the reference period, (4) the net additions (or subtractions) to the populations, by age, due to migration, and (5) the size and age distribution of the population at the end of the reference period. It is sufficient to have information to estimate four of these five components, since the fifth will then be determined.

If the survey is confined to interviews in a sample of households, coverage problems are introduced. It is clear from the list of items to be estimated that the universe of interest extends beyond the population residing in households during the entire reference period (or some part of the reference period) to include the population in institutions and other group

---

* Research supported by National Center for Health Statistics, Contract No. Ph 86-65-68.

1/ Ansley J. Coale proposed in 1961 that the Chandra Sekar and Deming procedure by used with sample registration areas and periodic household surveys.

quarters, as well. It can be argued that excluding the non-household population from the survey universe will not particularly affect the population estimates when it constitutes but a small part of the total population. In some countries, however, it is likely that a disproportionate number of annual deaths will occur to residents of long-stay institutions. For example, in the seven county North Carolina study (referred to previously) more than 12% of the total deaths occurred in institutions.

Some individuals in institutions are considered to be members of specific households, both by themselves and by the other household members. Others are definitely not attached to any household. Thus, it is not likely that an accurate estimate of total deaths in a designated period can be achieved if the survey universe is confined to the household population. This contention is supported in part by our North Carolina study in which the households associated with a sample of 66 individuals who died in long-stay institutions were interviewed. Only 70% of these deaths were reported in the household interviews, compared to 85% of a sample of 462 adult deaths not occurring in institutions.

The accuracy of the estimate of total births occurring in a reference period may also be affected adversely if coverage is confined to the household population. It is somewhat less likely that illegitimate births will be reported in household interviews in cultures where these births are socially stigmatized. For example, only 50% of a sample of 82 white infants registered as born out-of-wedlock in North Carolina were reported in personal interviews with the mother's household (i.e. the household to which the mother belonged at the time the birth occurred). This compares with 94% reported of a sample of 257 white legitimate births. It seems advisable, therefore, to supplement the household interviews by coverage of institutions and agencies which accept infants with illegitimate birth status in cultures where this is a problem. The household interviews can then be confined to reporting of infants (born in the reference period) who are currently members of the household or who have died.

It is expected that some type of area sampling frame will be used to select the households to be interviewed in population growth surveys and that the ultimate sampling unit will consist either of all of the housing units located within the boundaries of an area segment, or some defined portion of these housing units. Since information is required for a sample of all households (persons) resident in the region covered by the survey <u>at some time</u> during the reference period, special coverage problems are created.

Ordinarily, the households in the sample would be those which occupy the housing units associated with the selected set of ultimate sampling units. If this procedure were followed, however, households which are no longer resident in the survey region (i.e. those which have migrated out of the region or have dissolved due to death or other reason since the beginning of the reference period) would have no chance to be represented in the sample. It is necessary, therefore, to list for the sample the occupied housing units, plus those that are vacant and also those that existed at sometime during the reference period, but were demolished prior to the end of the reference period.

All households that were resident at any time during the reference period in the vacant housing units or in those housing units which no longer exist must be identified, presumably by neighbors. The households in this group which did not migrate out of the region or did not dissolve, but merely moved, intact, to some other location within the survey region are dropped from further consideration, since they are associated with some other sampling unit.

Similarly, it is necessary to identify all households which were resident at some time during the reference period in the occupied housing units and to exclude from the sample those that merely moved as a household to some other housing unit in the survey region. The procedure is somewhat involved, but necessary for appropriate coverage, particularly since death is one of the principle reasons for households to dissolve.

Interviews with neighbors were conducted in the North Carolina study concerning deaths known to have occurred in households previously in residence. The proportion of adult deaths reported by the neighbors in these cases was almost exactly the same (84%) as reported by respondents who were members of households where deaths had also occurred during the reference period. The proportion of births reported by neighbors for households that had moved was somewhat less (74%) than when the household itself was interviewed (92%).

Coverage in surveys is also affected by the proportion of sample households which cannot be contacted in repeated attempts or which refuse to be interviewed. Interviews for households in these categories in the North Carolina study were completed with neighbors whenever possible. In these interviews the proportions of births and adult deaths (known to have occurred in the households of interest) reported by the neighbor respondents were higher than for neighbors reporting these events for households which had moved away.

B. <u>Problems of Response</u>

It is generally recognized that births and deaths occurring in a specified reference period will be under-reported in personal interview surveys. Sabagh and Scott report 8% under-enumeration of births and 17% under-enumeration of deaths occurring in the 12-month period prior to interview in their methodological study in Morocco. In the Thailand project, analysis of the second quarterly round of household interviews showed an estimated 16% of the registered births and 17% of the registered deaths were not reported. The U. S. Bureau of Census 1950 Infant Enumeration

Study estimated 3.6% of infants born in the United States in the first three months of 1950 were not enumerated in the Census. Excluded from the data in the Census study are infants who died before April 1, 1950, illegitimate infants identified as such from birth records, and infants born in a State which was not the usual residence of the mother.

As indicated previously, the RTI study in North Carolina was directed toward investigating the accuracy with which known births and deaths are reported by respondents in the households involved. An initial sample of 2,777 addresses was selected from birth and death registration certificates for the period August 1964 through July 1965. This sample was later supplemented by 191 births and 106 deaths occurring in August and September 1965. The field work was carried out during October and November 1965. Two types of reference periods for vital events were employed with separate questionnaires and portions of the sample. The first covered events occurring during the 12 months prior to interview and the second covered events occurring since January 1, 1965. In addition to requesting the age, sex and marital status of each current member of the household, the questionnaires contained several direct questions concerning births, deaths in general, and infant deaths during the respective reference periods.

The sample data for births known to have occurred during the reference period were analyzed first. The percentages of these events reported in the interviews were approximately the same for both questionnaire versions. The combined results for both questionnaire versions for various segments of the population are shown in Table 1 by legitimacy status of the birth. Although not directly comparable, the estimated percentages of infants enumerated in North Carolina in the 1950 Census (based on the Infant Enumeration Study) are also shown in Table 1.

Except in one group (rural non-white), the proportions of births reported in the current study are less than estimated for the 1950 Census. Since the latter estimates excluded illegitimate infants, the results are not surprising. Under-reporting is seen to be quite severe for illegitimate births, particularly for the urban white group. The completeness of reporting is slightly higher for white births compared to non-white and for rural births compared to urban births. The latter comparison was influenced by the high reporting of illegitimate births by non-white rural households.

Additional analyses indicate that under-reporting tends to increase with age of head of household, but decreases with increasing income and education of the head of household.

Table 1

Births Reported in Household Interviews as a Percentage of Actual Births by Legitimacy Status; Estimated Percent of Infants Enumerated in 1950 Census in North Carolina

Group	RTI Study *			1950 Census **
	Legitimate	Illegitimate	Total	
	%	%	%	%
All births	92.7	84.8	91.6	94.5
White births	93.9	50.4	92.6	95.7
Non-white births	90.3	89.8	90.2	91.7
Urban births	91.5	65.8	88.3	95.0
Rural births	93.4	94.9	93.6	94.3
White urban births	93.6	37.3	91.7	96.2
Non-white urban births	86.1	71.9	81.9	92.1
White rural births	94.1	62.4	93.2	95.5
Non-white rural births	92.2	98.5	94.0	91.6

* Based on interviews in households selected from birth registration lists in 7 counties in North Carolina; 12 months reference period.

** U. S. Bureau of the Census Infant Enumeration Study: 1950, births in first 3 months of 1950 excluding illegitimate births, infant deaths and births in States other than usual residence of mother.

There was no indication that under-reporting increased as the interval between the interview date and the birth date of the infant increased. This is contrary to the Indian National Sample Survey which has noted that the number of births reported decreases as the time interval between month of interview and month of birth increases, a phenomenon referred to in the demography literature as "recall lapse." This could be observed, by the way, if a high proportion of the under-reporting was confined to those births which did not survive the first year of life. In this situation the decrease in reported births would hardly be detectable in countries with a low infant death rate, such as the U. S., but would be magnified in a country with a high infant death rate, such as India.

Table 2, which is based on interviews in households selected from death registration lists in 7 counties in North Carolina, shows the percentages reported of those deaths known to have occurred in the 12 months prior to interview for all deaths, infant deaths and other sub-groups of interest. Under-reporting of deaths was greater than for births. Again, there was no trend in the data by date of occurrence of the event in relation to the interview date. Under-reporting is heavier among non-white households than for white households. Only slightly more than one-half of the infant deaths were reported.

Table 2

Deaths Reported in Household Interviews as a Percentage of Actual Deaths

Group	Percent Reported	
All Deaths	81.6	
White urban		82.4
White rural		85.0
Non-white urban		75.0
Non-white rural		78.2
All Infant Deaths	53.4	
White urban		58.3
White rural		50.0
Non-white urban		45.8
Non-white rural		55.1
All Non-Infant Deaths	84.1	
White		85.1
Non-white		82.1
Urban		82.3
Rural		85.3

A greater degree of under-reporting of deaths occurred among unmarried decedents (27.0% vs. 10.4% for married decedents). In addition, decedents 65 and older were under-reported to a greater extent than those less than 65 (infant deaths excluded). Female decedents were reported to a lesser extent than male decedents, which is probably related to marital status at death and the greater likelihood of a surviving household following a male death than a female death. We have referred previously to the fact that deaths occurring in long-term institutions are less likely to be reported than other deaths.

The proportion of infant deaths reported may be related to the age of the child. For example, 46% of a sample 66 infants who lived less than one day were reported compared to 57% of 97 infants who lived at least one day. When the mother was respondent, two-thirds of the infant deaths (90 in sample) were reported, whereas only 37% were reported (67 in sample) when the respondent was a household member other than the mother of the deceased infant.

Since vital events which occurred prior to the beginning of the reference period were included in the sample, an estimate of over-reporting was obtained. In 168 interviews in households in which it was known that the birth of interest had occurred more than 12 months prior to the interview date, the respondents reported 5% of these births as occurring during the reference period. The percentage in this situation will vary, of course, depending on the proportion of the sample with birth dates near the end point of the reference period. The proportion over-reported was higher for non-white births. It was also somewhat higher for the questionnaire version which had January 1, 1965 as the beginning date of the reference period.

It seems clear from this North Carolina experience that the problem of response error in personal interview surveys designed to measure the population growth rate can be sizeable. The reporting errors in surveys conducted in other areas of the world will, in some instances, be further affected by factors other than those already noted, such as strong preferences for male offspring, superstitious fears associated with reporting the first born, and vague notions about dates and ages. The North Carolina study has indicated, to some extent, the nature and source of response errors in interview surveys designed to measure vital rates in this country. It is expected that the observed under reporting levels can be reduced through improved questionnaire construction and interviewing procedures. The possibilities in this direction remain to be developed and tested.

Tightening up the survey procedures may still be insufficient for providing the accuracy required for the estimates of population growth. In recent years, the increased concern with measurement errors in interview surveys has led to consideration of designs which permit measurement error components to be estimated from the data collected. This is merely an extension of the notion of using the observed data to assess the sampling error in probability samples.

Madow recommends the increased use of double sampling designs, with appropriate allocation of a portion of total survey funds, for reduction or elimination of response bias. We have referred previously to the procedures now in use in the

Pakistan and Thailand population growth estimation projects. These projects combine vital registration data with data collected in periodic household interviews to arrive at adjusted estimates of the birth and death components of population change. There are other examples of survey designs which permit estimation of adjustment factors for response errors in interview data. One such procedure, which is receiving increased attention for surveys of events in human populations, distributes the interviews over time with overlapping reference periods. The U.S. Health Interview Survey, conducted by the National Center for Health Statistics, uses a 12 month reference period for its weekly samples distributed over a year. This is a particularly effective procedure, since it provides the requisite information for adjustment of reported events (such as hospital admissions) for response and procedural errors associated with the elapsed time between the event of interest and the interview date. The Indian National Sample Survey, using annual samples, requests information on births and deaths which occurred in the 24 month period prior to interview, yielding a 12 month overlap for two successive surveys. Other techniques which could be incorporated into the survey design in order to achieve an increase in accuracy through data adjustment include re-interviews with sub-samples, interpenetrating samples and randomized questioning procedures. The latter permits respondent to give answers (sufficient for making unbiased estimates of the proportion in a given class) without revealing their own classification to the interviewer.

A systematic research effort is needed to determine the requisite combination of data generating procedures necessary to achieve relatively unbiased estimates in population growth studies. As in the design of any investigation, whether experimental or survey, the allocation of resources to achieve the specified accuracy should be based on considerations of the relative costs of the alternative measurement procedures and the relative magnitude of the components of error in each procedure.

REFERENCES

Brass, William. "Methods of Obtaining Basic Demographic Measures where Census and Vital Statistics Registration Systems are Lacking or Defective." Paper presented at the World Population Conference, Belgrade, 1965.

Cantrelle, Pierre. "Observation Demographique Repetee En Milieu Rural Au Senegal Methode et Premiers Resultats." Paper presented at the World Population Conference, Belgrade, 1965.

Coale, Ansley J. "The Design of an Experimental Procedure for Obtaining Accurate Vital Statistics." Proceedings of the International Population Conference, New York, 1961, 2, 372-376.

Chandra Sekar, C., and W. Edwards Deming. "On a Method of Estimating Birth and Death Rates and the Extent of Registration." Journal of the American Statistical Association, 44 (1949), 101-115.

Krotki, Karl J. "The Problem of Estimating Vital Rates in Pakistan." Paper presented at the World Population Conference, Belgrade, 1965.

Lauriat, Patience, and Anuri Chintakananda. "Technique to Measure Population Growth: Survey of Population Change in Thailand." Paper presented at the World Population Conference, Belgrade, 1965.

Madow, William G. "On Some Aspects of Response Error Measurement." Proceedings, Social Statistics Section, American Statistical Association, (1965), 182-192.

Sabagh, Georges, and Christopher Scott. "An Evaluation of the Use of Retrospective Questionnaires for Obtaining Vital Data: The Experience of the Moroccan Multi-Purpose Sample Survey of 1961-1963." Paper presented at the World Population Conference, Belgrade, 1965.

Som, Ranjan Kumar. "On Recall Lapse in Demographic Studies." Proceedings of the International Population Conference, Vienna, 1959, 50-61.

Vukovich, G. "The U.A.R. Project for Measuring Vital Rates in Rural Areas." Paper presented at the World Population Conference, Belgrade, 1965.

ACKNOWLEDGEMENT

The participation and assistance of Mr. Glenn Flinchum of the Public Health Statistics Section, North Carolina State Board of Health, Dr. James Abernathy and Dr. Abdel-Latif A. Abul-Ela of the Department of Biostatistics, School of Public Health, University of North Carolina and Mr. Walter Hendricks and Mr. William Grogan of the Research Triangle Institute in the research project referred to in this paper is gratefully acknowledged.

FIELD EXPERIENCE IN ESTIMATING POPULATION GROWTH

Patience Lauriat, Demographic Statistics Advisor
to the Government of Thailand (Bangkok),
U.S. Bureau of the Census*

INTRODUCTION

First, let me define the "universe", if you will, to which the following discussion relates. As a primary limitation, it should be pointed out that the surveys described here are designed to estimate growth rates -- not absolute numbers. The validity of attempting the latter by use of sample surveys has long been debated, but today the emphasis appears to have shifted to concern for making reliable estimates of rates of growth which would then, in turn, be used to develop more reliable estimates of total numbers. Obviously, for the countries with well-developed and accurate vital registration systems, the use of surveys to estimate growth rates is hardly necessary. But, in the developing countries, where so much of the recent work has been done, the point of departure for economists and others concerned with developmental planning is the growth of the population relative to growth in investment, national income, GNP, agricultural production, etc. And, generally speaking, reliable estimates of population growth are not available.

There is a long history of the use of sample surveys to obtain estimates of births and deaths, the major components of population growth for most countries under current immigration and emigration restrictions. However, very frequently the results have been rather poor, partly because the surveys were designed to measure several characteristics, of which fertility and mortality were only two, and partly because special precautions were not taken to assure accurate reporting of births and deaths, particularly prone to under-enumeration.

In an effort to avoid some of the pitfalls of previous studies, there is evolving what appears to be a promising technique whereby data on births and deaths are collected by two separate systems and individual events are compared to determine whether 1) the event was recorded in both systems or 2) recorded in only one. The end result of this matching operation is to obtain more reliable data than those obtained by one system alone. I shall not belabor the theoretical considerations behind this approach[1], as the concern here is to discuss field experience. But I should like to confine the following discussion to experience with the growth estimation studies which utilize creation and matching of two independent lists of events. This second limitation of the "universe" to be discussed is not to imply that other recent studies do not have a great deal of merit and have made some interesting methodological contributions to our growing body of knowledge.[2] Rather, it is dictated by the practical exigencies of time and the personal experience of the author.

There are currently three countries engaged in nationwide, continuing, single-purpose surveys designed to provide reliable estimates of growth rates by matching events recorded in a survey with those recorded in a registration system. They are: Pakistan, Thailand, and Turkey.

The alphabetical sequence of these three countries also reflects the historical progression of the surveys. Pakistan initiated its Population Growth Estimation Study (PGE) in 1962, Thailand its Survey of Population Change (SPC) in July 1964, and Turkey its Demographic Survey (TDS) in the spring of 1965. There are various documents relating to these surveys available[3] but, as yet, none has published a printed report. It is the experience of these three studies which the following discussion draws upon -- and most especially that of Thailand, with which the author is most familiar.

EXPERIENCE IN OBTAINING INDEPENDENT LISTS OF VITAL EVENTS

Survey and sample design

The three surveys are markedly similar in their basic design: all are single-purpose; all are continuing; all are nationwide; all use a fixed, area sample. The principal features of the surveys under discussion are set forth in Chart 1.

Sample size varies, of course, dependent on the level of detail to be obtained. The original Pakistan sample, consisting of 24 areas of about 5,000 people each, was designed to provide estimates for East and West Pakistan separately and for the nation as a whole. There is now some feeling that the sample may be sufficient to produce such rates, but is too small and inadequate to provide additional information useful for interpreting the birth and death rates obtained. A second phase of PGE, starting in April 1966, will experiment with different sizes of sample areas but will not attempt to obtain provincial and national estimates of growth rates.[4]

The Turkey survey started out with the largest sample of the three studies, and even after major cut-backs in the number of sample areas, due to methodological improvements, it remains the largest of the three samples. Since the Turkish study is designed 1) to provide annual estimates for five regions and three cities separately and 2) to measure the effects of the recently insituted Turkish Family Planning Program and rural health improvement program, a large sample is indicated. Originally, there were to be 150 urban and 150 rural sampling units in each of five survey regions and 150 urban units in each of the three largest cities. Because the first results indicated unanticipated response errors, the original enumeration, registration, reconciliation, and verification procedures were drastically overhauled and strengthened, at the expense of sample size. Instead of 150 sampling units in each of the urban and rural parts of the regions and cities, there are to be only 60 in each. Even so, the nationwide sample, when completely enumerated by the spring of 1967, will include about 500,000 population, or about 1.5 percent of the total population.

The Thai survey is not so ambitious. Since, from its inception, one of its secondary aims was to secure a measure of the degree of under-registration of the official system, we decided to select, basically, a two-stage sample of villages, the smallest geographic area identified on the official registration records at the local level. In the first stage, amphurs, or districts, were chosen with probability proportional to population. Within the selected districts, villages were chosen with equal probability, using a constant, over-all sampling rate of 1 in 150. For the municipal areas falling in the sample, blocks were delineated in supplementary field work and were selected using the same sampling ratio as for villages. For practical considerations, Bangkok and environs were excluded from the universe from which the sample was selected but we hope to use official registration records for those areas (where we believe registration is more complete) to adjust our survey data so that it will relate to the whole Kingdom.

The original survey sample in Thailand contained 331 villages and blocks in 60 districts. Nine sample villages in a remote province on the Burmese border had to be abandoned because of inaccessibility. Another three villages were inundated when a dam in the northeast region was completed. At the present time there are about 172,000 people in 29,000 households in the 319 sample areas -- 305 villages in rural areas and 14 blocks in municipal areas.

Our original intent in the Thai SPC was to do some experimental work with the relative merits of different types of sampling units.[5] We planned to depend primarily on a fixed sample of <u>areas</u>, enumerating and recording changes occurring to usual residents of those areas. In addition, we had hoped to work with a fixed sample of <u>persons</u>, recording changes regardless of where they maintained their usual residence. We thought such an experimental undertaking feasible because, from the available data, internal migration in Thailand appeared to be low. With a fixed sample of persons we hoped to eliminate biases arising from the association of vital events with change in usual residence. To this end we initially recorded the name and address of a nearby relative or friend who would know the address of destination if a sample household should move. After two intensive, unsuccessful attempts to obtain sufficiently detailed addresses of either individuals or households which had moved, we had to abandon the idea of a fixed sample of persons and limit ourselves to an area sample only. However, our tabulation plans will distinguish between events to full-time and part-time residents in an attempt to provide some information on the nature of the interrelation between change of residence and vital events.

The three surveys have the same basic approach: periodic enumeration of the sample areas coupled with maintenance of registers of births and deaths. In all three studies, the enumeration is conducted by persons with no responsibility for the registration work, in an attempt to preserve some degree of independence. However, the three surveys differ in intervals between enumerations, in recall periods, in overlap to pick up events possibly missed in a previous round, in defining the population within the purvue of the study, and in how the registers are obtained.

In Turkey, under the new design, a complete enumeration is carried out every six months relating to events over the last year and permanent, resident (local) registrars conduct complete monthly canvasses to record all changes in household composition due to births, deaths, and migration. The Pakistan study provides for quarterly enumerations with a reference period of a year and full-time, resident (local) registrars to obtain lists of births and deaths which occur in the same areas on a <u>de facto</u> basis. The PGE registrars obtain these lists through periodic visits to knowledgeable persons, such as midwives and village officials, as well as visits to about every 10th dwelling in their areas.

In Thailand, quarterly enumerations are carried out relating to the period since the last interview (usually about three months) and recording all changes in household composition. Every six months, vital events recorded in the official registration system are copied at the district offices. However, since late registration does not appear to be a major problem -- in other words, if a person registers an event at all, it is usually done within the required 15 days for births and 24 hours for deaths -- and a "special" matching round has successfully taken place for events recorded in Round 7 only, we are seriously considering copying registration records and matching with survey events every three months. Such a procedure has the decided advantage of permitting the field follow-up of registered-but-unsurveyed cases to be carried out more promptly than when copying and matching take place only once every two rounds of enumeration.

One element of the Thai SPC design deserves special note. From the inception of the study, its planners were greatly concerned about the effect of continual, quarterly enumerations on registration in sample areas. In order to get an accurate measure of under-registration, we felt it was vital to obtain some measure of the possible "conditioning effect" which the interview program might have on registration. That is, would the fact that an enumerator came around every three months asking questions about changes in household composition stimulate the sample population to become more aware of such changes and, therefore, incline them to register their births and deaths more often than if the interview program were not undertaken? In order to measure this possible source of bias, we drew a control sample of villages at the same time that we selected the original 331 sample areas to be enumerated quarterly, using the same sample selection factor. No interview work was done in the control sample areas, but our intention was to compare the registration totals for the control sample areas with those for the survey sample areas. If there were significant differences between the counts of events for the two sets of areas, the estimates of under-registration would have to be adjusted accordingly. Preliminary data from the first such comparison we have made, relating to the first survey year, indicate that no bias has been introduced by the interview program through "conditioning" the survey population to register their vital events more frequently than the population in non-interview areas.

It may also be of interest to mention one other feature of the Thailand SPC before getting into enumeration problems as such. We included questions, to be asked each round, as to whether 1) any woman in the household were known to be pregnant, and, if so, the number of months, and 2) any persons were known to be seriously ill and the nature of their illness. These inquiries were designed to ensure reporting of all births and deaths, a positive answer requiring the enumerator to make appropriate additional checks in subsequent rounds. Only about 70 percent of the births reported in the survey during the first year occurred to women having been previously reported as pregnant, and most of them had been reported pregnant only in the interview immediately preceding birth. The results for deaths during the first survey year were even lower -- only 4 percent of the persons who died had been previously reported as seriously ill. Further information is needed before we can determine that these questions do enable enumerators to report births and deaths they might otherwise have missed.

Enumeration problems

The troublesome areas with which the growth surveys in Pakistan, Thailand, and Turkey have had to contend are not necessarily unique to them. Probably, in large part, the enumeration difficulties and weaknesses are those facing any study carried out in a developing country where physical accessibility, means of communications, organizational structure, and control procedures are lacking.

One of the most basic problems is that of identifying the enumeration units. In Pakistan one of the essential duties of the PGE interviewers is to assign a house number to each dwelling in the sample areas. Such a procedure, in and of itself, is not a startling survey technique, but when you read the enumerator's instruction manual and find that "some 50 spare placards of galvinized iron sheet, black paint, two brushes, hammer and nails" are part of the enumerator's everyday work equipment -- well, one would usually concede, that is a little startling! As in some other eastern countries, there is no house-numbering scheme in Pakistan and so, in order to assure that the correct sample areas were identified and enumerated, the first job was to nail house numbers above the front door.

In Thailand, on the other hand, dwellings are assigned numbers by local administrative officials, but this system has its hazards too. Very frequently there is no particular order to the assigned numbers and quite constant renumbering goes on. Or, one often observes duplicate, and even triplicate, numbers because the remnants of several numbering systems are still in existence and used by the inhabitants.

For all three growth surveys, the basic enumeration schedule is simple and straightforward. After the usual identification items, it consists of a list of household members, with their relationship to the head, sex, age, and marital status. The Turkish Survey includes questions on age at first marriage and education. The PGE instrument has questions on caste and occupation. In the SPC we are asking no questions about characteristics beyond the basic demographic ones.

All three surveys maintain simultaneous sets of records in the field, for use in continuing interview and registration systems, and in the central processing office, to provide input for periodic hand and machine tabulations. All survey materials are generated manually up to the punch cards used as tabulation input. Obviously, as soon as a survey becomes dependent on several manual operations, sources of error are increased at each level. In the Thai SPC, for example, by 100 percent verification of all clerical operations and various checks and controls built into the survey operations, we feel we keep such manual errors down to a minimum. Even if an error goes undetected in verification, the system is circular enough that we will probably detect it in a subsequent processing step.

The most impressive feature of the three growth surveys underway is the elaborate system of controls which each has had to develop and without which none would enjoy whatever success it does. We all must admit that searching for a birth -- let alone a death -- in any population, regardless of size, has many elements of looking for a needle in a hay-stack! To ensure that all events are found and that found events are true occurrences taxes the leadership and imagination of all the technicians involved in the current growth studies. Shortly after we initiated our Thai SPC, we attended a seminar in Karachi to exchange experiences with PGE technicians first-hand. All of us from Thailand were very impressed with the elaborate procedures which had been developed there to ensure reliable data -- and promptly returned to Bangkok to establish our own!

When we started the SPC in July 1964, we hired nine temporary supervisors (graduates, usually in the liberal arts, of one of the local universities in Bangkok) and 60 part-time, temporary enumerators (usually 8th grade graduates referred by local government officials). Because of the lure of permanent appointments in other government agencies and because of the hardships involved in being in travel status outside Bangkok for extended periods, the turn-over rate of the supervisors became a very serious problem. By the end of the first year, only one of the original nine was still working on the SPC and we had had to hire an additional nine to keep the positions filled. The turn-over rate among the enumerators was also high -- about two-thirds were replaced during the first year.

An official administrative reorganization of the National Statistical Office of Thailand was put into effect coincidental to the 7th round of SPC enumeration (January through March 1966).

Provision was made for the establishment of 69 Provincial Statistical Offices, within the framework of the already-extant 12 Regional Offices. Each of the Provincial Offices is staffed by three or four full-time, permanent government employees (one or two supervisors and two or three enumerators) and is responsible for all the survey and census field work undertaken by the National Statistical Office in the province. The SPC supervisors and enumerators whom we wished to retain were given permanent positions in the Provincial Offices. The net effect on the SPC field staff of this organizational realignment was to increase the number of supervisors to 41 (the number of provinces in which the sample districts are located) and retain 59 enumerators, one for each district. Although these permanent supervisors and enumerators now work on more than just the SPC, this does not decrease the total time devoted to SPC field work because previously it was carried out every twelve weeks for six weeks of full-time work.

Apart from the enhanced prestige attendant on permanent civil service appointments for enumerators, morale was further increased by establishment of a far more realistic pay structure. Up until that time, enumerators had received a basic wage of $1.00 per day, including costs of transportation, lodging, and meals. Under the reorganization their basic wage was supplemented with per diem and transportation costs, bringing the total to $1.75 or $2.00 per day. Similar provisions were made for reimbursing supervisors for costs connected with carrying out official business.

The impact of these two reforms -- 1) a greatly increased ratio of supervisors to enumerators (from 9 to 59 to 41 to 59) and 2) sufficient remuneration so that our survey workers were not "out of pocket" in order to do their jobs properly -- was immediate and dramatic. Table 1 presents the number of events reported in each quarterly canvass for the first seven rounds of enumeration. It should be noted that whereas there were about 1,500 births and 400 deaths reported in each round for the second through the fifth, the numbers of events increased to 1,670 births and 511 deaths in Round 6 and, for births, increased even more sharply in Round 7 -- to 2,078. The Round 6 increases may be partially explained by the fact that for that round, we were able to employ an additional six supervisors, thus increasing the original 9 to 15, since it had become apparent within the first year of the survey that we had seriously overestimated the number of enumerators that could be effectively handled by one supervisor. After many discussions with supervisors and enumerators alike, we have concluded that the high numbers of events in Round 7 are the direct result of tightened supervisory controls and adequate payment for services rendered. Where previously the enumerator did not re-canvass every village in every round, and there was not sufficient time for our short-handed supervisors to visit every enumerator sometime in the course of each enumeration round, we now have an adequate number of supervisors to assure that the enumerators do go to <u>all</u> sample villages and, once there, <u>do</u> find more events, as well as increased numbers of in- and out-migrants. If the momentum seemingly set in motion in Round 7 continues to yield higher levels of events than were recorded in the first six rounds, careful analysis of the data obtained in the first rounds will have to be undertaken before any firm conclusions can be drawn regarding growth rates.

As a last point in this discussion of enumeration problems, mention should be made of one of the basic concerns of a demographer -- the <u>de jure</u> vs. <u>de facto</u> approach. In the three surveys under discussion, only the Pakistan PGE has taken the <u>de facto</u> approach; the Thailand SPC and the Turkish TDS utilize the <u>de jure</u> approach. Each has its own particular problems. Mauldin, in discussing the Pakistan PGE, outlined the salient difficulties and concluded that it is quite feasible to obtain vital events on both a <u>de facto</u> and a <u>de jure</u> basis.[6] This is the approach which, with modification, has been taken in the Thai SPC. However, it too has its pitfalls.

The <u>de jure</u> approach assumes that the concept of "usual residence" has a functional meaning to the respondent. In Thailand, this may or may not be a realistic assumption. The law provides for the maintenance of a Household Register and all population movement between households theoretically is to be reported to the proper local authorities. Furthermore, there is a system of individual Identification Cards, based on proper recordation in the Household Register. But there appears to be a good deal of shifting from one village to another that does not necessarily get recorded in the Household Registers and which, to the mover, is not necessarily a permanent, real "move".

As a consequence, SPC established an elaborate series of conditions to determine who was a usual resident of a sample household vs. those persons remaining there only temporarily. The latter were defined as visitors but, if found to be still there after six months, were redefined as usual residents. In the reverse case, persons reported as usual residents who were not present at the time of subsequent interviews were, after three rounds, deleted from the resident population if they had been absent for more than six months and <u>working</u> elsewhere, even if reported as intending to return. (Absent residents reported as <u>studying</u> elsewhere were retained as household members, regardless of length of absence.)

We have not attempted to compare our lists of household members with the Household Registers for our sample households. Nor have we conducted any intensive interviews to test the real meaning of "usual resident" and determine whether or not a respondent tends to answer in terms of persons registered as residing in the household or in terms of persons actually living there. It might be instructive if we did, but the comparisons we have made between SPC counts of households and counts reported by village headmen indicate that probably the <u>net</u> effect on the population-at-risk of an event would be neglibible. In a tropical country where living arrangements are fairly casual, basic food, shelter and clothing needs minimal, the extended-family system prevalent, and a very distinct seasonal work-pattern established by the major agricultural crop (rice), "usual residence" may be an academic nicety that does not adequately reflect reality.

The number of events reported as occurring to visitors is very small, and we are not planning to include them in our calculations of rates. Similarly, although we record the comings and goings of visitors in and out of the sample villages, they are excluded from our tabulations of the population-at-risk. Rather, our main concerns are coverage of events to residents, which may occur away from the place of usual residence or in households which move out before the next interview.

Registration problems

The second major piece of field work which the three population growth estimation studies under consideration carry out involves development of independent lists of events reported by a registration system. The Pakistan and Turkey study designs include the establishment of carefully controlled and executed registration systems for the sample areas covered by the studies. In Thailand, official registration records are utilized.

Regardless of the source of the registration records, the key issue here is the independence of the registration system from the enumeration phase. Hashmi has outlined the procedures incorporated into the Pakistan PGE to assure that collusion between enumerator and registrar was avoided or kept to a minimum:

> "Attempts were made to maintain the independence of the two systems in the experiment. Registrars were not allowed to keep the records of births and deaths for more than a couple of months. On the other hand, enumerators were collecting births and deaths which had occurred during the past twelve months. This minimized the scope of copying. The primary focus of the cross-sectional survey was to collect data on population size, household composition, age, marital status, occupation, caste and pregnancy. Obtaining information on births and deaths was at the tail-end of the schedule used and was a small part of the total survey. Since enumerators had to visit each dwelling to obtain most of the information, they would not turn to the registrars to copy a small bit of information on births and deaths. On the basis of our matching experience, and on-the-spot checks, we feel that there was little collaboration between registrars and enumerators."[7]

Seltzer and Hashmi, in evaluating the PGE mortality data, upon which Provincial Pakistan life tables for 1962 were based, raise the really critical issue of bias introduced because enumerated events will more likely be registered, and vice versa:

> ". . . Lack of independence can arise either because the enumerator and registrar overtly collaborated in the preparation of their reports or because the same types of deaths are likely to be omitted by both systems. Because of the possibility of overt collaboration, steps were taken from the very beginning of PGE to ensure that the work of the registrar and enumerator was as dissimilar as possible. . . .Lack of independence of the two systems due to other (non-intentional) sources remains to be properly evaluated."[8]

The Turkish experience in obtaining registration lists may be of particular interest. Whereas the original study design called for continuous registration in only a fifth of the sample areas and a comparison of distributions based on enumeration and sub-sample registration results, the revised procedure includes monthly enumerations in all sample areas by permanent, resident registrars and a case-by-case comparison with events recorded in the semi-annual, census-survey conducted by independent enumerators. Fisek, et al., include a recital of the findings which prompted them to make this radical change in their survey design:

> ". . . Shortly after the initial enumerations were begun, it became apparent that the number of enumerated vital events was short of expectations. For example, in one of the first series of rural enumerations, the Crude Birth Rate was reported as 45 per 1000 persons. The Crude Death Rate was 11 per 1000 and the Infant Mortality Rate was 122 per 1000 live births. Because of the questionable magnitude of these rates, particularly the mortality experience, a small enumeration was held in several sampling units in the same region in which vital registrars were previously established. When the registrars' records and the enumerators' reports were subsequently reconciled and verified, it was found that the Crude Birth Rate was nearer 46 per 1000, the Crude Death Rate 18 per 1000 and the Infant Mortality Rate 140 per 1000."[9]

While Pakistan and Turkey face the problems of assuring complete coverage within their specially-designed registration systems, Thailand's SPC takes a somewhat different approach. Since one of the secondary aims of the study is to measure the extent of under-registration of births and deaths in the official system, it was decided to use official registration records as the independent registration list, to be matched to events reported in the quarterly surveys. This approach was partly dictated by the difficulties of maintaining adequate controls over resident registrars in 319 areas spread throughout the Kingdom and by the fact that a direct measure of under-registration was needed before serious consideration could be given to improving the established registration system.[10]

This procedure of using official vital registration records is probably the greatest weakness of the Thai SPC, now that enumeration has been significantly improved. The clerks sent to the local district administrative offices have experienced great difficulties in finding the registration books for the communes in which the sample villages are located. In all district offices, maintenance of vital registration records is only one of many functions, and is often relegated to the end of the priority list of tasks to be accomplished by an inadequate staff, and may even be put on a "voluntary over-time" basis. Receipt of the registration books from the communes, where the registration actually takes place, is usually sporadic and there are no controls over the receipt (i.e., by date and/ or certificate number, etc.) nor any attempt made, in most district offices, for proper filing of books received and awaiting posting in the combined volume for the district.

Earlier this year (April to May), we conducted a special study in one-half of the sample villages regarding knowledge about, attitudes toward, and practice of (a "KAP" study) birth and death registration. The events covered were those occurring during calendar year 1965. Without indicating to the special study enumerator whether or not the office records showed an event to be registered, a direct question was asked of the respondent as to whether or not the event had been registered. If the respondent indicated that it had been, the enumerator asked to see the certificate and, in cases where it was available, recorded the number and date of registration. It came as something of a shock to find, from a special hand tally of cases we had originally classified as unregistered, that only 12 percent of the births and 7 percent of the deaths were admitted by respondents to be unregistered! And, for those reported as registered, 61 percent of the respondents produced birth certificates and 26 percent produced death certificates. In examining the reasons given by respondents who said the death had been registered but who did not have a certificate available, many indicated they had had to give it to the temple where the body is buried before its final cremation. Therefore, the proportion of deaths actually registered among those reporting them as registered is, in all probability, higher than the 26 percent indicated above.

As a consequence of this unexpected finding of our special "KAP" study of registration, we are instituting several measures, as of the current round of copying, in order to obtain more complete registration records. We are soliciting the cooperation of the Ministry of Interior, which has the direct responsibility for vital registration, to ask the appropriate officials in the 59 districts covered by the sample to keep the records for the communes which contain our sample villages separate and in good order. We are instructing the copying clerks to check serial numbers and dates from the end of one book to the beginning of the next for a commune. These and other measures, which we are still in process of developing, should, we hope, go far to strengthen what has now appeared as a bisic weakness in our survey procedures.

Matching experience

It would hardly be fair to leave a discussion of the experience of these population growth estimation studies without at least a brief discussion of there experience with matching the two independent lists of events so arduously and painstakingly obtained. And, since the nonmatches require some form of field reconciliation, such discussion seems within the purview of this paper.

To paraphrase Stein's immortal words -- "A match is a match is a match"! Were the problem so easy of definition, a great many of the headaches inherent in the current growth estimation studies would be eliminated. Mauldin has succinctly stated the basic considerations which any matching scheme must take into account:

> "Problems of matching are very difficult. There is no adequately developed theory for determining what is a 'match', and procedures developed to date in matching studies are not fully objective. If the matching criteria are too rigid (exacting), an event which has been picked up by both the registrar and by the survey will sometimes be counted as two events because the entries about this event differ slightly. If, however, matching criteria are too loose, different events may be considered as the same event. In general, if matching criteria are too rigid, one gets an inflated estimate of the total number of events, and if the criteria for matching are too loose, one gets an underestimate of the number of events."[11]

Since the Turkish TDS has only recently been redesigned, details of its matching definition and operations were not available to the author at the time of this writing. So the following discussion is necessarily limited to the Pakistan and Thailand experience.

The Pakistan PGE faces a rather unique problem in that new-born babies are not given names until they are some months old. Ergo, a simple name match for births is impossible. Rather, an elaborate codification of parents' names, as well as infants' where available, is done manually; cards with coded names and other requisite information punched; and machine listings of events arranged in sequence by name of mother, name of father, name of infant, date of birth, place of birth, and sex of the child obtained. All "apparent matches" (from three different sequential listings) are verified manually. All "possible matches" are divided, on the basis of manual inspection, into three types: 1) matches, 2) doubtful matches, and 3) nonmatches. Doubtful matches are referred to the field for additional follow-up by "third-party" enumerators.

As Seltzer and Hashmi have stated:

> ". . . While the matching procedure is subject to rigorous controls, including a reinvestigation in the field of many of the non-matching reports, errors do, of course, occur. They may be broadly classified into two types: false matches and false nonmatches. . . . No information is available to measure the relative frequency or the absolute effect of either type of matching error."[12]

In the Thailand SPC, the basis for matching is rather more straight-forward because it is, essentially, a manual match based on name of birth or death, names of parents and age of mother in the case of births, sex of birth or death, place of occurrence of the event, and place of residence of the parents or the deceased. Using the address given in the lists of registered events for each sample area, the clerks search the list of household members for the specified house number. If the house number cannot be found in the survey records, other households with heads having the same last name are identified from the house-listing sheets and residents of those households searched. If the name of the event, the names of the parents, or the registration address cannot be found, the case is classified as "non-surveyed - (preliminary)" and remanded to the field for follow-up. After all events on the registration lists have been processed, the sample area is searched for events reported in the survey but not included on the registration lists. On the assumption that all cases found in the quarterly surveys represent "true" cases, no field follow-up of surveyed-but-unregistered cases has been undertaken to date. In light of some of the enumeration and registration problems discussed above, we are planning to follow-up at least a sub-sample of some of these cases to determine that they are, in fact, cases within the defined scope of the survey.

Table 2 presents data available from the Pakistan, Thailand, and Turkish studies. Detailed comparisons between the three studies may not be indicated because of incomparabilities between the data, but some salient points should be noted:

1) In all three studies, the match rate for deaths is significantly less than for births, although the differences in the rates for births and deaths for both wings of Pakistan are much smaller than those for Thailand or Turkey. A lower match rate for deaths is probably indicative of the same factors that make reliable measurements of mortality difficult -- that death is a disruptive factor in the "life" of a household or reporting unit and is an unpleasant event which respondents may be reluctant to report. Also, in areas of high infant mortality, where deaths under one year of age account for large proportions of all deaths, non-reporting of infant deaths may be more frequent than in areas of low mortality.

2) As one might expect on methodological grounds, the original Thai SPC survey events are a consistently higher proportion of total events than the registered events. However, when the original numbers of unregistered events shown in Table 2 are adjusted to reflect the proportions of respondents in the special study who had registration certificates (60 and 30 percent, respectively, of the "unregistered" births and deaths investigated), the birth coverage of the survey is only slightly better than that of the official registration.

For deaths, on the other hand, based on either the original or adjusted figures, the survey includes a consistently higher proportion of events than registration -- up to 91 percent for the period July to December 1965. In Pakistan, on the other hand, the registration system appears to yield higher proportions of events than the survey. Apart from the fact that the Thai SPC is tied-in to the official registration records, it should also be remembered that the Pakistan survey, although conducted quarterly, has a recall period of one year, and only the fourth quarter survey results for 1962 were used to obtain the matching data shown in Table 2. The SPC conducts quarterly surveys, but the reference period is "since the last interview" (an average of about three months).[13]

3) The Turkish data shown in Table 2 are included to illustrate the kinds of "growing pains" to which all these current population growth estimation studies are subject. It was exactly on the basis of the figures in Table 2 that the major changes described above were made in the Turkish TDS design:

> "The results of this matching study /In 15 sampling units/, demonstrate that both the enumerators and registrars were individually missing vital events. Over 90% of enumerator errors were caused by under-reporting on the part of the respondents. Registrar errors were largely due to the mistaken conviction, on the part of the registrar, that they were thoroughly familiar with all the events that occurred in their village. . . . In addition /to a monthly canvass by registrars in all sampling units/, a semi-annual census-survey, with an overlapping recall period, will be conducted by central staff supervisors. Reconciliation and field verification of vital events reported on each form will be performed by an independent set of supervisors."[14]

As a final note on the actual field experience which we have been accumulating with these population growth estimation studies, I would like to draw your attention to Table 3. Lacking comparable data from the other studies, it presents only Thailand SPC data on the results of the field checks which were made of all registered-but-unsurveyed cases identified in the first three rounds of matching operations, covering events occurring between October 1964 and December 1965. In the case of Matching Round 1, the field follow-up was done between 3 and 9 months after occurrence of the event; for Matching Rounds 2 and 3, each of which covered six months of events, field follow-up was carried out 3 to 12 months after occurrence.

The improvement in the basic survey enumeration is indicated by the fact that only 52 percent of the births checked in Matching Round 3 should have been recorded by the survey and only 47 percent of the deaths in the same round. Although some may feel these proportions to be too high, they represent a distinct improvement over the proportions missed in Matching Round 2, when the enumeration (Rounds 3 and 4) obviously deteriorated -- 60 percent of the births and 56 percent of the deaths sent out for reconciliation in the field had not been properly recorded in the survey enumeration.

It is interesting to note the source of these errors. In the case of births, the major cause of missed events was due to enumerator failure to record the event itself -- 65 percent of the error in Matching Round 1, 77 percent and 83 percent in Matching Round 2 and 3, respectively. For deaths, on the other hand, although missed events were a major source of error, comprising 43 percent, 68 percent, and 65 percent, respectively, in the three Matching Rounds under discussion, the comparative contribution of events occurring to missed population was somewhat greater than amongst the births investigated.

CONCLUSION

This paper has attempted to describe some aspects of implementing the population growth estimation studies now being carried out in Pakistan, Thailand, and Turkey. The fact that the fore-going discussion is largely discursive, and not predominantly a detailed and sophisticated demographic and/or statistical analysis, is perhaps indicative of the state of development of these growth studies. They are currently in the position of "becoming". We are still improvising and revising procedures designed to refine and improve the results we are getting. The preliminary results are under constant scrutiny in an effort to identify and correct yet other weaknesses in procedures and methodology. Intensive analysis of the final results, now conspicuously absent, will be required to assure ourselves that the answers we obtain are valid and reliable.

The application of sample survey techniques to measure population growth is the extension of tested procedures to a new subject area. It has its own particular set of hazards. But, because the work under discussion here is being carried out in yet-to-be-developed countries, these growth studies must also contend with problems and issues, such as communications and trained personnel, that do not occur to as great an extent in developed countries. The burden of proving these studies to be reliable lies heavy and taxes all the resources and intelligence being devoted to them. The challenge is there and, with time, it appears that it will be met.

That we are on solid methodological ground is borne out by the three studies discussed above. We are demonstrating that sampling techniques can be applied to obtain accurate estimates of growth rates, that two independently derived lists of vital events do provide more reliable data than one system alone, that these techniques can be successfully applied in countries where vital registration is grossly inaccurate, or nonexistent. That the implementation process poses difficulties has been the burden of this discussion.

The opinions expressed in this paper are those of the author and do not necessarily represent those of the organization of which she is an employee, nor of the National Statistical Office of Thailand, with which she has been closely associated for over two years. The author is indebted to several of her Thai and American colleagues for their cooperation in providing information presented here and for their helpful criticisms of this presentation. However, she accepts full responsibility for the data and interpretations expressed herein.

[1] *See Chandrasekaran, C. and Deming, W.E., "On a Method for Estimating Birth and Death Rates and the Extent of Registration," Journal of the American Statistical Association, Vol. 44, No. 245, March 1949, pp. 101-115 and Coale, A.J., "The Design of an Experimental Procedure for Obtaining Accurate Vital Statistics," Proceedings of the International Population Conference, New York, 1961, Vol. II, pp. 372-375.*

[2] *See Cantrelle, P., "Repeated Demographic Observation in a Rural Area in Senegal: Method and First Results," paper presented to the World Population Conference, Belgrade, 1965, Doc. No. B.6/V/F/207; Majumdar, M., "Estimation of Vital Rates in the Indian National Sample Survey," paper presented to the World Population Conference, Belgrade, 1965, Doc. No. B.6/I/E/312; Sabagh, G. and Scott, C., "An Evaluation of the Use of Retrospective Questionnaires for Obtaining Vital Data: The Experience of the Moroccan Multi-Purpose Sample Survey of 1961-63," paper presented to the World Population Conference, Belgrade, 1965, Doc. No. B.6/V/E/56; United Nations, Guanabara: Demographic Pilot Survey (Population Studies No. 35), New York, 1964;-----, The Mysore Population Study (Population Studies No. 34), New York, 1961; and Vukovich, G., "The U. A. R. Project for Measuring Vital Rates in Rural Areas," paper presented to the World Population Conference, Belgrade, 1965, Doc. No. B.6/I/E/68.*

[3] *Pakistan: Ahmed, N. and Krotki, K.J., "Simultaneous Estimations of Population Growth: The Pakistan Experiment," The Pakistan Development Review, Vol. III, No. 1, Spring 1963, pp. 37-65; -----, "Second Report on the Population Growth Estimation Experiment," paper presented to the Asian Population Conference, New Delhi, 1963; Hashmi, S.S., "Estimating Vital Rates in Developing Countries with Special Reference to Pakistan," paper prepared for the Regional Cooperative Development Seminar on Family Planning, Karchi, 1966.*

Thailand: Lauriat, P. and Chintakananda, A., "Technique to Measure Population Growth: Survey of Population Change in Thailand," paper presented to the World Population Conference, Belgrade, 1965, Doc. No. B.6/V/E/507; National Statistical Office, Office of the Prime Minister, Government of Thailand, "Survey of Population Change: Study Outline," July 1964 (mimeographed).

Turkey: Fisek, N.H., Heperkan, Y. and Rumford, J., "The Role of the Turkish Demographic Survey in the Family Planning and Rural Health Programs," July 1965 (offset); -----, "The Evolution of the Turkish Demographic Survey," (unpublished manuscript).

[4] *Hashmi, op cit., pp. 15-16.*

[5] *National Statistical Office, op cit., pp. 7-8.*

[6] *Mauldin, W.P., "Estimating Rates of Population Growth," paper presented to the International Conference on Family Planning Programs, Geneva, 1965, p. 15.*

[7] *Hashmi, op cit., p. 14.*

[8] *Seltzer, W. and Hashmi, S.S., "A Note on the Limitations of Population Growth Estimation Data Used for Life Table Construction," (Appendix A in Aslam, N. and Hashmi, S.S., "Abridged Life Tables of Pakistan and Provinces by Sex, 1962," Research Report No. 48), Pakistan Institute of Development Economics), March 1966 (mimeographed), pp. 30-31.*

[9] *Fisek, et al., The Evolution . . ., p. 2.*

[10] *Lauriat and Chintakananda, op cit., p. 4.*

[11] *Mauldin, op cit., p. 14.*

[12] *Seltzer and Hashmi, op cit., p. 30.*

[13] *Discussion of estimating the numbers of events missed by both systems was deliberately excluded here on the basis that this presentation is confined to actual field experiences with growth estimation studies. The exclusion is not meant to imply that such an estimate is not an important element in any of these studies. On the contrary, it is an area of vital concern and it is to be hoped that final data from the studies under discussion here will yield new insights into methods of reliable estimation (see Chandrasekaran and Deming, op cit.).*

[14] *Rumford, J., "Progress Report, Month of October 1965," (Memorandum to M.J. Lieberman), November 3, 1965 (typed), pp. 1-2.*

Chart I - Principal Features of Surveys to Measure Population Growth in Pakistan, Thailand and Turkey

Item	Country		
	Pakistan	Thailand	Turkey
1. Name of survey.................	Population Growth Estimation Experiment (PGE).	Survey of Population Change (SPC).	Turkish Demographic Survey (TDS).
2. Start of enumeration...........	January 1962.	July 1964.	September 1965.[1]
3. Separate estimates for--.......	East and West Pakistan.	Whole Kingdom only.	Urban and rural for 5 regions, 3 largest cities.
4. Approximate sample size (number of persons)...........	120,000	170,000	450,000
5. Sample design..................	24 compact sample areas, averaging 5,000 population.	60 districts, subsample averaging 5 to 6 whole villages per district.	30 districts, 2 whole villages per district in rural part of each region; 60 blocks in urban part of each region; 120 blocks in total for 3 largest cities.
6. Type of registration system used..........................	Local survey registrars make "rounds" continuously to identify and record vital events.	Official birth and death registration.	Local survey "registrars" visit all households monthly.
7. Frequency and reference period of survey enumeration..	Quarterly, covers all events in preceding 12 months (overlap).	Quarterly, covers all events since previous enumeration.	Semi-annual, covers all events in preceding 12 months.[2]
8. In-scope events................	Events <u>occurring</u> in sample areas.	Events for <u>residents</u> of sample areas.	Events for <u>residents</u> of sample areas.
9. Frequency of matching and reconciliation................	Annual	Semi-annual[3]	Semi-annual

[1]Start of enumeration in Regions I and II. Other regions and cities are being introduced on a staggered schedule. Full coverage of Turkey scheduled for May 1967.
[2]First semi-annual enumeration in Regions I and II covered only the preceding 6 months.
[3]First and fourth matching rounds covered events for a single quarter.

TABLE 1 -- Sample Counts, Survey of Population Change (Thailand): Enumeration Rounds 1 to 7[1]

Item	Round 1 (July-Sept. 1964)	Round 2 (Oct.-Dec. 1964)	Round 3 (Jan.-Mar. 1965)	Round 4 (Apr.-June 1965)	Round 5 (July-Sept. 1965)	Round 6 (Oct.-Dec. 1965)	Round 7 (Jan.-Mar. 1966)
Beginning population.........	-	169,458	171,460	172,036	171,148	172,219	173,789
Changes:							
Births....................		1,460	1,579	1,539	1,487	1,670	2,078
In-migrants...............		3,072	2,382	3,692	2,775	2,896	5,546
Deaths....................		423	360	367	398	511	524
Out-migrants..............		2,107	3,025	[3]5,752	3,244	3,253	5,513
Other[2]....................					+451	+768	+1,814
Ending population............	169,458	171,460	172,036	171,148	172,219	173,789	177,190

[1]Usual residents only.
[2]Other includes persons missed in previous rounds as well as a small number of visitors reclassified as usual residents when they remained in the sample household for more than six months.
[3]Includes all persons in three villages inundated by a new dam.

Source: SPC Office Records.

TABLE 2 -- Events by match status: Pakistan, Thailand, and Turkish Growth Estimation Studies

Class of Event	Pakistan - 1962[1]		Thailand - Specified rounds[2]						Turkey - Pretests, 1965[3]	
	East	West	Oct.-Dec. 1964		Jan.-June 1965		July-Dec. 1965		10 villages 3 blocks	2 villages
			Original	Adjusted	Original	Adjusted	Original	Adjusted		
BIRTHS										
Total	2,469	2,140	1,732	1,732	3,704	3,704	3,385	3,385	316	42
Matched	1,678	1,306	1,045	1,307	2,330	2,901	2,111	2,743	232	22
Not Surveyed	394	521	220	220	406	406	198	198	29	16
Not Registered	397	313	437	175	951	380	1,053	421	55	4
Undetermined	-	-	30	30	17	17	23	23	-	-
Total Registered	2,072	1,827	1,265	1,527	2,736	3,307	2,309	2,941	261	38
Total Surveyed	2,075	1,619	1,482	1,482	3,281	3,281	3,164	3,164	287	26
Match rate	68%	61%	60%	75%	63%	78%	62%	81%	73%	52%
% reg. events of total	84%	85%	73%	88%	74%	89%	68%	87%	83%	90%
% surv. events of total	84%	76%	86%	86%	89%	89%	93%	95%	91%	62%
DEATHS										
Total	813	787	490	490	833	833	960	960	114	21
Matched	498	434	209	370	350	455	363	515	65	7
Not surveyed	206	267	63	63	117	117	88	88	23	10
Not registered	109	86	202	141	351	246	506	354	26	4
Undetermined	-	-	16	16	15	15	3	3	-	-
Total registered	704	701	272	333	467	572	451	603	88	17
Total surveyed	607	520	411	411	701	701	869	869	91	11
Match rate	61%	55%	43%	55%	42%	55%	38%	54%	57%	33%
% reg. events of total	87%	89%	56%	68%	56%	69%	47%	63%	77%	81%
% surv. events of total	75%	66%	84%	84%	84%	84%	90%	91%	80%	52%

[1]Figures relate to 8 areas in each wing where both registration and enumeration were carried out. For purposes of comparability with the Thai and Turkish figures, estimates of events missed by both registration and enumeration are not shown here. Source: "PGE Interim Report No. 5," Pakistan Institute of Development Economics, August 11, 1964, pp. 8-10.

[2]The undetermined cases shown here are those which occurred outside the SPC sample areas to residents of the sample areas but for which no special search of registration records could be made. The adjusted data assume that 60 percent of the unregistered births and 30 percent of similar deaths were, in fact, registered,(see discussion regarding "KAP Study" of registration). Source: SPC Office Records.

[3]Data in the first column are for 13 sampling units where a vital registration system had been in operation for 8 months prior to October 1965 when the matching pretest was carried out. Data for the second column are for two villages in the Kazan Health District where midwives' records for one year were used as registration lists. Source: Memorandum Rumford to Lieberman, "Progress Report, Month of October, 1965," dated November 3, 1965, Appendix Table.

TABLE 3 -- Results of field follow-up of events registered but unsurveyed, Survey of Population Change (Thailand): Matching Rounds 1 to 3

Item	Matching Round[1]		
	1	2	3
BIRTHS			
Total sent to field	561	1,233	936
Found to be outside sample areas[2]	272	495	448
Total missed by survey[3]	289	738	488
In missed households	66	74	45
Population to whom event occurred not enumerated	33	98	29
Event not enumerated	190	566	414
Percent of total missed by survey	52%	60%	52%
DEATHS			
Total sent to field	141	314	276
Found to be outside sample areas[2]	66	137	145
Total missed by survey[3]	75	177	131
In missed households	9	15	7
Population to whom event occurred not enumerated	34	40	39
Event not enumerated	32	122	85
Percent of total missed by survey	53%	56%	47%

[1]Matching Round 1 covers events occurring in Enumeration Round 2 (October-December 1964) and followed up in April-June 1965; Matching Round 2 covers events occurring in Enumeration Rounds 3 and 4 (January-June 1965) and followed up in October-December 1965; Matching Round 3 covers events occurring in Enumeration Rounds 5 and 6 (July-December 1965) and followed up in April-June 1966.

[2]Includes small number of events found to be out of scope and events for which a followup interview could not be conducted.

[3]These figures are higher than "Not surveyed" as shown in Table 2 because they include "late survey" and "late registration" events form previous rounds as well as events found in missed households.

Source: SPC Office Records.

USCOMM--DC

Discussion

W. Parker Mauldin, The Population Council

Introduction: The studies reported on by the Lauriat-Jabine team in Pakistan, Thailand, and Turkey are large in scale, serious in effort, important in purpose to economic planners, health personnel, and other government officials. Each of them applies the Chandrasekaran-Deming formula and procedures which were presented to us in 1949, and then lay dormant for more than 15 years. Such procedures are relatively expensive in terms of time required to collect and analyze the data, expensive in terms of funds, and, more important, expensive in terms of skilled manpower required. Those involved with these studies have reluctantly concluded that the present state of our knowledge of surveys is such that reliance cannot be placed on vital events collected only by retrospective surveys. This is the major issue of this session: Can valid estimates of vital events be obtained from sample surveys? Such surveys are useful for many purposes and they are relatively inexpensive. Must we seek other ways to get current information about vital events in those areas where registration is sorely inadequate, in areas where it certainly will require many years to produce good registration data throughout the country.

Linder and Simmons note that a very great deal needs to be done before adequately proven survey methods are a part of our statistical arsenal. They offer some principles as guidelines for research and for action. They suggest that one line of attack is intensified effort to develop and utilize more incisive survey procedures. They remind us that many methods must be tried because the current state of the art requires this.

A. Linder-Simmons: During the past 2-3 years I have had an opportunity to discuss this topic with Linder-Simmons and their colleagues on several occasions and to collaborate with them on aspects of the Pakistan program. I am happy to associate myself with the tenor of the Linder-Simmons remarks and to say that I have no major disagreement with them. I might make a minor quibble with reference to their statement that family planning programs, if they are to be effective, are necessarily expensive -- they cost only a few cents per capita; and at times I think that isn't so very much! I think they overstate the reliance of national programs on demonstration programs. And where they suggest that a family planning program developmental area must be visualized as including populations of the order of magnitude of 1,000,000 persons, I would prefer to say one should think in terms of a population of six digits more often than seven -- and thus we are only one digit apart! I would cite the studies in Taichung and Sun Dong Gu of Seoul, Korea, each with populations of 300,000 as very good examples of the type of research study that they have in mind. Also, I would want to question their statement that the universe of study will be the Family Planning Developmental Area rather than an entire nation (p.11). This seems to me to be applicable to India to a greater extent than most other countries. But these are relatively minor points. I agree with them on their major propositions.

B. Horvitz: The Triangle Research Institute project reported by Horvitz is most promising and is very unusual. Their design is so clean and pure and shining and complicated that it reminds me of one of our spacecraft -- highly sophisticated, enormously impressive, and the thing might actually work! Their design includes households in which births and deaths are known to have occurred; some of the births were illegitimate and, as expected, a much larger proportion of these were not reported to the enumerator. It is an anomaly that a larger proportion of illegitimate than legitimate nonwhite rural births were reported; the sample sizes were rather small, I believe, and it is doubtful that this differential is statistically significant (the figures are 98.5 and 92.2). The design also included deaths from socially stigmatized causes such as those from tuberculosis, venereal disease, mental diseases, cirrhosis of the liver, congenital malformations, suicide and homicide.

Recall lapse: The most surprising finding reported by Horvitz is that "There was no indication that underreporting increased as the interval between the interview date and the birth date of the infant increased." He notes that this is contrary to the Indian National Sample Survey which finds that memory decay increases with time. Horvitz notes that the differential noted by the Indians could occur if the underreporting were confined to those births which did not survive the first year of life. This assumed an Indian infant mortality rate of about 180/1000 as compared with a rate of 27/1000 in the U.S. He leaves it to us to make the inference that this difference in infant mortality might indeed account for the memory decay found in India but not in his study. Data from the 7th round of the NSS indicate that if the index for the birth rate at the point of origin is considered to be 100, respondents report decreasing proportions of births each month, with the index being only 83 in the twelfth month. Similarly, the index for reported deaths decreases to 67 for the twelfth month. I believe that a good deal more than infant deaths are involved in the memory decay observed in India. The Indian NSS also reports another interesting phenomenon, namely that by increasing the reference period to two years instead of one year, the number of events reported for the last year increases appreciably.

Persons in developing countries are not as well educated as those in the Horvitz sample; they are not as time- and number-conscious as are people in a developed economy. Perhaps the memory decay that has been observed in developing countries does not occur in ours. We shall look forward to a more detailed report on this at a later time.

Reference period: The Pakistan and Thailand surveys have adopted a quarterly survey procedure; in the Thai survey the reference period is for a quarter whereas in the Pakistan survey it is for a year. The purpose of inquiring about events during the past year in each quarterly survey is to avoid the boundary effect. If there is some telescoping of reporting events one gains by obtaining reports for a longer period than that for which data are required, provided you can later sort out the date of the event. Similarly, if respondents are not time-conscious, there would be some more or less random variation in reporting the date of a vital event, and if there is no overlap in the reference period one would tend to miss some of these events. The failure of the Thai survey to have an overlapping reference period tends to underestimate vital events, it seems to me. I don't have any data with which to quantify this observation, but as a matter of procedure it seems to me to be unfortunate.

We do not have definitive information about the optimum length of reference period, nor of the optimum frequency of surveys. I have argued with both our Pakistani and Thai colleagues that quarterly surveys are unnecessarily frequent; but they counter with the strong argument that enumerators will do a better job if they are in the field once a quarter collecting data on the same topic.

Open interval-pregnancy history: In a search for sensitive indicators of changes in level of fertility it has been suggested that use of the "open interval" -- the length of time since termination of last pregnancy -- would be rewarding. The idea is that if the distribution in months of time since termination of last pregnancy increases, women are postponing or reducing the number of births. Presumably this would occur first among higher parity women in a country where a family planning program is being launched. There are problems of obtaining sufficiently good dates of termination of last pregnancy, but this is an approach worth pursuing. There is also the possibility that well-trained interviewers could get good data on pregnancy history, with which one can reconstruct fertility rates for the women interviewed. I am personally pessimistic about the pregnancy history approach, but should like to encourage others to try it!

Randomized response technique: The TRI group has been experimenting with the randomized response technique suggested by Stanley L. Warner. This involves alternate wording of questions or statements, e.g., answer as true or false (1) I have never had an induced abortion, or (2) I have had an induced abortion. The questions or statements appear with a specified frequency, but the statement being answered is known only to the respondent. This approach is promising for sensitive subjects such as abortions, illegitimacy, crimes, use of drugs, cheating on exams, etc. A major disadvantage is that the sample size required increases sharply. I don't see any application of this technique for births and deaths on a large-scale survey, although if estimates of illegitimate births are needed, this approach might well be used.

Registration in sample areas: A number of persons have suggested that the least expensive and best single method of obtaining estimates of vital events in an area having poor registration data would be to establish a special registration system, or else "beef up" registration in a probability sample of areas. This is such an obvious technique that it is surprising that it hasn't been tried. One would have to supplement it with censuses, or head counts, but such counts need to be made only once in two years, I believe.

The Chandrasekaran-Deming approach: It involves collection of vital events by two independent systems, the comparison of those events on a case-by-case basis, and a classification of the events as:

1. Counted by both procedures, say, survey and registration
2. Counted by the survey procedure only
3. Counted by the registration procedure only

From this one can estimate the proportion of events missed by both procedures. They also suggest that if one can divide the population into homogeneous strata, the estimate can be improved.

Independence: Complete independence is difficult to achieve, but as Lauriat points out the studies cited apparently have done reasonably well on this score. I should like to note that the general effect of lack of independence is to underestimate the events that are being counted. If one of the investigators copies events from the other, or obtains them with the same biased procedure, e.g., talking with a few respondents rather than all those designated, then coverage is inadequate and some events are likely to be missed. Inadequate canvassing procedures are analogous to the Bureau of the Census Post Enumeration Survey in which roughly the same canvassing procedures are used as in the Census. The result is that the two systems are not wholly independent, and certainly there is an underestimate of non-white young adult males. Here the problem is not collaboration between two investigators but rather the difficulty of devising different procedures, each of which is adequate or better.

Matching: The C-D formula necessarily assumes that matching is carried out perfectly; not necessarily quickly or easily, but perfectly. One cannot escape this assumption, nor can one escape the consequences of a mismatch. If one sets the matching criteria so that uncertain cases are counted as separate events, then one is "sure" that all events counted as matches, are real matches. But the doubtful cases, some of which probably are matches, are treated as separate events, thus counting them twice, and thereby inflating the estimate. As a matter of fact, one counts them as slightly more than two cases -- one for the survey, one for the registration, and a fraction of this for the 4th cell -- those missed by both. This is true because NrNe/C has been inflated in the numerator, and deflated in the denominator.

On the other hand, if one adopts loose criteria, counting all doubtful cases as matches, thus insuring that each non-match is indeed a separate event -- by so doing one counts some separate events as the same event, and thereby decreases the estimate of total events.

The problems of matching are difficult indeed. In Pakistan, for example, babies frequently aren't given a name for weeks; there is a reluctance to speak, to use, the name anyway; in some societies where males are highly prized, a respondent may report the sex as female in order to ward off the evil eye -- but at a subsequent interview, possibly with the husband rather than the wife, the respondent may correctly report sex. The typical Pakistani village does not have house numbers or street names. As a consequence specific addresses are non-existent. In order to overcome this deficiency, the population growth estimation study numbered each house -- the enumerator carried with him metal placques and nailed one of these on the front entrance of each dwelling unit. There were also difficulties of transliteration of names into English.

I could go on in this vein for some time, but surely the point is clear -- matching is difficult; there is no theory of how to compare different records and determine what is a match; but collection of data by two different systems discloses with "certainty", if I may be permitted that term, that each system has missed some event; indeed, in studies undertaken to date it is abundantly clear that each system has missed appreciable numbers of events. Thus matching is a necessity.

We have devoted too little attention to these problems; and I hope that we can seek ways to add to our knowledge in this field. Let me advance one or two suggestions. What we need is an objective criterion of what is a real match. I believe it would be feasible in a methodological study for the registrar to take a footprint of the infant, and attach it to his report to headquarters. Similarly, the enumerator could take a footprint of the infant and forward it to headquarters. One could then carry out the matching procedure in the conventional fashion after which the foot prints could be read independently and match conclusively.

There are problems in this approach, of course.(1) Mothers would object -- but give them a gift of cloth, a chance at a prize, a small inexpensive painting, a special certificate, tobacco -- even money! I'm sure they would cooperate, and at a price that is not too expensive.

(2) Some babies would be missed -- those who die early; those who are seriously ill; those who move out of the area. But the residual group would be large and would contribute useful information. Indeed, one could take fingerprints of one or both the parents in some of these cases, and that would be helpful in the matching process also.

(3) You may also say that taking footprints is a tricky business, and that in the distant village of Pakistan where piped water and electricity are unknown, taking a clear footprint is virtually impossible. I know some of the problems; I have read some of the literature and talked with some of the experts. Perhaps I can dispose of this argument by saying that I know of experts who would be delighted to teach enumerators and registrars how to take footprints, in exchange for a trip to a distant land.

I do not mean to dwell overly long on the possibility of using footprints as a positive means of identification. The president of the ASA, Prof. Fred Stephan, was intrigued by this idea, and suggested one might also consider voice patterns, using a tape recorder to aid in the match. We know very little about voice patterns, and I could not resist asking Prof. Stephan whether it might be necessary to equip each investigator with a sterile pin, to stick the infant in order to insure a uniform "word", or its equivalent!

I do not mean to suggest a specific solution to this important problem -- footprints of infants, voice patterns of babies, fingerprints of parents, photographs of the family, a tattoo mark that would last for a few weeks only, etc. -- but I want to emphasize the very great need, the lack of an adequate statistical theory, and the opportunity to contribute to a fascinating and important problem.

X

FAMILY LIMITATION PROGRAMS AND TECHNIQUES

Chairman, DUDLEY KIRK, Population Council

MEASURING THE EFFECTIVENESS OF A FAMILY PLANNING PROGRAM: TAIWAN'S EXPERIENCE

John Y. Takeshita, University of Michigan Population Studies Center

Introduction

A large number of family planning programs under government auspices has come into operation in the last three or four years in developing countries in various parts of the world and many more are being planned for implementation soon.[1] In view of the fact that most of these programs are officially or unofficially tied in with their economic development programs and are driven by a sense of urgency, there is pressure for quick evaluation of the on-going operations even as there is need for efficient assessment of their long-term effects. This pressure and need must be met by the increasing number of statisticians, demographers, and social scientists who are being called upon to work with the administrators of the programs for such purposes.

The task of the statisticians, demographers, and social scientists involved in these programs would be to ensure the collection of appropriate data either routinely as part of the recording systems of the programs or by special studies to measure the impact of the programs on factors that would ultimately affect the birth figures and to analyze the data quickly and to interpret them within the framework of the program objectives. Where accurate routine systems of vital registration and population accounting are absent, they must also help in devising ways to obtain accurate estimates of vital events cheaply and quickly so as to be able to measure changes in their rates, when they occur.

Most family planning programs in the developing countries have as their objectives: (1) the increase in the effective use of contraception among a significant proportion of the population in the reproductive ages, and (2) the decline in over-all fertility to a level consistent with a population growth rate considered to be commensurate with their projected economic growth rate. The evaluation of the programs then must show the progress in the number of effective users of contraception and the change in fertility level and specify how these are related to program efforts.

I have been asked by the organizer of this session to describe the evaluation efforts of the family planning program in Taiwan in general terms with some reference to the statistical problems involved. Taiwan serves merely as an example. The program there has enjoyed relatively good success in the short time that it has been in operation and its progress has been unusually well documented. Furthermore, in their effort to evaluate the program effects, some challenging problems have come up. But before introducing these problems, let me briefly describe the family planning program itself and some of the major results to date.

Taiwan's Family Planning Program[2]

Taiwan's family planning program, launched on a large scale in 1964, aims to bring the island's population growth rate that was 3 per cent in 1963 down to about 2 per cent by 1973. To achieve this aim, which is believed by economic planners there to be consistent with their development plans, the birth rate which stood at 36.3 per 1,000 in 1963 must be reduced to about 24. The program, under the auspices of the Provincial Health Department and promoted ostensibly for health purposes alone, hopes to achieve this aim by having 600,000 married women in the ages 20-44 inserted with the new intrauterine contraceptive (IUCD), by 1969. This would amount to securing about 40 per cent of the island's 1.5 million married women 20-44 on the IUCD.

Some 300 field workers are employed full-time to promote the IUCD in local areas throughout most of this island of 12 million inhabitants. In addition to their educational efforts, they distribute coupons to interested women, who are instructed to go to a nearby doctor authorized by the program to make IUCD insertions. Doctors themselves and other health personnel also are authorized to distribute these coupons. The coupon entitles the woman to a 50 per cent discount on the insertion fee (US$1.50). The doctor in turn upon presentation of the coupon receives from the program a subsidy to make up the difference (US$.75). At present some 600 specialists and general practitioners, mostly in private practice, are authorized by the Provincial Health Department to make IUCD insertions under this scheme.

Monthly and annual targets are set; and, when insertions seem to be lagging, special "gimmicks" are introduced to boost them. For example, insertions have been offered free of charge for a limited time period in some of the areas and special fees have been offered to persons referring cases--with good effect.

While the program so far has relied almost exclusively on the IUCD, because of problems with respect to its retention rate which at present seems to be much lower than it was initially expected, plans are under way to offer the oral pill along with the IUCD.

Taiwan's program stands unique among family planning programs in the fact that it has been guided from the start by a strong evaluation team in the Taiwan Population Studies Center, which was established in 1961 under financial support from the Population Council of New York and operated in technical cooperation since 1962 with the University of Michigan Population Studies Center. It was through the facilities of the then nascent Taiwan Population Studies Center that in 1963 an experimental program to test the relative efficiency of different modes of communication and the acceptability of the IUCD was undertaken in Taichung City, the capital of this island province of the Republic of China.[3] The Taichung Study, as this experiment has come to be known, has served as a guide in planning and implementing the island-wide program which is now in its third year of operation. The Taiwan Population Studies Center, now equipped with its own IBM key-punch and sorter, effectively carries out the important task of evaluating the operations of this expanded program on a current basis.

Major Achievements to Date[4]

In 1964, the target was to have 50,000 women inserted with the IUCD. 93 per cent of this target was reached. In 1965, 100,000 insertions were aimed at and 99 per cent was achieved. The target for the current year is 120,000; at mid-year about 50 per cent of the target has been reached. From 1964 through mid-1966, more than 200,000 insertions have been made. This amounts to more than 13 per cent of the island's married women 20-44 years of age.

The crude birth rate declined 10 per cent from 36.3 in 1963 to 32.7 in 1965. The rate of natural increase slowed down somewhat from 3.0 per cent in 1963 to 2.7 per cent in 1965. (The crude death rate declined from 6.1 to 5.5 between 1963 and 1965.) It is especially noteworthy that the rate of decline in the birth rate has been accelerated since the start of the island-wide program in 1964: it declined 2.5 per cent between 1961-62; 2.9 per cent between 1962-63; 4.7 per cent between 1963-64; 5.5 per cent between 1964-65; and 6.6 per cent between 1965-66 for the first four months.

In Taichung, where the intensive experimental program was undertaken in 1963, the crude birth rate declined by 5.4 per cent between 1963 and 1964, more than twice the rate of decline (2.6%) during this period in the four other major cities combined and somewhat more than the rate of decline (4.7%) for the province as a whole. In the previous year, the decline in Taichung (2.6%) was lower than the decline in the other major cities (3.9%) and in the province as a whole (2.9%). The accelerated decline in 1963-64 in Taichung undoubtedly can be attributed to the large number of IUCD insertions made there during the intensive program. This advantage in Taichung, however, was not carried through to 1964-65 as the program was effectively extended into the other cities and counties throughout the island.

How the Program is Being Evaluated[5]

1. Fertility Measures from the Household Register

Taiwan, unlike most developing countries, has the advantage of an unusually accurate household registration system from which birth and death data can be obtained within a month or two of their occurrence in every one of her 22 cities and counties. What is more, not only crude birth rates but also age-specific fertility rates and hence general and total fertility rates for any given year are available for each of her 361 townships by the end of the first quarter of the following year. The easy availability of these rates permits quick evaluation of program effect in the most telling way. After all, whatever else may happen as a result of the program it must ultimately show in fertility decline if it is to be judged effective.

We now have at Michigan a deck of IBM cards on which are punched several demographic and socioeconomic characteristics of each of the 361 townships. We plan to add to this deck the IUCD insertion rates by age groups and a recent set of age-specific fertility rates, among other things, so that we may learn how the program efforts interact with pre-existing conditions in the local communities to affect their fertility levels. There is evidence from a recent study[6] that the fertility levels before the program varied systematically with the socioeconomic characteristics of these townships. The program presumably aims to facilitate fertility decline in all communities by having its effect cut across the pre-existing differentials. How this happens, if it does, needs to be specified for program evaluation and guidance, and fortunately in Taiwan this can be done rather easily.

2. IUCD Insertion Figures from the Coupons

The IUCD insertions themselves are tallied by each township within 10 days following the month in which they occur by the use of the coupons, which we referred to earlier as entitling the holders to a 50 per cent discount on the insertion fee. The coupons, on the basis of which the subsidies are administered, are collected monthly from the inserting doctors and sent to the Taiwan Population Studies Center by the 7th day following the end of the month for quick analysis.

The coupon contains not only the names and addresses of the case and the inserting doctor and the date of insertion but also such other items of information as: the woman's age, the number of living children and sons she has, the name and category of person by whom she was referred, the education attained by her, whether she wanted more children, the last method of contraception she and her husband used, and the date of her last child birth. These information are immediately punched on IBM cards and runs are quickly made for the monthly report that is prepared for evaluation purposes.

Here is a system that is built into the administrative routine of the program and at the same time provides data useful for evaluation on a current basis. The Taiwan Population Studies

Center has been able to make the following kinds of evaluation from this source:

a. Measure the progress of the program in the province and in each township in terms of the target set for the month. Problem areas are easily detected and because of the rapidity with which data become available corrective measures can be taken very quickly.

b. Obtain the number of IUCD insertions in a given month attributable to each worker, thus getting some measure of work efficiency.

c. Obtain the cumulative IUCD insertion rates per 100 married women 20-44 in the province and in each of the 361 townships to measure the progress of the program, which aims ultimately to reach 40 per cent of the women in the reproductive ages.

d. Measure the progress of the program in terms of the types of persons it is reaching: age groups, parity, past use of contraception, recency of child birth, and educational strata. Generally speaking, the program has tended to attract initially those past age 30, with 3 or 4 children and one or two sons, those who have already tried something in the past to check their family growth, and those who only recently had a live birth. Education has not shown too much difference. It is hoped, of course, that the program eventually will attract the younger women with fewer children and sons, those who have never before tried birth control, etc.--in short, to get married women to take up contraception earlier for spacing as well as for limitation purposes.

In computing the rates of IUCD insertions, we are confronted with the question as to whether it is realistic to use all married women in a given age range as base. In Taiwan, not a few women are already sterilized or using contraception that is satisfactory to them. If we can estimate the number of women so protected, we should be able to remove them from the base as ineligible. In Taichung, for example, where our target population was the married women 20-39 years of age, the acceptance rate from the start of the program to July 31, 1965, was calculated to be 20 per cent when all married women 20-39 were used as base but 27 per cent when those already sterilized or using satisfactory contraception were excluded from the base. The reason why education does not seem to make much difference is that more of the better educated are already protected either by sterilization or contraception. In fact, when acceptance rates in Taichung were computed on bases excluding the already protected, a direct relationship between education and acceptance did obtain. What we are saying in effect is that rates calculated on bases that are heterogeneous are not comparable with each other and interpretations can be misleading or unrealistic unless the heterogeneity is taken into account. From an administrative point of view, too, to assign a uniform target for all workers without regard to the composition of the base population of the area in which each works can be unfair and damaging to worker morale.

3. Island-wide Survey

In Taiwan fertility has been declining consistently from at least 1958. There is evidence that not a few women were already sterilized, using contraception, and even resorting to illegal abortion before the family planning program got under way. It is likely that even now not an insignificant number of women are taking up contraception or other forms of birth control on their own through non-program facilities. Some of these may be influenced by the program even if they may use facilities outside of the program. The full impact of the program then cannot be assessed merely by tallying the IUCD insertions reported by the doctors participating in the program. Besides, over-all fertility is going to be affected by birth control practices whether in or out of the program. The first of a series of biennial sample surveys covering the entire island population of married women 20-39 was taken in the fall of 1965, and the data are being analyzed currently to gauge, among other things, the prevalence of birth control practices in Taiwan irrespective of the program. The surveys over the years will permit an assessment of the changes in knowledge, attitude, and practice with respect to family planning as the program is intensified. In addition to providing data to trace the program influences on these changes, they will provide a basis for estimating the impact of the various practices on over-all fertility.

4. IUCD Follow-up Survey

A medical follow-up study of the IUCD cases in the Taichung program has revealed that as much as a half of the original cases fail to retain the IUCD after two years and some of them, albeit a small percentage, do get pregnant even with the device in place. Needed are estimates of both the retention rates and the failure rates for different periods for the cases getting the IUCD in the island-wide program. It is also important to know what proportion of those who discontinue do so voluntarily, what proportion of the discontinued cases take up other measures to protect themselves against further growth of their families, how the pregnancies that occur from device failure are disposed, and the extent to which there are demographic and socioeconomic differentials in these. To get these and other related information on those who take the IUCD in the national program, a follow-up survey of a sample of cases is being carried out on an annual basis. The techniques of measurement being developed by

Dr. Potter will be applied to the data from the sample surveys as they have been applied to the data from the Taichung medical follow-up study.

A Major Unsolved Problem

The most serious problem confronting the evaluation effort in Taiwan is that at present no one knows with any precision what fertility decline to expect given: (1) the large number of IUCD insertions in the program, (2) the considerable use of contraception, sterilization, and induced abortion outside the program, (3) the varying failure rates associated with the IUCD and other contraceptive methods, (4) the unusually high discontinuation rates of the various contraceptive methods, including the IUCD, that are adopted, (5) the fact that not a few women were already using various methods of birth control even prior to the program, and (6) the fact that birth control practice in or out of the program is variously selective with respect to age, parity, fecundity, social status, etc. The problem is complicated further by the fact that the effect of contraception on the fertility of a woman is not always easily differentiated from the effect of her fecundability, which is affected by her age, parity, recency of live birth, lactation experience, etc. What is more, not only is there selectivity with respect to these characteristics in acceptance of the IUCD and its retention but also this selectivity changes over time. The complexity of the situation would seem to call for a mathematical model that might be simulated on a computer. To bring all of these factors into a workable model obviously is no easy task especially when these factors are variable over time, but the need is there and the challenge stands before us.[7] Taiwan is one place where the necessary data can be and are being collected and where once the model is in working order it can be empirically tested. The implication for program planning is that such a model would enable us to determine the size of the program required to achieve a specified goal with much greater specificity than we are able to at present.

Footnotes

1. For a review of world developments in this field, see: Berelson, B., et al (eds.), Family Planning and Population Programs, University of Chicago Press, 1966.

2. For a detailed description of Taiwan's family planning program, see: L. P. Chow, "A Programme to Control Fertility in Taiwan," Population Studies, XIX, No. 2 (November 1965), 29-39.

3. For a description of the Taichung experiment and some early results, see: Bernard Berelson and Ronald Freedman, "A Study in Fertility Control," Scientific American, CCX, (May 1964), 29-39. For a summary of more recent results, see: Ronald Freedman and John Y. Takeshita, "Studies of Fertility and Famility Limitation in Taiwan," Eugenics Quarterly, XII, No. 4 (December 1965), 233-250.

4. Based on the monthly reports prepared by the Taiwan Population Studies Center, Taichung, Taiwan.

5. For a detailed description of the evaluation procedures used in Taiwan, see: L. P. Chow, "Evaluation Procedures for a Family Planning Program," in Berelson, et al (eds.) Family Planning and Population Programs, pp. 675-689.

6. Andrew Collver, Alden Speare, Jr., and Paul K. C. Liu, "Local Variations of Fertility in Taiwan," University of Michigan Population Studies Center. Submitted for publication.

7. For an interesting preliminary attempt along this line with some of these factors taken into account and applied to the family planning program in Korea, see: Byung Moo Lee and John Ibister, "The Impact of Birth Control Programs on Fertility," in Berelson, et al (eds.), Family Planning and Population Programs. pp. 737-758.

SOCIAL AND DEMOGRAPHIC CORRELATES OF IUCD EFFECTIVENESS: THE TAICHUNG IUCD MEDICAL FOLLOW-UP STUDY

R. G. Potter, Brown University
L. P. Chow, Taiwan Population Studies Center
A. K. Jain, University of Michigan
C. H. Lee, Taiwan Population Studies Center

Introduction

High hopes have been held for intra-uterine contraception as a method which is effective, inexpensive, and not requiring repeated action at an inconvenient time. The device has been assigned top priority in several national family planning programs. To measure adequately its use-effectiveness in a population, one would ideally like to know the proportions of acceptors still wearing the device at specified intervals after insertion and to have this information if not for a general sample representative of the population, at least for a clinic sample representative of a stage of the particular family planning program. As yet data even of the latter sort are scarce.

The present analysis is based on materials from the Taichung IUCD Medical Follow-up Study, in Taiwan. These data, relating to a sizable clinic sample, were collected during a three-year period 1962-64. Although like almost all IUCD studies the insertions were made in medical clinics, the sample of cases is more representative of the general population than is usually true for clinic studies, which tend to be highly selective even of those interested in family planning. The large number of cases came to the clinic following a large scale family planning program described in detail elsewhere.[1] The 4100 cases coming from the city of Taichung itself constitute about 11 per cent of all the married women in the ages 20-39 in the city. It will be shown later that retention of the intra-uterine device varies widely according to social and demographic characteristics of the wearer and it is a primary interest of this analysis to illustrate some of the differentials.

The present investigation represents a collaboration between the Taiwan Population Studies Center and the University of Michigan Population Studies Center. Special acknowledgement is owed Dr. Ronald Freedman who coordinated the work and to Claudia Ludvigh who wrote the computer programs. The present report is a brief abstract of a larger analytical report to be published next year as part of a monograph on family planning and fertility in Taichung.

Let us now briefly consider sample and follow-up procedures.

[1]R. Freedman and J. Takeshita, "Studies of Fertility and Family Planning in Taiwan," Eugenics Quarterly, Vol. 12, No. 4, (December, 1965), pp. 233-250.

Sample

The sample consists of 7295 women, two-thirds of them accepting one size of loop and most of the remainder other sizes of loop. Planned procedure was to follow up the women 6, 12, and 24 months after insertion. Each woman received a postcard to remind her of the scheduled visit to the clinic. If she did not come, a field worker went to her home, interviewed her, and asked her to return to the clinic if she might be still wearing the device. If she had terminated IUD, an attempt was made to ascertain circumstance and time of device loss.

Some women returned to the clinic shortly after an insertion because they had been told to come back at once if they experienced any problems. Roughly 30 per cent of the women came to the clinic without a home visit. For the rest, home visits were necessary and because of the heavy and uneven load of work, these visits were not always made on schedule.

Out of a total sample of 7295 women, 650 were not visited even once, nearly half of them because of the short interval between insertion and cut-off date of the study. Another 388 women were visited once or more but with a scheduled visit missing. It is believed that these missed visits reflect more staff limitations than patient resistance. Indeed, only 54 cases of actual failure of follow-up were reported, indicating a high level of respondent cooperation. No important differences were found between the social and demographic characteristics of women adequately followed up and those for whom one or more scheduled visits were missing.

It is believed, therefore, that bias from inadequacies of follow-up is small.

Attention will be restricted to first segments only, by which is meant the period of use from first insertion of an intra-uterine device to first interruption of its use or end of observation, whichever was earlier. The lengths of these first segments have been defined conservatively. Months of use by any woman classified as a continuing user are counted only up to her last clinic visit when it was certain that the device was still in place.

As time from insertion increases, the number of women exposed rapidly decreases. In the present analysis sample size is large enough to yield fairly stable rates for the first two years. For the third year the number of cases becomes too small for stability.

Methodology

So much for sample and follow-up procedure. As a last preliminary before considering results, let us comment briefly about life table methodology.

In any follow-up study of users of intra-uterine devices, for only some of the women will observation be complete in the sense of observing when and under what circumstance the device was lost. Other women, usually comprising a majority of the sample, will be classified as continuing users at time of last visit. Naturally one wants to be able to use these incomplete histories as well as the more complete ones to derive an unbiased picture of retention and loss of IUD as a function of time from insertion. This is what the life table approach is designed to do.

The useful wearing of an intra-uterine device may terminate for any of several reasons - pregnancy, expulsion, or removal, which reasons themselves may be subdivided. To be fully satisfactory, then, the life table methodology must provide for competing risks of termination. At the University of Michigan Population Studies Center, in collaboration with Dr. Christopher Tietze of the National Committee on Maternal Health, we have been developing such a procedure; more precisely, we have been adapting the multiple decrement life table to the analysis of IUD effectiveness. Recently we have called upon Dr. Bernard Greenberg and his associates at the University of North Carolina to assist us with the derivation of formulas for estimating standard errors. Hence our methodology is still in the process of refinement and figures in the hand-out tables may not be final.

Two types of rate will be used. What Tietze has called "net rates" allow for the presence of competing risks. For example, a net cumulative rate of expulsion allows for and is slightly reduced by some women becoming pregnant or removing the device before they have had a chance to expell. Net cumulative rates are additive. The net rates of pregnancy, expulsion, and removal add to the rate of termination for the three reasons combined. Because of this additivity, net cumulative rates are appropriate for studying the relative frequency of different types of termination in a single sample.

However a problem arises when one uses net cumulative rates for comparing the relative frequency of a particular type of termination such as expulsion in two different samples. For example, suppose that in sample B, the monthly rates of expulsion - that is, the probabilities of expelling during the first, second, and so on month after insertion if retaining the device up to the beginning of that month - are lower in sample B than in sample A. Now if the levels of competing risks are also lower in B than in A so that fewer women in B are lost to pregnancy and removal and therefore are exposed on average longer to the risk of expulsion than in A, then it is possible for sample B to show a higher net cumulative rate of expulsion despite its lower monthly rates of expulsion.

To cope with this problem one may use what Tietze calls "gross rates." Gross rates are predicated on there being only a single cause of device loss and all competing risks eliminated. A gross cumulative rate of expulsion is a pure function of the monthly rates of expulsion and formally independent of the levels of competing risks. Hence in the example above the gross cumulative rate of expulsion of sample B with its lower monthly rates of expulsion would always be lower than in A and in this manner leads to sample comparisons that are more easily interpreted. However gross rates, exceeding net rates, are not additive and therefore not appropriate for studying the frequencies of different types of terminations in a single sample.

To summarize these comments on methodology, we will use the additive net rates to study the frequency of different types of termination in a single sample and gross rates to assess the relative frequency of single risks in different samples.

Results: Total Sample

Turning now to results, consider Table 1 which gives cumulative net termination rates per 100 women at the end of 12 and 24 months by type of termination. Looking at the last column on the right which gives terminations for all causes, we see that about one-third of the women have terminated by the end of one year and about one-half by the end of two years. Even allowing for the fact that many of the women accept reinsertions, these loss rates are high absolutely and disappointing relative to the high expectations entertained by many persons for IUD. Preliminary returns from an island-wide follow-up study indicate that termination rates in this more general Taiwan sample may be even higher with perhaps two-thirds of the first segments terminated within two years. It may also be noted that the termination rates reported in Table 1 are not much higher than those reported by Dr. Tietze in his December, 1965 Sixth Progress Report of the Cooperative Statistical Program for the Evaluation of Intra-Uterine Devices, based on an assemblage of clinic samples predominantly from the United States. While the apparently high termination rates after one or two years in either Taiwan or the United States may be discouraging to some, they should be considered in comparison to experience for other types of contraception in use in large populations. It is doubtful, for example, that anyone can show at present continuation rates as high as fifty per cent after two years for any other contraceptive or combination of contraceptives in a

TABLE 1

Cumulative Net Termination Rates per 100 Women, at End of 12 Months and 24 Months of Use, by Type of Termination, Based on Data from the Taichung Medical Follow-up Study of Users of Intra-uterine Devices

Months of use	Pregnancies			Total expulsions	Removals				Total terminations
	Device in situ	Device undetermined	Total pregnancies		Medical reasons	Personal reasons	Nonrelevant reasons	Total removals	
12	3.3	1.2	4.5	12.2	15.1	.7	1.9	17.7	34.4
24	5.9	2.3	8.2	14.9	22.0	1.2	5.0	28.2	51.3

large sample for a developing country.

Looking at other columns of Table 1, the following points may be made:

(1) Removals are the predominant cause of device loss and most of these removals are attributed to medical reasons (e.g., cramps, staining, menstrual irregularity, and the like).

(2) Expulsions are the next most common cause and occur mainly in the first year.

(3) Slightly fewer than 10 per cent experience pregnancy during the initial 24 months, usually with the device in situ.

In Graph 1, we consider whether monthly probabilities of pregnancy, expulsion, and removal start at high levels and then progressively decline as time elapses from insertion. What is being graphed are conditional monthly probabilities, that is, the probability of terminating for a specific reason during the next month given that the device has been retained up to the start of this month.

Only expulsions follow the pattern of starting at a high level and then progressively decreasing as time elapses from insertion. Depressed during the first few months by postpartum amenorrhea, the monthly pregnancy rate soon stabilizes at a fairly low level, while after the first three months the removal rate stabilizes at a surprisingly high level. Why women who have worn the device for 12 or 18 months should continue to show such a high rate of removal is puzzling. The most common reasons cited are medical and one would expect physical side-effects to subside after the first few months. Relatively high removal rates after 12 or 18 months also appear in the data for the Tietze samples. Further study of these late removals is planned.

Differentials

Now finally let us turn to Table 2 which offers an illustrative set of differentials. The five column headings NPERT stand respectively for subsample size, pregnancy, expulsion, removal, and total terminations. The rates in the three columns headed PER are gross cumulative rates, representing the cumulative proportions of women terminating IUD for a specific cause under the assumption that the particular cause is the only one operative in the sample and all competing risks are eliminated.

In Table 2 four differentials are especially worthy of notice.

First, cumulative rates of removal and expulsion are strongly negatively correlated with mother's age and number of previous pregnancies before insertion. See variables #1 and #2. Young women or women with few pregnancies have relatively high rates of removal and expulsion. Perhaps their physical systems tolerate the device less well and physical side-effects are more severe. Possibly too, being at an early stage of their family building, they are less willing to accept temporary discomfort from IUD.

Secondly, concerning variable #3, women coming from outside Taichung have similar pregnancy and expulsion rates but lower removal rates than do Taichung residents. The difference

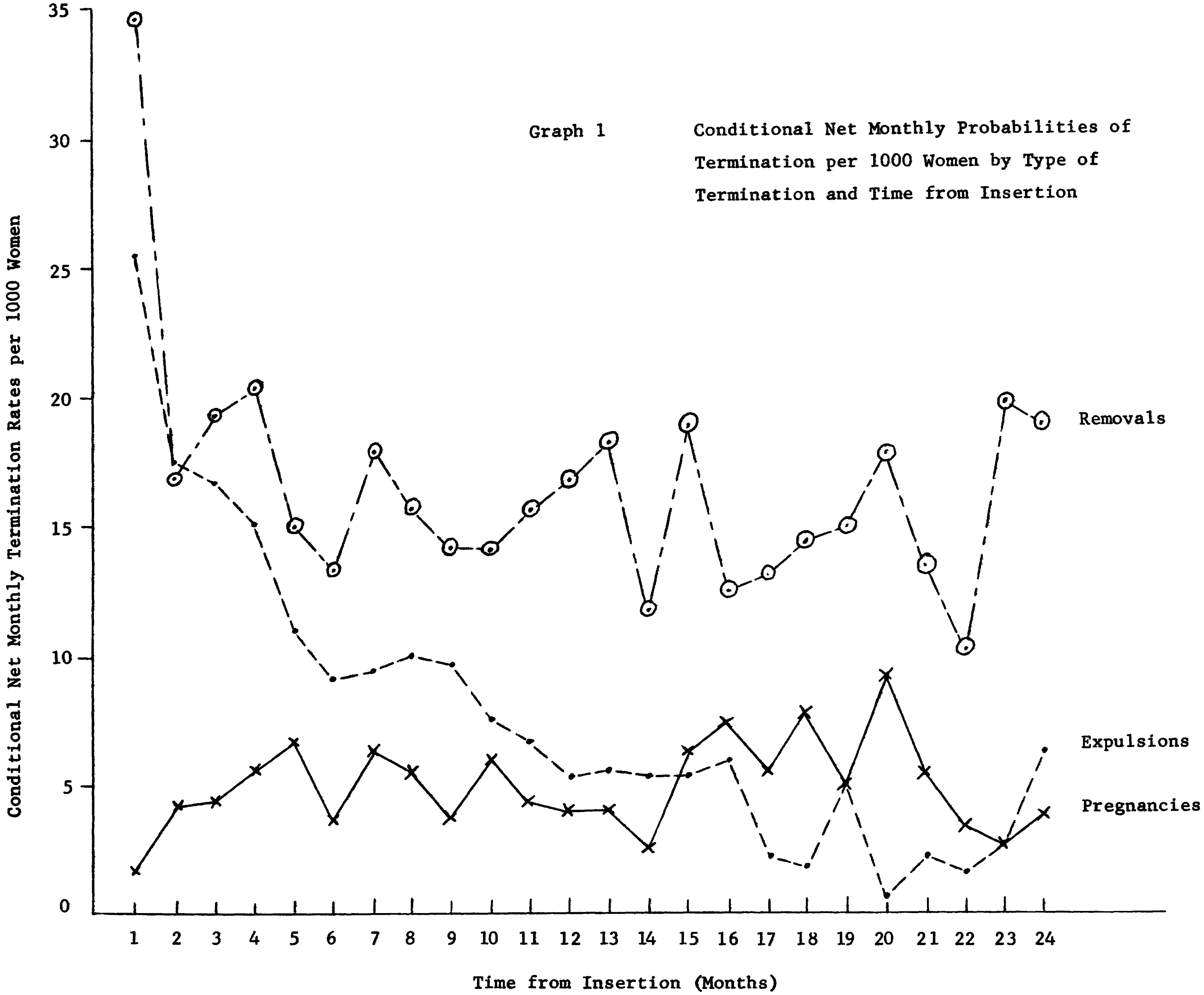

Graph 1
Conditional Net Monthly Probabilities of Termination per 1000 Women by Type of Termination and Time from Insertion
Conditional Net Monthly Termination Rates per 1000 Women
35
30
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Time from Insertion (Months)
Removals
Expulsions
Pregnancies

TABLE 2

Cumulative Gross Termination Rates per 100 Women at the End of
12 Months and 24 Months of Use by Type of Termination
and Social and Demographic Characteristics

Characteristics	12 months of use					24 months of use			
	N	P	E	R	T	P	E	R	T
1. Age of wife at 1st insertion									
13-24	895	6.6	26.0	34.8	54.9	14.9	32.0	58.1	75.8
25-29	2112	6.7	17.0	22.6	40.1	15.0	22.5	40.1	60.5
30-34	2109	5.7	10.3	15.6	28.6	12.3	13.9	26.5	44.6
35 and over	1521	3.7	7.3	14.1	23.3	6.2	9.5	23.9	35.4
2. No. of pregnancies preceding 1st insertion									
Less than 2	162	3.9	31.8	53.7	69.7	19.9	37.4	89.0	94.5
2	566	7.5	30.0	33.7	57.0	17.3	32.9	58.2	76.8
3	953	6.2	19.4	22.8	41.6	14.2	26.0	40.9	62.5
4	1187	6.4	14.9	22.3	38.2	14.6	20.7	38.2	58.1
5 or more	3774	5.1	9.0	15.3	26.8	9.8	12.0	25.9	41.2
3. Areal location									
Taichung - Urban	2769	6.4	15.2	23.3	39.2	12.4	17.5	38.9	55.9
- Rural	1331	5.3	13.3	20.3	34.6	13.0	17.9	33.4	52.4
Outside Taichung	2456	5.1	11.9	14.6	28.6	8.4	18.0	24.2	43.1
4. Outcome of pregnancy preceding 1st insertion									
Live birth	4994	5.6	15.2	19.7	35.7	12.5	19.7	34.4	53.9
Induced abortion	1271	5.5	8.3	18.9	29.7	8.4	10.5	29.6	42.3
Other	253	6.6	11.9	21.6	35.5	7.5	14.9	33.1	47.4
5. Interval between last* preceding live birth and 1st insertion									
0-2 months	592	2.8	14.1	19.6	32.8	18.6	20.0	39.6	60.7
3-6 months	856	3.2	14.3	23.4	36.4	12.2	21.0	38.3	57.2
6-12 months	1349	7.5	16.4	20.3	38.3	11.5	21.1	34.7	54.3
12 or more months	2154	6.1	15.1	18.0	34.6	11.7	18.6	31.5	50.8
6. Type of device at 1st insertion									
Loop 1	4351	6.8	12.9	18.1	33.5	13.4	16.7	31.4	50.5
Loop 2-4	922	1.9	11.5	25.1	35.0	1.9	14.1	43.4	52.3
Coil	1372	3.4	17.2	23.6	38.9	7.0	21.2	37.5	54.2

TABLE 2 (CONTINUED)

Characteristics	12 months of use					24 months of use			
	N	P	E	R	T	P	E	R	T
7. Contraceptive methods used prior to 1st insertion									
None	4422	5.6	15.0	19.4	35.4	12.1	19.2	33.3	52.7
Ota Ring	685	5.0	8.5	14.5	25.6	8.8	10.6	28.4	41.6
Other methods	1364	5.3	11.6	22.5	35.2	10.9	15.0	35.2	51.0
8. Husband's education									
No formal education	784	6.1	11.8	11.3	26.5	8.4	15.0	22.8	39.9
Primary education	3213	5.3	12.9	17.7	32.1	12.2	17.8	29.8	49.3
Junior high	748	5.9	15.0	23.1	38.6	11.3	16.5	38.3	54.3
Sr. high or more	1816	5.7	15.6	25.0	40.3	11.9	18.2	41.4	57.8

N = Number of women
P = Cumulative net pregnancy termination rate
E = Cumulative net expulsion termination rate
R = Cumulative net removal termination rate
T = Cumulative net total termination rate

*For women whose last preceding pregnancy ended in a live birth.

persists when examined within strata classified by number of previous pregnancies. We conjecture that the relatively greater distance traveled to the clinic by women from outside Taichung selects for higher motivation which expresses itself in a greater toleration of side-effects.

Third - see variable #6 - users of loop 1, a small size of loop, have a higher pregnancy rate than users of loops 2-4, representing larger sizes of loop. This pregnancy differential, substantiated in other work, has been the reason why in the Taichung program the small loop has been abandoned in favor of larger size loops.

Finally, with respect to variable #8, wives of highly educated husbands show barely higher pregnancy and expulsion rates but substantially higher removal rates than wives of less educated husbands. The same results obtain when classification is by wife's education. Perhaps because well educated persons have more family planning alternatives available to them, they are less disposed to persist with IUD in the face of side-effects. Preliminary results from the island-wide follow-up study indicate that the well educated more often turn to other methods after discontinuing IUD than do the less educated.

To summarize: in the present Taiwan sample, the rate at which first segments of IUD are terminated has proven high in absolute terms though not relative to the results of other follow-up studies. Over half the terminations have taken the form of removals and these removals show important differentials among groups classified by social and demographic characteristics. Presumably underlying these differentials is an interplay of physical and motivational factors which our analysis has only begun to explore.

FAMILY PLANNING EXPERIMENTS: A REVIEW OF DESIGN

W. Parker Mauldin and John A. Ross, The Population Council

The last few years have seen a number of field experiments of relatively large size, all dealing with highly sensitive matters, and carried out mainly in underdeveloped areas. They have been inspired primarily as guides to larger family planning programs, but they attract interest for other reasons as well:

-- They have produced a large body of cross-cultural data on the sociology of fertility and its reduction.

-- Their before and after sample surveys give insights into changes in knowledge about, attitudes toward, and practice of family planning; and also give information about demographic changes, both secular and planned, past and present.

-- They have illuminated diffusion dynamics, especially with reference to adoption of the new intrauterine device (IUD).

-- They have proven the feasibility of such field experimentation.

This paper focuses upon another aspect, their character as experimental designs.

The India-Harvard-Ludhiana Population Study: This project was a pioneer effort to apply the methods of epidemiology to the population problem. The project was designed to determine whether the population of rural India would practice a simple method of contraception sufficiently to make a significant change in the rate of population increase. In order to evaluate the effectiveness of the program, researchers needed dependable figures on population, births, deaths, and migration. In the course of securing reliable base-line figures, deaths were examined for cause, births for factors favoring survival, and migration both to aid in getting base-line figures and to learn the reasons for movement. The underlying purpose of the program was a quantitative evaluation of the factors affecting fertility and population movement.

The study was done in 16 villages in a rural area in Ludhiana District, Punjab State, in northwest India. The 16 villages had a population of about 16,000 in January 1960. The action program was run in three stages: an exploratory stage in which the contraceptive method to be offered subsequently was selected; a pilot study in which the acceptance of the method was tested; and "the definitive test" designed to measure the effect on the population, and particularly on the fertility rate. During the exploratory phase the foam tablet appeared to be the most acceptable simple method of contraception, and this single method was offered in the action program which continued for a period of four years.

The over-all program was designed along the lines of a laboratory experiment: the exploratory study was to determine the most acceptable of five simple contraceptive methods, the pilot study made a preliminary test of the acceptance and effectiveness of the method under field conditions, and in the definitive study a test population was measured against both a working control population (to isolate the effect of the contraceptive advice and materials) and a "blank" control population (no project workers except for the collection of demographic data from village officials). In order to test the effectiveness of the contraceptive program (and the medical influence in the working control villages), accurate and continuing figures on births, deaths, and the rate of natural increase were required in the three types of population -- test, working control, and "blank" control.

In each village, an area survey was made at the outset -- maps of the villages showing all residences -- followed by a census by residence. In addition to such demographic data, the staff monthly collected information on attitude, knowledge, and current practice of family limitation. And during the course of the study there was an annual census of all villages, except those in the "blank" control where the census was taken only at the end of the study, in 1960.

The results of this carefully designed study were disappointing. After two and a half years of a sustained program, 17% of couples established themselves as contraceptive users. Furthermore, acceptance and use rates followed a characteristic pattern: although 80% reported in the preliminary survey that they wished to learn a contraceptive method, only about 45% of the couples used some method of family planning at some time in the study period. Perhaps 30% at any given point was the largest participation in the program, and that figure then fell to 17%. Observed efficiency among regular users of foam tablets was 63%.

The experimental design was excellent; but in our view it was over-elaborate for the program input. I think it is not hindsight to say that dependence on a single method was unnecessarily restrictive, and in this case apparently doomed the experiment to program failure, however successfully the design was adhered to and however accurately the data were collected. A more fundamental criticism is that the informational and educational aspects of the program were limited in scope, in theory, and in application. Frequent home visits where there was face-to-face contact were wrongly regarded as frequent "intimate" contacts. Each worker was left to his or her own devices as to what was said at each visit; there was no systematic plan of what messages were to be imparted at each visit, no systematic recording of what transpired at such meetings other than the collection of demographic data. Long before the experiment had run its course it was clear that the impact of the program on the fertility rates would be minimal.

The Singur Study: Singur is in West Bengal, not far north of Calcutta, and is the site of the demonstration and research center of the All-India Institute of Hygiene and Public Health. This study is an outgrowth of the Harvard-Ludhiana study and was designed to offer several simple methods of family planning, using local field workers rather than high-level workers from outside the area, and to provide more systematic information about family planning.

The study covered a group of 8 experimental villages with about 7,500 population, and 15 control villages with about 13,000 population. Both male and female field workers were used, usually a husband-wife team. Preliminary contacts were made with village leaders, elders, and medical practitioners. Community meetings followed, separately for men and women, with a maximum attendance of 200 for women and 250 for men. The major part of the action program consisted of small group meetings with an average attendance of four to six, lasting an hour or more, and individual meetings between field workers and married men or women at their homes approximately once every two months. At a typical group meeting the field worker gave a prepared talk covering (1) introduction to the subject as part of the activity of the Singur Health Centre, (2) economic and health reasons for family planning, (3) physiology of human reproduction including especially the cycle of ovulation and menstruation as a basis for the use of rhythm, and (4) methods of family planning, initially rhythm, coitus interruptus and foam tablets -- at a later stage the condom was used, then vasectomies, and still later the IUD's. After the prepared talk there was free discussion within the group. The teaching was reinforced with visual aids: flip charts, flannel boards, and flash cards, each telling a story related to family planning. Wall calendars were distributed as a reinforcer and reminder of the program, the first with a portrait of Tagore, the Indian poet and Nobel prize winner, and a quotation from him that condemned having more children than could be well cared for.

The results from this study showed that fertility rates fell more rapidly in the experimental than in the control area; more important, this was perhaps the first program in a rural, underdeveloped area to demonstrate that fertility rates could be brought down using conventional methods of contraception. Periodic surveys demonstrated that although couples were learning about methods of family planning, in the early months they often were misinformed about physiology of reproduction, about rhythm, and about the mechanism of action of foam tablets. Information collected in this manner was necessary for the continuing revision of the educational program.

The Koyang Experiment: In 1962, near Seoul, Korea, 14 villages became the target of two birth control programs. One program was actually nationwide in scope, but was studied by investigators only in a group of seven villages called the Kimpo area. On the far side of a large river, in the Koyang area, seven other villages were exempted from the national program and received instead a more intensive local program. Altogether, some 3,200 married women in the childbearing ages were affected. Before, during, and after measures were obtained on a sample of 1,200 women. Preliminary results were quite clear that both programs succeeded, with the local, intensive program being more effective than the less intense nationwide program in raising levels of contraceptive use, in reducing pregnancy rates, and in spreading favorable attitudes toward family planning.

Dacca, East Pakistan: This project illustrates the complexity and difficulties that are encountered in a sizeable field experiment. A principal aim was to compare the effectiveness of four treatments for spreading the practice of family planning. The four treatments were executed in four different areas of the city, each involving visits by trained field workers to the home. The four treatments differed according to who was contacted:

1. Both men and women
2. Men only
3. Women only
4. No one (the control group was served by a pre-existing and quite low key government program)

The hypothesis was that the treatments would be effective in the order listed.

Each treatment was brought to bear on one of four government "housing colonies." These were selected from a larger group of colonies and were matched on a number of demographic characteristics. Each colony contained 150 or more apartments, and all colonies were filled with clerical or lower scale government workers. The four colonies were well insulated from each other spatially, and residents of each colony worked mainly in the same governmental agency.

A before-survey of demographic characteristics, fertility, and family planning attitudes, knowledge, and practices was conducted in each of the three areas receiving stimuli. It was recognized that in theory the questioning involved in the pretest might change behavior, but the design did not permit a measurement of this effect.

Several features of the experiment were calculated to permit generalization to the sort of groups that would become future targets. First, residents of the four colonies were not representative of Dacca's population. Instead, they constituted a homogeneous type which would receive high priority in family planning efforts. Second, the treatments applied called for cost and personnel requirements that were realistic for general application. Third, the records systems used greatly improved the action aspects and were realistic for wider application.

Strenuous efforts were made to keep field procedures uniform over all treatments. Workers of about equal qualifications were obtained and underwent careful, uniform training and similar supervision. In addition:

1. Home visitors were rotated among the three stimulus colonies.

2. A clinic was set up in each of the four areas to dispense contraceptives and information; the four staffs were rotated quarterly.

3. All action procedures were checked constantly to keep them uniform, i.e., initial home visits, follow-up visits, group discussions, large group meetings, etc.

4. Clinic and field records were kept to monitor all inputs to the colonies, including some stimuli not part of the family planning program. All this required supervision and painstaking maintenance so that, insofar as possible, the only difference between the treatments would be the sex channel used: men, women, both, and neither.

The clinics afforded a measurement of the results. Attendance at the clinics and the flow of contraceptive supplies were two quickly available measures. Diffusion could be gauged by the number of persons from outside the colony coming to the clinic for information and service. The after-survey gave additional measures of program effectiveness.

The Taichung Study: One of the most ambitious and successful field experiments in the social sciences ever conducted occurred during 1963 in the city of Taichung, Taiwan. The population of 300,000 was allocated to a twelve-cell design and exposed to nine months of stimuli to encourage family planning. Four different types of programs were instituted:

1. In all areas posters were displayed and mass meetings held in various public halls.

2. In some areas in addition to the above, letters and pamphlets were sent to married couples with at least two living children and to newlyweds.

3. As an additional program input in some areas field workers also visited every married woman aged 20-39 in order to inform her about family planning services; small group meetings were also held.

4. Finally, husbands were also visited in some areas.

These treatments were mixed in different proportions in three more or less similarly constituted sectors of the city. In the first sector, one-half of the neighborhoods received home visits; in the second sector, one-third; and in the third sector, one-fifth. "Mail" and "nothing" treatments were distributed equally in the remaining neighborhoods in each sector. For political reasons, the use of mass media was kept to a minimum and the posters and letters made no reference to the IUD. The target population was about 36,000 married couples, with the wife being between ages 20-39. This study is notable for its size, its multiple design features, the inclusion of IUD's as a method of contraception, and the fact that the study was undertaken in an area of relatively high literacy and, for an underdeveloped country, among a population with moderate levels of living.

From its start in February, 1963 to April 1, 1964, 5,454 couples accepted family planning guidance in the program. Nearly 80% chose the IUD; less than 2% the oral pills; and about 19% traditional methods, mostly condoms. The proportion choosing IUD's was least among those exposed to home visits, presumably because field workers gave explanations about various methods whereas diffusion of information by word-of-mouth was concentrated almost entirely on the IUD. The home-visit treatment yielded the highest returns, but additional visits to husbands did not yield much. The mail campaign was ineffective. There was not much difference between the light and medium density sectors, but the heavy sectors attracted appreciably more contraceptors than did medium and light sectors. This study is continuing in modified form and the proportion of users continues to rise.

Other Studies: This is not an attempt to list the various important family planning pilot or demonstration projects that have been undertaken. If we were to compile such a list we would include the large Sungdong Gu study in Seoul, Korea, modeled along the lines of the Taichung study; the Barbados experiment which combines an IUD program with an attempt to eliminate cervical cancer; various studies in Latin America designed to reduce the high incidence of abortion; the highly successful Photharam project among a population of about 70,000 people in rural Thailand; the Bogue studies in Chicago and the rural South; the comparative study of oral tablets and IUD's now underway in Puerto Rico; the studies in Taiwan and Korea of acceptance and continued use of oral tablets at different cost levels; and the comparative studies of what is called the hospital postpartum program -- a program to determine to what extent family planning can be spread through large maternity hospitals. This study involves about 20 hospitals in about 15 countries, and includes several of the largest "baby factories" in the world, those having 35,000 to 50,000 deliveries a year.

These experiments have been designed to determine (1) whether a given contraceptive method will be accepted and used and have an effect on fertility rates, and/or (2) whether a given treatment brings about a significant change in fertility. There are intermediate objectives, such as an increase in knowledge about physiology of human reproduction, contraceptive methods, location of clinics, etc., but the ultimate objective is to affect fertility.

Program inputs are almost infinitely variable, and are a composite of: (1) contraceptive methods, (2) informational content, (3) type of media, (4) frequency of stimulus, (5) intensity of stimulus, etc.

Some of the experiments have been designed to measure whether or not a given program has an effect, and typically differences between a control and the experimental population are used as measures of effectiveness. In other experiments two or more matched population groups are selected, different programs are introduced, and differences in the responses and behavior of the various populations are assumed to measure the relative effectiveness of the programs; in such cases control populations are sometimes, but not always, used.

It is difficult to identify and describe many of the stimuli in a manner that they can be replicated. A program may include posters, occasional radio "spots," group meetings, home visits, and information plus service at health clinics. One can describe and replicate the posters, the radio "spots," but not the group meetings or home visits. One can substitute another group meeting, another home visit; one can obtain information about how the group meeting was conducted, but it cannot be reproduced with precision. Indeed, records are not kept in sufficient detail to tell what went on at group meetings, although the general purpose and approach are known. In practice, generalization is also limited by poor specification of the services offered; for example, where clinics are located, what are their procedures, how are clients received and treated, etc. Even if this were known, an important source of variation would remain obscure. This is the set of unique personalities administering the program and supervising contacts with the populace. One could design an experiment which would largely remove this as a variable, but it would be time consuming and expensive, and of questionable worth.

Criteria: The measures by which stimuli are judged to be effective or not are mainly of three types: (1) acceptances, as indicated by IUD insertions, sterilizations, the practice of rhythm, and the taking of supplies of ordinary contraceptives; continued use is equally important, though less well reported; (2) fertility levels and change; and (3) KAP surveys designed to give information about knowledge, attitudes and practices of family planning. Information about acceptances within the program is readily available from service statistics, but special follow-up studies are necessary to give information about continued use. Locating initial acceptors who do not return for supplies or service is often very difficult. Sample studies of the population included in the program area gives information about persons who receive their information and services from outside the program -- from friends, private physicians, etc. -- and also can give one information about "dropouts" from the regular program. Very few of the developing countries have good vital statistics, and the typical survey procedure does not produce consistently valid estimates of birth and death rates. Several of the studies have used annual surveys or censuses supplemented by reports throughout the experiment from midwives and other knowledgeable people in the area. Some of the experiments have attempted to collect information from two independent systems, one a census and the other a modified registration procedure. Some thought is now being given to use of the "open interval," or length of time since termination of last pregnancy, as an early indicator of change in fertility -- if the distribution of months since termination of last pregnancy for women in the reproductive ages shifts so that the average interval is longer than before, particularly for high-parity women, this would indicate lower fertility rates.

Experimental Groups: Units assigned to the experimental groups have varied widely, but they have always been areas, not individuals. They have included villages, housing colonies, neighborhoods, and counties. The ideal of creating experimental groups by random assignment has been realized partially, even impressively, given the magnitude and difficulty of operations involved. The principle of randomization has been followed in order to equalize the experimental material acted upon. For example, within each sector of Taichung and Sungdong Gu each of the thousands of neighborhoods was randomly assigned to treatments. In Dacca, each of the housing colonies was randomly assigned to one of the information channels.

Contamination: There are many sources of contamination in experimental programs. One might worry about a change in secular trend due to more widespread education, to a rising level of living, to more news articles about what other countries and groups are doing about population. One might reason that a before-survey which asks questions about family planning, and in the process almost necessarily gives some information about it -- one might reason that this is a contaminating influence. Some of the experiments have been designed to take advantage of natural channels of communication and to encourage diffusion of information.

Experiments have given some thought to such matters, but have not been greatly worried about most of them. In the first instance, it has proven to be rather difficult to change attitudes and behavior, particularly behavior so fundamental as having or not having children, of spacing or not spacing births. The problem has been to find a composite program that will have any effect at all! Further, the program specialists have been looking for things that make a considerable difference, not those that make a small difference and are hard to measure. Even so, attempts have been made to choose experimental and control populations that are geographically separated and which do not normally have much communication with one another. Care has been exercised to insure that the initial stimulus, the program stimulus is applied only to the population in the area specified; but no attempt has been made to decrease diffusion of information across experimental boundaries.

There is considerable evidence that diffusion of information, particularly about the IUD, is rapid and large in magnitude. It is likely that such diffusion works in the direction of reducing difference between the control group and the experimental group, or between the experimental groups with the least and with the most effective programs. Consider, for example, the differential effectiveness of a program that includes home visits, group meetings, well-located clinics with courteous personnel, posters, radio "spots," and all the rest. Call this the "everything" group. One may wish to determine whether a less intensive campaign is effective, say, a campaign that includes a few clinics, radio "spots" occasionally, and posters -- but no home visits, no group discussions. One may be concerned about the contamination of information given by people in the "everything" area to people in the "clinic" area. If the results of the experiment show that there is no difference between the two groups, and that the number adopting family planning is low, one's concern is not about lack of differences, but about lack of an effective program. If the differences are large, and in the expected direction, one is safe in concluding that the "everything" program is more effective than the "clinic" program. If the effects of both programs are large, and the differences are small, one would be considerably worried about contamination. The administrator would have to decide whether to invest funds for the "everything" program over a larger area; and if funds are tight he probably would try the less intensive approach, watch the results carefully, and plan to include more program input if initial results are disappointing.

Conclusions:

We are impressed with the size and complexity of these undertakings. The early studies seem to us to have worried too much about the measuring instruments, too much about contamination, too much about the possible influence of mild stimuli, and to have been too little concerned with major program inputs designed to inform substantial numbers, to answer their questions, doubts, fears, to offer a service economically, courteously, conveniently, privately. The design features of some of these studies have been very good indeed; most of them have been only fair. Some very sophisticated studies are needed to answer questions about the long-range effects of new contraceptive methods, about differences in effectiveness and side effects of the more promising devices and materials. But where the interest is to provide program information to an administrator who must make decisions about the number of field workers per unit of population, how they will go about their work, etc., it seems to us that the experimental programs we have referred to have given us substantial information. They could give us more information; but the aim is to provide useful information early, rather than more detailed information much later.

DISCUSSION

Henry S. Shryock, Bureau of the Census

I have found it helpful to restructure somewhat one of Parker Mauldin's classifications when considering the evaluation of family planning programs. The evaluation can be conducted by examining the presumed effects at at least two levels. First, there are the effects upon the knowledge, attitudes, acceptance, and persistence of use within the given population. Second, there are the effects upon population growth and particularly upon its fertility component. For brevity, we may call these the "KAP" effects and the "demographic" effects, respectively.

The KAP effects can be analyzed separately for individual methods and techniques of family limitation or of communication. These effects may be the most important concern of personnel whose responsibilities are limited to carrying out an action program in the field of family limitation. To the demographer, however, these KAP effects are essentially means to demographic ends; and the economic planner, in turn, looks ahead to the effects on such things as income per caput. With respect to the demographic effects themselves, moreover, the interest of economic planners tends to center on overall population growth, whereas the demographers are more likely to want to know what is happening to fertility and mortality -- and these programs do have effects on mortality. After the level and trend of the crude birth rate have been studied, the demographer soon begins to ask what is happening to completed size of family and even to child spacing and how those measures of fertility have been changed by a given program.

Expenditures on a given action program are reflected at both the KAP and demographic levels. The fact that not all of the change--or maybe none of the change--in the KAP or demographic variables is attributable to the given program, experimental or otherwise, is perhaps better recognized in the case of the former than the latter. The knowledge of contraception methods in a demonstration area may be received simultaneously from a broad general program and from a localized demonstration or experimental program. Since the demographic effects are causally further removed from the action program, the explanation of observed changes is even more complex, although administrators are sometimes quick to credit a decline in the birth rate, for example, to a particular program, even when there has obviously not been enough elapsed time between the presumed cause and effect. Changes in the birth rate, of course, may also be attributable to such factors as the changing structure of the population by age, sex, and marital status, to mention only one example.

As part of its second Five Year Plan, the Economic Planning Board of the Republic of Korea set as one of its goals the reduction of the annual rate of population growth from 2.9 percent, in 1955-60, to 2.0 in 1970. The Ministry of Health and Social Affairs was first asked to arrange for the insertion of 1 milion I.U.C.D.'s in calendar year 1966. Possibly there was some sort of crude calculus involved here as to what was required to produce the target rate although I have not seen the model, or even the scratchpad calculations. At any rate, the Health Ministry argued EPB down to 400,000, pointing out that there were not enough medical and paramedical personnel available and that the task of persuasion becomes progressively more difficult once the more desperate cases have come forward.

In a country with good vital statistics, such as Taiwan or Japan, it is feasible to think about estimating the effects of a particular action program upon fertility. This is a problem in imputation, factoring out the important external factors.

In most of the other East Asian countries, the prior problem is that vital statistics are either very deficient or totally lacking. Hence, it is necessary to improve them or to institute stopgap measurement systems--sample surveys, demonstration registration areas, "supra" registration systems on a sample basis, matching studies between vital events registered and reported in household surveys, etc. In this situation, concern with just how the birth rate is being reduced, if it is being reduced, is obviously quite subordinate.

Just as funds are used to publicize an action program in the family planning field, so funds could appropriately be used to propagandize on behalf of the registration of vital events, in order to improve the evaluation of the ultimate success of the action program. Presumably, funds for these two purposes could often come from the same source, or, in other words, some part of the funds used to publicize the action program could be used to publicize the importance of registering vital events.

What happened in South Korea shortly after the coup d'état of 1961 is instructive. As the result of underregistration in a decaying system of civil registration, the number of births registered in 1960 had fallen to 908,000, the equivalent of an unbelievably low birth rate. With a vigorous new administrator in temporary charge of the Bureau of Statistics and a modest expenditure of about $3,000 on a nationwide publicity campaign, registered births shot up to 2,387,000 in 1961. Many of these had occurred in earlier years, of course. Here again, it is difficult to say what part of the rise was attributable to the publicity campaign since there was no designed experiment.

Admittedly, complex analyses can be made of the KAP effects without regard to the demographic effects. Here vital statistics are not necessary. The double-decrement life-table technique described by Potter and his associates is an interesting example of this kind of analysis. Incidently, I wish we could develop a more self-explanatory terminology for what Tietze calls "net" and "gross" rates in this connection.

On the whole, however, I was a little disappointed in this afternoon's papers in that they were mainly descriptive in nature and paid limited attention to the technical problems of evaluation. Perhaps the American Statistical Association represents an audience so unfamiliar with the programs and problems in this field that a broad introduction was needed; but, on the other hand, it is certainly a good forum for technical discussion. I should have liked to have heard more, for example, along the lines of research by Tietze 1/ and by Lee and Isbister 2/ on the impact of birth control programs on fertility with explicit attempts being made to analyze the causal sequences, both theoretically and empirically. Mauldin discussed the use of a control population as a tool of such analysis, and it is certainly a useful tool.

Lee and Isbister also discussed very briefly the economic effects of a given reduction in the population growth rate. Obviously, this is a very complex problem in cost-benefit analysis because of our lack of knowledge about the interactions involved. As Takeshita suggests, the economic planners merely assume that a given drop in the rate of population growth is "consistent with their development plans".

In Taiwan it can be seen that good statistics breed good statistics. Not satisfied with a civil registration system that produces vital statistics and local population figures, promptly, accurately, and in considerable detail, Taiwan has gone ahead to design several imaginative sample surveys that yield valuable supplementary information on fertility. Freedman and Takeshita have been leading spirits in the formulation of that research.

A final comment on Takeshita's paper is that the acceleration of the decline of the crude birth rate in Taiwan somewhat antedated the provincial action program introduced in 1964. Perhaps then we need to know whether the acceleration itself was accelerated.

1/ Christopher Tietze, "Pregnancy rates and birth rates" Population Studies 16 (1): 31-37, July 1962.

2/ Byung Moo Lee and John Isbister "The impact of birth control programs on fertility" in: Bernard Berelson et al., editors Family Planning and Population Programs (Proceedings of the International Conference on Family Planning Programs, Geneva, August 1965), Chicago, University of Chicago Press, 1966: 737-758.

XI

COORDINATION OF FEDERAL, STATE AND LOCAL STATISTICS

Chairman, RAYMOND T. BOWMAN, U. S. Office of Statistical Standards

DATA REQUIREMENTS FOR ADVANCE FISCAL PLANNING*

Selma J. Mushkin

State and local governments have begun applying scientific inquiry and new technology to the affairs of government itself. They are improving decision-making processes by use of quantitative analytical techniques and by focusing conscious attention on the policy values that are implicit in program choices. New hardware permits the rapid processing of vast amounts of information; new budgetary methods clarify the types of information required. But the usefulness of these tools and methods depends upon acceleration of research and the consequent strengthening of the conceptual framework of the theory of public expenditure and of investment in human resources. And it relies on the availability of large amounts of relevant data.

At this stage of the scientific era in state and local governmental affairs, the technology of the computer outdistances budgeting methods, the theoretical framework for model building, and the data base. Thus, three of the four components lag behind. The work that is done toward more rational decision making will encourage application of second-best proxies when directly relevant information is not available. In some instances, far less precision may be achieved than is essential and as a result the effectiveness of the analytical budgeting process could be greatly impaired.

Measurement of outputs of public services and their inputs requires quantification, and therefore statistics and other precise information to be used in such quantification. I have indicated elsewhere that caveats and definitions of limitations are no substitute for quantification and that the quantification must be designed to provide answers to precisely defined questions so that the measurements can furnish a guide to choice among methods of meeting a program objective, or among programs. (1)

Quantification is needed for planning not only for single years but for other appropriate time spans. Advance fiscal planning, or planning for budgetary allocations, over, let us say, a five-year period creates additional data requirements that range over almost the entire spectrum of economic, demographic, fiscal, and social statistics, as the informational feed-in for projection and evaluation purposes. Accordingly, a complete assessment of data requirements would be a large undertaking, extending over a wide range of sources and data gaps. The task that was assigned to me was more restricted; namely, to review with you the data gap as it became apparent during the course of a study on state and local finances projected to 1970, a study initially undertaken for the Federal Interagency Committee on Economic Growth, with the multiple objectives of:

1. Determining the future impact of state and local governmental expenditures on the national economy.

2. Measuring the impact of national policy directions on the expenditure programs of the states.

3. Experimenting with a new measure of fiscal capacity, a measure that would help answer the question: What are the differences in the capacity of the states to meet their future expenditure requirements?

More than a research objective was in mind: we planned that the study would be conducted in such a way that the states would be involved in the processes of projection and would be encouraged through this involvement to engage in their own advance fiscal planning and programming.

In a unique experiment, state officials designated by their Governors participated in carrying out the project work. More specifically each was asked to develop 1970 projections of his state's tax system within a specified framework of economic and demographic assumptions. Reactions were sought on the 1970 projections of population, personal income, labor force, wages and salaries, and so forth, and adherence to the specified assumptions for each state was requested.

Demographic and economic data

Advance fiscal planning in states and communities necessarily rests on a data base adequate to anticipate population changes and changes in economic activity both in the aggregate and by detailed components. In the simplest formulation of the advance fiscal planning problem the size and composition of the population affects the workloads of public programs and in a more complex formulation the level of performance of the economy affects both the revenue-take and the cost per workload unit.

Much research and collection of data are going forward in an effort to improve the economic and demographic information available to states and communities for programming purposes. We are a substantial distance from having even in broad outlines the types of information required. This paper does not undertake to deal

* This paper draws on the work of a project of research and education on state and local finances supported by a special grant from the Ford Foundation to The George Washington University, Washington, D.C.

with the basic demographic and economic information required for formulating a model or series of models in order to begin the processes of advance fiscal planning. While these are vital to an advance fiscal planning exercise they are dealt with so often and in so many different contexts that it seems desirable here to limit the discussion by excluding any detailed examination of such general types of data.

Two types of issues I should like to mention, however:

1. Population projections can be carried out in one central place, or independent population projections can be developed by states or local communities. Economy of operation and the dovetailing of projections so as to make a consistent set of aggregates point to national responsibility for carrying out this work. But familiarity with the local scene and awareness of forces making for local movements suggest that local governments carry out such studies. Between the two extremes lies a series of combinations of approaches to demographic studies that would involve the cooperation of a national population statistics center with local communities. Much more serious consideration should be given to such a cooperative effort than has been given in the past, which would provide more current population estimates for local areas and recurrent population projections for states and localities. [As an aside may I indicate that when we started our research on state-local finances for 1970 (2) there was not available a set of even statewide projections of total population that reflected the findings of the 1960 Census of Population.]

2. Similarly, projections of personal income in the states can be carried out by the individual states and communities or through central research undertakings. The advantage of a central undertaking in this instance, where the basic data requirements are substantial, is even greater than it is in the case of population studies, and the gains possible through linking state projections to national economic models are even larger. However, the carrying out of economic studies in a state or community can contribute to a process of understanding, an understanding of the direction in which a state or community's economy is moving and of the public programming that is required in the light of economic change. Such a process is an integral component of advance fiscal planning and program development. State and community economic studies, accordingly, should be encouraged. But immediate steps should be taken to extend the work of the Regional Economic Division of the Office of Business Economics so that projections of personal income in the states and counties become available. (3) National agencies or organizations, such as the National Planning Association, should be encouraged to experiment with other measures of income and with income distributions both for states and for local areas. (4, 5)

The national economic backdrop for the 1970 state-local finances study was provided by the Federal Interagency Study on Economic Growth and Employment Opportunities. The assumptions for the fiscal study were based on one of the several preliminary models of the national economy developed for the Interagency Study. While the state-by-state projections of personal income that were made within this national economic model were approximate, the national aggregates served as a checkpoint at each phase of the work. The projections were based on estimates concerning the labor force, by state, and on wage and salary trends, with a separate projection of each of the following additional income components: farm income, interest and dividends, nonfarm entrepreneural income, and three classes of transfer payments.

The study of 1970 state-local finances sought on a highly disaggregative basis to evaluate approximately 100 expenditure components for each of the 50 states and the District of Columbia and approximately half as many revenue items. Nationwide data that would assure comparability for all the states were used, rather than state or local data. The summary of data requirements for advance fiscal planning drawn from this study is presented below, starting first with several of the more important tax revenue sources, and then turning to selected expenditure components.

Property taxes

Property taxes continue to account for the major share of local tax revenue. In the fiscal year 1965, 87 percent of the funds collected through local taxes were derived through property taxation. The property tax is essentially a combination of taxes on different classes of assets, each with its own characteristic responsiveness to changes in economic activity and income.

As viewed from the perspective of the local taxing jurisdiction projection of property tax yields at constant effective rates involves a knowledge of the assessed values of classes of property on the tax rolls, probable new construction activity in the jurisdiction, changes in market values of existing property as a consequence of changes in construction prices and in land values and land use, and of property depreciation. These are the items of information, accordingly, that are required for beginning an analysis of the impact on the property tax base of local conditions -- economic activity, family formation, and land use patterns. A considerable body of information is available locally. Records of tax assessments are kept by the local jurisdiction, building permits are issued by it, land use plans are formulated by it, and in some localities studies are routinely made of the changes in market value and ratios of property assessments to market values.

However, in a survey that was conducted as an extension of the work of Project '70, only about 30 of the approximately 200 counties reporting indicated that they made long-range projections of tax revenues. (6)

At least a partial data base for state projection of property taxes exists in approximately half of the states. Those states conduct regular statewide studies, on an annual or biennial basis, of assessment ratios or the ratios of assessed valuations to market prices. The studies involve comparison, on a sampling basis, of the sales price of sold properties with their assessed valuations. In some states supplementary appraisals are made for classes of property with low turnover rates.

Both the 1957 and 1962 Censuses of Governments took important steps toward achieving the types of information needed for projecting property tax yields at constant effective rates. The information that has been published suggests more data collection, more current collection, and tabulation of information collected in the course of the censuses of business as supplementary data on taxable wealth.

1. Producers durables and inventories--The Census information on market sales is limited to data on taxable real property. (7, 8) Producers' durable goods and inventories, however, are taxed under the property tax in jurisdictions in which over half of the population of the United States resides. (9) Year by year information is collected on purchases of producers' durable equipment by manufacturing firms in the states; (10) except for mining equipment, for which comparable data are gathered in Census of Minerals years, (11) there are no state statistics on other purchases of producers' durable equipment.

Inventory data are even more scant on a state-by-state basis. Though much of this inventory information is collected in the course of the censuses of business, the information is not tabulated. (12-15)

2. Industrial, commercial and public utility properties--Data on taxable real estate are limited by the technical feasibility of obtaining market sales information on classes of property with low turnover rates, the foremost example of which is public utility property. As is emphasized in the 1962 Census of Governments volume, Taxable Property Values, the problem of limited sales volume for calculation of assessment ratios from data on transfers of property is most acute for industrial property. (8) Ways must be found to improve the data on real property of public utility and industrial firms by obtaining supplementary assessor and appraiser information through cooperative arrangements with State Boards of Equalization. Even though the cost of carrying out the work would be considerable, the multipurpose use of such information suggests that a central collection or compilation of the information would pay off. Moreover, compilation of data collected by the states through their sample surveys while falling short of a national picture of taxable real property would help in gaining indicators of movement of such wealth in relation to changes in economic activity. On a more nearly current basis than is now possible, some tests for judging projections, by comparison with reality would become available. Such a routine compilation should be considered by the Governments Division of the U. S. Bureau of the Census.

3. Farm property values--Two sources of data on farm property values are now available, through the Census of Governments and the Department of Agriculture. Neither source yields a wholly adequate measure of change in market value of farm property. There is reason to believe that the Census ratios of assessments to market values of farms, derived from a sample of sales of properties, are lower than such ratios for farm property generally. A disproportionate share of farm properties sold are those with greater acreage values, undergoing conversion from farm to nonfarm uses. Application of such data results in relatively high estimates of market values of farm property. Department of Agriculture data on the other hand tend to understate the effect of potential conversion of farm land to urban uses. (16, 17)

4. Current indicators of taxable property--One set of problems in developing advance projections of property tax yields concerns how to obtain data required for analysis of determinants of changes in property values. Another concerns how to obtain indicators of current changes in such values. Construction statistics compiled by the McGraw-Hill Publishing Co. and F. W. Dodge Company are helpful as such indicators. Local community information on building permits and land use is another source of indicators. To evaluate adequately current movements in relation to property tax yields, moreover, more and more current data are required on changes in land prices, as well as possible indicators of such land-price movements.

Income and sales taxes

While the property tax is the principal tax source for local governments, there is no one tax that predominates in the revenue base for state governments as a whole. However, sales taxes and gross-receipts taxes, together with income taxes, account for well over three-fourths of the tax revenue of state governments. In the administration of those taxes, a considerable body of data is made available that is useful for economic planning purposes as well as for analyses of the interaction between economic change and tax yield change.

In the administration of the individual income tax laws, states imposing the tax collect information on adjusted gross income, taxable income, number of exemptions, and so forth, for the individuals required by the statutes of the states to file tax reports. Similar data of course are collected by the U. S. Internal Revenue Service, but the alternatives available to the federal taxpayer in the place of filing limit the usefulness of the state information on the federal returns. Matching studies are needed to determine the "biases" involved in use of Internal Revenue data for state tax-base information. (18)

About one-third of the states that levy an income tax do not tabulate the income information from their tax returns; information on total adjusted gross income is not available. (19) In contrast, some states tabulate the income tax information by income size class and publish the statistics derived from the tabulations. (20)

In the administration of the gross-receipts taxes and sales taxes, sales data are collected that not only are a guide to analysis of tax-base changes as income grows, but also to market-type sales data. National information on cigarette consumption, by state, for example, is derived from the tax reports of the states and is compiled by the Department of Agriculture. (21) A considerable amount of the state-by-state data on alcoholic beverage consumption, compiled by the trade associations, also is based on tax reports. (22, 23)

Distribution of revenues from the general sales tax, by type of business, is reported by many states. The information by type of business is not uniform and reflects differences in sales tax records and in tax law definitions. In its annual report, Detail of State Tax Collections, the Governments Division of the U. S. Bureau of the Census reports revenue by type of business classification on a selectively abridged basis. (24)

To improve the data base for long-range projection of income and sales taxes a series of steps might be considered. Some of these are:

1. Provision by the states for a routine sampling of income tax returns would make data available on the changes in distribution of adjusted gross income in the states so that these changes can be related to growth in personal income and other variables. The sample tabulations would also provide the data required to estimate the potential impact of changes in tax law on revenue yield. Internal Revenue data on income need to be compared with such state data.

2. Additionally, a continuous small sample of taxpayers' returns that follows the same taxpayer from year to year not only would yield, over the long run, important information on the operation of specific tax provisions of the law, for example, on deductions for medical expenses, but would also become an important source of economic statistics.

3. Corporate income reported in the state for tax purposes, if tabulated, would provide the information required to analyze the responsiveness of corporate income as defined in each state to changes in national economic activity and to changes in corporate income in the nation. Tabulations of corporate income by industrial classification would further enhance the usefulness of the material for projection purposes.

More use of the information obtained on corporate income tax returns would improve assessments of past trends and past relationships between corporate income, by industry, in the state and in the nation, and would help provide better perspective on the factors that can be used for projection or predictive purposes.

4. Unless gross receipts or general sales information is tabulated, by business or industrial groups, long-range projection analysis tends to be deficient. As a minimum taxable sales of producers' durable equipment, building materials, contract construction and services should be separately tabulated.

As the material presented on property tax projections and on sales and income taxes suggests, the deficiencies are in the stock of statistics that can be used as dependent variables in studies of the factors determining tax-base changes and as indicators of the accuracy of the projections. Tax collection data, the data most frequently available, combine changes in the tax rate and the tax base (statutory and administrative); and variations in collections may reflect either rate or base modifications. While rate of tax may affect the value of the volume of the base, we need at the outset to separate those components in order to understand their longer-run movements in response to economic trends.

Public welfare data

The gaps in statistical materials for analysis of tax revenues, impressive though they may be, are fewer and less restrictive than are the gaps in information on public programs -- gaps in data on the requirements, on the extent to which they are met, on the groups receiving the benefits and services, and so forth. To suggest the informational gap, illustrations are given for several expenditure categories. Projections of public welfare payments by states or communities, for example, requires:

1. Information on numbers in the population at risk--The number of poor in each of the assistance categories can at best be only approximated. But the necessary income information is not available by state and county except for the past two decennial census years, 1950 and 1960, and then from special surveys. Data on income distribution are needed in order to cost alternative programs for all of the population at risk or for some segments of it.

Even projections that ignore the overall need of the groups at risk, and are based on actual recipient counts, are made difficult by the lack of a uniform standard count within a state. Transfers of caseloads from one categorical program to another, in response to "most favorable grant-in-aid matching", limit the usefulness of recipient statistics, by state, published by the U. S. Department of Health, Education, and Welfare, Welfare Administration. (25-27)

Data on the number in the population in certain categories, such as the blind or the disabled, are not available on a statewide basis. And, in general, projections of population by broad age groups, by state, are available only intermittently.

For past periods, the data now available by state do not yield answers to such questions as: What portions of the number of persons in need have received assistance? Has the proportion of poor assisted by the welfare programs increased or decreased? Has the number of public assistance recipients changed with the changing number of persons in need? For purposes of program projection, the best available data are the administrative data flowing from program operations, and these are subject to the limitations of incomparability due to administrative or legislative revision.

2. Information on determinants of payment levels--When one turns to the problems involved in projecting the amounts of payment per recipient the data become even thinner. What has been the experience in the past regarding changes in average levels of assistance per person aided? What has been the change in the margin or difference between "need" and "resources", that is, in the two factors used within the program to set payment levels?

At present, each state defines need, or its assistance standard, by determining the quantity, quality, and price of the expenditure items in the family budget for which it will make provision under its assistance program. The number of persons with income below the defined standard depends upon the proportion of persons with low income in the state and upon such eligibility requirements as (a) limitations on the amount of property and income that a recipient may possess, and (b) the length of his residence in the state. Average amounts of payments per recipient are affected by the level of a state's assistance standards and the resources (income and property) that recipients have. In some states average payments fall below those set by its assistance standards as a consequence of (a) a statutory maximum on the amount of assistance that can be paid to an individual recipient, and (b) inadequate finances that compel reduction in assistance payments across the board for all recipients.

Assistance levels, accordingly, are affected by a variety of factors, including changes in income distribution in the state, in the cost of minimum budgets, in price movements, in taxable capacity. Data on income distribution are not adequate. Cost of living indexes are not available.

The Welfare Administration of the U. S. Department of Health, Education, and Welfare has made sample surveys of the characteristics of public assistance recipients, but the surveys are far too infrequent. (28) Routine collections of data are needed on sources of income of assistance recipients by age, family size, and so forth, that with other things can clarify the supplementation of other public payments through assistance. (29) Moreover, a continuous sample of assistance recipients and low income families would help the decision maker in assessing the progress made toward eliminating poverty. The present statistics on reasons for the closing of cases fall short of providing guidelines for policy decision. (30)

Health and hospital outlays

The paucity of data on state and community levels that could provide a basis for formulating health care programs and assessing existing or perspective public hospital and health expenditures contrasts sharply with the body of statistics and estimates now available nationally. (31) Some illustrations may help to clarify the types of information required for program projection and analysis. We need information on:

1. Health status of the population by state--The National Center for Health Statistics, which has developed out of the 1956 legislation authorizing continuing collection of data on morbidity, has filled the void in national information on the health status of the population. However, except in a few places, comparable data are not available regionally for health planning purposes. Mortality data by cause of death are not a good proxy for data on health status. For one thing, while the fact of death is unambiguous, the cause (or causes) of death is not. For another, many of the illnesses in a population, especially those illnesses affecting days lost from work, do not cause death.

Methods must be found to foster collection by states and communities of data on the health status of their residents if orderly health programming is to become possible. We need to know how health status varies with age, sex, education, race, income, and so forth. The National Health Survey has experimented with methodology through two types of surveys; in one the sample data are collected through family interviews, in the other through sample physical examinations. National health legislation must take account of differences among states and communities in health program needs, but the data required for that are not available. The National Health Survey samples are not large enough to provide state data.

2. Health resources--Data on the physical facilities available for the provision of health care and on the personnel -- physicians, dentists, nurses, physical therapists, speech therapists, nutritionists, dental hygienists, psychologists, and so forth -- are of varied adequacy. Population Census data on health manpower yield a set of figures on state distributions that do not correspond with information from professional associations. Much of the available regional information on health manpower has been compiled and published by the Public Health Service. Such critical information as number of physicians is not current and the necessary data to adjust the information available to yield estimates of the number of physicians in each state providing services to patients are even more scanty. Data on physicians' offices and the types of equipment and paramedical personnel available to provide the range of health services generally are lacking both for the nation and the states.

3. Health program expenditures--A major gap in data for state and community planning purposes and for long-range projection is the lack of health-expenditure statistics that can serve to throw some light on what is purchased in the private health sector, so that one may judge the public sector's role in undertaking to compensate for the deficiencies in the private sector. (32) Where special studies have been made of health expenditures that permit the public agency expenditures to be fitted into the overall picture in the community, the findings have been most illuminating. In New York City, for example, it was found that nearly a third of the bill for personal medical services in the city is paid out of public funds. (33)

When one turns from the expenditures for health generally to those for the segment of health services that is termed "public health" the data problems are complex. Two sources of data are now available; the expenditure data collected from the state health agencies by the Public Health Service, and those collected by the Governments Division of the U. S. Bureau of the Census. (34, 35) The definitions of the two series appear to be fairly similar and the national aggregates are no longer far apart, but the state-by-state figures are more disparate for some unexplained reason. The only source of information on public-health work by activity is the Public Health Service's compilation from state reports, but even this series lumps all expenditures of local health agencies so that the distribution among the multiple activities -- the many splintered activites of public health work -- is essentially restricted to state public health expenditures.

A growing segment of public expenditures is represented by payments to private hospitals and other providers of care for the medically indigent. Data on such payments to all hospitals under the public assistance program's vendor payments provisions are collected by the Welfare Administration; and data on vendor payments, additional to the public assistance program payments, (36) are part of the informational reports compiled by the Governments Division of the U. S. Bureau of the Census, but essentially there is no source of information on the total public payments to private hospitals, by state. Moreover, the data on the number of persons receiving assistance with their medical bills, and the share of the bills that is financed in this way are inadequate.

An exception to the "data are inadequate" rule is found in the statistics on mental hospitals and institutions for the mentally retarded. As an example of what can be achieved through federal-state cooperation by way of comparable statistics the data for the mental institutions should be carefully studied. A result of the 15-year effort by the National Institute of Mental Health and the states is a data base for trend and regression analyses. (37)

Some states and local communities have begun the process of compiling information on the health resources in their jurisdictions, and a few have sought to gather the data necessary to evaluate public expenditures for health care in the context of the total health outlays. A few places have undertaken to collect sample data on health status. Programming for health work and costing of this work over a period of years require those types of data.

Public investments for physical facilities

Advance planning for public works has been achieved more quickly than has long-term budgeting for current operations, and is applied more extensively. The fiscal pressure to schedule various projects has led to this concern of communities and states with public-works planning. Moreover, federal agencies, by their grant programs, have encouraged long-term planning of expenditures by states and communities for physical facilities.

Planning imposes certain requirements for information. It points to the need for data on existing stocks of capital assets: for example, the number of hospital beds, the number of miles of paved roads, the number of classrooms, the number of libraries, and the number of playgrounds. It points to the need for data on the age of existing stocks, together with data on the length of effective life of the existing facilities. It requires an engineering type of assessment of additions and expansions as a consequence of emerging patterns of population and industrial growth. And it necessitates some knowledge about new technology, and about population and industrial movements that will create obsolescence. Furthermore, no assessment of needs for physical facilities is free of value judgment about what is desirable: for example, what the desirable size of a class should be.

Those informational requirements define the nature of the statistics needed.

1. Stocks of public facilities--Figures on stocks of public facilities are far from adequate on a regional or a national basis. Inventories have been made of certain types of public assets. For example, in projecting school construction as part of the state-local finances study we used an inventory of school facilities made by the Office of Education. (38, 39) Data on hospital beds are reported each year in the "Guide" issues of Hospitals, published by the American Hospital Association. An inventory of municipal water facilities, made by the Public Health Service, is available. (40, 41) Data on existing mileage of surfaced roads, by type, are collected by the Bureau of Public Roads. (42)

In not all instances are inventories of public facilities available, and frequently inventories that are taken are not kept current. For example, the number of classrooms required for alleviating or eliminating overcrowding was based on a 1962 inventory updated for 35 states to the spring of 1964. A routine updating procedure is needed for all public facilities to provide, by state, information on what exists, which can then be related to what ought to exist.

2. Data on age and on depreciation of public assets--Construction requirements for replacement purposes can only be quantified if information is available on a realistic basis, of the remaining functional life of existing facilities. Some inventories of public-facility needs do not note the age of existing facilities. For example, from the Office of Education Inventory of School Facilities we know the number of classrooms built prior to 1940, but we know nothing about the upkeep of the schools or their current stage of repair or disrepair.

3. Capacity of facilities to meet workloads--To project the need for public facilities information is needed on the extent to which the added number in the population or industrial growth affects the requirements for such facilities. How much by way of dormitories, for example, is needed to accommodate the growth in college and university enrollments? A range of data problems is involved in answering such a question, even after the size of the enrollment increase in a state has been established and the number of out-of-state students estimated. What proportion of the students are likely to be married? What proportion to be single? What share of the number of students will live at home? What are the alternative student living arrangements? What share of the increased number of students can be accommodated in each of various ways? To suggest some of the turnabouts that the real world produces in terms of statistics after-the-fact, we may ask: What influence does the availability of dormitories have on enrollments? On enrollment of male and of female students? Assumptions were made in Public Spending for Higher Education about dormitory space in relation to enrollment increases, based on previous experience, and square footage of construction was costed, by state. (43)

Similarly, as is indicated in Transportation Outlays of States and Cities, highway construction often creates its own use so that a road constructed to relieve congestion may itself contribute to further congestion by making automobile use more convenient than it was earlier. (44)

4. Public facility plans--A number of states and communities plan at least some components of their capital outlays in advance, and others have well-established capital budgeting. Compilations of such expenditure plans by states, and on a sample basis for localities, would, over a period of years, provide information for making projections and testing techniques. Sample surveys are now made of prospects for industrial capital outlay, but similar surveys have not been made of public sector outlays. The Standard Land Use Coding Manual prepared by the Urban Renewal Administration and the Bureau of Public Roads make possible the collection of information on public-facility plans on a standard basis. (45) The use of this manual would simplify the classification of facilities and would fit this classification into a general land use pattern that is projected for communities and states. Not only would the work of projecting be improved, but the public facilities could be planned more effectively in relation to other land use patterns in the community.

5. Price indexes--Information on interstate variations in construction prices and on changes in such prices over time is needed to help understand changes in amounts spent for capital outlays.

We need to develop for classes of public services and facilities standard bundles of goods and services, and to provide for the routine collection of price data so that appropriate price indexes can be developed. The collection of price data should be carried out state by state so that a more precise statistical basis may become available for interstate comparisons. A beginning should be made on the fact-gathering that would as a minimum permit correction for price variation of dollar measures of revenue collections, expenditures, and of income. In summary, what we require is: (a) a set of price indexes for state and local public expenditure and revenue components and (b) state-by-state indexes. The research of Hurwitz and Stallings on interregional differentials in per capita real income change and the recent continuation of this research by the Office of Business Economics demonstrate the usefulness of cost of living indexes in interpreting income variations. (46)

Summary and conclusions

We have sought in this paper to draw on the experience of Project '70's study of state-local finances to describe some of the data limitations and gaps as they were encountered in the conduct of the study. We noted earlier that national data sources were used, that is, nationally collected data, or data compiled by national agencies from reports from the several states. Comparability of definition and of reporting for "state's own" fiscal planning purposes would at first glance appear to be less important than it is in a nationwide study that undertook to project the finances for 1970 for all the states. However, clearly this is not the case. Interstate comparison is one yardstick frequently used in formulating programs and tax policies. Questions raised frequently are: "What are neighboring states doing?" and "What are the program levels in jurisdictions of similar industrial development or size class?"

The steps taken by the Governors' Conference to gain greater comparability of statistical information for state and community planning purposes underscore this concern. Intergovernmental cooperation in the carrying out of a program for greater comparability will give more weight to the type of data collection now carried out by the Governments Division of the U. S. Bureau of the Census, and by the Welfare Administration, the Office of Education, and the National Institute of Mental Health, all three of the U. S. Department of Health, Education, and Welfare.

As more of the states and communities adopt advance fiscal planning and longer-term programming, data requirements will be enlarged. Some of the administrative-type statistics now kept by states and localities may prove useful, others not. Increasingly "requirement" data will be emphasized that can help to assess what the public programs are doing to meet a problem or a program objective. It is clear even at this early stage of work on intergovernmental aspects of program budgeting that far more serious attention will have to be given to direct federal and direct state public-program operations in cities and other localities. Local action on harbor development cannot proceed without knowledge of the expenditures and work planned by the Army Corps of Engineers; local action on assistance for the aged cannot be formulated without knowledge of the number of persons receiving old-age insurance benefits and the prospects for the immediate period ahead of growth in the insurance rolls. Needs for general relief cannot be quantified without knowledge about the state unemployment compensation benefits. Public hospital planning has to proceed from adequate information on Medicare benefits for the aged and the number of persons in the locality that are eligible. In the new setting of computer technology, program budgeting, economic model building, and informational systems, the regional distribution of direct federal activities becomes an important datum for state and community planning purposes.

References

(1) Dorfman, Robert: Measuring benefits of government investments. Washington, D.C., The Brookings Institution. 1965.

(2) Mushkin, S. and Lupo, G.: Project '70: Projecting the state-local sector. Review of Economics and Statistics. (Future publication)

(3) U. S. Department of Commerce: Survey of Current Business. (August issues)

(4) National Planning Association, Center for Economic Projections: State projections to 1975: A quantitative analysis of economic and demographic change. Washington, D.C. The Association, 1965.

(5) National Planning Association, Center for Economic Projections: The economies of fifteen metropolitan areas; historical and projected employment, output, population, and personal income 1950, 1957, 1960, 1962 and 1975. Washington, D.C., The Association, 1965.

(6) The George Washington University, State-Local Finances Project. (Unpublished)

(7) U. S. Bureau of the Census, Governments Division: Taxable property values in the United States. 1957 Census of Governments Vol. V. Washington, D.C., GPO, 1959.

(8) U. S. Bureau of the Census, Governments Division: Taxable property values. 1962 Census of Governments Vol. II. Washington, D.C., GPO, 1963.

(9) The Advisory Commission on Intergovernmental Relations: Measures of state and local capacity and tax effort. Washington, D.C., The Commission, October 1962.

(10) U. S. Bureau of the Census: Annual Survey of manufacturers. Washington, D.C., GPO, (Annual).

(11) U. S. Bureau of the Census: Census of mineral industries, 1939, 1954, 1958, 1963. Washington, D.C., GPO.

(12) U. S. Bureau of the Census: Census of manufacturers. Washington, D.C., GPO, 1963.

(13) U. S. Bureau of the Census: Census of business, retail trade and wholesale trade. Washington, D.C., GPO, 1963

(14) U. S. Bureau of the Census: Monthly wholesale trade reports. Washington, D.C.

(15) U. S. Bureau of the Census: Annual retail trade report.

(16) U. S. Department of Agriculture: Economic Research Service. Farm real estate market developments. Washington, D.C.

(17) U. S. Bureau of the Census: U. S. Census of Agriculture: 1959, Washington, D.C., GPO.

(18) U. S. Internal Revenue Service: Statistics of income, individual income tax returns. Washington, D.C., GPO, Annual issues.

(19) Replies to inquiry from the Federation of Tax Administrations on behalf of state-local finances project. The George Washington University (Unpublished).

(20) University of Wisconsin Tax Study Committee: Wisconsin's state and local tax burden. Madison, Wisconsin, The University of Wisconsin, 1959.

(21) U. S. Department of Agriculture: Economic Research Service: Tobacco situation. Washington, D.C. (Annual series)

(22) Distilled Spirits Institute, Inc.: Apparent average annual consumption of distilled spirits. The Institute (Annual series).

(23) Wine Institute: Wine consumption. The Institute (Annual series).

(24) U. S. Bureau of the Census, Governments Division: Detail of state tax collections. Washington, D.C., The Bureau (Annual series).

(25) U. S. Department of Health, Education, and Welfare, Welfare Administration: Trend report: Graphic presentation of public assistance and related data. Washington, D.C., December 1963.

(26) U. S. Department of Health, Education, and Welfare, Welfare Administration: Welfare in Review (Monthly).

(27) U. S. Department of Health, Education, and Welfare, Social Security Administration: Social security bulletin, annual statistical supplements.

(28) U. S. Department of Health, Education, and Welfare, Welfare Administration: Characteristics of families receiving aid to families with dependent children, November-December 1961, April 1963.

(29) U. S. Department of Health, Education, and Welfare, Welfare Administration: Tabular release on concurrent receipts of public assistance and OASDI by persons aged 65 or over, early 1964. December 1964.

(30) U. S. Department of Health, Education, and Welfare, Welfare Administration: Release entitled Reasons for opening and closing assistance cases, July-Dec. 1963, June 1964.

(31) U. S. Department of Health, Education, and Welfare, Public Health Service (National Center for Health Statistics): Medical care, health status and family income, Series 10, No. 9. May 1964.

(32) Reed, L. S. and Hanft, R. S.: National health expenditures, 1950-64. Social security bulletin, Vol. 29, No. 1, January 1966.

(33) Piore, Nora K.: Metropolitan medical economics. Scientific American, January 1965.

(34) U. S. Bureau of the Census, Governments Division: Census of Governments: 1962, Vol. IV, No. 4.

(35) U. S. Department of Health, Education, and Welfare, Public Health Service: State and local expenditures for public health. Washington, D.C., PHS (Annual).

(36) U. S. Department of Health, Education, and Welfare, Welfare Administration, Public Assistance: Vendor payments for medical care by type of service, fiscal year ended June 30, 1964. Washington, D.C.

(37) U. S. Department of Health, Education, and Welfare, National Institute of Mental Health: Patients in public mental hospitals and also patients in public institutions for the mentally retarded. Washington, D.C., GPO (Annual series).

(38) Mushkin, S. and McLoone, E.: Local school expenditures: 1970 Projections. Chicago, Council of State Governments, 1965.

(39) U. S. Department of Health, Education, and Welfare, Office of Education: National inventory of school facilities and personnel, Spring 1962. Washington, D.C., 1964.

(40) U. S. Department of Health, Education, and Welfare, Public Health Service: Statistical summary of 1962 inventory of municipal waste facilities in the United States. Washington, D.C., 1964

(41) U. S. Senate, Select Committee on National Water Resources, Water Resources Activities in the United States: Future water requirements for municipal use, 86th Cong. 2d sess., 1960.

(42) U. S. Department of Commerce, Bureau of Public Roads,: Highway statistics. Washington, D.C., (Annual).

(43) Mushkin, S. and McLoone E.: Public spending for higher education: 1970. Chicago, Council of State Governments, 1965.

(44) Mushkin, S. and Harris, R.: Transportation outlays in states and cities. Chicago, Council of State Governments, 1965.

(45) Resources for the Future: Land use information - a critical survey of U. S. statistics including possibility for greater uniformity. Baltimore, Johns Hopkins Press, 1966.

(46) Hurwitz, A. and Stallings, C. P.: Interregional differentials in per capita real income change. (In) National Bureau of Economic Research, Regional income. Studies in income and wealth, Vol. 21.

INFORMATION FOR URBAN TRANSPORTATION PLANNING

G. E. Marple, U.S. Bureau of Public Roads

A continuing comprehensive transportation planning process in which the State highway departments and the local communities are cooperating is now underway or is being organized in all of the 230 urban areas over 50,000 population. It is necessary to include the qualification "or is being organized" only because 5 of the 230 passed the 50,000 population mark only recently.

Much of the recent emphasis on transportation planning stems from the Federal-Aid Highway Act of 1962. A section of this Act declared it to be "in the national interest to encourage and promote the development of transportation systems embracing various modes of transport in a manner that will serve the State and local communities efficiently and effectively." It instructed the Secretary of Commerce to cooperate with the States in the development of long-range highway plans and programs which are properly coordinated with plans for improvements in other affected forms of transportation. It also specified that the plans be formulated with due consideration to their probable effect on future development.

This directive establishes the broad scope of the information required for transportation planning. While the basic problem is to project growth of transportation, to do this it is first necessary to project growth and shifts of the population and economy. To assess the probable effect of transportation improvements on future development requires knowledge not only of how transportation serves the land but also how it can itself influence land use. And the coordination of plans for all modes requires knowledge of the factors that affect the division of travel between the various modes.

Once future travel has been forecast, the demand must be measured against the capabilities of existing facilities. The planning to overcome deficiencies must seek the best balance among the three choices of new construction, upgrading of existing facilities, and traffic engineering improvements. And, insofar as roads and streets are concerned, there must be a division into systems for purposes of administrative, operational, and financial responsibility.

The urban transportation planning process may be characterized in four general phases: inventories; analyses of existing conditions and the development of forecasting techniques; the forecasting of future conditions; and, a systems analysis which provides for the essential feedbacks between the transportation and land use elements. The inventories form the base upon which the entire process stands.

The Bureau of Public Roads has suggested that inventories and analyses are required for 10 basic elements: economic factors affecting development; population; land use; transportation facilities; travel patterns; terminal and transfer facilities; traffic control features; zoning ordinances, subdivision regulations, building codes, etc.; financial resources; and social and community value factors. I shall discuss these very briefly in order to identify more specifically the informational requirements.

The first three are closely related. The economic activity and population measures are key determinants of travel demand. The land use inventory and forecast pinpoints the location of activities within the urban area.

An analysis of the existing economic structure is needed as well as estimates not only of the total level of probable future economic activity but also of its probable character. The specific products of economic studies that are inputs to the transportation planning process are forecasts of employment, labor force by industry category, income distribution, and car ownership. You will recognize there are wide variations not only in the scope of the studies but in the techniques employed including the judgment decisions that must be made.

Answers must be obtained to such questions as, "How many people live in how many households in the study area now? How many will there be in the future? What is the breakdown of the population with respect to age and sex?" The inputs to population studies are, of course, data on births, deaths, and migration. The outputs--future population and households--establish the total potential trip makers at some future time.

Land use studies as we define them are concerned with the spatial distribution within the study area of both current and future population and economic activity. We need to know the type of activity and the intensity or density by small geographical areas. While the current techniques utilize land use categories at the one-digit level of generalization, we strongly urge that all field inventories provide for the listing of specific activities and on a parcel by parcel basis. There is a growing need for the collection of land use information in a form that will allow comparability between localities and regions, and that will also permit the study of trends over time. To assist in this the Housing and Home Finance Agency (now the Department of Housing and Urban Development) and the Bureau of Public Roads jointly issued in January of 1965 a manual for identifying and coding land use activities at the two-, three-, and four-digit levels.

It is probably realistic to state that the land use forecasting process is the least advanced simulation tool in the transportation planning process. Data descriptive of the changing locational patterns of urban activities are extremely difficult to come by and exhibit a substantial degree of qualitative and quantitative variation nationwide. Seemingly in contradiction to this is the fact that land use forecasting is becoming more and more important in the overall planning process. Since the introduction of mathematical models into the forecasting procedures, planners and policy makers have the opportunity to examine the effects of alternative planning and policy decisions on the likely activity structure, and to evaluate and choose between alternative structural arrangements.

Four of the basic elements are directly concerned with transportation facilities and travel patterns.

Information is needed on the physical features of each link of the major street system such as right-of-way width, roadway width, roadway type and condition; parking regulations, and traffic control regulations and devices. And, in order to evaluate the existing system, it is necessary to have information on the operational characteristics--the capacities of the roadway and the major street intersections; the volumes of traffic on each segment of the system; the speed of traffic movements at different volumes; and, the frequency and location of accidents.

Consideration of public transit requires information on route location, passenger counts, passenger fare distribution, and such operational data as: revenue vehicle-miles; average seating capacity by type of service; route-miles and terminal-to-terminal running time; headways; and regularity of service.

Since the effectiveness and efficiency of the urban transportation system is dependent to a large measure upon the availability of adequate terminal and transfer facilities, the present supply of parking spaces in critical areas must be inventoried, and the parking, loading, and unloading requirements for terminals must be determined.

And, of course, urban transportation planning requires specific knowledge of the current travel patterns. Information is needed on the location and amount of travel by the various modes, and on such trip characteristics as purpose, length, time of day, and land use activity at the termini.

The task of travel forecasting is usually handled in two steps. Utilizing the land use and travel inventory data, trip generation and trip attraction rates are established for various types of activities and trip ends are computed for each analysis zone. In the second step, the trips originating in each zone are distributed to other zones in the study area.

Two of the elements that we have classified as essential to the transportation planning process relate directly to plan implementation. Zoning ordinances, set-back requirements, subdivision controls, building codes, and the official map together with licensing powers are basic techniques used to control community development. Existing laws and ordinances should be analyzed in the light of the objectives for future development. Deficiencies should be carefully documented, and recommendations for needed revisions or additional regulations should be prepared.

Financial resources are, of course, one of the more critical factors influencing the selection of an urban transportation system and the programs devised to implement the system. In addition to determining the estimated costs of proposed facilities, the transportation planning process should also survey and analyze the abilities of the various governmental units to finance the needed improvements.

The last element, social and community-value factors, is perhaps the most important, but it is also extremely elusive when we attempt to establish quantitative measures. Some community values are measured in dollars, such as the amount of funds that a community will raise through self taxation for public services. Many others are measured in distinct terms by the number of votes cast for candidates for public office. Some, however, involve the full range of attitudes, opinions and even the emotions of the citizenry, and are not subject to precise measurement. Undoubtedly we need to expand our knowledge of environmental effects. But this additional knowledge will not eliminate the conflicts in the choices to be made. Hopefully it will assist in the development of bases upon which the necessary trade-offs can be evaluated more objectively and consistently.

Basic to the development of any plan is the establishment of goals and objectives by both planners and policy makers. To assist in the difficult job of determining what people value, some of the transportation planning studies have utilized attitude surveys. The Minneapolis-St. Paul study, for example, asked 4,600 area residents their attitudes about housing, employment, transportation, government, recreation, and other aspects of their environment. The results were used in judging alternative growth patterns.

In this brief discussion I hope that I have made clear that we believe transportation planning requires lots of information. It should be of interest to you that the work programs of the State highway departments for the 1966 fiscal year included over 35 million dollars for urban transportation planning.

I should like to close with the reminder that there has to be a big pay-off if those of us who are concerned with the collection and analysis of data can find ways of utilizing the regular administrative records of governmental units more effectively. We need also to make greater efforts to tailor our inventories and surveys so that the data will serve multiple users. We are anticipating substantial benefits from the current studies being made by the Bureau of the Census which hopefully will result in their making available many items of information from the 1970 census by small geographical areas.

NEW YORK STATE'S STATISTICAL COORDINATION PROGRAM

Harold Rubin, New York State Division of the Budget

New York State's statistical coordination program is two years old. It was started at about the same time that the Governor's Conference adopted a proposal for its Executive Committee to make "appropriate investigation and recommendations to the states for the standardization of statistical data in reporting, analyzing, and evaluating governmental services." The Executive Committee made a study and recommended that each state establish an office of statistical standards and that a national conference on the subject be called. This past February the first National Conference on Comparative Statistics was held in Washington D.C.

However, the background of New York State's program goes back more than a decade, and the Albany Chapter of the American Statistical Association played a role in founding New York's program.

Background

During the 1950's New York had an organization of research directors -- the Interdepartmental Committee on Research -- which met periodically to discuss matters of pertinence to those in the research and statistical field. One concern of the Committee was the lack of data in certain areas, particularly in the field of population. A Subcommittee on Population was established to investigate the problem and in 1956 issued a report documenting the need for better and more current data. A major recommendation called for establishment of an Office of Population Research within the Executive Department, but no action was taken at that time. (Population projections, now the responsibility of the Office of Planning Coordination, Executive Department, have been made for the State for the next 30 years, with breakdowns by county, age and sex.)

During the same period, the 1950's and early 1960's, the Albany Chapter of the American Statistical Association was also concerned with the lack of data and the lack of a focal point for statistics within the State. A number of meetings were held between representatives of the Chapter and of the Division of the Budget to discuss the problem and possible solutions. The Chapter then recommended establishment of a central statistical office in a central staff agency.

Thus, there was pressure for a central statistical agency from both within and without State government.

Factors Leading to Need for Central Statistical Office

What were the "gaps" in statistics that required establishment of some sort of central statistical agency?

1. In some cases basic statistics were not available. There was no focal point within the State government to pinpoint the need for data.

2. New York State has some 40 agencies, many of which engage in research and statistical activities in the social sciences. There was no formal mechanism for an exchange of information among these agencies. In some cases, different agencies were working on the same problem, unaware of the activities of others.

3. The statistics produced varied widely in terms of coverage and quality. Some figures were simply by-products of a routine operation with no attempt made to modify them for general use. In other cases, the conditions under which statistics were originally produced had changed, but no comparable change had been made in the basic data.

4. The growth of State government, in terms of programs, people and cost, has increased its complexity and has added to the decision-making alternatives for State executives. What was needed was an improved data base for decision-making for operating programs, planning and budgeting.

Factors Considered in Establishing the Central Statistical Office

In establishing a central statistical office, two basic decisions had to be made -- the location of the office within State government, and its functions.

1. Location

In deciding where to locate the function, a number of factors were considered. Several State agencies, such as Commerce and Labor, are heavily involved in the production of statistics of a general or economic nature. Some thought was given to establishing the central statistical agency within such an operating department. However, since each of these is equal in terms of hierarchy, it was felt that a central statistical office within an operating agency might not command the cooperation of other line agencies. Consequently, it was decided that the proposed agency should be located within the Executive Department.

Of the staff agencies in the Executive Department, the one with statutory authority, continuity, and good communications with all State agencies was the Division of the Budget. Accordingly, the central statistical function was established within the Division of the Budget.

Establishment of the Office of Statistical Coordination

The specific impetus for establishing the Statistical Coordination section came in 1964 when the Division of the Budget instituted an integrated Planning/Programing/Budgeting system. This section was charged with providing the statistical base for the system. Both budget planning and coordination were placed under an assistant director reporting directly to the Budget Director

Functions

In broad terms the four principal functions of the central statistical office are as follows:

1. To improve statistical output so that the data necessary for decision-making, budgeting, forecasting and planning are available.

2. To promote the exchange of information within State service, and between the State service and outside organizations.

3. To act as a clearing house.

4. To provide technical assistance.

(The specific functions of the office -- current and prospective -- are enumerated in an appendix to this paper.)

Improve Statistical Output so that Data Necessary for Budgeting, Forecasting and Planning are Available

(1) As a part of New York State's long-range planning program, each agency is required to plan on a detailed basis for the next five years, and on a less detailed basis for a longer period, up to 20 years. In order to make sure that the data base necessary for Planning/Programing/Budgeting is available, we have asked each agency for information on:

(a) Data they are now using as a base for their plans; and

(b) Data they would use if it were available.

We plan to organize the various data needs with a view to setting up research projects to gather the needed data, either by one or more State agencies or by an outside consultant. Preliminary indications are that the basic data needs are details of population projections -- white vs. non-white, income distributions and commuting trends.

Incidentally, the lack of an adequate data base for decision-making is most evident in the current concern in New York State over the cost and coverage for the State's new medicaid program. Eligibility under this program is based on income, family size, savings, and life insurance coverage.

However, there is no good current information on population, distributed by income, family size, and savings. Income figures are available from the 1960 Census of Population which reports 1959 income. There was a nationwide survey in 1963 which provided figures on resources by income class, but there were no breakdowns by age or family size.

(2) As a by-product in helping to establish a training course in the research field, we published Guidelines for Chartmaking, a pamphlet designed to help researchers, statisticians and others do a better job in presenting statistical data in a graphic form. This little booklet has proved to be a useful guide. Accordingly, in the future we plan to prepare a comparable guideline for tabular presentation.

(3) New York State is divided into 12 planning regions. For administrative and other purposes, each agency uses its own regions. Obviously, the Banking Department's regions differ from those of the Division of Parks. To permit each agency to maintain its own regions and still make full use of the data, we are using the county as the basic building block.

Many statistical series in New York are now compiled on a county basis -- financial data of localities, per capita income, and so forth. Others are compiled according to the geographic area of the collecting agency, as for example, by district office and municipality. County data can be obtained in some cases by addition, as for example, welfare statistics; in other cases more detailed analysis is required, such as for employment and unemployment data.

Promote the Exchange of Information Within the State Service and Between the State Service and Outside Organizations

(1) As a means of eliminating interagency knowledge gaps of the statistical and research studies conducted by other agencies, we publish bimonthly the New York State Statistical Reporter to report on completed, current and prospective studies. Also included is other information of interest to statisticians and persons in related activities, as well as announcements of personnel changes and positions available.

Our format has apparently met with some favor. Last year New York City's Office of Administration initiated publication of its Research Reporter along similar lines as our own publication. This year the State of Oklahoma started publication of a Statistical Reporter, again along similar lines to ours.

(2) We have conducted an inventory of all State agencies to obtain information on their statistical series. The report, which provides a description of the series, frequency of issue, where published, etc., is now at the printers.

(3) As another means of improving information flow among State agencies, we helped last year to reactivate a dormant Interdepartmental Committee on Research. The Committee meets three times a year to exchange information and to hear speakers on topics of mutual interest. In addition, the Committee provides a means of meeting some of the collective needs of research directors (training courses for employees, etc.). The Office of Statistical Coordination acts as the secretariat for the Committee.

(4) Another responsibility is to bring together functional people in different agencies who are concerned with research and statistics. Last year, as a prelude to setting up a new reporting system of local expenditures for the administration of criminal justice in New York State, we called a meeting of representatives of the diverse agencies in the field. These people were given an opportunity to make their data needs known so that the reporting system would be as comprehensive and useful as possible. The system will provide current financial data from the more than 1600 localities within the State. We recently held another meeting, of essentially the same group, to discuss the research and statistical studies underway and projected in the field of administration of criminal justice. Among those who attended were representatives from the Correction, Probation, Parole, Mental Hygiene and Social Welfare agencies, the Judicial Conference, and the Dean of the new Graduate School of Criminal Justice, State University of New York.

(5) To afford State agencies an opportunity to make their data needs known concerning the 1967 Census of Governments, we called a meeting in Albany of State people. Mr. Allen Manvel, Chief, Governments Division, described the forthcoming census while the New York people made known some of their needs.

(6) As part of our secretariat function for the Interdepartmental Committee on Research, we published one of the talks given before the committee on "The Impact of Population Changes on State Programs." We plan to continue this practice, perhaps using the instrument of "Occasional Papers."

(7) To make sure that we are in the best position to assist State agencies in learning about Federal data developments, a member of our staff spent a week in Washington last year with the Bureau of the Census to learn of its operation. We urged the Bureau to institute a regular program so that State and local employees can be brought together to learn of current and future plans of the Federal statistical agencies. The Bureau established such a program this year, and some ten states sent representatives to Washington last month. Our man participated in this program which he said was most worthwhile. Mr. Bowman was one of the Federal participants in this program.

(8) Among future programs to promote the exchange of information, is publication of a statistical yearbook. We are in the process of hiring a person for this job, and hope to issue a yearbook in 1967. We plan to emphasize county data, since we feel that the county is the most useful building block for State purposes.

(9) Another planned publication is a directory of statisticians, economists and others in the research field, probably somewhat along the lines of the comparable Federal publication.

Act as Clearing House

(1) We act as a clearing house for inquiries both on an intra and interstate

level. I expect this function to grow as the news of our existence spreads. Since the National Conference on Comparative Statistics in Washington last February, we have received requests from other states contemplating setting up a comparable office -- on what we are doing, for examples of our publications, and for suggestions in establishing statistical coordination programs.

(2) We assist State agencies in locating qualified research and statistical employees. Our "want-ads" in the Statistical Reporter is one aspect of this operation. I am on the Governor's Manpower Committee for Statisticians. The Committee has proposed that the Office of Statistical Coordination act as a clearing house for statisticians, economists and research analysts. (The very existence of our Office has stimulated various groups to assign responsibilities to us in terms of employment, publications, standards, etc.).

(3) The Office has acted as a clearing house for requests originating both within and without the State. A current example is the request for cooperation made by the recently established New York Regional Statistical Center. The goal of the Center, established by the New York City Chapter of the American Statistical Association, is to encourage production of comparable data for the 22 counties in three states comprising the New York Region. One plan of the Center is to test the feasibility of putting local data (health, welfare, police, etc.) on a block basis so that they can be aggregated by different districts (welfare, police, etc.). (The street index guides that will be forthcoming from the 1970 Census of Population and Housing should facilitate organizing data on a block basis.) The Center is also interested in coordinating local data systems so that they are compatible.

Provide Technical Assistance

(1) We have assisted some smaller agencies without professional statistical personnel, either to obtain data or to establish plans for data collection.

Thus, when the State Office of Economic Opportunity was established, it needed figures to measure poverty. We compiled the figures, on a county basis, for various items measuring education, income, welfare cases, etc. We ranked the counties by the specific measure to obtain some idea of relative poverty. We also prepared a pamphlet of the data sources which was made available to localities to assist them in applying for grants.

(2) Our Office has assisted a number of agencies in establishing training courses in statistics. In addition, under the impetus of the Interdepartmental Committee on Research, we helped initiate a course in sampling for research directors.

(3) Skilled statisticians are scarce. To make the maximum use of such personnel, we are considering retaining such persons to be available to operating agencies on a consulting basis for specific assignments; they would not be concerned with operating functions.

Conclusion

The problems of statistical coordination are many; our period of existence is short and our staff is small. Hopefully we have made a beginning. I know that we have a long way to go.

The need is such that I am sure that our work will expand, particularly with regard to basic data -- such as production of more and better economic indicators, revision of existing statistics to be more comprehensive, increasing the availability of local area data, etc.

The establishment of statistical coordination agencies in other states should prove to be a stimulus to the existing ones. We should all grow together so that better data will be available, and hopefully better informed decision-making.

Appendix

New York State's Office of Statistical Coordination

Current and Prospective Functions

Improve Statistical Output so that Data Necessary for Budgeting, Forecasting and Planning are Available:

1. Review and analyze statistical output to encourage agencies to improve the quality of their series, initiate series not now produced, and eliminate duplication.

2. Determine needed data not now collected and promote collection.

3. Establish standards in reporting (graphs, charts), in terminology, bases, areas, etc.

4. Promote more current reporting of statistical data.

5. Provide, or encourage production of, the basic data needed for agency forecasting.

6. Review department questionnaires addressed to the public to eliminate duplication, meet the needs of other agencies, conform with State policy, etc.

Promote the Exchange of Information Within State Service Agencies and Between State Service Agencies and Outside Organizations:

1. Act as secretariat for reactivated Interdepartmental Committee on Research, to facilitate interchange of research ideas and data.

2. Hold periodic conferences of State statistical and research personnel, both with and without their counterparts in private industry and other governmental jurisdictions, to discuss topics of interest to the State -- economic, financial, population forecasts, new techniques, operations research, etc.

3. Publish bimonthly New York State Statistical Reporter containing news of studies under way and research reports available.

4. Conduct inventory of State statistical series and publish compilation in Statistical Series of New York State. This publication could list -- by agency and by function -- series that are available, how often compiled, and where and when available.

5. Compile State statistics for publication in annual Statistical Abstract of New York State.

6. Act as liaison with the U.S. Bureau of the Census to determine availability of data and the needs of State agencies to obtain the maximum amount of information and at a minimum of cost (e.g., eliminate duplicate purchases of unpublished data and make purchased data available to other interested agencies).

Act as Clearing House:

1. Handle inquiries to the State (from both within and without the State) concerning sources of data, availability of statistics, publication dates, etc.

2. Assist research and statistical units in locating qualified employees and in placing employees with agencies.

Provide Technical Assistance:

1. Assist budget examiners in determining needs of statistical and research units in staffing and related aspects.

2. Upon request, counsel agencies in planning research projects.

3. Recommend to Budget Director and Governor what the State's official position should be on statistical issues in representations to the Federal government and other outside organizations; upon request, represent the State before such bodies.

Discussion

Maurice I. Gershenson, California Department of Industrial Relations

Each of these excellent papers points up interesting problems of coordination, standardization, and comparability. In the short time allotted to me I can touch on only a few of them.

First of all, I want to take issue with Selma Mushkin's proposal for centralizing the making of demographic and economic projections for states and local areas. The states will not give up responsibility for such projections for their respective jurisdictions. Highly sophisticated work is going on in many of the states, and the familiarity of local people with local trends should not be lost.

I would go for partnership or a cooperative arrangement between a national agency or organization and the states in developing projections for the states. I can see many advantages to such an arrangement. But I am certain the states will not be receptive to accepting projections handed down to them from above.

Interstate comparability

Selma Mushkin's paper points up the problem of lack of interstate comparability of many statistical series. I would like to touch briefly on the difficulties in those fields where laws and administrative practices vary from state to state, and where there are no effective inducements to achieve comparability.

Most of us would agree that it is highly desirable to have good comparability of federal, state, and where applicable, local statistics. In some fields, comparability may not be essential to the particular operation, but the administrator likes to know how his jurisdiction compares with others. It may help him in gauging how well he is doing, or suggest modifications he may introduce to improve efficiency.

Governor Bellmon stressed this point in his address to the National Conference on Comparative Statistics held in Washington, D. C. in February 1966.

> *"We looked for statistics to indicate the degree of progress our state was making under my administration. We sincerely wanted measuring sticks by which we could compare our state with other states. In short, we wanted a standard by which we could judge our performance."*

Workmen's compensation is a good example of a field where interstate comparability of statistics may not be needed to operate the program in any one state. But ever since the beginning of this social insurance program more than a half century ago administrators have decried the lack of comparability of statistics. The international organization of workmen's compensation administrators--the International Association of Industrial Accident Boards and Commissions--has, for years, urged that something be done about this.

The Committee on Statistics of the IAIABC wrestled with this problem for a number of years. Valiant attempts were made to bring about a small measure of comparability by focusing on those items which were generally common to most of the jurisdictions.

One of the first things we did was to develop a "Glossary of Terms for Workmen's Compensation Statistics," which attempted to standardize terminology, define terms and present some elementary rules for standardizing tabular presentation. This glossary was officially adopted by the IAIABC at one of it's annual meetings and all workmen's compensation agencies were urged to use the standard definitions and terms. So far as I know, very few states or provinces have made any serious attempt to do so.

We developed a standard form for use by employers in reporting work injuries sustained by employees. It was our hope this universal form would lead to greater comparability of both injury and workmen's compensation statistics. To date only a few of the states use this form.

This has been the discouraging experience in one field.

I turn now to another field where we have had better success.

The members of the Association of Labor Mediation Agencies found they had difficulty in comparing their respective state statistics. A committee appointed to work on this problem asked several of us from the labor statistics field to serve as consultants.

We developed what we called "A proposal for uniform mediation statistics." The proposed standard was adopted by the Association, and many states have moved far toward converting their statistics to the recommended basis. Differences in laws relating to the functions and responsibilities of the mediation agencies as well as differences in administrative practices make it difficult to conform strictly to the standard on uniform mediation statistics. Nevertheless, there is greater comparability now than there ever has been, and I look for further improvement in the years ahead.

It should not be very difficult to achieve reasonably good interstate comparability of statistics in fields where there are Federal-State cooperative arrangements such as the BLS current employment statistics program. But how can we best achieve the greatest degree of interstate comparability in those cases where there is no federal-state relationship, and where laws and administrative practices differ, and where no effective machinery now exists to achieve standardization?

Coordination and quality

Harold Rubin's fine description of New York's new program points up the fact that statistical coordination agencies are now being established in many states. This is a development long overdue. We have had a coordinating unit in California for 21 years. Governor Earl Warren established the California State Interdepartmental Research Coordinating Committee in 1945, and it has functioned actively and continuously ever since.

One of the important functions of all statistical coordinating organizations is the improvement of the quality of statistical work.

I would like to ask what are the specific mechanics for accomplishing this objective? What powers with respect to quality do the state coordinating agencies and the Federal Office of Statistical Standards have?

More particularly, I would like to know what has been done, or is being done, to police the quality of work done *for* federal, state and local agencies by outside contractors. I have seen some pretty bad statistical work performed by outside consultants for government agencies and, in some cases, for pretty fat fees.

How can we best guard against this? I hope we can stimulate some discussion from the floor as to how coordinating agencies can best function in carrying out responsibility, not only for the quality of statistical work performed *by* government agencies, but also *for* government agencies.

Vertical coordination

I now turn to a problem of coordination that has bothered me for some time. It relates to what I call "vertical coordination" and concerns the release of two different figures by federal and state agencies which purport to measure the same thing.

Here are a few examples:

The Census Bureau regularly publishes estimates of current population by states. The figures for California are close to, but not the same as, those published by the California agency responsible for population estimates.

The Bureau of Labor Statistics recently issued state estimates of union membership in 1964. The figure for California was close to, but not the same as, the one my agency published for the same year.

The President's Manpower Report for 1963 and 1964 carried labor force projections by states, prepared by the Bureau of Labor Statistics. A total labor force of 9,660,000 was projected for California for 1970. Our published projection for that yeat was 9,001,000, a difference of 660,000. In this case, unlike the population and union membership estimates, the figures were not close, but were quite far apart.

Statistics released by the U. S. Public Health Service indicate that the number of non-white live births in California *decreased* by 1,000 between 1963 and 1964; those released by the State Department of Public Health show an *increase* of 1,000.

The Federal Bureau of Narcotics reported 691 new narcotic addicts in California in 1964. The California Bureau of Criminal Statistics published a figure of 2,310 newly reported "addict users" for the same year--(more than 3 times the federal figure).

The user who finds the two figures asks: "Why the difference? Which one shall I use?"

In most cases there may be a simple explanation of how the difference came about. But I submit, this may not be a good reason why two different figures which purport to measure the same thing need to be published. I want to make a plea that we attempt, wherever possible, to eliminate these situations.

It has been done quite successfully in some fields. Let me cite a few examples:

The California Division of Labor Statistics and Research and the U. S. Bureau of Labor Statistics have worked out an arrangement whereby the statistics of work stoppages for California, published by the two agencies, are identical.

Some years ago we found that the statistics of net and real spendable earnings of manufacturing production workers for the San Francisco-Oakland and Los Angeles-Long Beach Metropolitan Areas, published by our agency and the regional office of BLS, differed slightly. The reason for this was that we deducted State income tax while BLS, in line with the national computation, did not. We took this matter up with the regional director of the Bureau of Labor Statistics who immediately recognized that the two sets of figures could cause confusion. It was agreed that only one set would be published--that which makes allowance for State income taxes.

The current estimates of personal income in California, prepared by our Department of Finance, are tied directly to the previous year's state figures published by the Department of Commerce. As soon as the official Department of Commerce figures are released they become the official California figures. Our state agency believes a better set of California figures can be produced, and is working with the Department of Commerce to bring about improvements.

I am convinced that if earnest attempts are made, the issuing agencies can get together and agree upon a common set of figures. I know that in some cases it's not going to be easy, but I believe the effort would be well worthwhile.

It has been done--it can be done!

Summary

We are in a period of intensive action to set up "coordination" and "standardization" machinery. I predict that much good will come of this effort, and that the greatest achievements will be in the states, along the lines of horizontal coordination *within* the states.

More difficult to attain will be coordination between different levels of government and between states, particularly where machinery for interstate coordination does not now exist.

We should set up a coordinating organization of coordinating agencies--federal, state and local--and work together toward such goals as the improvement of statistical output, the filling of important data gaps and the elimination of duplication of effort.

XII

TIME SERIES VII: MANPOWER APPLICATION

Chairman, J. E. MORTON, W. E. Upjohn Institute for Employment Research

LONG-TERM ECONOMIC GROWTH

A New Census Report

Julius Shiskin, Bureau of the Census

S U M M A R Y

A new Census Bureau report, Long-Term Economic Growth, presents in convenient form the principal annual time series needed by students of economic growth. It is intended to simplify the task of analysts in this field, whatever their explanations of economic growth and standards for judging performance happen to be, by providing a broad base of information related to economic growth and relieving those concerned with theoretical issues and economic policies of a large part of the laborious task of compiling basic data and making computations from them.

The new report provides annual data over a long span of years for each series, often back to 1860. In addition to almost 400 basic time series and almost 800 component series, the report contains numerous charts, growth rate "triangles," and scatter diagrams to facilitate the summarization, analysis, and interpretation of long-term trends in the U.S. economy. This compendium is the third phase of the Census Bureau work on economic fluctuations, which also includes the seasonal adjustment program and Business Cycle Developments.

I. Objectives

The Census Bureau will soon publish a new statistical report, Long-Term Economic Growth. This report is designed to show in convenient form the principal annual economic time series needed by students of economic growth. It represents a response to the increasing interest in expanding economic welfare, both in developed and developing countries; the economic competition among countries with different economic systems; and the establishment of economic growth as a major policy objective of the U.S. Government. It supplements many descriptive studies and causal analyses on this subject that have been prepared in recent years. It is expected to simplify the task of students in this field, whatever their explanations of economic growth and standards for judging performance happen to be, by providing a broad base of information related to economic growth and relieving those concerned with theoretical issues and economic policies of a large part of the laborious task of compiling basic data and making computations from them.

There is, at present, considerable uncertainty regarding the appropriate measures of economic growth, the methods of compiling the measures, and the accuracy of the historical records. While there is some agreement about the factors which affect long-term economic growth, there is less about their quantitative importance. In fact, there is only one comprehensive series of estimates of the quantitative importance of these factors--that by Edward F. Denison. Denison's study has had a major impact on investigations of economic growth, with one of its many contributions being the demonstration of the tenuousness of many of the estimates that are available and the need for more basic information. Another major objective of this Census Bureau publication, therefore, is to encourage and facilitate the development of better estimates by providing a convenient framework for such work and by bringing the statistical gaps out into the open.

Thus we hope that this report will provide an information base that will facilitate judgments on economic performance, aid in the formulation of economic policy to accelerate growth, contribute to development of the theory of economic growth, and point up some of the gaps in the statistical intelligence system.

The objective of this paper is to describe this new report and invite suggestions for improving it. Limited resources and experience have confined this first edition to those data most readily available. For this reason and because of the large task of inspecting and appraising all the series that could have been included, it is recognized that this issue will have to serve as a working document to break the ground and set a pattern for subsequent reports.

Our plan is to issue a revised edition in about a year. Experience with similar new reports indicates that substantial changes may be expected as a result of suggestions made by those making practical uses of such material. We, therefore, welcome the comments and criticisms of those who make use of our report. As in the case of many other Census reports, we expect future issues to be considerably different and more useful.

Before discussing the new report itself, however, I would like to make a few observations on the relations of our seasonality and business cycle work to this new "growth" report.

ACKNOWLEDGEMENTS: During the past two years the technical work for this new Census report has been under the immediate supervision of Allan H. Young, who was assisted by Gerald Donahoe, Norman Bakka and Robert Taylor. The early planning work was under the immediate supervision of Samuel L. Brown, who contributed substantially to the basic organization of the report. He was assisted during this period by James Hines and Judith Faust. Mr. Young also contributed substantially in the preparation of this paper, especially Part III, Problems of Measurement. Murray D. Dessel provided valuable technical advice throughout.

For the many other persons in other Government agencies and private research organizations and universities who also contributed, see the acknowledgements to the Census Bureau publication, Long-Term Economic Growth.

II. The Census Bureau Program for the Analysis of Economic Fluctuations

This new statistical report on economic growth may be considered as the third phase of our research and development work on economic fluctuations, conducted over a period of more than 10 years. The first phase was the development of computer programs for analyzing seasonal, trading-day, and irregular fluctuations in economic time series. The second phase was the development of a set of statistical tools, including Business Cycle Developments, for analyzing intermediate fluctuations lasting from about 3 to 8 years. Each of these three projects should be considered in relation to the others, not as independent undertakings. As a result of this continuing research program, the Census Bureau can now provide facilities for studying nearly all types of economic fluctuations in the United States.

The first of these facilities is our time series analysis program--Census Method II--designed for the intensive study of short-term movements. The latest variant of this program, X-11, has greater generality and scope than any of the earlier programs. It has a separate routine for quarterly as well as monthly series, and for series with negative and positive numbers as well as those with positive numbers alone. The X-11 version not only measures and adjusts for seasonal variations, but also for trading-day variations. Further, it computes many summary and analytical measures of the behavior of each series and includes various techniques--such as spectral analysis, F-tests, and variance analysis--for use in extending the scope of time series studies.

The second of these Census Bureau facilities, our Business Cycle Developments (BCD) report, permits the timely yet comprehensive study of intermediate economic movements. This monthly report brings together several hundred monthly and quarterly "economic indicator" series for the analysis of short-term economic trends and prospects. These particular series have been selected, tested, and evaluated, after half a century of continuing research, as the most useful and reliable for this purpose. The publication not only provides the basic data, but also various charts and analytical tables to facilitate studies of intermediate-term fluctuations. In addition, a time series punch card file, a diffusion index program, and a separate summary measures program are available for those who wish to carry on further research in business cycle analysis.

The third and latest facility is this "growth" report, modeled after BCD, and designed specifically for the study of long-term economic movements. Since the remainder of this paper is concerned with the content of this new report, I shall defer discussion of it for a few moments. Suffice it to say here that the experience we have already had with the Census Method II seasonal adjustment program and with BCD indicates that the new report on economic growth will be widely used by government, business, and research organizations.

The capabilities that have been developed for this "time series analyzer facility" are available to the public in various forms: (a) periodic publication of the basic data required for studies of economic fluctuations; (b) published computer-generated charts and analytical measures which present and summarize conveniently the underlying trends of the basic data; (c) computer programs (written in a simplified computer language, Fortran IV) which permit further analysis of the fluctuations; and (d) data files in the form of punched cards and computer tapes, which provide the statistical raw material for these computer programs and publications.

Taken together, this Census Bureau "system" will help to improve and extend the techniques used by economic analysts in their study and understanding of economic fluctuations. This "system" makes it possible for the academic or business economist, who has a computer available, but not a research staff or programmers, to carry out extensive research in the field of economic fluctuations.

III. Problems of Measurement

Many conceptual and statistical problems beset the measurement of economic growth and analysis of its sources. Some of them are briefly reviewed below. The purpose of this review is only to indicate the nature of the problems and the many uncertainties that now surround them. More comprehensive statements of these problems, the alternative solutions and their implications, especially for data compilation, appear in the references.[1]

1. Concepts for Judging Economic Growth

Economic growth is usually considered to be growth in the output of the economy. Such growth can be measured in terms of output either on a total, a per capita, or a per worker basis, with the choice depending on the problem at hand. Alternatively, economic growth is sometimes defined in terms of per capita consumption or personal welfare. Another alternative view is in terms of changes which take place in the economic and social structure of a nation as it undergoes economic growth, for example, the changes in the rate of population growth and the amount of the labor force in agriculture which a nation about

---

[1] For the most part this review is based on more detailed discussions of the same problems in The Sources of Economic Growth in the United States and the Alternatives Before Us, by Edward F. Denison, Supplementary Paper No. 13, Committee for Economic Development, January 1962, and "The Measurement of Aggregate Economic Growth" by George Jaszi, Review of Economics and Statistics, November 1961. Also, see The Study of Economic Growth by Solomon Fabricant, Thirty-Ninth Annual Report, National Bureau of Economic Research, pp. 1-13, May 1959, and Six Lectures on Economic Growth by Simon Kuznets, The Free Press, 1959; and the additional references given in the bibliography to the new Census Bureau publication.

to begin economic development may experience. All the above definitions are directed to the long-term, that is, to the changes or trends which occur over several years, perhaps a decade or longer, and sometimes a century.

2. Definition and Measurement of Output and Related Economic Processes

There are many problems in defining and measuring total output and the other economic activities presented in this report. Some of the principal ones concerning total output are indicated below. Similar problems affect many of the other types of measures presented in the report.

All growth analysts consider real gross national product, as distinguished from money gross national product, as the appropriate measure of output. However, money data are sometimes used as a proxy for data on the physical volume of output because of the difficulties of compiling "real" data, either directly or through price deflation. For the most part data on real output are derived through price deflation. In many areas there is a paucity of actual output data so that physical volume measures cannot be built up directly. This is particularly true for the service industries and government services. Therefore, the indirect way of measuring output is used, that is, dollar volume figures are divided by price deflators. In some sectors where physical volume data are available the advantages of the price deflation method are illusory, because price data are no more abundant nor any more accurate than physical volume data. However, some direct measures of physical volume are included in this report, for example, the Federal Reserve index of industrial production.

Total output as compiled in the U.S. National Income and Product Accounts, prepared by the Office of Business Economics, is the market value of the final output of goods and services produced by the Nation's economy. In addition to the sales of final products to their ultimate consumer, the value of total output includes additions to business inventories and the value of force account construction.[2] The services of housewives and similar nonmarket items are excluded. The effect of this may lead to some overstatement in the long run growth of output since many services which were previously performed in households and excluded from GNP are now included. A similar problem is inherent in international comparisons, where in many countries a larger portion of productive activity occurs outside the market economy than in the U.S.

There is also the point of view, held most notably by Simon Kuznets, that the concept of total output should be less inclusive than that used by OBE. Kuznets defines total output as final output intended to satisfy wants of individual consumers. Under this definition he excludes those government expenditures which represent services to business enterprises and many national defense expenditures.

In addition there are the conceptual and practical problems of taking quality changes into account. While there is general agreement that improvements in product quality should be considered as increases in the quantity of output, quality changes cannot be fully taken into account in practice. It is generally believed that the price deflators do not completely reflect quality changes, since the relative quality of new products must be higher than their relative prices for them to replace the old products in the market place. Consequently, there is a tendency for the rate of growth to be understated in the output measures.

Several related problems may be mentioned. One is that of deflating the output of the construction industry. The present price deflators measure in general the costs of inputs rather than the outputs of the construction industry. The result is generally an understatement of the rate of growth of construction, since productivity increases are not adequately allowed for. Another problem is that the output of government is not directly measured, but is based on compensation of government employees. The deflated value of government output, obtained by adjusting for changes in the government wage level, does not include productivity changes. Similar methods are used to obtain the "output" of domestics and nonprofit institutions. As is well known, GNP is often used in place of net output because of difficult conceptual and measurement problems in arriving at the capital consumption allowance; that is, the amount of capital used up in the production process, especially when the replacement capital embodies newer technology.

Still another problem is that of weighting the components of aggregate output. Since relative prices change over time, the selection of the base year determines the weighting of the various components of national product and affects its trend. Studies show that those output components growing most rapidly tend to show the smallest price increases while those growing least rapidly tend to show the largest price increases. Thus, a recent price base gives greater weight to the slowly growing components than does an earlier price base, and vice versa.

Finally, earlier data are less comprehensive and less accurate than recent data, themselves still subject to important limitations. From 1810 to 1899 industrial censuses were decennial, and from 1899 to 1919 they were quinquennial. Also, relatively fewer data were compiled on activities other than manufacturing in the early years of the period covered by the report and these are

[2]Also, imputations are made for four nonmarket items. They are: (1) employee compensation received in kind; (2) food and fuel produced and consumed on farms; (3) services derived from owner-occupied residences; and (4) the services rendered by financial intermediaries without explicit charge. The resulting net addition is about 7 percent.

still inadequate in various respects. World Wars I and II and the depression of the 1930's demonstrated the need for more information, and the passage of the Employment Act of 1946 stimulated further interest in statistics and their uses. In addition, the increasing interdependence of economic activities and the growth of the economics and statistics professions led to the development of improved methods of statistical compilation. In many cases the government has taken over the series and methods of private investigators and provided better current statistics through the use of more comprehensive and more accurate underlying data it is able to collect.

In this connection it is to be noted that the effects of estimating errors are reduced as the span of comparison is extended. Thus an error in the figures involved in a comparison, which affects the year-to-year percentage change by 5 percentage points, will affect the average annual percentage change over 50 years by only one-tenth of 1 percentage point. Similarly, the longer the period over which the comparison is made, the smaller the effects of cyclical and irregular factors. Because there may be persistent biases in some measurements of change, however, and because significant differences in trends may take place during a nation's economic history, a single measure of the average long-term trend must be used with caution.

3. Selection of Statistical Indicators

The selection of statistical indicators useful in studying the sources of economic growth is beset with many difficulties. One is that a comprehensive theory of economic growth is at an early stage of development and does not yet provide adequate guidelines. A second is that despite the relative abundance of our statistics, there is a paucity of data in certain key areas. For example, our national wealth data are piecemeal, particularly on the age and efficiency of capital. Also, few data are available on quality of education or quality of labor. A third difficulty is that many of the series available cover only a relatively short span of years. This point is true of our series on capacity (which start in the late 1940's) and research and development (which start in the 1930's).

The series included in this report as measures of the sources of economic growth represent a selection which several experts in the field of economic growth now consider most relevant. To a large extent the selection relies on the list of 31 factors presented by Edward F. Denison which potentially could affect the rate of growth (some to a much greater degree than others). Many of these factors are presented in Parts I and II of the report. Several, however, are not directly presented in this report because data are not available. They include the elimination of several types of institutional barriers to the most efficient use of resources, the increased mobility of labor, the reduction of crime, and an increase in the advance of knowledge. Some studies emphasize other sources of growth such as the availability and utilization of natural resources and energy; or the intangibles such as the role of the innovator and risk-taker and our method of economic organization, dominated by free markets and competition. In general, series for such additional factors have not been included in this report principally because adequate relevant data do not now exist.

4. Separation of Long-Term Growth from the Business Cycle

Since 1834, the American economy has experienced 31 business cycles from about 3 to 8 years' duration. These cycles have been characterized by alternating periods of expansion and contraction. In addition, there have been four wars with major effects upon the pace of economic activity. The measurement of economic growth and long-term trends in many of the series is greatly complicated by the presence of fluctuations associated with business cycles and the types of irregular movements caused by wars.

For example, from 1919 to 1965, the annual percentage changes in total real GNP ranged from -14.7 to +16.1 a year. These changes primarily represent the year-to-year effect of the business cycle as the economy shifts from high to low level operation or vice versa. Such shifts do not represent growth in output in the sense that we are concerned with in the report. Rather, growth is represented by various types of measures which "adjust" for business cycles and long-term irregular movements. Thus, year-to-year changes in measures of potential GNP, that is, estimates of GNP assuming reasonably full employment, range from -0.2 to 6.5 with most measures concentrated in the interval from 0.1 to 3.9 as can be seen from the table on the following page.

Four techniques are used in our report to show measures of long-term trends as distinguished from cyclical and irregular fluctuations.

(1) Potential GNP estimates made by the Council of Economic Advisers and by the staff of the Joint Economic Committee of Congress (Knowles) are presented. These measures show estimates of GNP assuming reasonably full employment.

(2) A new technique was developed to distinguish rates of change which may be taken as "true" measures of growth from those that are biased from this point of view. This technique, suggested by Denison, is used in the presentation of the growth rate triangles in Part V. The total unemployment rate is used as a measure of how close the economy is operating to its potential output in selecting appropriate years for comparison. Comparisons between years with similar unemployment rates are taken as more valid measures of economic growth than (1) comparisons between years of relatively high unemployment rates and years with relatively low rates, or conversely, (2) between years of relatively low unemployment rates and years with relatively high rates.

Distribution of Year-to-Year Growth Rates in Actual and Potential Real GNP

Interval of percent change	Actual GNP 1909 to 1965		Potential GNP (JEC,Knowles) 1909 to 1964		Potential GNP (CEA) 1952 to 1965	
	No. of measures	Percent	No. of measures	Percent	No. of measures	Percent
All intervals....	56	100.0	55	100.0	13	100.0
-4.0 and lower...	8	14.3				
-0.1 to -3.9.....	8	14.3	1	1.8		
0.0..............	1	1.8	0	0.0		
0.1 to 3.9.......	11	19.6	34	61.8	13	100.0
4.0 to 7.9.......	15	26.8	20	36.4		
8.0 to 11.9......	6	10.7				
12.0 and greater.	7	12.5				

Note: Source of actual GNP data is OBE for the years 1909 to 1965.

(3) An averaging technique was used to combine annual data into measures of the average level of activity over each business cycle. These business cycle averages then provide the basic data in computing growth rates and in showing the relative importance of geographic divisions and industries in Part III. They minimize the effects of the varying cyclical amplitudes of the geographic divisions and industries. These cycle averages, unlike the comparison of selected years in which the unemployment rates are equal, measure the average level over the business cycle, thus reflecting an "output" rather than a "capacity" concept of growth.

(4) Growth rate comparisons of U.S. geographic divisions and industries and of the U.S. and foreign countries are presented only for long spans where the terminal dates have been picked carefully in order to minimize the effect of cyclical fluctuations. In general, growth rates were computed from one cycle average to another or between years of approximately equal unemployment. In some instances, the standards have been relaxed a little to include comparisons based on the current period which does not include a complete business cycle. Therefore, current comparisons may be influenced more than longer, historical comparisons by the business cycle and other short-term effects.

Although it is highly useful to separate the short-term from the long-term fluctuations in measuring economic growth, as is done in this report, the two types of economic movements are interrelated to some extent. For example, cyclical fluctuations often influence business and government decisions concerning the timing and scope of long-term investment commitments. In the 1930's, they also affected the birth rate with a consequent effect on today's labor force. Likewise, expected long-run increases in economic activity, foreshadowed by such indicators as population, affect the patterns and magnitude of cyclical fluctuations.

5. Selection of Growth-Rate Formulas

A growth rate can be defined as the slope of the trend line of a historical series. A constant rate of growth over a period of years is usually expressed as the "average percentage increase per year." A trend line with a constant rate of growth appears as a straight line on a ratio scale chart. Two widely accepted alternatives for computing such growth rates are (1) the method of selected points, and (2) a linear trend fitted by least squares to the logarithms of the data.

The method of selected points, the most frequently used technique, does not take account of intervening values; it estimates the growth rate by simply connecting with a straight line the logarithms of the beginning and terminal values of the period of years considered.[3] It is not influenced by the particular pattern of cyclical variations which occur between the initial and terminal years.

A linear trend fitted by least squares to the logarithms of the data minimizes the sum of the squared deviations of the logarithms of the data from the logarithms of the trend and equates the sum of the logarithms of the data with the sum of the logarithms of the trend. Thus, it is influenced by the particular pattern of cyclical variations between the initial and terminal years.

---

[3] The trend line is given by the compound interest rate formula which in logarithms is $\log X_t = \log X_1 + n \log (1 + r')$ where X_1 is the initial value and X_t the terminal value of the series, n is the span of years, and $r = r' \times 100$ is the percentage rate of growth. To calculate the rate of growth the formula is rearranged $r = (\sqrt[n]{X_t/X_1} - 1.0) \times 100$.

There are several alternatives to the more common technique described above of fitting a linear trend to the logarithms by least squares, which involve fitting an exponential curve directly to the data themselves. The advantage of these alternatives is that they equate the sum of the data with the sum of the trend values rather than with the sums of the logarithms (sums are more meaningful for economic data than products, i.e., sums of logarithms). However, the results are usually quite similar to those obtained by the standard technique.[4]

In estimating growth rates, the time period to be covered should be carefully selected. If the period is too short, say 5 to 10 years, the estimated growth rate may be greatly influenced by transitory conditions in the economy. In such instances, the estimated rate will not actually represent the long-term trend of the series. On the other hand, a growth rate can be computed over too long a period. The path of development of some series over long periods cannot be approximated by a trend line representing a constant percentage rate of increase. In such cases, it may be more meaningful to compute growth rates for various sub-periods or to fit a trend line which does not have a constant rate of growth. In addition, the time period should be selected in such a way that short-term cyclical fluctuations do not bias the calculated growth rate, particularly for a relatively short period where the effect of the business cycle may be large.

Trend lines for GNP in the U.S., derived by various methods of computing growth rates, are shown for selected periods in Chart I of this paper.

It seems appropriate to close our section on Problems of Measurement with a quotation from Simon Kuznets, an outstanding authority in this field:

". . . the conceptual and other difficulties of measurement do not justify the refusal to measure and the substitution of a cavalier treatment of uncontrolled impressions . . . for the strenuous task of empirical corroboration and testing. Despite the limitations resulting from a scarcity of basic, underlying data and from concepts that are outmoded because of a serious cultural lag, much can be learned by a determined scrutiny of the data--provided that one looks at them with significant questions in mind and is sufficiently familiar with the characteristics of both the data and the underlying processes. Whatever mistakes one may make in the process--and they will be many--can at least be corrected by others; cumulative improvement and learning are possible so long as the data are mobilized to serve as a basis of one set of generalizations and as a check on another." */

[4]Two methods of fitting an exponential trend to the actual data are discussed by Neville L.Rucker and Dudley J. Cowden in Tables for Fitting an Exponential Trend by the Method of Least Squares, Technical Paper 6, University of North Carolina School of Business Administration. Other procedures for fitting an exponential trend directly to the data are described by Boris P. Pesek in "Economic Growth and Its Measurement," Economic Development and Cultural Change, Vol. IX, No. 3, April 1961.

*/ Simon Kuznets, Six Lectures on Economic Growth, The Crowell-Collier Publishing Company, 1961.

IV. Description of the Report

1. General Plan

Long-Term Economic Growth brings together almost 400 aggregate annual economic time series and almost 800 additional component series that seem useful for this purpose. The report carries each series far back in history--sometimes to 1860--and will update them in subsequent editions. Future issues will also incorporate all revisions of source data as they become available. The adequacy and appropriateness of particular series are undergoing a continuing review by the Census Bureau research personnel, in consultation with specialists in the field of long-term economic growth. It is expected that new series will be added to future editions, while some of the present group may be dropped after further review. Annual publication is planned until the expected suggestions of users are incorporated and the report is stabilized in this sense. Subsequently, less frequent publication may suffice.

The report is organized into five major parts. The first part presents about 150 annual time series, measuring aggregate output, input and productivity. These are the basic measures of economic growth.

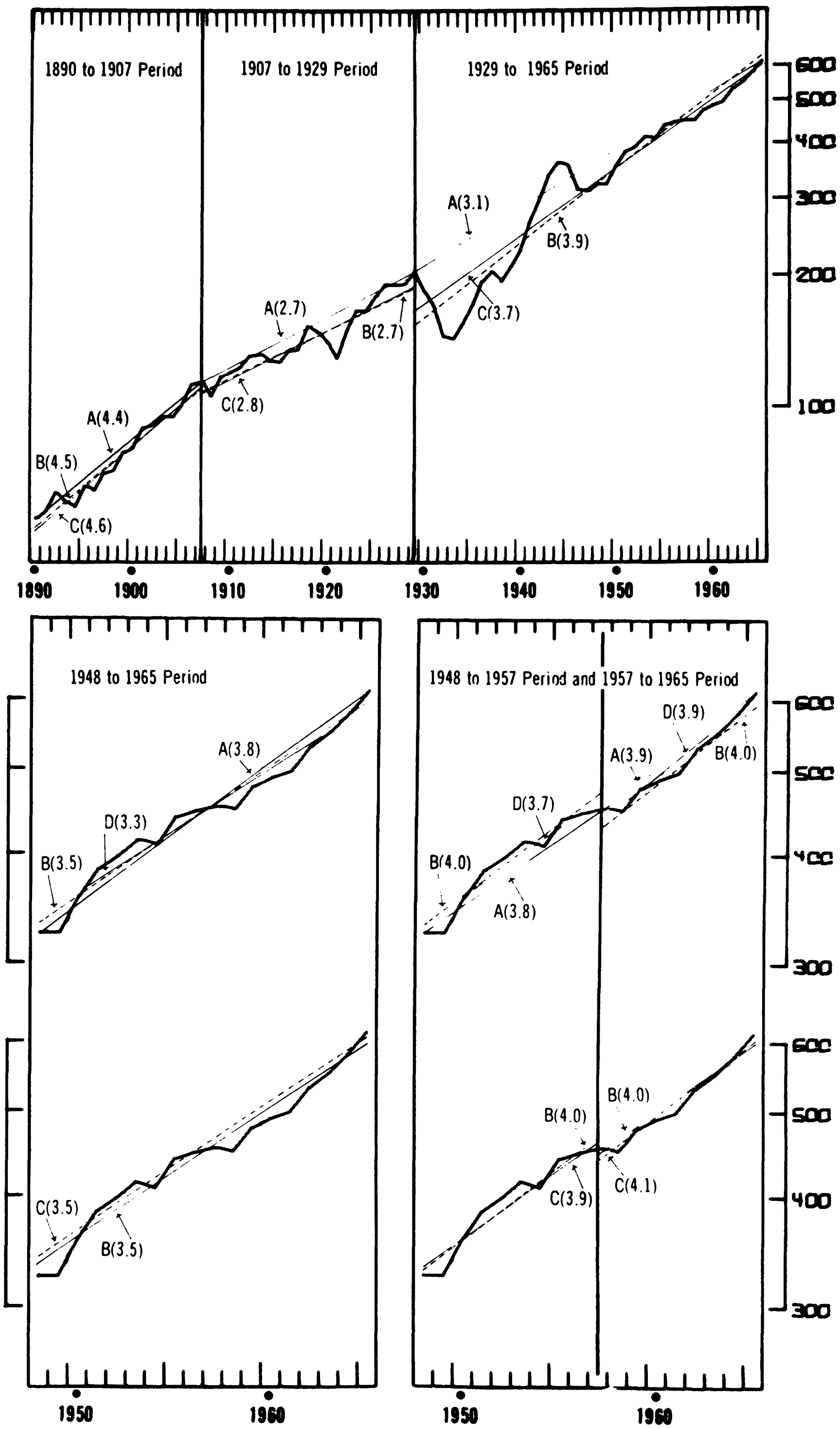
First, various measures of the growth of actual output of goods and services along with measures of potential output are presented. These are followed by measures of the growth of inputs of various human and material productive factors. The input measures indicate the changing levels of economic resources which have been used, or are available, over the time period covered. Finally, measures of productivity, obtained simply by dividing the volume of output by the number of units of input, are presented.

The second major section covers economic processes importantly related to economic growth. In some cases the relation to economic growth is clear. This is true for the series on education, health, and research and development. Other series

Chart 1

COMPARISON OF ALTERNATIVE GROWTH RATE FORMULAS, U.S. GNP, 1890 TO 1965

Charted below are the trend lines fitted by four alternative growth rate formulas. The growth rates are shown in parentheses after the letters designating the formulas.



A. Trend line calculated using initial and terminal years of annual data as selected points.
B. Trend line calculated by fitting an exponential equation to logarithms of annual data.
C. Trend line calculated by fitting an exponential equation to annual data -- Pesek.
D. Trend line calculated using initial and terminal business cycle averages as selected points.

represent background economic activities which certainly affect long-term economic growth, though how is less clear. These include data on prices and interest rates, savings and debt, the assets of financial institutions, the balance of payments, and monetary gold stock. The measures of the intensity of utilization of labor and capital resources and of the magnitude of seasonal and cyclical forces which are also included in this section, provide quantitative information which furnishes a perspective against which the measures of long-term growth can be better appraised.

In the third section are measures below the aggregate level which can be used to understand and interpret economic growth more effectively. Both regional and industry series are shown.

The fourth section shows measures of output, input, and productivity for six foreign countries. The countries are United Kingdom, Canada, West Germany, Italy, France, and Japan.

Various analytical aids are included in Section V and the appendixes: (1) Growth triangles, which make it possible to compare growth rates in the United States for any pair of years between 1890 and 1965, for GNP, manhours, and productivity. Criteria are provided to help in making judgments regarding the comparability of any two years used in the comparison. (2) A growth rate conversion table, which facilitates similar computations for the many other series in which the user may be interested. To use this table all that is needed is the ratio of the values for last and first years to be compared. The growth rate can then be found in this table. (3) Basic data and brief descriptions with references to more detailed explanations.

Since growth is essentially a long-term phenomenon, it cannot be considered in terms of developments since last year, the year before, 5 years ago, or perhaps even 10 years ago. Consequently, data in this report go back many years, wherever possible, to 1860.

In order to observe such long-term trends, we have had to build up series from different sources. Official Government series on gross national product extend back only to 1909 and the components only to 1929. However, various research students, particularly those at the National Bureau of Economic Research, have provided estimates extending back to the beginning of the industrial history of the United States, and these have been brought together with corresponding official government figures. Even when they are intended to measure the same thing, these series, being estimates, are often somewhat different. In addition, since different investigators were involved, there are some differences in concepts. Thus the series are not strictly comparable. In order to indicate the extent of differences, an overlap of about 10 years is provided, and a detailed description of each series and references to the author's original discussion are given in the descriptive appendix.

For this first edition we have not been able to consider data prior to 1860. We may do so for the next issue.

2. <u>Aggregate Output, Input and Productivity</u>

Altogether 58 output series are included in the first section. Gross national product data are used to measure output throughout this report, because of the difficulties of taking out depreciation. However, a single series on national income is shown so that we do not lose sight of the fact that this is the more ideal measure. Then some of the principal breakdowns of gross national product are presented--e.g., the gross private domestic product, gross nonagricultural product, gross manufacturing product, gross farm product, personal consumption expenditures, gross private domestic investment, and so on. Series on industrial production and personal income are also included. Finally, various income distributions are provided.

Next we turn to the input factors. These are viewed in broad terms and cover the supply and utilization of labor and capital. Two basic sets of total input estimates are available, one prepared by the National Bureau under the direction of John Kendrick and the other prepared by Edward Denison. The principal difference is that Denison allows for changes in the quality of labor. Unfortunately, the record for these series is not so long as that for output. Kendrick has decade estimates for 1869-78 and 1879-88, and then provides an annual series beginning in 1899. His series extend only to 1957, but we understand that he will bring these series up to date in the not too distant future. Denison's series start in 1909 and extend only to 1958.

In addition to these comprehensive measures of input, separate series for labor and capital input are also shown, not only at the aggregate level, but also for major components. Thus, series for total private manhours as well as manhours in nonagricultural, manufacturing, and agricultural industries are shown. Similar breakdowns are also shown for total employment. An occupational distribution of the labor force shows on a percentage basis the number of farm workers, manual workers, and white collar workers. Lebergott's early series for the labor force are included along with recent BLS data. Next, total population, the farm and nonfarm population, and the age distribution of the population are shown. These are followed by series for the birth, death, and immigration rates. Finally, Goldsmith's estimates of the civilian tangible wealth and many of its components (for example, the net reproducible private business wealth, the net stock of private residential nonfarm structures, and the stock of private inventories) are shown. The estimates available from the Office of Business Economics for the stock of fixed business capital, on alternative service lives, and business depreciation schedules for the period 1929-61 are also included.

The final part of the first section shows indexes of productivity. Here are included Kendrick's and Denison's series on output per unit of total input and details for labor input and capital input. These are followed by various series on output per employee and output per manhour.

3. Economic Processes Importantly Related to Economic Growth

The next section of this report presents measures of processes that appear to be strategic in determining the rate of productivity, that is, the factors which explain why output has grown more rapidly than input. Many scholars in this field hold the view that it is not a matter of one, two, or even three key factors, but rather that a large number of different factors have been responsible for the high productivity in the United States. Unfortunately, data are not available for many of them, and we are able to present information for only a few of the most strategic--in particular, education, health, and research and development. For education, such series as school enrollment, the average length of the public school term, and total expenditures in the education system are shown. Improvements in health represent another way of expanding the input of human resources, both in terms of quantity and quality. Under health, there are data for public expenditures for medical research, days lost by employed persons due to illness, and average life expectancy at birth. Research and development has increasingly been looked to as a way of improving the quality of capital; for this area, data on funds for scientific research and development, and applications for patents are given.

This section also includes a large number of series which provide a broad background of information which is helpful in making judgments of past and prospective performance. These include data on the money supply, both narrowly defined to include currency and demand deposits and broadly defined to include also time deposits. Two series on the velocity of the money supply are also shown here. Prices of commodities, money, and equities are included; and the implicit price deflators for total GNP and its major components. Series on profits, savings, the balance of payments, and the monetary gold stock follow.

Third, data on the utilization of resources, both of labor and of capital, which show how close to capacity the economy actually operated in particular periods of our history, and measures of the magnitude of cyclical fluctuations are also given here. These data are expected to contribute to good judgments about the validity of growth estimates over various time periods.

4. Regional and Industry Trends

The presentation up to this point is at the aggregate level and provides some relatively simple guidelines of overall performance. It is commonly recognized, however, that an aggregate is only a convenient summary of a large variety of activities that take place below this level, and detailed inspections of the pattern of events beneath is required for a thorough appreciation of factors affecting economic growth.

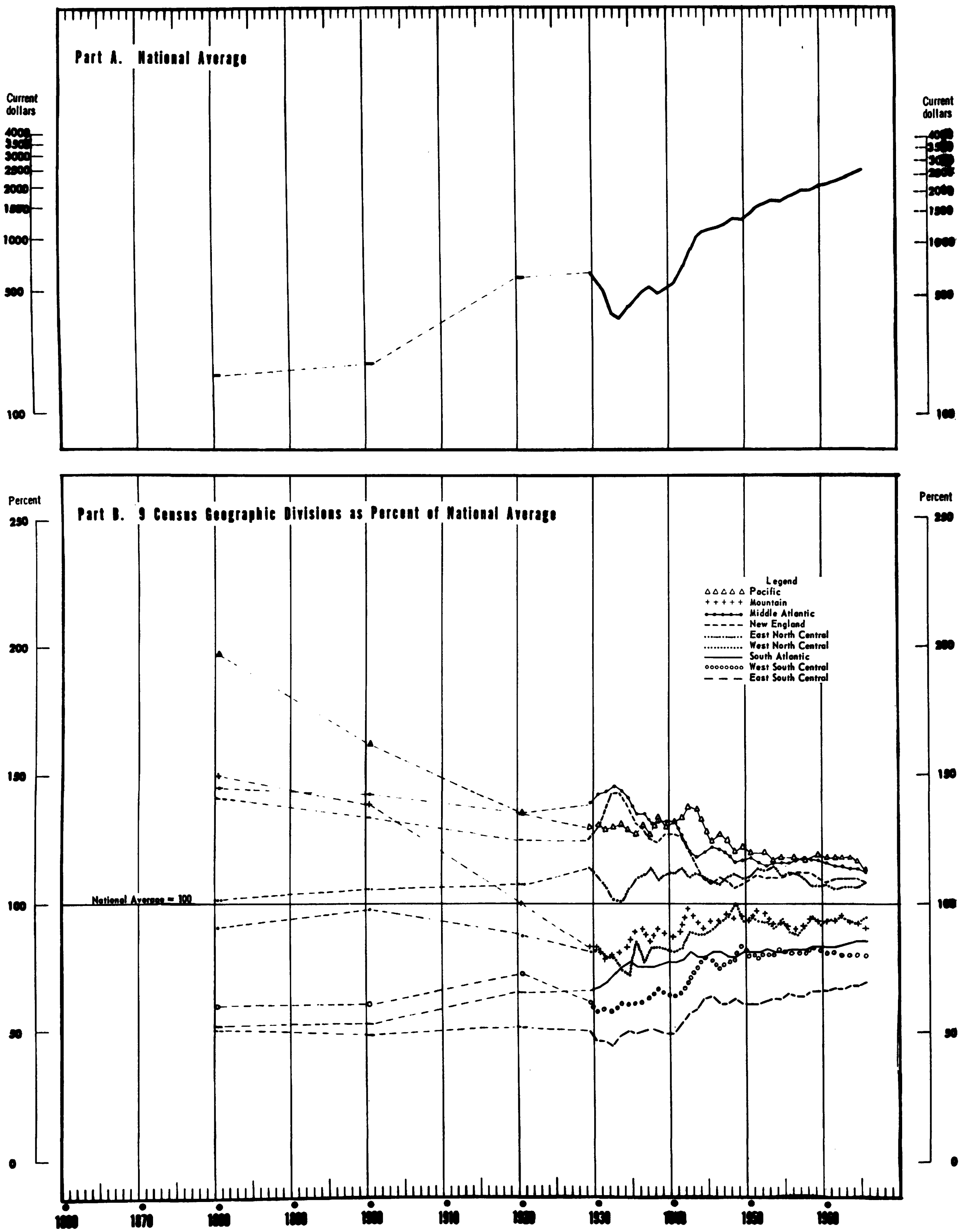
There are, of course, great volumes of U.S. data for regions and industries. To provide all such information in detail would swamp this whole report. Therefore, in order to bring out the principal regional and industrial developments without taking an undue amount of space in this volume, two presentation techniques have been employed in the third section of the report:

The first is the familiar method of plotting all the data for all the regions on the same time scale, and such charts for the 9 Census Geographic Divisions are shown for several measures including population, per capita personal income, and value added per employee in manufacturing. To indicate the relative importance of the development of the different geographic divisions, however, the data for each are shown as a percentage of the national average. As a result, most of the charts are quite similar to that illustrated below (Chart II) for per capita personal income, with fairly broad gaps among the divisions in the earlier periods of our history, 1880 and 1890, gradually being narrowed over the years until they are fairly close together now.

The second technique is a special type of scatter diagram. Here the growth rate for one period is plotted against the growth rate for another. For example, the growth rate for each State and Census Division for the period from 1929 to 1965 (vertical scale) is plotted against the growth rate for the period from 1880 to 1929 (horizontal scale). In this kind of chart the national average for the latter period is shown as a line drawn parallel to the horizontal scale, and the national average for the earlier period as a line drawn parallel to the vertical scale. For States falling in the area above the diagonal, the recent growth rate has been greater than the earlier growth rate. For States (or Divisions) shown in the upper left-hand boxed-off portion of the chart, the recent rate of growth has been greater than the national average in the recent period and below the national average in the early period. For States falling in the upper right boxed-off portion, growth rates were above the national average both in the recent and early periods. States shown in the lower left-hand portion were below the national average in both periods. States shown in the lower right-hand portion were below the national average in the recent period, but above in the early period. Thus this chart shows that, on a per capita basis, Florida, North Caroline, Texas, and West Virginia fared well in both periods. South Carolina, Arkansas, and Georgia did especially well in the recent period, but not so well in the early period. California was below the national average in both periods. In considering this statistic, it is to be borne

Chart II

PER CAPITA PERSONAL INCOME, UNITED STATES AND 9 CENSUS GEOGRAPHIC DIVISIONS, 1880 TO PRESENT



From 1929 to 1965, data are annual. Before 1929, data are available only for 1880, 1900 and an average for the period 1919 to 1921.

in mind, of course, that it is the changes which have been plotted. In terms of the level of per capita income, California, of course, rates very high.

Similar charts are shown for other comparison periods and for the various manufacturing industries. Thus, our chart for the manufacturing industries will show that, compared to total GNP, the best growth record since 1948 has been for the transportation, communication, and public utilities industries; the services industries; and the finance, insurance and real estate industries both since 1960, and also from 1948 to 1960. Construction has done relatively poorly. Among the individual manufacturing industries the recent record of the chemical industries, electrical machinery, and rubber is especially good.

5. International Comparisons

The interest in economic growth has come to the fore in recent years partly because of the greater awareness of the importance of this factor in determining the welfare of our own population and in resolving many of the difficult social problems affecting the poor, but also because economic growth has become an international issue. Thus accelerating economic growth has become a principal objective of economic policy in many of the under-developed countries. Adversary nations have pointed with pride to their rapid rates of economic growth and challenged our economic system to demonstrate that it can match theirs. In addition, the relatively poor economic performance in the U.S. during the later years of the fifties and the first few years of the sixties, compared to economic performance in Japan, Germany, France, and other Western nations, has been a cause of considerable concern here, and led to a careful re-examination of our own economic policy. For these reasons, a section showing the rates of growth in the United States and the principal industrialized countries with which we trade is included--United Kingdom, Canada, West Germany, Italy, France, and Japan. The number of countries covered in this section has been limited partly because fewer historical data are available for foreign countries than for the United States, partly because there are serious problems of comparability, and partly because of our own staff resources. In later editions we hope to add other countries to this section.

6. Analytical Measures

In this publication we depart from the more familiar types of statistical publications in several respects. First, the basic data are supported by computer-generated charts. Today charts are, of course, a common feature of many statistical publications. The fact that they are computer generated means that they can be used in much larger quantities. Indeed, they have become the primary method of presentation with tables occupying a relatively minor role. Most charts in this new publication are the familiar time series charts. Others are special types of charts, such as the scatter diagrams which provide a great deal of information in a small amount of space.

But, in addition to charts, we have included special "analytical" tables to facilitate studies of economic growth. The first of these are growth triangles. Growth triangles, now a familiar tool in growth presentations, show the same years along the horizontal and vertical scales. The growth rate between any two years can be found at the point of intersection between two lines perpendicular to the dates. Thus it is possible to find growth rates in GNP for any pair of years from 1890 to 1965 in our first growth triangle.

We have also introduced a new type of criterion, suggested by Edward F. Denison, in this table. Because of differences in the extent to which resources are utilized, or in other words, differences in the stage of the business cycle, every pair of years is not comparable from the point of view of measuring economic growth. For instance, the growth rate computed from a business cycle trough year to a business cycle peak year will be higher than the true rate of growth. Similarly, if we start with a business cycle peak year and end up with a business cycle trough year, the growth rate computed between these two years will be lower than the true rate of growth.

A measure that would be suitable for this purpose would be percent of total capacity with appropriate comparisons being those periods in which the economy operated at about the same rate of capacity. But such data are not available. The unemployment rate may be considered a measure of the extent to which the labor force is utilized and, therefore, when inverted, can serve as a proxy for a measure of capacity operation. Since data on the unemployment rate are available back to 1890, it has been used to call attention to those spans of years that will result in biased growth rates and to indicate years that are essentially comparable. Growth rates between years for which the unemployment rate is about the same are printed in black on a white background. Growth rates for which the unemployment rate in the initial year exceeds the rate in the terminal year are printed in black on a shaded background; these growth rates are likely to be greater than the true rate of economic growth. Growth rates for years in which the unemployment rate in the terminal year exceeds that in the initial year are printed in brown on a shaded background; for these the growth rate shown is likely to be less than the true rate of economic growth.

Two different standards have been used in preparing these tables. In one case we have had fairly exacting standards and in another more relaxed standards. Consequently, 13 percent of the 2850 possible comparisons in the first table show growth rates which are comparable under our assumptions. In the table with the more relaxed standards, 35 percent of the 2850 possible comparisons show growth rates that are comparable under our assumptions. In addition to these two growth triangles for GNP, there are also included growth triangles for total manhours and gross private product per manhour. In these four tables the compound interest rate formula is used to compute the growth rates between the initial and terminal year. As an alternative the growth rate computed with a linear trend fitted to the logarithms of the data is shown for total output.

One of the principal requests made by those who reviewed earlier editions of our new report was for more growth rate triangles. Since these are very space consuming and we could have had one for just about every series in the book, we sought a simple way of meeting this interest without unduly expanding the volume. Our solution was the preparation of a growth rate conversion table. Here the familiar compound interest rate table is modified so that the user no longer has to interpolate between tabled values. To use this new type of table, three simple steps are necessary: to (1) compute the ratio of the value in the later year to the value in the earlier year; (2) check the stub of the table to find the number of years over which the comparison is being made; and (3) search on that line for the two values between which this ratio falls. The rate of growth is then given on the top row between these two values. For example: GNP was $452.5 billion in 1957 and $614.4 billion in 1965. The ratio of 614.4 to 452.5 is 1.35779, and the number of years spanned from 1965-1957 is 8. The average annual growth rate is then found by locating the interval within which 1.35779 falls on the 8-year horizontal line, i.e., 3.9%.

This table covers 70 years and the growth rate is shown to one decimal. We have also prepared, and can make available on a cost basis, similar tables showing the growth rate to two decimal places or growth rates above the 10% limit in the present table. In addition we have provided a formula for computing growth rates for periods longer than 70 years but less than 140.

The growth rate conversion table is useful for computing the growth rate for any series between any pair of historical years. For extrapolating growth rates we have also provided a standard compound interest rate table for periods from 1 to 20 years. More detailed compound interest tables can be obtained from other sources.

V. Long-Term Projections

The report to be published next month is a statistical history of economic growth in the United States. Such a history is of interest for its own sake, but the information it presents also may reveal important knowledge that can be helpful in stimulating growth in future years. A related use of these data is to provide the basis for forecasts of future growth in the U.S. These in turn are helpful in a large variety of necessary long-range planning projects, such as the aggregate demand for goods and services, urban development, transportation facilities, educational requirements, and so on.

To close this paper, I thought it might be helpful to provide one illustration of how this new report can be used by presenting a few representative long-term projections.

Chart III and Tables 2 and 3 show two types of projections of GNP to 1980. First are analytical projections, which attempt to allow explicitly for factors that may affect future economic growth. They have been prepared by various Government agencies and private planning organizations.[5]

Two major assumptions underlie all these analytical projections: (1) there will be no deep or prolonged depressions, and (2) the unemployment rate will fall in the range 4.0 to 4.5 percent in the terminal year.

The second type are "naive" projections, which assume that the trend of a given historical period will continue into the future. They do not take into account in a systematic way prospective policy changes and structural shifts in the economy, and for this reason are not forecasts in an economic sense. But they do provide a broad perspective for judging future prospects. At a minimum they provide a standard against which analytical projections can be judged, by establishing a range within which an analytical projection would be expected to fall, if past conditions do not change much. Conversely, the "naive" projections can help to indicate the impact of any major change in past conditions assumed in preparing an analytical projection.

The figures used to make up these projections are provided in Tables 2 and 3 and, in addition, corresponding projections for labor input and productivity are shown. First, an observation about the relations between the analytical and "naive" projections--the "naive" projections fall over a wider range than the analytical projections. If our recorded history is used as the basis for projecting, the range of possibilities in the future would appear to be greater than if the analytical projections are used. Most experts believe the analytical projections will prove to be more accurate than the "naive" projections. One reason is some of the underlying conditions, particularly the future population of working age, can be fairly accurately estimated on the basis of the present population. The "naive" projections implicitly allow for more variation because the population of working age has grown at different rates in different historical periods. However, we have learned from experience that it is very difficult to make accurate projections. One danger of the analytical projections is that most forecasters are heavily swayed by the conventional wisdom of the day, and base their work on similar assumptions. This may be part of the explanation why the range is smaller than that of the "naive" projections.

---

[5]The Government agencies are the Council of Economic Advisers and Joint Economic Committee of Congress and the non-Government agencies are the Committee for Economic Development, National Planning Association, Resources for the Future, McGraw-Hill, and National Industrial Conference Board.

Chart III

GROSS NATIONAL PRODUCT IN 1965 DOLLARS PROJECTED TO 1980

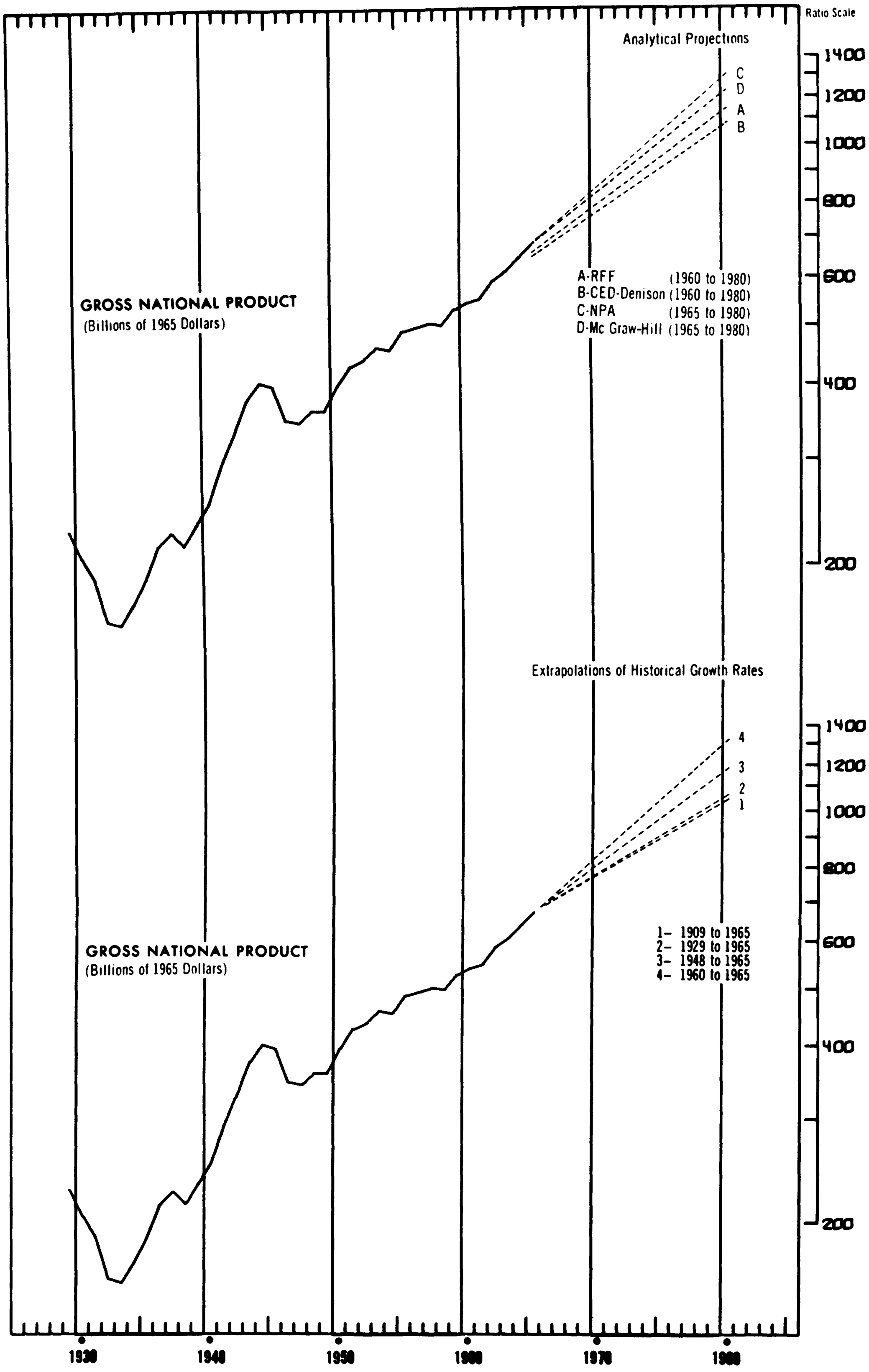


Table 2.--Analytical Projections to 1975 and 1980 Compared With Historical Growth Rates

PART A.--Analytical Projections -- Average Annual Growth Rates

	CEA 1964 to 1970	CED-Denison 1960 to 1975	CED-Denison 1960 to 1980	JEC-Knowles 1959 to 1975	NPA 1965 to 1975	NPA 1965 to 1980	RFF 1960 to 1980	NICB 1964 to 1975	McGraw-Hill 1965 to 1980
Total Labor Force.....	1.7	--	1.7	1.7	1.8	1.7	1.7	1.8	1.7
Total Employment......	--	--	1.68	1.7	1.9	1.8	1.8	1.9	1.9
Average Weekly Hours									
Private............	--	--	--	--	-0.4	-0.4	-0.3	-0.4	--
Total..............	--	--	-0.53	-0.5	-0.4	-0.4	--	-0.4	-0.5
Man-Hours									
Private............	--	--	--	--	1.3	1.2	1.1	--	--
Total..............	--	--	--	1.2	1.5	1.4	--	1.5	1.3
Output Per Man-Hour									
Private............	--	--	--	--	3.3	3.4	2.7	3.0	--
Total..............	--	--	--	3.5	3.0	3.0	--	2.8	2.75
Output									
Private............	--	--	--	--	4.7	4.5	3.8	--	4.1
GNP									
[1]Actual...........	4.7	3.55	3.52	4.7	4.5	4.4	3.8	4.35	4.1
[2]Potential........	4.0	3.30	3.33	4.0	4.3	4.3	--	--	--

PART B.--Historical Average Annual Growth Rates

	1909 to 1965	1929 to 1965	1948 to 1965	1960 to 1965
Total Labor Force................	1.4	1.3	1.3	1.4
Total Employment.................	1.4	1.3	1.3	1.6
Average Weekly Hours				
Civilian.......................	-0.4	-0.5	-0.3	0.0
Total (NPA)....................	-0.5[3]	-0.5[4]	-0.4	-0.1
Man-Hours				
Private........................	0.6	0.3	0.5	1.3
Total (NPA)....................	0.9	0.6	0.9	1.6
Output Per Man-Hour				
Private........................	2.3	2.7	3.4	3.5
Total (NPA)....................	2.1	2.5	2.9	3.1
Output				
Private........................	2.9	3.1	3.9	4.9
GNP............................	3.0	3.1	3.8	4.7
Per Capita Disposable Income (1965 $)......................	--	1.6	2.1	3.3
Industrial Production (Index: 1957-59 = 100).................	3.8	3.7	4.4	5.7

[1]Four percent unemployment rate assumed in terminal year, except for NICB which assumes 4.5 percent.

[2]Potential defined as the GNP which would be produced if unemployment were 4.0 percent in initial and terminal year.

[3]Initial year is 1910.

[4]Initial year is 1930.

Table 3.--Levels in 1975 and 1980 Implied by Analytical Projections and by Extrapolation of Historical Growth Rates

(The levels shown below were calculated by extrapolating the currently published data with the growth rates shown in Table 2. For Part B the extrapolations were made from 1965 and for Part A from the years enclosed in parenthesis.)

PART A.--Analytical Projections

	1965 Value	1975				1980			
		JEC Knowles (1959)	CED Denison (1960)	NPA (1965)	NICB (1964)	RFF (1960)	CED Denison (1960)	NPA (1965)	McGraw-Hill (1965)
Total Labor Force (thousands).	78,357	94,220	--	93,930	93,000	102,445	102,445	101,400	101,400
Total Employment (thousands)..	74,901	89,226	--	90,280	89,000	98,862	96,557	97,500	94,900
Average Weekly Hours									
Private (hours)(NPA)........	38.86	--	--	37.28	--	36.8	--	36.52	--
Total (hours)(NPA)..........	38.69	35.8	--	37.17	36.7	--	35.0	36.43	35.5
Man-Hours									
Private (index: 1965=100)...	100.0	--	--	113.7	--	116.6	--	118.8	--
Total (index: 1965=100)(NPA)	100.0			115.8	117.6	--	--	122.6	121.6
Output Per Man-Hour									
Private (index: 1965=100)...	100.0	--	--	139.0	138.4	144.8	--	164.1	--
Total (index: 1965=100)(NPA)..	100.0			134.4	135.5	--	--	155.8	150.2
Output									
Private (index: 1965=100)...	100.0	--	--	158.0	--	168.0	--	194.9	--
GNP (index: 1965 = 100).....	100.0	162.9	135.0	155.6	159.3	168.7	159.8	191.0	--
GNP (bil. of 1965 $)........	681.2	1101.4	913.0	1057.9	1027.5	1140.6	1080.6	1299.5	1244.6

PART B.--Extrapolation of Historical Growth Rates

	1965 Value	1975				1980			
		1909 to 1965	1929 to 1965	1948 to 1965	1960 to 1965	1909 to 1965	1929 to 1965	1948 to 1965	1960 to 1965
Total Labor Force (thousands).	78,357	90,045	89,160	89,160	90,045	96,526	95,108	95,108	96,526
Total Employment (thousands)..	74,901	86,073	85,228	85,228	87,786	92,269	90,913	90,913	95,037
Average Weekly Hours									
Civilian (hours)............	40.5	38.9	38.5	39.3	40.5	38.1	37.6	38.7	40.5
Total (hours)(NPA)..........	38.69	36.81	36.81	37.18	38.31	35.90	35.90	36.44	38.11
Man-Hours									
Private (index: 1965=100)...	100.0	106.2	103.0	105.1	113.8	109.4	104.6	107.8	121.4
Total (index: 1965=100)(NPA)	100.0	109.4	106.2	109.4	117.2	114.4	109.4	114.4	126.9
Output Per Man-Hour									
Private (index: 1965=100)...	100.0	125.5	130.5	139.7	141.1	140.6	149.1	165.1	167.5
Total (index: 1965=100)(NPA)	100.0	123.1	128.0	133.1	135.7	136.6	144.8	153.5	158.1
Output									
Private (index: 1965=100)...	100.0	133.1	135.7	146.6	161.3	153.5	158.1	177.5	204.9
GNP (index: 1965 = 100).....	100.0	134.4	135.7	145.2	158.3	155.8	158.1	175.0	199.2
GNP (bil. of 1965 $)........	681.2	915.5	924.4	989.1	1078.3	1061.3	1076.8	1191.9	1356.7
Per Capita Disposable Income (1965 $)....................	2411	--	2826	2968	3336	--	3059	3293	3924
Industrial Production (Index: 1957-59 = 100)	143.3	208.1	206.1	220.4	249.5	250.7	247.1	273.4	329.1

The sources of the projections shown in Tables 2 and 3 are shown at end of text.

To consider some of the prospects, I have selected three different projections. One is the highest among them, the other is one of the lowest among them and the third is the median. These all turn out to be "naive" projections, but similar conclusions could be drawn from the analytical projections.

The implications of the recent improvement in economic growth and stability are staggering to the imagination. A continuation of recent trends will carry us to unbelievable levels of economic activity in our own lifetimes.

The divergence of these various curves as they approach 1980 indicates how important relatively small difference in annual growth rates can be when cumulated over longer periods of time. However, even if we repeat the experience since 1929, one of the slowest growth rates projected, we shall have by 1980 a 58 percent growth in gross national product in constant dollars and 27 percent growth in per capita disposable personal income. A continuation of the record since 1948 will yield an increase of about 75 percent in GNP and 35 percent in per capita disposable income. If we have, indeed, conquered the business cycle, we shall do far better. The increase in gross national product in constant dollars will be almost double and the increase in per capita disposable personal income about 60 percent. Industrial production could increase even more rapidly, 70 percent on the most unfavorable assumption and 130 percent on the most favorable. It seems most unlikely, however, that consumers would want to take so much of their increased income in terms of goods. What would we do with all of them? More likely there will be substantial shifts from goods to more services and from goods and services to more leisure.

Thus there is in sight, within our own lifetimes, the prospect of another vast improvement in economic welfare. This is not to say that by 1980 we shall have enough to meet all our economic aspirations. But we shall have a great deal more than we have now, even though we encounter many unexpected pitfalls which impede our progress.

This projection exercise illustrates one important way of exploiting some of the data brought together in this report. We are hopeful that it will facilitate the preparation of new and better projections and that it will be put to many different additional uses. We shall be very glad to hear of your experiences with it.

The sources of the projections shown in Tables 2 and 3 are listed below.

CEA	Council of Economic Advisers, Annual Report, January 1965.
JEC-Knowles	James W. Knowles, The Potential Economic Growth in the United States, prepared for the Study of Employment Growth, and Price Levels, Joint Economic Committee, Congress of the United States, January 30, 1960.
CED-Denison	Edward F. Denison, The Sources of Economic Growth in the United States and The Alternatives Before Us, Committee for Economic Development, 1962.
NPA	National Economic Projections to 1976-77, National Economic Projections Series, National Planning Association, to be published in September 1966. The 1980 figures were taken from NPA worksheets.
RFF	Hans H. Landsberg, Leonard L. Fischman and Joseph L. Fisher, Resources in America's Future, Resources for the Future, Inc.
NICB	Supplied by the National Industrial Conference Board. See also, "Economic Potentials of the United States for the Next Decade," reprinted from The Conference Board Record, December 1965, NICB.
McGraw-Hill	American Prospects For Growth Through 1980, McGraw-Hill Economics Department, McGraw-Hill, Inc.

ARTICULATING THE TIME DIMENSION FOR ECONOMIC ANALYSIS

Gerhard Bry, New York University, Graduate School of Business Administration

and National Bureau of Economic Research

I. The Need for Articulation

This paper deals with various ways in which the time variable can be handled in the statistical analysis of economic behavior. In our rapidly changing society, one thing is certain: at any point in time economic conditions are different from those at any other point. Similarities and regularities, not to speak of constancies, must be discovered and can rarely be presumed. It is true, of course, that the search for regularities and differences must not be blind. We have the right to expect peacetimes to differ from wartimes, inflations from deflations, fair deals from square deals, cyclical expansions from contractions, and so forth. Conversely we may hope that all contractions, all periods of steady growth--or perhaps some of them--have features that permit qualitative and quantitative generalizations as well as identification of unique experiences.

Analyses of individual time series are usually well attuned to the need for describing, and distinguishing between, the historically unique, the cycle specific, and the systematically pervasive.[1] They permit visual and quantitative distinction between early and late segments; long-term, cyclical, seasonal, and irregular behavior; and behavior during subperiods that may be homogeneous with regard to specified characteristics. Maintenance of historical integrity by unsummarized descriptions, as well as the summarization of like subperiods, is at least possible on the basis of the conventional Harvard technique of time series desegregation; it is a major concern of the various forms of business cycle analysis developed at the National Bureau of Economic Research.[2]

However, in describing economic relationships among several activities--whether by regression analysis, econometric model-building, or cross-spectra--we tend to proceed as if economic behavior were a basically stable process. This is reflected in the fact that most applications of these approaches describe economic behavior by a fixed selection of variables to which constant coefficients are attached and presumed to be valid over extended time periods without much regard to the characteristics of subperiods. The presence of an explicit time variable and the behavior of "residuals" reflect some aspects of changing dynamics. The time variable, however, veils rather than reveals the long-term functional relationships between the included variables. And the residuals do not, of course, describe changes in relationships.

It is the thesis of this paper that treating time summarily is neither desirable nor necessary, and that relationships should be described and tested for relevant subperiods, however they may be defined. Any specified segmentation of the time scale will be termed its articulation.

II. Articulations Currently in Use

The idea of subdividing the time period of analysis is certainly not new to theorists, econometricians, or others. Formulations are found in the literature that refer only to narrowly specified time periods, specified cycle phases, specified time series components, and so forth. A brief sampling of recent publications will show the wide variety of ways in which the time dimension has been handled in economic analysis.

At one extreme there is the undifferentiated treatment of time; parametric stability is assumed throughout the period of observation and frequently beyond. Many of the newer analyses that use econometric models are highly sophisticated with regard to inclusion of variables, degree of desegregation, statement of interdependencies, use of optimal or of distributed lags, expansion of strategic subsectors, analysis of residuals, and so on. But they tend to postulate the basic stability of the systematic relationships during the time period covered.

At the other extreme are formulations that stress the historical uniqueness of economic relationships during narrowly defined time periods. Economic historians can be expected to be most conscious of this uniqueness. The informal model

---

Acknowledgment: It gives me pleasure to acknowledge my indebtedness to Charlotte Boschan, who prevented me from overestimating the justness of my cause; to Geoffrey H. Moore, who always makes me work a little more; to Judy Feins, whose assistance permitted me to work a little less; and to H. Irving Forman, who drew the charts. The electronic computer time needed for the calculations was provided by a grant from the International Business Machines Corporation.

constructed by Conrad and Meyer to investigate the profitability of the slave system before the Civil War is an example.[3] An obvious implication of their approach is the belief that analytical problems, strategic variables, and of course most parameters can differ vastly from period to period. Hence there is a need to construct models that mirror the conditions of the specific regime. Prices of slaves, costs of their sustenance and reproduction, yields per slave and per acre, prices of cotton--these are the variables necessary to understand the dynamics and test the viability of economic institutions during that particular period.

But it is not necessary to hark back that far in history to establish the need for restricting the validity of economic models to a small number of years. Rendigs Fels devised a special model to explain the 1948-49 recession. It is highly pertinent, for our purposes, to listen to his reasons: "Since the circumstances preceeding the 1948 downturn were unusual, none of the general-purpose models of the business cycle fits the facts adequately, and nobody has devised a model for this special purpose."[4] Fels's model is "special issue," both with regard to variables and, of course, to parameters. Arguments about the uniqueness of historical circumstances could of course be advanced for any contraction, expansion, or, for that matter, any other subdivision of economic history. Fels is quite aware of the broad methodological significance of his procedure, for, in writing about historically unique models in connection with business cycle research, he concludes that "it is surprising that this approach did not long ago become standard operating procedure for those like myself who study business cycles as historical individuals...it seems to me quite possible that we shall solve the riddle of the business cycle only by constructing tailor-made explanations for each one and then examining them for the common elements running through them all."[5] Fels asks for rigorous formulation of relationships during narrowly defined periods and for conclusions drawn from comparative analysis. In view of the theory-bound nature of models, this approach may lack the objectivity that is sometimes associated with induction. However, in the context of the present concern, his arguments for formulations during contiguous historically and cyclically defined time periods, with comparisons and summarizations thereafter, are of great relevance.

Theoretical and empirical justification for historical segmentation is not hard to find. Even basic mechanisms change their character--temporarily or permanently. Gardner Ackley writes: "Fixed investment has generally been viewed as the main source of instability in the economy. Theoreticians have devised several mechanisms which account for its unstable behavior. The experience of the mid-1950's reinforced our preconceptions on the subject. Then why didn't it happen this time?"[6] The implication is that other formulations of economic interrelations may be required for recent years than for the fifties.[7]

Between the extremes of undifferentiated and historically articulated treatment of time, there are attempts to combine what may be called dynamically related time segments. These may be groupings into short- or long-term movements, cyclical expansions and contractions, or other subdivisions. One of the oldest distinctions is that between long-term and short-term relationships. Franco Modigliani,[8] in his analysis of the saving-investment ratio, distinguishes a secular movement that "carries real income per capita above the highest level reached in any year," from a cyclical movement, up or down, that "leaves real income per capita below the highest previous peak." As a result of this distinction, he obtains a measure of cyclical marginal propensity to save of .23 which is strikingly different from a secular propensity of only about .10 or .11. Similarly, Paul Boschan,[9] in his analysis of steel production, shows that a crude correlation with industrial production for the period 1919-40 leads to a regression equation that describes neither cyclical nor trend relationships. Boschan distinguishes between a capacity-determined long-term relationship (represented by an associated function of capacity estimates) and an inventory-demand-determined short-run cyclical relationship (represented by capacity utilization rates). He also distinguishes between demand conditions during the twenties and thirties. These distinctions are not only analytically superior but they result in closer relationships and in improved projections, both for the short and the long term. Note that Modigliani and Boschan solved their problem of articulation not by explicit segmentation of the time scale but by a formulation that allows for differential responses depending on whether there is much slack or little slack in the economy. These time periods were delineated by the behavior of a specific variable.

More recently, Milton Friedman[10] differentiated between the consumption effect of permanent income and that of transitory income. This distinction is closely related to the problem of articulating the time dimension. Friedman not only measures the effects of the permanent income component by relating income to consumption over extended periods but also explains the systematic tendency of the income elasticity of consumption to rise with the extension of the measurement period. He stresses the different responses of consumers during extraordinary periods such as the two world wars and the Great Depression. The following observations are of particular relevance: "Human beings are more flexible than the particular mathematical equations we used to summarize their behavior; they recognized, as the equations could not, that the Great Depression was something exceptional and special to be taken into account in a different way than the run-of-the-mill up and down of economic activity."[11] This raises the question whether such extraordinary periods should perhaps be excluded from the material that serves as the basis for broad generalizations and the economic mechanisms under such circumstances be described separately.

A cycle-phase oriented articulation was used in Duesenberry, Eckstein, and Fromm's[12] simulation of recession experiences for different mixes of automatic stabilizers. The experiment was conducted for a recession starting in the third quarter of 1957. The parameters were based on historical evidence for the postwar period, and sometimes for earlier years. A clear distinction between recession and nonrecession experiences was made: "For several of the equations... the system is assumed to behave differently in recessions than at other times. Data drawn only from recession periods were used to estimate these functions. As a consequence the model is appropriate only for recessions. It is not designed to explain the upper turning point in the business cycle, nor is it appropriate for periods of general prosperity....A model fitted to a wider range of business cycle conditions might not have been able to reflect the stability properties of the system."[13]

Several other authors distinguish between expansions (or early, late, and full expansions) and contractions. They present parameters separately for every cycle phase, for all corresponding cycle phases,[14] and for long-term relationships. A number of analyses differentiate time periods or single years by introducing dummy variables into the descriptive equations. This may well be used to characterize time periods and years as similar to others, as dissimilar from others, or as unique. The analysis may suggest the existence of differences in structural relationships during different time periods. However, the coefficients of the dummy variables do not describe the differential relationships themselves, if they exist.

Let us conclude our report with a particularly interesting instance, in which cyclical articulation did not only affect the parameters but also led to a radically dichotomous formulation of cyclical dynamics. Meyer and Glauber[15] hypothesize that under conditions of full capacity utilization, a type of accelerator mechanism determines investment behavior. By contrast, during slack periods in the utilization of existing capacity, investment is strategically influenced by cash inflows from current operations. Regardless of the validity of this hypothesis, its mere existence emphasizes the possibility of the dominance of different economic mechanisms at different times, and the loss of insight that may result from a failure to articulate the time dimension.

III. Some Experiments with Articulation

This section demonstrates the wide variation in statistical measures that may result from alternative articulations of the time dimension, and the benefits that may accrue from such articulation. First we shall experiment with alternative time segments, then with alternative time units and time spans.

In the first group of experiments, correlation and regression analysis is applied to various time segments of four variables--the dependent variable corporate profits, and the independent variables corporate product, unemployment rate, and time. Corporate product is meant to reflect industry growth and possibly the associated external and internal economies of scale; it is expected to vary positively with profits over the short and long run. The unemployment rate is introduced as a stand-in for capacity utilization. On a gross basis it should be inversely related to profits, particularly for cyclical movements. At very low unemployment rates (very high capacity utilization), the relationship may be expected to weaken. Finally, the time variable is introduced to relieve the other independent variables from explaining long-term relationships so that they may reflect shorter-term responses more effectively. Several combinations of the three independent variables are used, but no lags--mainly to avoid undue proliferation of alternatives. Let us begin our experiments with bivariate analysis.

Chart 1 shows corporate profits plotted against corporate product. The long line of average relationship, identified by its regression coefficient b = .13, is the least-squares approximation for the full period 1946/I-1966/I. Note that the line provides only a moderately good fit: it is steeper than the scatter configuration during the first five years, flatter than the configuration during the last four or five years, and it fails to represent either the long-term or the short-term responses during the middle period. Nevertheless, as column 2 of Table 1 shows, the correlation coefficient is as high as .93, the adjusted coefficient of determination .86. The reason for the high correlation is, of course, the pronounced upward trends in both series.[16] These create a huge total variance, of which the unexplained variance constitutes only a small fraction. For our purpose, this means that the correlation coefficients may be unsatisfactory criteria for comparing the effectiveness of subdividing the period into major time segments. This will become apparent shortly.

Chart 1 also contains regression lines for three subperiods which are identified.[17] Obviously the fit for the early and late segments is very close, with correlation coefficients of .99 in both cases. The slope of the regression line for the middle segment expresses the long-term tendency of this segment better than the line for the full scatter. The correlation coefficient, however, is only .53, with an adjusted R-square of .27. This merely reflects the fact that the short-term fluctuations of profits are inadequately explained by the regression. Visual inspection of the scatter supports the contention that subdivision of the time period led to markedly improved representation, particularly for the first and last segments. The difference between the regression coefficients (.33, .06, and .22 respectively) seems to indicate that responses were substantially different in the three segments.

It was noted that the short-term response patterns during the middle period are not well represented by its regression line. Let us see what further subdivision of the time variable will do.[18] Chart 2 reflects a subdivision of time into segments that show similar response patterns. The scatter suggests that early profit contractions are relatively mild and inversely correlated, uncorrelated, or perhaps positively correlated with the continuing rise in corporate product.[19] By contrast, during late profit contractions and the subsequent expansions, a sharp positive response pattern is found. The similarity among the responses is well described by regression coefficients within the narrow range of .34 to .42. These results are important, and suggest that proper articulation of the time dimension may permit discovery of homogeneous response patterns in a number of (cyclical or other) subperiods that lend themselves to common inquiry and effective generalizations.

It may be said that this periodization was "ex post" and thus subject to methodological objections. Although pragmatically discovered similarities are legitimate starting points for analytical inquiries, it is not necessary to depend on them. Systematic regularities in responses of corporate profits to corporate product can also be found within a predetermined framework of time periods. The following tabulation shows the regression coefficients (profits versus product) for expansion and contraction periods of general business activity, as delineated by the business cycle chronology of the National Bureau (shown in years and quarters).

Expansions	Regression Coefficients
1946/1 to 1948/4	+.34
1949/4 to 1953/3	+.14
1954/3 to 1957/3	+.09
1958/2 to 1960/2	+.26
1961/1 to 1966/1	+.22
Contractions	
1948/4 to 1949/4	+.55
1953/3 to 1954/3	+.21
1957/3 to 1958/2	+.56
1960/2 to 1961/1	+.67

A given change in corporate product seems to evoke a smaller change in corporate profits during business cycle expansions than during contractions. The observed difference may be largely due to the long leads of the profit variable at peaks. Whatever the reason, the systematic difference in response patterns is apparent after segmentation of the time period into business cycle phases.

So far the emphasis has been mainly on differences in the parameters of simple regressions during different time periods. Inclusion of proper additional variables would not only give a fuller explanation of the dependent variable but might also modify the parameters of a simple explanatory variable (say corporate product) in such a way that the sharp differences between the subperiods would disappear. However, this should not be taken for granted. Table 1 shows what happens if time and unemployment are added to the independent variable. The addition of time adds to the explanation and has indeed an equalizing effect on the regression coefficient of the product variable. In the bivariate analysis the regression coefficient for corporate product varies between .06 and .33 (see column 2); after addition of the time variable, the variation is confined within the narrow limits of .39 and .45 (column 3), which, incidentally, is close to the values of the short-term responses shown in Chart 2. In this sense, the addition of the time variable is highly effective. However, addition of variables does not necessarily have such homogenizing effects. Column 7 of the table shows the coefficients for the full system of four variables. Note the enormous variation of the regression coefficients, for any of the independent variables, from subperiod to subperiod. Perhaps a more circumspect selection of variables could bring about a degree of structural stability which would make periodization unnecessary. It might, if the structural changes of the system are caught by the variables and the expression of their functional relationships. But in order to demonstrate either that temporal subdivisions are necessary or that they are not, we must analyze the time segments--in short, we must articulate the time dimension.

There are obvious limitations to the disaggregation of time periods. If the time segments chosen are short and the independent variables numerous, multivariate analysis tends to break down. We run out of degrees of freedom, we increase multicollinearity, we lose significance--all aspects of the same problem.[20] The first variable introduced will tend to explain most of the explainable variance. Thus, if we wish to have separate measures for short subperiods, we may have to consolidate similar periods and handle the problem of time trends statistically (for instance, by expressing data for each subperiod as relatives of the average for that period;) alternatively, we may have to confine ourselves to bivariate analysis. As a basis for judging the homogeneity of subperiods and for related research purposes, a systematic investigation of bivariate relationships might prove to be of value. A multidimensional matrix showing gross relationships (in the form of simple regression functions or of elasticities) between the hundred or so strategic variables that are commonly included in macroeconomic systems would constitute a research tool to supplement similar systems of basic measures that exist or are being developed for single time series.[21] The cycle stages and cycle averages of the National Bureau's business cycle analysis might provide a convenient chronological framework for the computation of such response patterns. I shall not elaborate on this suggestion with a discussion of details. Response measures for each variable to each of the others, for all conceivable time periods, would surely not be necessary. But a data bank and a program that would permit computing such measures for any specified combination of variables and periods might well be wanted. That is another possibility of articulating the time dimension in measuring economic relationships.

In the second group of experiments, concern is not with the segmentation of the time scale into subperiods but with the effects of modifying the time units and time spans used in the statistical analysis of economic data. Analytically it does not make much difference whether production of coal is measured in units of short tons or million tons, and whether prices are given in cents or dollars. But it may make a lot of difference whether the time units are months, years, or longer periods. The aggregation into longer units averages intraunit fluctuations, and this affects the measurement of these fluctuations as well as any statements of the relationship between fluctuations of several activities. Obviously, correlation between annual time series will not reflect seasonal and shorter-term covariation, and may reflect cyclical covariation to a limited degree only. Similar observations could be made about changes covering different time spans. Since all this is well known, it is puzzling why the effect of different units and different time spans is not more frequently determined in quantitative economic research.[22]

To dispel any possible thought that these effects are really negligible, correlation and regression analysis were made for the same set of variables, using different time units and time spans. Throughout this experimentation, the average work week, as the dependent variable, is related to employment, the unemployment rate, and time--singly or in combination, for the period 1929-65. Employment is here conceived as a measure of effective demand for labor input. It is expected to vary positively with the duration of the average workweek over the short run,[23] though not necessarily over the long run. The unemployment rate is introduced as a measure of the tightness of the labor market on the supply side. Over the short term, low unemployment (limited additional supply) tends to force employers to increase labor input by lengthening the workweek, which leads to the expectation of an inverse gross relationship between weekly hours and unemployment rate. Over the long run, the declining secular trends in the average workweek and in the unemployment rate contribute toward a positive relationship. The time variable is designed to relieve the labor market factors, at least in part, of the task of describing long-term relationships.

Table 2 shows the results of the analysis for monthly, quarterly, annual, quinquennial, and decennial data. There is not much point in recapitulating the information contained in the table. Suffice it to say that correlation coefficients as well as regression coefficients are sensitive to modifications in time units, smoothing terms, and time spans. Characteristically, the correlation increases gradually with increasing size of unit and span. In the case of employment (column 2), the increase is substantial. This is readily understood in view of the gradually decreasing importance of random elements and the increasing importance of long-term trends. On occasion, reversals in the direction of change of the correlation coefficients are observed. This usually occurs when the cyclical and long-term relationships are in opposite directions, so that first the cyclical but eventually the long-term forces dominate the relationship. Similar changes and reversals in direction are also observable for the various regression coefficients, sometimes accompanied by reversals of sign (column 7, smoothing terms, employment). In evaluating the sensitivity of the coefficients to changes in time definitions, it must be noted that smoothing terms and time spans vary only from one to sixteen quarters, while the time units include five-year, and sometimes even ten-year, nonoverlapping averages. Furthermore, the number of observations available differs from shorter to longer time periods. The difference is most radical for the upper panel, where 444 observations are available for the monthly correlations, but only three for the decennial ones.

Table 3 deals with variations in correlation and regression coefficients as different cycle stages and phases are used as input for analysis. Some of the differences are startling. (See particularly the differences between the coefficients of employment versus the average workweek at peaks and troughs of business cycles.) They may have been caused partly by extreme values at the 1929 peak and the 1933 trough of the Great Depression, particularly since the number of included cycles, and therefore of observations, is small (six or seven). These large variations, as well as those in the correlation coefficients for expansion and contraction amplitudes, deserve more analysis than can be afforded here. For the purposes of this paper, the decisive finding of these experimentations is that measures of relationships can vary considerably with alternative time periods and time measures. This has consequences for the design of economic research and analysis.

I do not suggest that research workers should go through all the versions of time partitioning for any combination of reasonable variables. But I suggest that they find out whether systematic similarities or differences exist during subperiods relevant to their research objective. I am sure that more attention to the articulation of the time dimension will generate valuable insights.

IV. Programmed Articulation

Experimentation with alternative articulations of the time dimension involves additional computational work. However, in the age of electronic computers, the additional costs are usually not prohibitive. Under certain conditions, it may be advantageous to make the choice of alternative time periods and time units a part of the computer program.

It might be said that the proposed disaggregation of the time dimension leads to increased subjectivity of research results. In addition to the discretion of chosing variables, functions, and over-all time coverage, there is now the discretion of choosing time segments. No doubt, any proliferation of options increases the opportunities for selectivity. However, this bane could be converted into a boon if it became an accepted rule to report the effects of relevant alternative selections of time periods, time units, time spans, etc., on research results. Programmed analysis makes this feasible.

In the regression analysis of profits, one articulation involved the determination of cyclical turning points in the profit variable (see Chart 2). It is desirable to facilitate this process and remove turning-point-determination, as far as possible, from the predilections of the research worker. Criteria for the determination of specific turning points have been available for a long time,[24] but the process of implementation left much to personal knowledge, experience, and judgment. Attempts to specify and program a process of turning-point-determination that meets the specified criteria are now under way. Chart 3 shows the turning points picked by a program under development at the National Bureau. There are still obstacles to be overcome and refinements to be made, but the goal is in sight. In the context of the present paper, the significance of such a program is that it would permit the partitioning of the time dimension in accordance with the cyclical behavior of any component variable of a system.

Notes

1. Analyses of periodicities, such as spectral analysis, do not fall into this category.

2. Measures for individual cycle fractions and summary measures for corresponding cycle fractions are part of the National Bureau's Standard Business Cycle Analysis. The Recession-Recovery Analysis is mainly designed to compare rather than summarize cyclically corresponding experiences, but the latest programmed version of Recession-Recovery Analysis also gives averages over several cycles.

3. Alfred H. Conrad and John R. Meyer, "The Economics of Slavery in the Ante-Bellum South," Journal of Political Economy, April 1958, pp. 95-130.

4. Rendigs Fels, "The U.S. Downturn of 1948," American Economic Review, December 1965, p. 1061.

5. Ibid., pp. 1072-73.

6. Business Economics, Winter 1965-66, p. 15.

7. Several of the "fixed-variable, fixed-parameter" models continue to be tolerably effective forecasting instruments, in spite of the change of environment. This raises highly interesting questions; however, they are beyond the scope of the present concern.

8. Franco Modigliani, "Fluctuations in the Saving-Investment Ratio: A Problem in Economic Forecasting," Studies in Income and Wealth," New York, National Bureau of Economic Research, 1949, pp. 379-382.

9. Paul Boschan, "Productive Capacity, Industrial Production, and Steel Requirements," Long-Range Economic Projection, Princeton University Press for NBER, 1954, pp. 233 ff.

10. Milton Friedman, A Theory of the Consumption Function, Princeton for NBER, 1957, pp. 125 ff.

11. Ibid., p. 152.

12. J. Duesenberry, O. Eckstein, and G. Fromm, "Simulation of the United States Economy in Recession," Econometrica, October 1960, pp. 749-809.

13. Ibid., pp. 752-753.

14. See, for example, Edwin Kuh, "Cyclical and Secular Labor Productivity in United States Manufacturing," Review of Economics and Statistics, February 1965; or Peter Eilbott, "The Effectiveness of Automatic Stabilizers," American Economic Review, June 1966 (also his bibliography reporting previous analyses that distinguish behavior during expansions and recessions).

15. John R. Meyer and Robert R. Glauber, Investment Decisions, Economic Forecasting, and Public Policy.

16. The correlation coefficient of profits against time is .88, that of product against time .99!

17. The little squares on Charts 1 and 2 designate the quarters at which the period is partitioned.

18. Computing regressions separately for "all expansions" or "all contractions" will not help, since the line of average relationship continues to reflect both cyclical and long-term changes.

19. The positive correlation during 1959 has, perhaps, elements of a fluke. The correlation is very low and is based on four quarters only. The terminal quarter is lower than the initial, but the intermediate quarters give the least-squares line a positive tilt.

20. The problem may be seen in the top panel of Table 2. For the ten-year period there are as many variables as observations, but fewer degrees of freedom. Thus, multiple regressions could be provided only up to five-year periods.

21. See Julius Shiskin, "Long-Term Economic Growth, A Statistical Compendium," published in these Proceedings.

22. There are, of course, exceptions. For example, Arthur F. Burns and Wesley C. Mitchell, in Measuring Business Cycles, New York, NBER, 1946, deal in Chapter 6 with the effect of the time unit on cyclical measures, and in Chapter 8 with the effects of smoothing. Recently, Milton Friedman and Anna J. Schwartz measured differential rates of monetary growth over increasing time units and time spans. (See NBER, Forty-Sixth Annual Report, New York, 1966, pp. 47-48.) A forthcoming National Bureau study, "Variable Span Diffusion Indexes," by Geoffrey H. Moore and Julius Shiskin, explores the effects of varying the time span.

23. The leads of cyclical turning points in the workweek over those in employment require some qualification of this statement. See Gerhard Bry, The Average Workweek as an Economic Indicator, New York, NBER, 1959, p. 15.

24. See Burns and Mitchell, Measuring Business Cycles, Chapter 4.

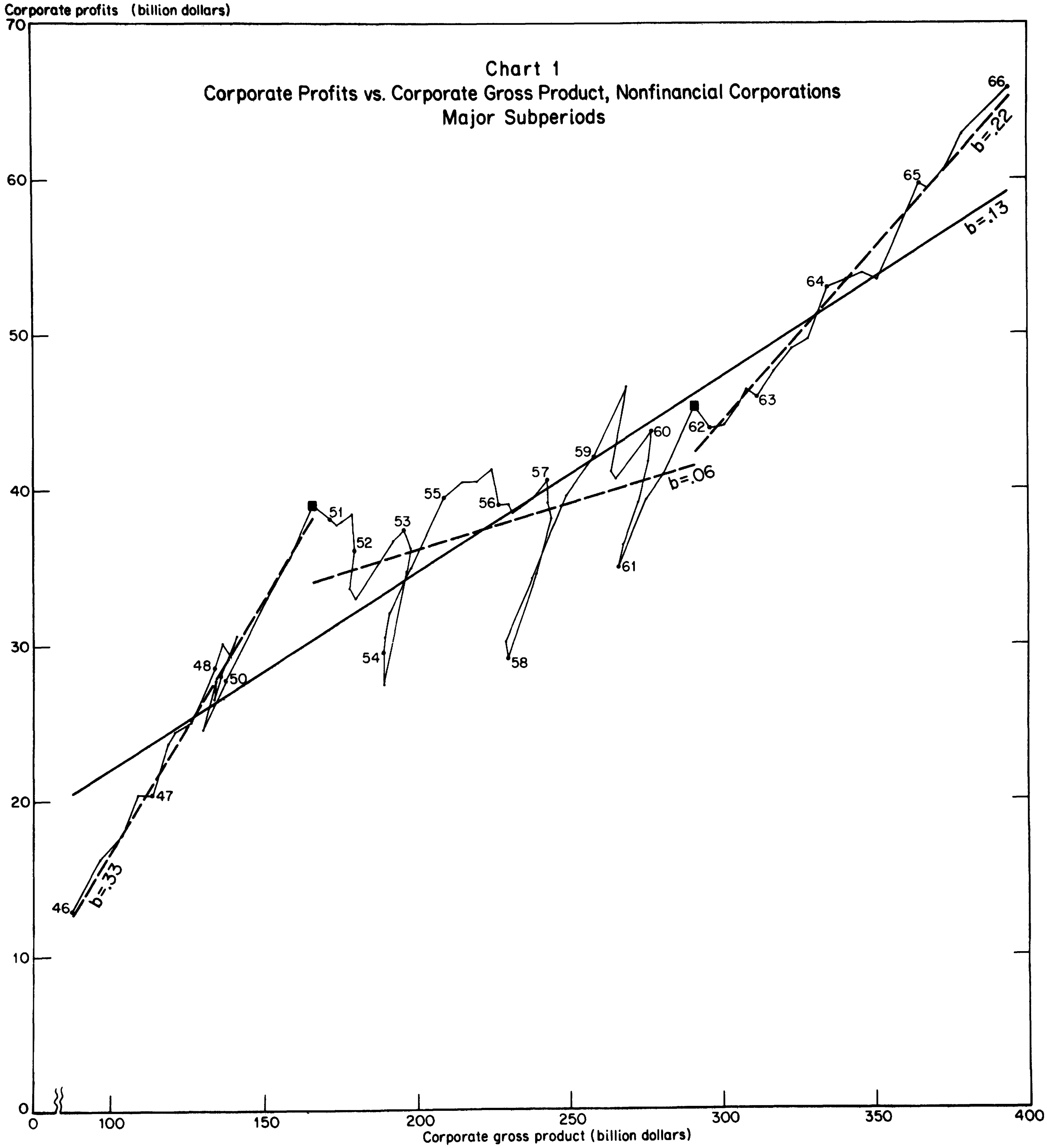

Corporate profits (billion dollars)
Chart 1
Corporate Profits vs. Corporate Gross Product, Nonfinancial Corporations
Major Subperiods
b=.33
b=.13
b=.06
b=.22
46
47
48
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
0
10
20
30
40
50
60
70
0
100
150
200
250
300
350
400
Corporate gross product (billion dollars)

Chart 2
Corporate Profits vs. Corporate Gross Product, Nonfinancial Corporations
Cyclically Analogous Segments

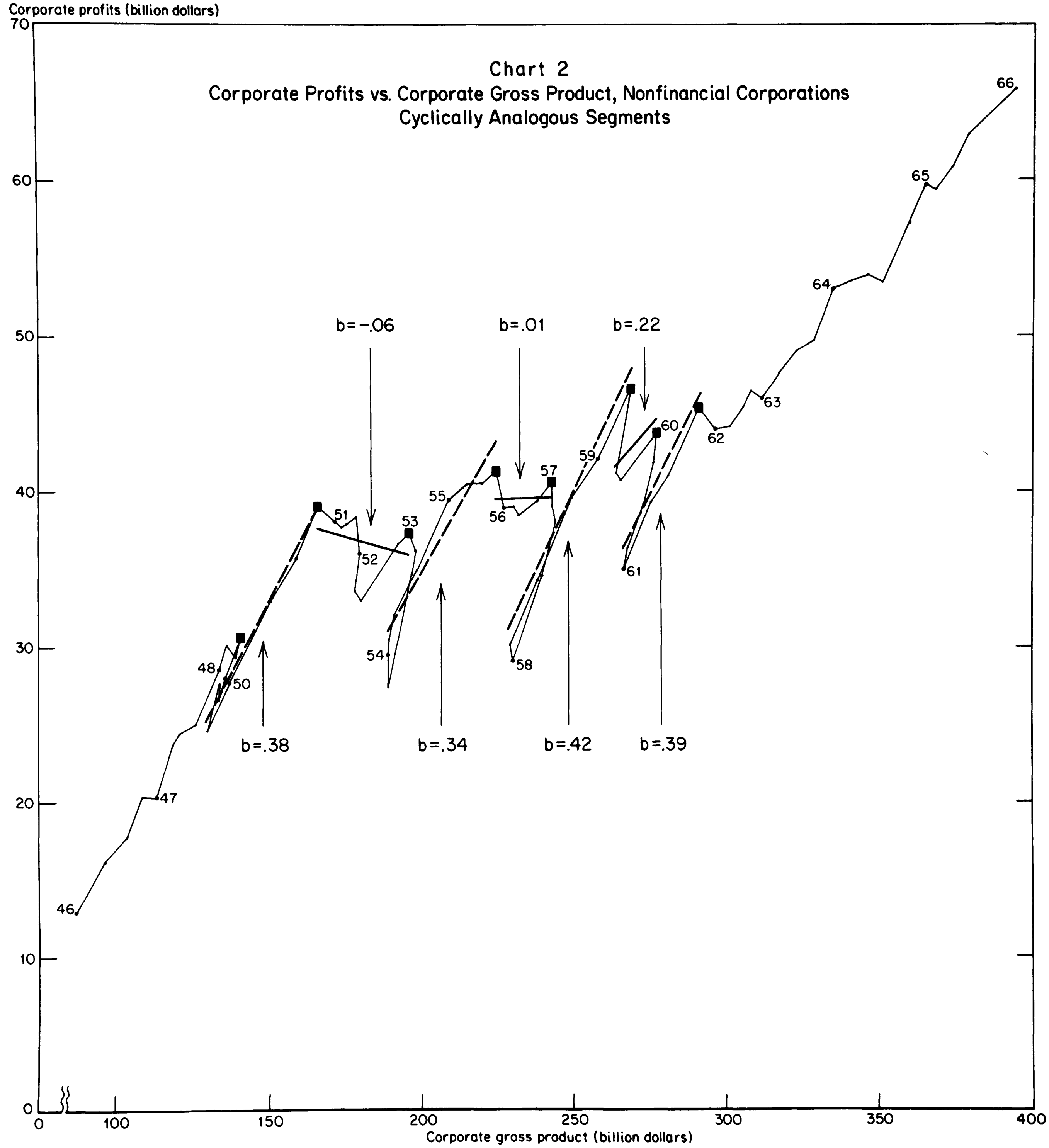
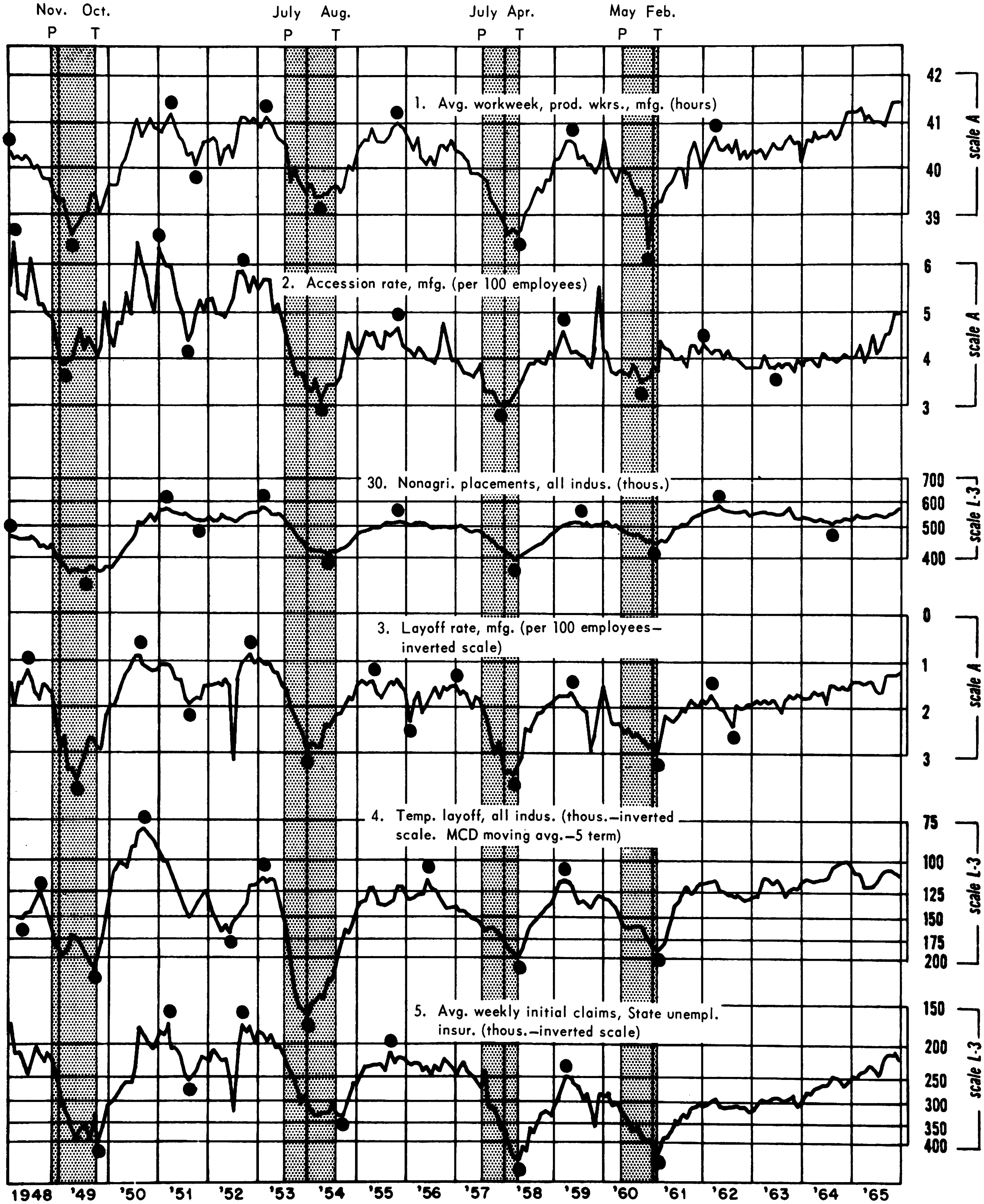


Chart 3
Programmed Turning-Point Determination



Source: chart, Business Cycle Developments, June 1966, P 10; turning points, program under development at NBER.

TABLE 1

Corporate Profits vs. Corporate Product, Unemployment Rate, and Time
Correlation Analysis, Major Periods, 1946/1 to 1966/1

	Independent Variables						
Years and Quarters	Time (1)	Product (2)	Product Time (3)	Unemplt. (4)	Unemplt. Time (5)	Product Unemplt. (6)	Product Unemplt. Time (7)
Full Period 1946/1 to 1966/1							
Correlation Coefficients	.8779	.9296	.9697	.0855	.9277	.9520	.9698
Adjusted R-Squares		.8624	.9388	-.0053	.8570	.9039	.9381
Y-Intercepts		9.4320	-12.4685	34.1309	33.5244	16.7912	-14.3688
Regression Coefficients							
Product		.1262	.3740			.1350	.3867
Unemployment				.7680	-2.9533	-1.9353	.1845
Time			-.8268		.4532		-.8720
Subperiod 1946/1 to 1950/4							
Correlation Coefficients	.8733	.9915	.9962	.0804		.9957	.9963
Adjusted R-Squares		.9821	.9916			.9903	.9911
Y-Intercepts		-15.6713	-22.3874			-13.6859	-23.9178
Regression Coefficients							
Product		.3253	.3997			.3305	.4114
Unemployment						-.5676	.1088
Time			-.2676				-.3135
Subperiod 1950/4 to 1961/4							
Correlation Coefficients	.3799	.5339	.8821	-.0021	.6072	.7080	.9512
Adjusted R-Squares		.2685	.7675	-.0232	.3386	.4775	.8779
Y-Intercepts		23.7450	-17.0148	37.5793	35.6776	22.4543	-55.5388
Regression Coefficients							
Product		.0615	.4507			.1079	.7610
Unemployment				-.0064	-2.3481	-1.9152	3.2400
Time			-1.1088		.3106		-2.2161
Subperiod 1961/4 to 1966/1							
Correlation Coefficients	.9679	.9881	.9930	-.7818		.9898	.9957
Adjusted R-Squares		.9749	.9842			.9770	.9895
Y-Intercepts		-21.6296	-5.8700			-33.6868	-19.5515
Regression Coefficients							
Product		.2203	.3925			.2390	.4359
Unemployment						1.1040	1.4150
Time			-1.0137				-1.1276

TABLE 2

Average Workweek vs. Employment, Unemployment Rate, and Time, 1929-65

Variation of Correlation and Regression Coefficients
Over Different Time Units, Smoothing Terms, and Time Spans

	Independent Variables						
	Time (1)	Emplt. (2)	Emplt. Time (3)	Unemplt. (4)	Unemplt. Time (5)	Emplt. Unemplt. (6)	Emplt. Unemplt. Time (7)
<u>Time Units</u>[a]							
Correlation Coefficients	.13 to .55	.55 to .83	.70 to .70[a] (.72)	-.78 to -.80 (-.71)	.87 to .87[b] (.91)	.82 to .78[b] (.86)	.87 to .87[b] (.90)
Regression Coefficients							
Employment		.578 to .387 (.333)	1.091 to .695[a] (1.096)			-.550 to -.598[b] (-.628)	-.053 to -.228[b]
Unemployment				-.322 to -.195	-.436 to -.368[b] (-.439)	-.512 to -.462[b] (-.537)	-.450 to -.450[b] (-.475)
Time	.003 to .189[a]		-.015 to -.541[a]		-.011 to -.585[b]		-.010 to -.517[b]
<u>Smoothing Terms</u> (Moving Averages)[a]							
Correlation Coefficients	.13 to .42	.56 to .82	.71 to .90	-.79 to -.85	.89 to .92	.84 to .85 (.86)	.89 to .92
Regression Coefficients							
Employment		.581 to .642	1.096 to .995			-.560 to -.118 (-.561)	-.061 to .249 (-.063)
Unemployment				-.323 to -.267	-.439 to -.398	-.517 to -.314	-.455 to -.307
Time	.009 to .022		-.044 to -.031		-.033 to -.029		-.031 to .030
<u>Time Spans</u> (Changes)[a]							
Correlation Coefficients	-.11 to -.13 (-.21)	.56 to .83	.57 to .90	-.46 to -.92	.46 to .93	.58 to .93	.58 to .95
Regression Coefficients							
Employment		1.487 to 1.559 (1.352)	1.491 to 1.719 (1.357)			1.202 to .400 (.398)	1.239 to .725 (.585)
Unemployment				-.355 to -.486 (-.491)	-.351 to -.484 (-.486)	-.127 to -.393 (-.401)	-.112 to -.314 (-.337)
Time	-.002 to -.014 (-.018)		-.003 to -.038		-.001 to -.010		-.002 to -.021

Note: The two top figures in each cell denote the extremes of a monotonic change, from first to last of the indicated alternatives. When there are reversals in direction of change, the value showing the largest deviation from the initial coefficient is given in parentheses.

[a]The time units considered are monthly, quarterly, yearly, five-yearly, and ten-yearly periods. The smoothing and the time spans considered are one, two, three, four, six, eight, twelve, and sixteen quarters.

[b]Up to five-year unit only.

TABLE 3

Average Workweek vs. Employment, Unemployment Rate, and Time, 1929-65

Variation of Correlation and Regression Coefficients
Over Different Cycle Stages, Cycle Phases, and Cycle Amplitudes

	Independent Variables						
	Time (1)	Emplt. (2)	Emplt. Time (3)	Unemplt. (4)	Unemplt. Time (5)	Emplt. Unemplt. (6)	Emplt. Unemplt. Time (7)
Expansion Stages							
(I,II,III,IV,V)							
Correlation Coefficients	.67 to -.68 (.74)	.92 to -.32 (.98)	.98 to .70 (.99)	-.98 to -.59 (-.99)	.999 to .91 (.81)	.99 to .86 (.85)	.999 to .93 (.98)
Regression Coefficients							
Employment		.969 to -.470 (.780)	1.635 to .339 (.902)			-.638 to -1.003 (4.207)	-.002 to -.419 (3.602)
Unemployment				-.314 to -.498 (-.226)	-.382 to -.515 (-.260)	-.502 to -.741 (1.760)	-.382 to -.612 (1.364)
Time	1.256 to -1.183		-.133 to -.154 (-.0312)		-.051 to -.121 (-.013)		-.050 to -.088 (-.026)
Contraction Stages							
(V, VI, VII, VIII, IX)							
Correlation Coefficients	-.68 to .64	-.32 to .92 (.93)	.70 to .96 (.62)	-.59 to -.98 (-.68)	.91 to .995 (.85)	.86 to .995 (.81)	.93 to .997 (.89)
Regression Coefficients							
Employment		-.470 to .951	.339 to 1.366			-1.003 to -.643 (-2.005)	-.419 to -.375 (-1.524)
Unemployment				-.498 to -.311 (-.295)	-.515 to -.362 (-.347)	-.741 to -.503 (-1.287)	-.612 to -.451 (-1.152)
Time	-.118 to .095		-.145 to -.072 (-.060)		-.121 to -.032 (-.032)		-.088 to -.018 (-.016)

(continued)

TABLE 3 (concluded)

	Independent Variables						
	Time (1)	Emplt. (2)	Emplt. Time (3)	Unemplt. (4)	Unemplt. Time (5)	Emplt. Unemplt. (6)	Emplt. Unemplt. Time (7)
Cycle Phases (Expansions, Contractions)							
Correlation Coefficients	.44 to .04	.88 to .50	.93 to .71	-.82 to -.71	.86 to .88	.92 to 92	.97 to .94
Regression Coefficients							
Employment		.588 to .399	.823 to .864			1.752 to -1.536	1.966 to -1.084
Unemployment				-.205 to -.276	-.270 to -.449	.437 to .985	.431 to -.867
Time	.355 to .045		-.380 to -.768		-.289 to -.695		.376 to -.363
Cycle Amplitudes (Expansions, Contractions)							
Correlation Coefficients	-.63 to .79	.90 to .81	.90 to .84	-.99 to -.97	.996 to .98	.993 to .994	.999 to .994
Regression Coefficients							
Employment		1.614 to 4.194	1.673 to 2.588			-.491 to 1.452	-.375 to 1.639
Unemployment				-.515 to -.565	-.582 to -.501	-.648 to -.454	-.676 to -.468
Time	-1.499 to 1.921		.108 to .940		.423 to .359		.372 to -.158

Note: The two top figures in each cell denote the extremes of a monotonic change, from first to last of the indicated alternatives. When there are reversals in direction of change, the value showing the largest deviation from the initial coefficient is given in parentheses.

DISCUSSION

Kenneth B. Williams, Federal Reserve Board

Back in the fall of 1961, when Shiskin was preparing to begin regular publication of his Business Cycle Developments I was asked, among others, to give him my recommendations. I wrote him a long, carefully reasoned and well-documented letter recommending that BCD not be published. He, of course, went ahead and published it anyway and it has proved not only to be popular -- which I knew it would -- but also a great contribution to current economic understanding. Each issue brings together each month up-to-the-minute tables and charts on many of the most important, generally-watched current series. Moreover, these are organized in a highly useful format, whether or not one is addicted to the NBER business cycle.

So, apparently knowing my dyspepic nature, Pete Morton arranged to have me comment on this new work of Shiskin's -- hoping, no doubt, that if I were strongly against it, its publication also would be a resounding success.

However, I learned my lesson in 1961 and so this time, I am not going to recommend against publication of Long-term Economic Trends. Indeed, I am highly in favor of its publication,if for no other reason than to provide background for the St. Louis Federal Reserve Bank monthly releases of growth triangles. Consequently, the success of this venture will have to depend entirely on its own merits, of which I shall point out only a few.

Shiskin's paper is a preview and explanation of his soon to be published volume. That volume is like a dictionary -- it is full of good words, short words, long words, easy words, hard words, etc., but, except for saying it should have more words or less words, there is not a lot to say about it, assuming the words are all spelled right and defined correctly. The volume does not claim to, nor does it, contain any great new theories of growth, nor does it discuss at length the causes of growth. Rather it focuses on being a useful compendium or catalog of economic and demographic facts in table and chart form, with growth rates calculated on several bases, plus a do-it-yourself kit with which to calculate your own growth rates.

Even better than most catalogs, however, the volume warns you not to expect too much from the merchandise. Shiskin has developed an ingenious table which by coloring and shading warns you visually not to compare growth from depression years to prosperity years -- as nearly all political candidates like to do when they want to point with pride -- or, conversely, to calculate from peak to cycle trough when they wish to view with alarm. This table serves as a warning against that easy kind of misinformation.

More importantly, use of similar unemployment rates as the selection criterion is far from a really adequate guide to comparable periods of economic growth. Basically, the chart encourages comparing years in which unemployment rates are about the same. Unemployment, however, is an inadequate proxy even for the available labor supply alone, as Arthur Ross pointed out so well in his speech Tuesday night when he differentiated active job seeking from full labor resource mobilization and from full realization of human potential The current unemployment figure ignores the changing pool of eligible workers outside the work force. Similar unemployment rates can occur in a variety of economic environments and for quite different reasons, and can have quite different implications for over-all economic growth.

There is probably more "hidden" potential labor supply currently available than in other past periods of comparable over-all unemployment rates. This is due not only to the structure of the population increase, but also to the changing propensities of the population to work. There are more teenagers available to fill labor gaps. Technology in the home is releasing the housewife from the kitchen. She is less burdened with children and more in need of income to meet payments on consumer credit and mortgage debt. With more white collar and clerical jobs available, she is choosing labor force participation more often than before. Simply fitting a trend line through real GNP in past years of comparable unemployment rates might well tend to underestimate the potential future economic growth rate of the economy.

When Shiskin started publishing his BCD in 1961 we were not far above the bottom of the last recession. Hence, there was widespread and continuing interest in business cycles, which in this country -- though not abroad -- had been coming along more or less on schedule ever since the War. Their timing and their cure were matters of great public and political, as well as professional, interest. Ever since then we have not had a business recession So, the BCD publication might be thought to be less and less useful as we become more and more successful in dodging recessions. Meanwhile, highly-placed economic authorities have said we can have continued expansion without recession if we just apply appropriate fiscal and monetary policies -- hopefully, however, in somewhat more balanced proportions than recently.

Thus, interest in cycles -- until just recently -- has lagged and interest in growth has greatly increased for all kinds of policy purposes. No longer can we speak only about "record" levels or absolute changes in levels. As economic statisticians we must now always talk about changes in relation to rates of growth and conduct our debates mostly within the confines of 3 to 5 per cent per year for real GNP -- at least in this country. Facetiously, I hope that publication of this volume will not have the same effect on continued economic growth as the publication of BCD did on cycles and that, consequently, growth will stop or be much slower from here on.

It may be that interest in growth and comparisons of growth rates themselves are subject to cyclical fluctuations. Certainly, interest in trends and adjustment of time series for trend were keenest in the 1920's and dropped nearly to zero during the 1930's when growth was negative or so slow as to warrant the application of the term "stagnation" to the economy. Then leading economic theorists began to point to the loss of the Western Frontiers, the absence of new large capital-using innovations such as electric utilities and autos, the failure of new capital investment to increase at all for several years, and the curtailed rate of population growth as causes of the stagnation and to recommend newly-formulated fiscal policy measures as the cure for these conditions.

Since the 1930's, of course, much has happened. Fiscal policy as then outlined has become an accepted -- and in the view of many is the major instrument to be used in moderating economic fluctuations and stimulating the economy to grow at a rate more nearly consistent with full employment growth. Monetary policy too, has been adapting to the new world in which it operates and the Government generally has accepted far greater responsibility than before for employment and levels of living for all of its citizens.

I do not mean to suggest that all of our goals have been reached or that applications of the instruments of economic policy have been, or now are, perfectly timed or administered in just the right dosage. But great progress has been made and it is highly appropriate that the Census Bureau should now be racking up the score on our long-term growth experience and providing the framework and many of the facts needed by scholars and policy makers to formulate the policies and programs for further progress and, hopefully, for rapid and sustained economic growth. I suppose, if nothing else, the volume will be available and widely assimilated in time to provide ample ammunition for both parties in the next Presidential election.

Now I should like to make a few more specific comments:

I find the use of growth rates for projections of future developments some 5, 10, or 15 years ahead a highly dangerous business if not done with care. To be sure, economic policy must live in the future and measurement must come out of the past. But the future is never like the past. Hence, any mechanical or automatic projection of past trends into the future can have unfortunate results. As Shiskin points out in the section of his paper that he did not read, "analytical" projections, which allow explicitly for factors expected to affect future developments, have many advantages over "naive" projections which just carry forward past trends.

We hope in fact that the future will not be like the past -- including the wars, the deep depressions and the inflations. If we use our old, as well as our new, knowledge expertly we should be able to achieve and generally sustain much higher rates of over-all growth, although maybe not quite so high as those in the past 3 or 4 years when the unemployment rate was being significantly reduced.

Second, the selection of series is always a debatable matter but I have a bias in favor of physical measures such as employment and against dollar measures which embody varying prices. Prices, for example, can and do go down as well as up but over time may not and, on average, should not go up along with employment and physical output. But in so many value series the price component is hidden or difficult to really isolate statistically, as we learned in the Firestone session on prices and costs in research and development. As a result series relating to a variety of fields are often used without adequate regard for the price components. This is frequently the case with respect to monetary, credit, and debt figures where increases and decreases in values do not have the same meaning for living standards at home, for military power or for world prestige as those relating to physical measures.

Third, in spite of my bias in favor of physical measures, I think that the Shiskin volume needs to include many more financial series in it and also more price series. While they may require more care in their use than employment or tonnage figures, they do have great relevance for economic policy. I hesitate to express this view, however, because Shiskin is likely to ask me to provide the series.

All in all, as I think I have indicated, the Shiskin growth volume, and Shiskin's paper today advertising it, are both worthy of very serious attention by professional economists and statisticians and all those concerned with public and private policies. The volume does not answer all the questions you will have but, knowing Shiskin's willingness to accept suggestions and criticisms, I am sure we can look forward to steady improvement as each succeeding volume is issued. Whether publication should be annual, or less frequent or more frequent, I do not know. But judging by the frequency and extensiveness of the revisions of many of the basic figures, I would be inclined to a short period rather than a long one.

I should also like to congratulate Shiskin and the Census for giving attention to presenting data in an analytical form conducive to use by statistical consumers. Too often producers of data satisfy their professional consciences by reference to the purely statistical pedigrees of their series and surveys without adequate regard for how they will be used and misused and the implications they may have for policy. I am convinced there is great merit in having producers and users of economic statistics be embodied as much as possible in the same persons -- or lacking that in organizational structures which require both consumers and producers to work closely together in the collection process and in the final analyses. These publications of Shiskin's, while not ideal in this respect, are a major step in focusing the attention of the Census data producers on some of the major uses and interpretations of their data.

XIII

STATISTICS FOR THE POVERTY PROGRAM

Chairman, HERMAN P. MILLER, U. S. Bureau of the Census

Page

EVALUATION OF OFFICE OF ECONOMIC OPPORTUNITY PROGRAMS -- A PROGRESS REPORT

Robert A. Levine, Office of Economic Opportunity

The obvious first question to be asked concerning the evaluation of any operating program is what is it that is being evaluated--what are the program objectives which we are supposed to measure? On one level the answer to this question is obvious. Any program must be evaluated for the achievement of its immediate objectives; a health program must measure the effects on the health of those who participate in it, an education program the educational achievement of those who pass through it and so forth. This type of evaluation is necessary but it is relatively routine, and it is not the sort to which I will be primarily addressing myself in this paper.

Rather, in reporting on the progress of Office of Economic Opportunity programs I will be reporting on the evaluation of their effects in helping to eliminate poverty. Even this is not sufficient however; the objectives of the programs I will be discussing are somewhat more complex than that. The primary objective of the War on Poverty is to change people from "poor" to "non-poor" through their own efforts--to enable them to support themselves above the poverty line through their own earnings. This is the meaning of an "opportunity program." Only secondarily should the objective be to change the state of the poor by giving them money, thus making them less poor or not poor at all. Let me make clear that the overall War on Poverty must do both. There are too many people who are incapable of being reached by opportunity—the aged and the mothers of large families who should be taking care of their families instead of working, for example. If we mean it when we say that we want to eliminate poverty in this country, we must do something for them through direct income maintenance as well as providing for those who can take advantage of opportunity programs. OEO is only a part of the War on Poverty--a part measured by the fact that in fiscal 1967 the President's Budget calls for $24 billion of anti-poverty expenditures of which $1.75 billion are funded through OEO. This $1.75 billion is properly devoted entirely to opportunity programs and this paper will concern itself only with the opportunity objective of increasing the capability for self-support above the poverty line. It will be concerned primarily, although not exclusively, with OEO programs.

This is a progress report and it is therefore interim.

It is a report on what we are doing to evaluate rather than the results of evaluations. The OEO and its programs are not yet two years old. Programs like Job Corps are just now turning out graduates in substantial numbers. Other programs like Community Action could not conceivably work to make measurable changes in poverty communities in these first several years of buildup. This is particularly true if, rather than the effects of an anti-poverty program being proportional to expenditures, below which effects are not visible at all. Expenditures on programs like Community Action are still quite low relative to total needs and we might in some cases consider the current period one of organization in which we should expect nothing measurable. Thus, this is a report on the evaluations we are setting up rather than what our evaluations have told us about our current programs although a bit of the latter will be brought in.

The programs of OEO and associated War on Poverty programs can be divided into three categories according to the evaluation problems they present. In order of increasing difficulty, these are: vocational training and other programs directly connected to jobs and earnings; educational programs directly connected to jobs and earnings but only over the long-run; and programs of the comprehensive Community Action variety which we are convinced are vital to opportunity objectives, but work through a complex system to ultimately affect opportunity.

TRAINING

Of the three, training is conceptually the most simple to evaluate. The objective of an opportunity program is to increase earnings, and the connection of vocational training to earnings is obvious and straightforward. To evaluate a training program our primary data needs are for characteristics of the trainee before entering the program and after passing through it. Such characteristics include both items of the demographic type such as age, race, and education as control variables; and also jobholding experience and earnings as variables of direct interest.

It is, of course, not proper to evaluate the effectiveness of a training program by comparing earnings after to earnings before, because change may be an effect of lapsed time or of other factors outside the program, as well as a direct effect of the program. Just as obviously, even though such before-after comparison is not really legitimate, we do it because we frequently have nothing else to fall back on. But the really important and legitimate form of evaluation compares the before-and-after experience of trainees to the similarly time-phased experience of non-trainees in a carefully selected control group. This makes things a lot less simple, particularly when we talk about youth training. Youth control groups are particularly difficult to come by, both because so many young people between the ages of 16 and 21 have passed through OEO programs, and because the self-selection process for our youth training programs means that the possible members of a control group may differ from the trainees in certain specific (but undefined) characteristics which led them to avoid the training program. Because of these difficulties we are proceeding rather carefully in setting up the youth control group which is vital to our program evaluation. We are, however, already collecting substantial data on the pre-training characteristics and some data on post-training characteristics of our program participants and are in the process of setting up a youth control sample.

A system has now been established for the collection of individual characteristic data for all enrollees in both Job Corps and Neighborhood Youth Corps. These data include information not only on the enrollees' own education and previous work experience, but also on family characteristics such as family size, number of working members, occupation and work status of family head, estimated family income plus a number of binary items on such things as housing and welfare. The Job Corps also regularly reports some initial aptitude and achievement level tests and the NYC is in the process of experimenting with such instruments.

There have been a few very small sample followups (about 200) of Job Corps graduates and a slightly larger sample followup is to be reported late this summer. There are presently no routine followups of NYC graduates, but a followup procedure is in the developmental stage.

A pilot study is now underway to determine the feasibility of setting up a large national sample of poor youth, 16-21, which could be tested and then reinterviewed and retested two or three years later. The objective is to create a sample which is large enough to allow the establishment of groups of youth which are statistically similar in terms of relevant socio-economic variables, some of whom have participated in training programs, Job Corps, NYC and MDTA, and some of whom have not. Hopefully, such a sample would allow both cross-section and longitudinal comparisons which would provide some of the sort of evaluative information, admittedly crude, which is necessary for realistic judgements about reasonable program mix. Elements of this sample would serve as a control group, so that program effects could be separated to some degree from effects due simply to growth of the youths and to changes in general economic and social conditions. In addition, it may be possible to determine which types of programs are best suited for which types of poor youth and, within programs, which program characteristics seem to be related with favorable results. We are fully aware of the limitations of such an approach to the problems of evaluation -- and this is one of the reasons for proceeding first on a pilot basis -- but it seems to us that the logic of the program evaluation effort leads inevitably to this type of an approach. If the limitations of this type of evaluative method prove too severe, I suppose we must re-examine the logic of program evaluation as it now stands.

Even lacking the control sample at this time, we already have some capability to do the less legitimate sort of before-and-after evaluation, and we have made some still quite hypothetical cost-effectiveness comparisons of our youth programs. It should be pointed out in discussing the cost and effectiveness of training programs, however, that we try to avoid the sort of cost-benefit analysis which states that a training program is justified if discounted future earnings are greater than costs. The program implication of such a cost-benefit calculation is that if discounted earnings are less than costs, transfer payments would be more economical than the training program, but given the hierarchy of objectives mentioned above -- that it is more important to get people out of poverty through their own earnings than it is to give them money -- War on Poverty training may be justified even if costs are higher than estimated benefits. For this reason, we avoid the sort of benefit calculation which compares something to nothing, and rather use such calculations to compare programs with some similarity.

Table 1 shows such a calculation -- highly hypothetical, it should be re-emphasized -- comparing Job Corps and the out-of-school Neighborhood Youth Corps. What was done was to estimate (on the basis of very sketchy data on before-and-after job-holding and earnings characteristics) the time period which it would take for Job Corpsmen to pay back to the economy with their increased earnings the cost of their Job Corps training. This came to about 5 years on the basis of these data plus some rather conservative assumptions, and the question was then asked what assumptions for the substantially cheaper Neighborhood Youth Corps program would equalize the pay-back period. We came to the obvious conclusion that it takes less change in earnings to pay back the NYC costs simply because these costs are lower than the Job Corps costs.

We also, however, came to the less obvious conclusion that for the higher costs of Job Corps to really justify themselves unambiguously, the Job Corps program should be capable of reaching and helping a group of hard-core youth that Neighborhood Youth Corps cannot reach at all. If a group of youth does exist which cannot be helped by NYC then the pay-back period on NYC costs for them would be infinite; if Job Corps can help these youth with its more intensive programs, then its pay-back period would at any rate be less than infinite. It is this group which Job Corps is trying to reach.

TABLE I

		JOB CORPS*	Out-of-School NEIGHBORHOOD YOUTH CORPS
I.	Costs:		
	a. Steady State Costs Per Graduate	$6,980	$1,000**
	b. Assumed Success*** Rate Per Graduate	80%	50%
	c. Cost Per Success (Cost per graduate ÷ success rate)	$8,725	$2,200
II.	Computation Equalizing Payback Period for These Costs:		
	d. Equalized Payback Period****	5.1 years	5.1 years
	e. Annual Earnings Gain = Costs/Payback (Rounded)	$1,700	$ 430
	f. Estimated Hourly Wage Gain Due to Program	$ .60 = $ 1.60-$1.00	$ .25 = $ 1.25-$1.00
	g. Assumed Annual Hours Worked Previous to Program	1,500 hours	1,500 hours
	h. Annual Hours Worked After Program (Computed to equalize payback period)	2,000 hours	1,544 hours

*Job Corps data estimated from preliminary sample.

**Because NYC is a Work Experience Program rather than a training program like JC, it is assumed that all enrollees are graduates having received some benefit, therefore the costs per graduate are the same as the costs per enrollee.

***Success is defined as holding a good steady job. A good job is defined as semi-skilled or better.

****Neighborhood Youth Corps payback period set equal to calculated Job Corps period.

This example clearly illustrates the importance of making certain alterations in the conventional form of cost-benefit analysis when it is applied to poverty-type programs.

It is our feeling that in the case of poverty programs the attempt to quantify benefits beyond those directly represented by changes in the lifetime income stream -- sometimes called external or spillover benefits -- is quite important. Most importantly, when looked at from the governmental decision point (where decisions are subject to the governmental budget constraint) the effects of the programs on other social expenditures need to be taken into account. If youth training programs not only raise incomes but also result in lower requirements for public expenditures for such things as welfare, unemployment compensation, health, crime prevention, then an attempt should be made to take these additional "benefits" into account in making program evaluations. We are currently attempting to develop some means of quantifying, at least in order of magnitude, these additional benefits.

EDUCATION

The evaluation of education programs is less straightforward. Remarkably little has been done along the lines of systematic comparative evaluation of different educational techniques for reaching the underprivileged. The Office of Education which disposes of a billion dollars or more a year for the education of the poor under Title I of the Elementary and Secondary Act of 1965 is beginning a comprehensive evaluation program, although this will not be simple because the Office of Education statutorily provides monies without being able to control them. Up until this effort, virtually nothing has been done on a systematic basis.

Tables 2 and 3 bring together an inventory of Federal programs which are intended to contribute to the compensatory education effort or which are or can be used to contribute to the compensatory education effort. Table 3 inventories direct program or financial assistance programs that provide support to such programs, i.e., train teachers, provide facilities, fund research and development of new methods, materials, curricula, equipment, etc. In addition to identifying Federal programs, and administering agencies, Tables 2 and 3 estimate the number of beneficiaries and the amount of funding through 1967.

The two left hand columns briefly summarize the kind of evaluation information available about each program. It is important to note that while there are several inventories of Federal educational programs for the poor, this Table represents the first attempt to capsule what is known about the impact of such programs on the amount or rate of learning whether this be measured in cognative, behavioral or attitudinal terms.

Obviously, the findings reflected by this Table do not pretend to be definitive. It is important, however, that it represents the first time evaluative findings to date have been set down, cheek by jowel, with measures of the numbers to be served and dollars to be spent in the Federal effort. We think this Table drives home the need for a major effort to evaluate the essentially experimental programs being funded. We do not think we should stop experimenting or providing programs, but that we should take immediate steps to provide that every program and every experiment should include an evaluation design from the start. Only then will we increase our capacity to know what works, or what works best in compensatory education. As we now stand, the state of evaluation of educational programs is such that we cannot even be sure that when favorable program results are obtained, they are the result of good program design or merely a Hawthorne effect. Given the sums of money now being spent by Federal and other agencies for compensatory education of the poor, this is slightly shocking.

Table 2

SUMMARY OF FEDERALLY FUNDED COMPENSATORY EDUCATIONAL SERVICES AND FINANCIAL ASSISTANCE TO THE POOR

KIND OF PROGRAM	FEDERAL AGENCY AND AUTHORITY	BENEFICIARIES (thousands) 65	66	67	FUNDING (millions) 65	66	67	IMPACT ON LEARNING OF DISADVANTAGED: Experimental (Controls)	Operational (Norms or Judgments)
Pre-School (3-5)									
Head Start	OEO-CAP	561,000 20,000	500,000 100,000	500,000 210,000	$85	$180	$310	Deutsch--modest gains of pre-school year were retained and expanded in K and 1st grade. Control groups suffered cumulative slippage in K & 1st grade.	1% survey showed gains, 5 points on non-verbal tests of intelligence. No retention data available yet.
Other Pre-School programs	OE, Title I, ESEA Pre-School only Pre-School and Kindergarden		60.0 348.0	See below	---	See below	See below		
In-School Children (5-15)									
Compensatory, remedial tutorial programs in schools and complementary to schools	OEO, CAP Education Components (including Upward Bound)			339.0	---	---	77.0		Scattered evaluations, largely subjective. Tutorials and Upward Bound projects show most observers think these experiences beneficial.
OE ESEA	OE-ESEA Title I	---	6,600.0	7,000.0	---	775.0	1,070.0	Systematic evaluation in planning stage. No info. as yet.	Preliminary analyses of 500 project sample showing types of projects only. No big cities included.
In-School Youth (16-21)									
Counseling and Guidance	OE-NDEA, Counseling and Guidance	(2,800.0)	(3,400.0)	(3,600.0)	(156.2)	(209.4)	(209.7)	N/A	Descriptive and judgemental evaluations and periodic administrative assessments.
Vocational Education	OE -Voc. Ed. Act of '63								
NYC in-school work programs	OEO-Labor-EOA	102.2	115.0	125.0	28.4	75.0	81.2	N/A	Increased attendance and retention in school, but slight negative impact on academic achievement.
Work Study -- Voc. Ed.	OE-HEA	114.4 15.0	115.0 85.0	125.0 ---				N/A	---
Opportunity Grants (HE)	OE-HEA	---	115.0	220.0	---	60.8	122.0	N/A	No information.
Higher Ed. Loans	OE-NDEA	---	(400.0)	(375.0)	---	(179.3)	(190.0)	N/A	No systematic evaluation by income level of recipients.
Guaranteed Student Loans	OE-HEA	---	(132.0) 110.0	(775.0) 150.0	---	(9.5)	(46.0)	N/A	No systematic evaluation showing impact on entry, retention, or achievement of low-income students.
Work Study, Higher Education	OE-HEA		60.0 150.0	90.0 210.0		[illegible]	134.1		
Out-of-School Youth (16-21)									
Job Crops	OEO-EOA Men Women	15.6 1.3	325.8 4.2	39.0 6.0	183.0	310.0	228.0	N/A	Info. on basic educational attainment of enrollees not available yet.
NYC out-of-school program	OEO Labor (little basic ed.)	(61.7)	(60.0)	(64.0)	($44.8)	($97.0)	($138.6)		
[1]MDTA Institutional	OE-Labor	10.4 (125)	10.4 (125)	---					See below MDTA
OJT	MDTA-Labor	(---)	( )	( )					
Adult Education--(Basic Education Only)									
CAP Adult & Adult Basic Projects	CAP-Sections 206, 207 largely basic literacy or remedial subjects, taught alone as prerequisite to vocational education	23.9	87.5	117.0	---	15.3	47.0		No information.
Adult Basic (literacy) Education	OE-OEO Title II B	---	75.0	100.0	---	21.0	30.0		Preliminary findings by Greenleigh not yet available.
Adult Education -- (Basic Education Combined w/Voc. or other adult Education)									
Work Experience	OEO Welfare Administration, Title V (provides adult basic in absence of II B program).	(88.0)	(109.3) (No information on % receiving basic compensatory education)	(105.0)	(110.0)	(150.0)	(158.7)		No information on Basic Educational attainment of enrollees.
[2]MDTA Institutional Training	OE-Labor	14.5 (125)	14.5 (125)	---	(No information on funding of programs basic education components)				Four experiments showed average gain of 1.5 grade levels in basic subjects in 15-20 weeks. No systematic information on basic educational attainment of enrollees.

Table 2
(Continued)

SUMMARY OF FEDERALLY FUNDED COMPENSATORY EDUCATIONAL SERVICES AND FINANCIAL ASSISTANCE TO THE POOR

KIND OF PROGRAM	FEDERAL AGENCY AND AUTHORITY	BENEFICIARIES (thousands) 65	66	67	FUNDING (millions) 65	66	67	IMPACT ON LEARNING OF DISADVANTAGED: Experimental (Controls)	Operational (Norms or Judgements)
OJT	Labor								
Vocational Education	OE Voc. Ed. Act of '63 etal	(2,281.0)	---	---	(18% of total funding for adult classes)			None	None

NOTE:/ Numbers in parenthesis are total of beneficiaires or total funding, when it is not known how many enrollees actually receive literacy or or other compensatory education, or are poor.

1/ MDTA institutional programs. About 42% of enrollment is under 21 years of age. Twenty percent of these receive compensatory basic education. All of these latter number were counted as poor.

2/ Adults over 22 years of age accounted for 58% of enrollment in MDTA institutional training programs Again about 20% of these are receiving significant amounts of basic education. All these have been counted as poor in enrollment figures w/o parenthesis.

3/ Does not include basic compensatory. Enrollments often for single evening course.

Table 3

SUMMARY INFORMATION ON DEVELOPMENT OF EDUCATIONAL RESOURCES FOR PROVIDING COMPENSATORY PROGRAMS TO THE POOR

PROGRAM	FEDERAL AGENCY AUTHORITY	FUNDING (millions) 65	66	67	PARTICIPANTS % of Prog. of Poor	(thousands) 65	66	67	PERFORMER	EVALUATION OF IMPACT ON LEARNING OF DISADVANTAGED Control Groups: Experimental	Operational
I. Staff Development-Inservice										(Control groups)	(Against norms)
a. Professional teachers											
Institutes for in-service teachers Disadvantaged K-12	OE-NDEA Title XI				100		3.0	---	Colleges and Univ. Selected colleges and univer. Teachers Colleges	None	No information how affects poor learners Subjective type evaluation of teachers institutes based on questionnaires to participants.
Subject matter Institutes for teachers, the Sciences, social science, English, other humanities and arts	OE - NDEA Title XI	---	(34.0)	(40.0)	?		24.0	28.0	Colleges and Univer.	None	"
	NSF - general authority		(---)	(---)	?		---	---	Colleges and Univer.	?	?
Advanced Fellowships for Exp. teachers and impr. teacher ed. programs	OE HEA, Title VC		(7.5) 5.0	(7.5) 5.0	60		2.3	5.8	Colleges and Univer.	New	New
Teachers' Corp Training	OE, HEA, Title VB		13.2	31.4	100		3.7	0.08	Colleges and Univer.	New	New
Institutes for Junior Coll. faculty, science, math, social sciences	NSF, general educ. authority		?	?	?		?	?	Universities	No info.	None
Training of community service staff	OE, HEA, Title I		(10.0)	(20.0)			(100.0)	(200.0)	Urban universities	New	New
Training for educational professional staff	OEO - CAP				100					Greenleigh No significant difference between results with prof. & non prof. teachers.	

Table 3
(continued)

PROGRAM	FEDERAL AGENCY AUTHORITY	FUNDING (millions) 65	66	67	PARTICIPANTS % of Prog. for Poor	(thousands) 65	66	67	PERFORMER	EVALUATION OF IMPACT ON LEARNING OF DISADVANTAGED Control Groups Experimental	Operational
Staff Development-Inservice											
a. Profess. Teachers (con't) Training for JC prof. teacher staff	OEO,-JC		?	?	100	?	?	?	In-house program, 3 staff training centers	?	?
Training grants, teachers of handicapped	OE, Mental Retarda. and Conservation Act		(19.5)	(24.5)	?	(6577)	(6577)	(9132)	Colleges and Univers.	?	?
Teacher Trng for ABE	OE-OEO, Title II B	---	---	1.2	70		(1800)	(2160)	?	?	?
b. Aides	OE, Title I, ESEA	?	?	?							
Institut. for teachers Aides	Title XI, NDE		?	?					Colleges and Univers.		Greenleigh study showed effect of Volunteer teachers
c. Volunteers	VISTA				100				Colleges and Univers.		
d. Administrators	OE, Title I, HEA '65	--	?	?					Urban Universities		No information
e. Researchers											
Trng. Grants for Advance study for Ed. Researchers	OE, Title IV, ESEA	--	(6.8M)	(8.1)	?	---	(.03)	(20)			No information.
Institutes, Counsellors and guidance personnel	OE-NDEA	---	(7.0)	(7.0)			(.072)	(.07)	Colleges and Univers.		No information.
Institutes for Media specialists	OE, NDEA, Title III	--	0	(2.5)	?		0	(0.9)	Colleges and Univers.		No information.
II. Staff Develop. Pre-Service											
a. Prefessional teachers											
Graduate fellowships to improve or expand prog.	OE, NDEA, Title IV		(59.0)	(81.0)	10	0	0	0	Universities		
Trng. grants for advanced study in ed. rsh.	Coop Rsh Act.		(?)	(?)					Universities		
Inexperienced teacher Fellowships	HEA, Title VC	--	---	---	?						
III. Curriculum Development											
Compensatory curricula development (including math and science)	OEO Cooperative Research (curriculum projects)	(.05)	(17.5)	(20.0)	?	(Not available)			Universities and pro. for learned societies	No controlled experiments with poor learners.	No general assessment of curricula produced, or effect on learning by socio-econ. deprived.
	NSF Science & Math	(15.9)	(18.2)	(21.0)	?						
	OEO - Job Corps	?	?	?							
	OEO - NYC	?	?	?	100						
	OEO-CAP components	?	?	?	100						
	PHS health education curricula research	(5.0)	(7.5)	(8.9)	?				Universities	No information.	No information.
	OE-OEO II B Adult literacy	---	.093	1.4	100				No information.		
Materials Development											
Adult Basic Education	OEO-OE IIB Ad. Bas.	---	2.3	2.8							
Cooperative Research (compensatory remed. materials)	OE-Coop Rsh & Title IV	?	?	?							
Media Research	OE, NDEA	?	?	?							
IV. Combined Support Services	OE, ESEA, Title III Supplementary Ser-			75.0	1.45	(Not available)			Local School Systems & Cooperating Colleges and Universities.		
V. Facilities & Equip.											
a. Libraries, E&S	OE-ESEA Title II	---	(100.0)	(105.0)	20						
Libraries, Higher Ed.	OE, HEFA & HEA '65		(---)	(---)	10						
b. HE Instruc. Facilities	HEFA - '63										
Grants to community colleges		(47.3)	(103.9)	(100.7)	10						
Grants to other HE institutions		(177.0)	(359.6)	(357.2)	"						
Grants grad. Facil.		(60.0)	(60.0)	(60.0)	"						
Equipment, minor remodeling, E&S Schools	OE-NDEA, Title III	(69.9)	(19.2)	(54.2)							

NOTE: (Figures in parentheses are amounts not disaggregated to show facilities, development or curricula or compensatory type programs directed to needs o poor.)

The major education program run directly by OEO as such, Head Start, is easier to evaluate than some other education programs because it is easier to compare something to the nothing which previously existed in the field of pre-school programs for poor kids than it is to evaluate marginal additions to in-school education.

The basic evaluation data for Summer 1965 Head Start were collected by the Bureau of Census on a representative 1% sample of the children. These data cover parent participation, worker evaluation, medical history, family characteristics information, pre-and-post-testing of cognitive and behavioral gains, and staff information. Planning Research Corporation was contracted to prepare the summary evaluation report on the summer program, incorporating results from a variety of sources, including the Census data.

Other data available include that collected by the National Opinion Research Corporation (NORC) on social history and experience by interviewing 2,500 Head Start families. OEO consultants in the fields of health, education, psychology, sociology, and nutrition visited Head Start projects to evaluate program components. Independent research and evaluation studies were done by 36 private contractors, and PRC also drew upon results of local evaluation projects. These contracted and local evaluation studies investigated such areas as instruments for measuring achievement gains, differential effects of various teaching methods, effects of Head Start experience on social and emotional behavior of children, value of different program mixes, demographic information on children served, follow-up on achievement of children served, follow-up on achievement of children in the first two grades, designs for new techniques in the teaching of disadvantaged pre-school children, and various studies of impact of Head Start on children as compared with non-Head Start control groups.

Evaluation results and statistical data received to date are those on the Summer 1965 program. The highlights of these results are as follows:

1. Head Start children showed definite gains when pre-and-post-tested with instruments to measure cognitive achievement and aptitude. The extent of these gains varied according to the testing instrument.

2. Generally, when pre-and post-testing was administered to a Head Start group and to a non-Head Start group drawn from a similar population, the Head Start group registered significantly higher post-scores than the control group.
3. When Head Start children were compared to a control group of non-Head Start middle-class children, the Head Start group registered greater gains but did not reach the level of the control group even after the Head Start program.

4. Kindergarten and first-grade teachers reported that generally Head Start children began their first school year better adjusted, less shy and withdrawn, more self-confident, more attentive, and more socially oriented than comparable non-Head Start children.

5. Parent participation was particularly heavy. Parent meetings were held in 61% of the roughly 1,000 Centers visited by educational consultants, and teachers were responsible for helping to solve family problems in 74% of the Centers. Ninety-six percent of the programs provided for helping parents with child rearing, and 49% included homemaking education programs. Eighty percent of the parents expressed a new awareness of community concern for their problems. These statistics indicate that a definite national program of parent participation (such as the proposed Adult Head Start program) would meet with favorable response and could easily be organized and expanded.

6. Over 90% of workers were enthusiastic about their experience in Head Start and over 80% expressed interest in participating again. Head Start staff included 46,000 paid non-professionals (25% of total) and 97,520 volunteers (53% of total).

7. Medical testing showed that 31.35% of the children had physical defects which would have gone undetected without the medical examinations conducted through Head Start.

The Bureau of the Census has conducted a data-gathering program this spring from a sample of the centers in the annual 1966 program. The data will cover parent participation, worker evaluation, family characteristics information, medical and dental information follow-up, and staff member information. In addition, PRC has been giving behavioral and cognitive tests to a sub-sample of the Census sample (about 870 children in 72 centers) by varying program lengths. This comparison by length of program should serve as some basis for comparing the Summer and Annual programs. Test results between the Annual and Summer 1966 programs will be compared.

All this shows substantial progress on evaluation of Head Start for its effects on educational and child development achievement. It is not, however, an anti-poverty evaluation. To compare Head Start to the other programs of the War on Poverty, it is necessary to create a long hypothetical chain going from pre-school programs to in-school achievements to probability of successful completion of school programs at age 18 or thereafter to further earnings. This is something we are now studying, but is not worth carrying through on a quantitative basis until more of the Head Start data are in.

COMMUNITY ACTION

Finally, we come to Community Action. In many ways, this must be considered the most important OEO program over the long haul. We can, as has been suggested, get rid of poverty through basic income maintenance at any time we want. But it seems very unlikely that we can get rid of the fundamental obstacles to opportunity without breaking up the communities of poverty -- the urban and rural slums where bad housing and bad health facilities, lack of intellectual stimulation and just plain injustice conspire to keep the poor down. Such comprehensive programs for ending the various aspects of community poverty are far and away the most difficult to evaluate. What we are out to do is to change the total environment of poverty for the people who now live in these communities. Environmental change has no direct connection to earnings in the way that training and education do. Nonetheless, it is a necessary indirect element supporting all the other elements. A child benefiting from Head Start and from improved in-school education under Title I is far more likely to fail in his total life effort if he has to return to the same slum home and the same depressing family situation every night.

The ultimate measure of effectiveness of Community Action programs is the change in the number of people in poverty who come originally from the slum communities which are CAP's prime target areas. (The change in the numbers living in these communities is not by itself sufficient because such a change could indicate merely dispersion of the poverty problem rather than cure.) To get such an ultimate measure of change we have a number of evaluation projects underway. We have augmented the Current Population Survey taken by the Census Bureau with an additional sample of 30,000 people whom we asked additional questions helping us to get at the root of the difficulties between the poor and the non-poor. The total augmented CPS sample of 80,000 enables us to get further information never before obtained. Of particular interest in evaluation of Community Action programs, it will give us an up-to-date record of the number of poor in pre-selected Census tracts designated as poverty tracts according to their characteristics. Taken each year it will give us a measure of the effectiveness of Community Action in reducing number of poor without merely shifting them into non-poverty tracts. The 80,000 sample is not large enough to give us information on any specific geographic areas with the possible exceptions of New York, Chicago and Los Angeles, but we hope in 1968 to have taken a much larger sample Census with enough observations to make specific area-by-area comparisons with the 1960 Census and ultimately with the 1970 Census. This will give us some real measure of progress in individual Community Action programs as well as in Community Action taken as a whole.

While measures of this sort which count changes in the numbers of people in poverty are the ultimate necessity, however, they do not suffice for short-run evaluation. Community Action programs are bound to be slow-acting because of their attempt to change fundamental conditions which have existed for centuries. In order to measure current progress, we will need proxy variables measuring changes of conditions which precede the ultimate change in the poverty count in the subject communities. Such variables can include employment in the slum areas, participation of residents in Community Action programs, and other factors such as health improvement, etc. But to tie them together into a comprehensive picture of the change taking place in these communities is an extremely complex matter. We are trying to measure this change at several levels.

The CAP monitoring function, as part of the overall Grants Management operation, has two major aspects, (1) operational evaluation of the programs of CAP grantees, and (2) a general and continuous overseeing of programs funded under Sections 204, 205 and 209(b) of the Economic Opportunity Act.

The objectives of this effort are to learn about the performance of individual grantees, to systematically collect and examine information about the individual grantee programs so as to determine general problems and concerns which may be applicable to the entire CAP program, to spot potential trouble areas, and to learn of modifications which CAP might make in its own operations to facilitate program improvement. In examining the performance of individual grantees, questions such as the following are covered: the overall objectives of local Community Action Authority as originally preceived and as modified, relationships of the CAA to established institutions in the community and coordination with other Federal programs, relative program priorities of the CAA, internal management mechanism of the CAA and overall, the eligibility of the CAA for refunding.

In accomplishing its task, the monitoring function relies on a variety of sources ranging from internal reporting forms to on-site visits by in-house observation teams. It is the on-site visit which provides the most comprehensive effort describing the depth and quality of service rendered by the CAA during its period of service.

The major problem which the monitoring function now faces is the development of a system for effective use of all the monitoring inputs.

In addition to the CAP monitoring evaluations being carried out, OEO now has comprehensive evaluations underway in seven Community Action Agencies choosen to represent different types of communities (Knox County, McDowell County, Atlanta, Kansas City, Seattle, San Diego, Baltimore, Austin). These are being carried out under contract with universities and private research firms. The contractor independently designs and executes a research plan. He studies the organizational structure of the CAA and seeks to determine the impact of CAP on poverty. As such the research will look at the CAP and its individual programs; analyze not only the components themselves but their relationship to one another; study the non-CAP elements which influence the CAP and the influences which the CAP makes on these elements. These evaluations will be interdisciplinary efforts, utilizing research teams composed of economists, political scientists, sociologists and other social scientists.

The final question, of course, is how do we evaluate the overall War on Poverty. Here I am afraid that nothing but the Census and survey method will do. If we are serious about ending poverty in the United States we can do so -- we can do so by 1976. If we are serious about measuring this end to poverty -- and the Director of the Office of Economic Opportunity as well as its Office of Research, Plans, Programs and Evaluation is quite serious about it -- then we can do so. We expect to be able to carry on such measurement from year to year and we think that in the final analysis this sort of measurement is the ultimate evaluation of the Office of Economic Opportunity and the War on Poverty.

"WHAT'S HAPPENING, BABY?"--ESSENTIAL RESEARCH FOR THE WAR ON POVERTY

Sar A. Levitan
The W.E. Upjohn Institute for Employment Research*

The question in the title of this paper has been asked by the Office of Economic Opportunity in a much-publicized, but not universally praised, venture in the visual arts. Though the poverty warriors raised this question nearly two years ago, they have not yet delivered a clear answer. This Association, dedicated to facts and numbers, is a proper forum to ask again: "What's happening, baby?"

Official OEO Reporting About Its Achievements

Official public reports by the OEO leave much to be desired. First, there exists a credibility gap in some of its statistics. Expenditures per enrollee in the Job Corps, for example, are often the subject of conflicting reports. I am sure you have favorite examples of your own.

Second, public statements made by OEO officials are often sprinkled with disturbingly imprecise words such as "reached," "affected," and "served." For example, Sargent Shriver, in recent testimony before the Senate Committee on Labor and Public Welfare on the accomplishments of his agency, asserted that his program has "affected the lives of four million impoverished Americans in the slums of 800 urban and rural communities...." Shriver failed, however, to particularize the ways in which the poverty program has "affected" these people. Thus, "affected" could mean anything from giving a word of encouragement to providing a job or shelter.

Third, OEO interprets its statistics in the most favorable light possible. The OEO claimed in one study, for instance, that, of 399 work-experience trainees in nine states who had completed their assignments at least three months prior to the study, two-thirds were employed at an average monthly wage of $258. Before their selection for work experience, 60 percent of the trainees were or had been public assistance recipients for an average period of 26 months. The conclusion drawn by OEO and reported to Congress was that work experience had resulted in preparing relief recipients to obtain employment and in significantly reducing the relief rolls. The report failed to note, however, that, in a period of increasing labor shortage, the number of relief recipients is likely to decline anyway. Furthermore, the first people to withdraw from relief are likely to be the same ones who would participate in and receive most benefit from a work experience program. Thus, similar results might have been obtained even in the absence of a work experience program.

On a more significant and broader issue, Sargent Shriver recently exhorted Congress to eliminate poverty by 1976, the 200th anniversary of the Declaration of Independence. Few would quarrel with this laudable goal. Shriver failed, however, to tell Congress that the achievement of the goal would require the addition of at least $20 billion to annual expenditures in aid of the poor. Little good is done for the body politic by official pronouncements of lofty aspirations without an indication of their costs, of their prospects of implementation, and of a realistic appraisal of the chances of success.

The Need for Objective Reporting and Evaluation

Unsupported claims of achievements and exaggerated official promises for the federal war on poverty regrettably have serious repercussions. Unfulfilled promises create frustration and disappointment among those who hope to benefit. Opponents have been quick to publicize unrealistic claims as evidence of the program's shortcomings.

Despite the deluge of inflated claims and the concerted attacks of the detractors, the war on poverty has actually enjoyed remarkably sustained public support--as evidenced by diverse public opinion polls covering the population at large and more sophisticated segments. According to an opinion survey conducted by the Chase Manhattan Bank last April, 9 of every 10 academic economists, my favorite group, supported the idea of a federal effort, and a majority approved the direction the program had taken. Although business economists, understandably, showed greater reserve, 76 percent supported the concept of an anti-poverty war and 44 percent approved its operations.

---

* The Institute shares no responsibility for the views expressed here. The paper is part of a study, The Great Society's Poor Law: A New Approach to Poverty, devoted to an evaluation of the Economic Opportunity Act and financed by a grant from the Ford Foundation.

If public support of the program is to be sustained, more reliable information than exists at present is urgently needed about the operations of the several measures comprising the anti-poverty package. Such information would allow the public and Congress to rally behind programs that prove themselves and to drop activities that do not pass muster. It is not likely that the information necessary for evaluation will be forthcoming from government--either from Congress, the Office of Economic Opportunity, or other executive agencies.

Congressional hearings frequently illuminate program operations, partly through testimony by advocates and opponents, but more significantly through testimony of expert witnesses. With hardly any exceptions, the annual hearings on the Economic Opportunity Act have been devoid of the latter. Testimony before the appropriate Congressional Committees on EOA has been restricted almost exclusively to governmental witnesses and a few ideological supporters or opponents. As a result, the hearings in 1965 and 1966 offer very little meaningful information concerning program activities. To supplement the information obtained at the formal hearings, the House appropriated funds last year to the Committee on Education and Labor (the Committee responsible for the legislation) to study the anti-poverty program. The results of this investigation have never been revealed to the public and apparently not even to the members of the Committee.

Open critical appraisal of program operations is also not forthcoming from the executive agencies. These agencies necessarily advocate ongoing programs. Until a decision is made to scuttle--a rare occurrence--or modify a program, shortcomings revealed by internal research are normally classified as "administratively restricted," which means that the documents are not made available either to Congress or to the public. An expanding practice fraught with danger is the government contracting with private consulting firms and academic institutions for the survey and evaluation of public programs. The products of the outside experts become the property of the contracting agency and are frequently not published.

OEO Research

No adverse reflection is intended on the research staff of OEO and the quality of its work. Indeed, the official statistics released by OEO bear little resemblance to the products developed by its Office of Research, Plans, Programs and Evaluation. The research staff of the Office of Economic Opportunity, first headed by Dr. Joseph A. Kershaw and now by Dr. Robert A. Levine, has taken the lead among federal agencies in the application of systems analysis techniques to welfare efforts. Drawing on the vast supply of pertinent statistics, the Office of Research, Plans, Programs and Evaluation in OEO has classified and quantified the various sub-universes of the poverty population, analyzed the applicability of existing welfare programs to these groups, and prepared complementary and alternative plans for combatting poverty.

Thus far, the findings of the OEO research staff remain largely in the files of the "Poverty House," the name by which the headquarters of OEO is known, though some of it has been transmitted to the Bureau of the Budget. All that we know about this significant work is based on sketchy newspaper reports, the result of some "leaks," inadvertent or perhaps contrived. And it is very doubtful that the product of OEO research will ever become public property, unless the recommendations are adopted as official government policy, not a likely event. Neither Congress nor the public, therefore, may ever have an opportunity to assess knowledgeably the merits of the proposed multibillion dollar programs. This is unfortunate because the product of the OEO research staff deserves public attention and consideration.

Planning-Programming-Budgeting System

Better public understanding of government programs should result from the emphasis placed upon the new Planning-Programming-Budgeting System (PPBS). Closely related to the systems analysis approach, PPBS requires program planners not only to estimate budgetary costs but also to analyze their effectiveness, to examine alternative approaches, and to compare expected benefits in relation to anticipated cost.

A significant element in this approach, pioneered in the federal establishment by the Defense Department, is to plan program budgets over a longer period than the customary one year interval. Congress has steadfastly insisted that appropriations for federal programs normally be limited to one year. Accordingly, executive agencies have budgeted their programs for the same period. In practice, federal administrators have even a shorter lead time to implement proposed activities since Congress rarely makes the necessary funds available before a new fiscal year starts. Shriver and his associates still do not know today the amount of money they can commit or spend during the current fiscal year which started on July first. This fact has led to considerable confusion in administering programs; and it has proved a serious constraint on efficient administration since appropriated funds must be committed, if not spent, within the year for which appropriated. A scramble is experienced at the end of each fiscal year, a rush to commit all appropriated funds lest some be lost to the program.

There is no guarantee that advance planning over several years by executive agencies will deter Congress from insisting that the nation's federal business be run on a year-to-year basis. But the hope is that advance planning by executive agencies will also prompt Congress to make efforts to run the government on a more businesslike basis. This assumes, of course, that executive agencies will learn to plan their programs on a more sophisticated basis than previously and that they will develop techniques which would convince Congress of the desirability of adopting the aspects of PPBS applicable to its own work.

Cost Effectiveness

The basic objective of PPBS is to get the optimum return for the buck. Cost effectiveness measurement, a major component of rational program planning, seeks to determine the cheapest way to accomplish defined goals or to get the maximum advantage from a stated expenditure. As applied to the Economic Opportunity Act, the approach might be used to provide answers as to the most economic means to motivate and train disadvantaged youth, to equip them with job skills salable in the open market. Since comparable data are available on the costs of the Neighborhood Youth Corps, the Job Corps, and related programs, the determination of the cost effectiveness of the youth employment and training programs would, at first, appear a matter of simple calculation. One might too hastily conclude that the Job Corps is a more expensive program than the Neighborhood Youth Corps, for it costs about five times as much to maintain a youth in the Job Corps than to provide him with employment under the Neighborhood Youth Corps. The products of the two youth programs, however, are not necessarily interchangeable. To motivate and train certain youths it may be necessary to remove them from their environment, as the Job Corps does, and to provide them with continuing care and supervision. If that is the case, then the Job Corps, though much more expensive, may be the only way to help some disadvantaged youths.

Determination of cost effectiveness may also raise questions about the composition and direction of specific efforts. Thus far, the Neighborhood Youth Corps has concentrated upon providing employment and income to disadvantaged youths and the nature of the work is too often of dubious quality, reminiscent of old-fashioned work relief. The theory presumably is that as the youths mature they will get accustomed to the world of work and will become able to shift for themselves. Available statistics indicate that as youths mature their level of unemployment declines. But some critics have advocated the need for "enriching" the Neighborhood Youth Corps program by providing enrollees not only jobs but also basic education and more meaningful training. In view of the limited resources available to the administrators of the Neighborhood Youth Corps, enrichment would necessarily reduce the number of enrollees. A question that must be answered, therefore, is whether the effectiveness of the program for society would be raised by limitation of enrollees but with more intensive preparation for the world of work.

The above illustrations suggest that the quantifying of expected output can be elusive since it involves qualitative elements and that the pursuit of standards of quality seriously affects cost.

If a cost effectiveness analysis does come up with persuasive evidence that an alternative to an existing program is preferable, would the responsible officials be able to admit failure of their past efforts? Past experience has shown that such humility is rarely found. Despite the questionable value of some of the anti-poverty programs inaugurated in the last two years, none have been discontinued. Each program has attracted advocates within the federal establishment and a clientele outside of the government, and administrators find it most difficult to drop a program once initiated. And even if internal obstacles to changes within the executive establishment could be overcome, approval of changes in established programs or the substitution of alternatives still requires Congressional approval. Each program has its Congressional sponsors and supporters who may present insurmountable impediments to change.

On the other hand, there is no guarantee that programs whose cost-effectiveness is proven will be adopted. The Office of Economic Opportunity has concluded that expenditures for family planning is probably the most cost-effective anti-poverty measure. Nevertheless, OEO has been most sparing in funding birth-control projects--only a third of one percent of total Community Action Program funds were allocated to this activity.

The Cost-Benefit Precedent

While great hopes are expected for cost effectiveness, it might be useful to recall the lessons of cost-benefit analysis, which has been practiced by the government in the field of public works for three decades. It might appear comparatively easy to add up the total costs of a public works project, but even if the reckoning is "clean," the decision whether a given project should be undertaken still involves value judgments and guesswork. In addition, there are political considerations which cannot be ignored. It is a relatively simple task to determine the costs of labor and materials to be used on a project. But in calculating social costs it makes a considerable difference whether these resources would have been employed elsewhere in the absence of the project. Thus, it may be argued that the employment of idle labor should not be included as part of the cost of a project--at least not all of the labor cost, since idle workers may be collecting unemployment insurance or relief payments in the absence of work provided by the public works. Experts also disagree about the interest rate which should be applied to discount future benefits. The contingent and remote benefits from the project are even more difficult to calculate, and the estimates require arbitrary assumptions and projections. In the final analysis, it may be impossible to assign the dollar value benefits accrued to the various classes of consumers from a project and also to calculate losses to others, now and later. The current debate about constructing a dam in the Grand Canyon offers an excellent illustration. What cost is to be assigned to marring one of the outstanding tourist attractions in the United States as against the benefits resulting from adding a water resource?

In more recent years the government has also sponsored cost-benefit analyses in the field of manpower training programs. These studies have generally indicated an excess of benefits over costs. The conclusions may be valid, but they are based only upon certain explicit costs. In the field of training, as well as related activities, where training resources--counselors, testers, instructors--are scarce, a realistic cost-benefit analysis should include the impact of the newer programs upon the price and utilization of the scarce resources and their impact on education and other activities competing for the same manpower. A new training course may, for example, deprive the public employment service or the school system of part of the limited supply of counselors. A true cost-benefit analysis would have to consider the negative impact upon the latter institutions resulting from the expansion of demand for a limited supply of needed technicians. There is no easy way to measure this type of cost, especially if it is ignored! The studies which have concluded that the benefits of governmental training programs exceed costs may be useful to sell the desirability of funding these programs to Congress and the public. It can hardly be claimed, however, that the studies supply definitive answers to the questions they purport to study.

The Responsibility of the Academic Community

We are therefore forced to the uncomfortable, but nevertheless realistic, conclusion that PPBS and related approaches are not going to provide a complete blueprint for rational public policy and, in most cases, the results of analysis will not be made available to the public for independent appraisal. Political considerations remain potent: they are likely not only to determine the outcome of controversial undertakings, but also to prevent public airing of the questions raised by the analysis.

There is, however, an urgent need in a free society for the public and Congress to be better informed than they now are about the operations of publicly funded programs. At the very minimum, the public is entitled to frank discussions and interpretations of program operations prepared by detached experts without vested interests. The needed interpretation and evaluation of public programs can be supplied by the academic community and related private institutions, provided government agencies reveal needed information. Evaluation is particularly important in the case of the Economic Opportunity Act and related anti-poverty programs. The momentum created in favor of these programs by the inauguration of the Great Society is diminishing, partly against the background of our expanding military involvement in Southeast Asia. Greater public awareness about the achievements of successful programs will provide the necessary support for continuing and perhaps expanding effective anti-poverty efforts and for dropping those which are of questionable value.

Congress acknowledged the inadequacy of public information concerning governmental operations by passing the "Freedom of Information" Act of 1966, which curbs the power of executive agencies to withhold information about their activities. The new legislation, according to President Johnson, will no longer allow government officials "to pull curtains of secrecy around decisions which can be revealed without injury to the public."

Thus, impediments to the study and evaluation of government programs by independent researchers, if they have existed at all, in principle no longer apply. The neglect of meaningful academic research of government programs has not been due to the inaccessibility of information. The reasons for the neglect must be found elsewhere. A prime reason, in my opinion, is that institutional study has fallen into disrepute, at least in the field of economics. The emphasis in recent years on quantitative analysis has often led economists to build models without vital organs, to use Professor Jacob Viner's bon mot. Preoccupation with quantitative techniques, devoid of substantive issues, precludes controversy, attracts funds under the guise of objective scientific analysis, and is convenient for an age of consensus. Descriptive reporting, analysis and interpretation of institutional operations can lead to controversial conclusions and offer few brownie points to the aspiring academician seeking status in his profession.

Another serious impediment to the study of ongoing government programs is the trend toward greater government support of academic research. This support has been available for some years in the physical sciences and is becoming increasingly the source of funds for research in the social sciences. Government support of social science research provides no special incentive for critical evaluation of a sponsor's ongoing work, if publication of the results is also contemplated. Universities with faculties engaged in critical evaluation of government programs may find that federal spigots eventually run dry. Academic communities dependent upon government largess for support of faculties often enjoy greater prestige and acquire greater material rewards by working on grants than by teaching students. Expanding government support of research has its insidious aspects.

If the academic community is to discharge its responsibilities to the public by attempting to evaluate ongoing and proliferating government programs, researchers must not be burdened by risk of retribution, subtle or direct. As long as the rewards are found elsewhere, an adequate number of researchers will not be interested in evaluating controversial government programs. Unless universities and foundations assume a more active role than they have in the past, in encouraging the needed research, the vital job will be left undone. The major responsibility rests with university faculties which possess the expertise to do the work.

In any event, the product of the research must be freely available and the researcher must be independent of thought controls. Whether university administrators will live up to the challenge of recognizing the value of such research remains to be seen. The need for the research is indisputable.

DISCUSSION

Carl H. Madden, Chamber of Commerce of the United States

Sar Levitan's courageous paper raises the important social issue whether the bureaucratic forms of scientific procedure, without its historic spirit of detachment, may not result in developing a society that, almost absent-mindedly, imprisons the individual in an enveloping strait-jacket of cooked statistics.

To some, this formulation may seem strained and pretentious. To those it has to be said that the crusader rhetoric of the "war on poverty" threatens to stifle criticism by placing disinterested observers in grave peril of being accused of "being against people" or "favoring poverty." Levitan, with impeccable credentials as a concerned social scientist, stands on a different footing from self-interested critics. Nevertheless, he says some rather harsh things about OEO reporting, public statements, research-contracting procedures, congressional evaluation, and use of new benefit-cost techniques. If he is correct, the rather shrill metaphor of my opening remark may well be justified.

It is the tone and import of his observations that raise the most important questions. They raise the general question of the relationship between social reality and statistical procedure. We live in a more and more sophisticated environment, in which painstaking critical analysis -- especially marked in modern statistical analysis -- has become a powerful instrument of discovery and as powerful a molder of life styles. In this environment, the crudity of the poverty war rhetoric and the absurd naivety of slogans like "ending poverty by 1976" is, suggests Levitan, essentially corrupting.

Levitan itemizes the potential impact of such corruption on the research community of the government agencies, unable to publish any except approved reports; on the Congress in which independent expert evaluation of OEO results have been lacking; and even, he suggests, on some members of the academic community itself who are sworn by the terms of research grants to silence except upon approval of the contracting agency.

It is, of course, a familiar phenomenon to those who, like the Chamber of Commerce of the United States, hold that there is a real issue in the comparison of government intentions and results. Still, we are known to support the doctrine of limited government and, to that extent, confess to stated bias. To the statistician, however, the problem of stated bias is in principle easier to handle than the problem of unstated bias; and the next step, toward concealed bias, is anathema to him.

So Levitan's important paper, in this interpretation, strikes to the heart of issues close to the statistician's loyalties. Any growing profession, especially one as powerful and relevant to complex social analysis as modern statistics, cannot afford to neglect its institutional configuration. The innocent and heady idea of a neutral and always progressive science that characterized the past has been buried by the twentieth century. The medical profession in this country now wrestles with the consequences of confusion about its institutional mandate. Levitan, to me, suggests the urgency of the need for distinguished leaders of social analysis and statistics to turn at least some of their attention, as the physicists have had to do, to the design of institutional frameworks that will best permit their subject-matter to enlighten and not corrupt the public interest.

By this I mean that statisticians might ask themselves and their colleagues in the intellectual community how to apply in their contributions to operating institutions the historic criterion of detachment and disinterest that is the heart of the scientific spirit. Is it sound procedure for operating agencies, such as OEO, with vested interests in the justification of political programs, to serve as its own evaluator? What institutional arrangements would be needed to achieve sympathetic but disinterested evaluation of broad social programs?

These are powerful questions that pervade the intellectually honest and straightforward observations Levitan makes. They have application to a far wider range of programs than the war on poverty. The OEO, which after all is admittedly only a modest part of the Federal government's anti-poverty efforts, is itself in the almost impossible posture of "coordinating" Federal research on welfare questions at the same time as it conducts its own. And it is a symptom of the times that OEO's public information staff not long ago outnumbered its research staff by more than two to one.

Levitan also raises technical questions that space limits deny extended comment on. One that should not go unmentioned is the critical theoretical question whether benefits can be compared among alternative programs increasing welfare. This question plunges the analyst into interpersonal comparisons of utility which have plagued theoretical economics for generations. Yet the unsophisticated application of benefit-cost analysis is only more cooked statistics that obscures the issue of alternative uses for resources, at the heart of the economic problem.

Robert A. Levine's paper is subject to many of the comments and criticisms which Levitan directs to poverty research. It is interesting to contrast the OEO research effort as reflected by Levine's paper and Levitan's observations with the U.S. Office of Education report, Equality of Educational Opportunity. That study, marked in conception and execution by the influence of modern epidemiology and its statistical ideas, strikes expert reviewers such as Christopher Jenks as "the most important piece of educational research in recent years" (The New Republic, Oct. 1, 1966, p. 21). Of unprecedented scale for educational research, the study explores with care the effect of different school characteristics on what individual students actually learned. Its results are far-reaching for educational policy. Why should such a contrast exist between the piecemeal and confused OEO approach and the powerful and general OE approach?

The contrast points up how science can be trivialized by what Levitan suggests is self-serving. It lends force to not only Levitan's concerns but to others, such as Dael Wolffle and Henry Heald, about the impact of government grants on universities. It suggests that the academic community, in its growing role of service to public and private agencies through university research, may have to grow more sensitive to moral choices implicit in seeking contract funds or face the threat of breeding a new species of Organization Man -- the Scientific Organization Man.

DISCUSSION

Jacob J. Kaufman, The Pennsylvania State University

These two papers include one written by a government economist, who is seriously concerned with the evaluation of governmental expenditures which are designed to reduce and eventually eliminate poverty from our society. The second paper represents a rather cynical view of the ability of any governmental agency, or for that matter, any academic institution to carry out such an evaluation program. What I am not certain about is whether or not Mr. Levitan includes private research institutions in his concern of bias in evaluation.

I think one of the most significant points made by Mr. Levine is his suggestion that "every program and every experiment should include an evaluation design from the start". I would only stress that not only should an evaluation design be included, but care should be taken that the action program should itself be planned so as to provide for program evaluation. This does not mean that the action program should be subordinated to the evaluation efforts, but it does mean that with some minor adjustments in an action program we can develop a better evaluation design. In other words, both the program and the evaluation design should be evolved jointly between the director of the action program and the program evaluator.

I am also impressed by the fact that Mr. Levine is concerned with what we might call external or spillover benefits. This is an area which requires a good deal of work. I might point out that currently we at The Pennsylvania State University are conducting a cost-benefit study with respect to vocational education. We are also taking into account the benefits which may accrue to society because of vocational education programs. We are interested in determining whether or not the benefits "spill over" and reduce, say, unemployment, alleviate public assistance, etc. This analysis requires further investigation.

It is, of course, quite easy to be critical of cost-benefit studies. It is easy to set up the ideal type of study, but the real problem is to carry out an actual study. One encounters many problems in administering a project of this type. There is not a single cost-benefit study conducted by economists with which I am familiar, where one does not find a full awareness and recognition of the limitations of such a study. All we can hope for is that each successive project will make an additional contribution to the methodology and procedures. I would, however, like to make two comments with respect to Mr. Levine's reference to cost-benefit studies. He indicates that certain types of training programs may be justified even if the costs are higher than estimated benefits. What he implies is that there are other objectives besides earnings which should be taken into account. All I would suggest is that we be very conscious of these costs in relation to the benefits, to be sure that these costs justify the non-economic objectives. Surely, we would not advocate infinite costs if we wanted to achieve these non-economic objectives. I think it is possible to take care of every poor person in the United States by giving him personal training, personal guidance, and personal assistance in obtaining a job. What we must consider is the alternative use of these funds for other programs, rather than simply to compare the costs and benefits of a given program.

In this connection, I would point out that a comparison of the costs and benefits for youth enrolled in the Job Corps and those enrolled in the Neighborhood Youth Corps is still not complete. It is true that we could justify the heavy investment in youngsters enrolled in the Job Corps on the assumption that the Neighborhood Youth Corps could not meet his needs, but we might examine the question of whether or not extra expenditures in the area of public assistance, which would permit a better home life, might not in the long run prevent youngsters from becoming eligible for enrollment in the Job Corps. In other words, it might be cheaper to spend more money for public assistance, which in turn might be more costly than the Neighborhood Youth Corps, but still less costly than the Job Corps. What I would suggest here is that the Office of Economic Opportunity consider not only the alternatives within its own program, but also the alternatives with other programs which are outside of its jurisdiction.

In connection with community action programs, Mr. Levine indicates the difficulty in evaluating expenditures for this type of activity. I would agree that it is much too early to make any evaluation of this type of program, and that for the first few years we shall be confronted with vast costs involved simply in the creation of an administrative agency.

We, at Penn State, are just completing a study of a community action program involving 31 small communities in the so-called Mon-Yough Region of Allegheny County in Pennsylvania. The final report is to be submitted to the governmental agency which financed this study by September 1, 1966. We find that there are really four stages in the development of a community action program. The first can be designated as the aspirational period, when a number of people define an area's economic problem and set forth certain aspirations in connection with it. The second stage is really the mobilizing stage, when the founders of the community action program begin to include other groups in the community to participate in a community action program. The third stage, which can be called the mobilizing phase, is really the incorporation or establishment of an actual agency which will conduct various community action programs. The fourth phase can be described as the synthesizing phase, namely, that stage during which the community action program unites the various organizations in the community for the purpose of developing specific programs.

It might be pointed out that the area with which we were concerned consists of 31 small communities and is not concerned with a large city. It is clear that the process by which community action programs can be developed is unusually slow. Although Mr. Levine indicates the necessity of on-sight inspections of these programs, I would suggest that it would take a team of skilled specialists in economics and sociology for evaluation to maintain close contact with these programs and study them almost on a day to day basis for long periods of time.

Mr. Levine reveals, and correctly so, a concern for the weakness in program evaluation in the field of education. One of the reasons he ascribes for this weakness is the decentralization of the operations of the various educational programs. In Pennsylvania, we have begun to develop some close relationships with the Department of Public Instruction, and we are in the process of developing a series of research projects which will be concerned primarily with program evaluation of the types that the O.E.O. is conducting in the field of training and education. I would like to suggest that the development of a relationship between a University and a state agency, and even local school districts, may be the way in which we can develop better procedures for evaluating educational programs throughout the country.

Needless to say, everyone would agree with Mr. Levitan's concern over the need for making objective studies of governmental programs. The only problem is that on the basis of his standards it would be exceedingly difficult to find any institution, public or private, which is devoid of any influence by a granting agency. The fact is that if we are to carry on program evaluations of the types that Mr. Levine suggests, such projects become very costly. The alternative is to allow an independent scholar in an academic community, using essentially his own funds, to conduct such evaluations. But this type of research is not particularly fruitful, as it lacks any scientific basis. I think that Mr. Levitan would look favorably on this type of research, but I would enter a strong dissent. In fact, some of the suggestions of Mr. Levitan would not meet the current requirements for adequate program evaluation. I think he has a misconception of the role and use of model building and quantitative analysis. I am in favor of good research that is done by good researchers. I do not care to defend bad research carried on by bad researchers.

One final comment. I think that Mr. Levitan's concern about too great a reliance by academic researchers on government funds is considerably exaggerated. None of us is pure. We are all subject to various kinds of influences, but I must say that our substantial experience during the past several years, in which we have carried on research projects under grants from the Federal government, reveals that the influence is virtually nil.

I would suggest that in view of the fact funds for research are available from a variety of sources and as long as several scholars tackle the same problem, we can minimize the influence of the grantor on the grantee.

XIV

MEASUREMENT OF VOTING BEHAVIOR

Chairman, MARY POWERS, Fordham University

THE ELECTORAL SERIES OF THE SURVEY RESEARCH CENTER*

Donald E. Stokes

Survey Research Center
The University of Michigan

Modern governments keep records of voting on so grand a scale that no other electoral data are likely to rival the official returns. The public tally of votes will continue to be the primary material of electoral analysis, and the inferences to be made from variations across electoral units and over time can be very revealing of influences on mass political behavior. As a means of seeing into the electorate's mind, however, the analysis of election returns suffers two related disabilities, which are none the less severe for being so obvious. On the one hand, we cannot penetrate many aspects of electoral behavior without data on individual voters, whereas the official returns are tallied by precincts, wards, counties and other aggregate units. On the other hand, we cannot probe some influences on voting without measuring a much richer set of explanatory variables than could ever be gotten from official sources, even when the voting returns are augmented by census and other data aggregated by election units.

It was therefore natural that the interview survey should be applied to the study of elections, and as survey methods have developed over the past thirty years they have indeed played an increasingly important role in this type of political analysis. The first newspaper polls, in addition to showing their skills in forecasting, supplied a much surer description of the social composition of the vote. And the first academic studies, although they were confined to limited geographic areas, brilliantly displayed the versatility of the survey interview in terms of the range of information which might be collected and used. In some respects, however, survey studies began to provide basic time series on American national elections with the advent of the studies undertaken by the Survey Research Center of the University of Michigan in the years after the Second World War.

SOME CHRONOLOGY AND DESIGN

For anyone who values the chance development in research the entry of the Survey Research Center into election studies must be a gratifying story. The Center indeed became involved in voting research largely by accident in the year of Mr. Truman's surprise victory. Some weeks before the 1948 presidential election one of the Center's economic surveys asked a sample of Americans a pair of very simple questions about their voting intention. When Mr. Truman confounded the forecasters, as well as the Republicans, these data were lent a quite unexpected interest in two respects: first, this small but carefully designed probability sample by no means gave Mr. Dewey a long pre-election lead--in fact those who had formed a clear view gave Dewey no lead at all; second, and more important to the Center's growing involvement, the sample's design permitted a reinterview on the same individuals after the election. Such a follow-up survey was promptly undertaken as part of the Social Science Research Council's inquest into the difficulties in which the polling agencies had found themselves.[1]

From this modest start a very considerable program of research has in fact emerged. In the Eisenhower-Stevenson election of 1952 the Center undertook a more ambitious study, one which exploited with a nationwide sample for the first time the real possibilities of an intensive survey of voting.[2] The 1952 study took a first round of interviews in September and October and a second round with the same sample in the six weeks following the election. Such an interview-reinterview design has become a standard element of subsequent studies, and the Center has now interviewed a national sample before and after each of the last five presidential elections.[3]

This program of research has encompassed the mid-term congressional elections as well. A moderate-sized sample was interviewed before the 1954 election, somewhat larger samples after the elections of 1958 and 1962. Plans are now in hand for such a study following the congressional election this year. With this additional work, a sample of Americans will have been interviewed in every national election from 1948 to 1966, excepting only the congressional election of 1950.

---

*This brief sketch of research owes a good deal to the friendly stimulus of participants in the 1966 summer program of the Inter-university Consortium for Political Research.

[1]Elements of the Center's findings are incorporated in the volume issuing from the S.S.R.C.'s enquiry. See Frederick Mosteller et al, The Pre-Election Polls of 1948 (New York: Social Science Research Council, 1949). The Center's 1948 study is reported more fully in Angus Campbell and Robert L. Kahn, The People Elect a President (Ann Arbor, Michigan: Institute for Social Research, 1952).

[2]The findings of the 1952 study are reported in Angus Campbell, Gerald Gurin, and Warren E. Miller, The Voter Decides (Evanston Illinois: Row, Peterson and Company, 1954).

[3]The findings of the 1956 study, with reanalysis of the 1952 data, are reported in Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes, The American Voter (New York: John Wiley and Sons, 1960). Selected findings from the subsequent studies are given in Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes, Elections and the Political Order (New York: John Wiley and Sons, 1966).

The samples interviewed in these nine elections have been independently drawn, with one main exception. The sample interviewed twice in 1956 was transformed into a longer-term panel and as many of its members as possible were interviewed a third time after the congressional election of 1958 and a fourth and fifth time before and after the presidential election of 1960. The resulting individual-level data on political change during the full four years of a presidential election cycle have been an invaluable complement to the time series which can be formed from the independent samples interviewed during this lengthening span of contemporary electoral history.

TIME DATA AND EXPLANATORY MODELS

It would be difficult to exaggerate the importance to this work of its extension through time. The dividends of this extension are partly those accruing to any research which takes the investigator more than once over the same ground: the refinement of concepts and measures, the extension of theoretical focus, greater efficiency in the reduction of data--all of these came naturally enough as the research extended to additional elections. In some respects, however, repeated measurements of the same population, and even of the same individuals, have been indispensable in solving several main analytic problems of the research itself. Let me give a very few examples of how the longitudinal character of these studies has shaped the development of correct analytic or explanatory models.

Party identification and electoral choice. Only a very undiscerning student of American politics could miss altogether the importance of party loyalties in our elections. The stability of the American party system over the past century has allowed partisan identifications to become deeply ingrained in the traditions of families and other social groupings, providing those who are socialized into these groups a partisan standard that is entirely capable of guiding the voter's thought and action through an entire lifetime. Prior survey studies of voting had not missed the importance of party identification; indeed, evidence of the extraordinary persistence of these "brand loyalties" had forced a basic revision of the pioneering study by Lazarsfeld and his associates in Erie County, Ohio, in the presidential election of 1940.[4]

In many respects, however, the understanding of party identification could move beyond the common wisdom only if it were explicitly measured and brought under close empirical analysis. Accordingly, the Center's earliest studies sought to assess these enduring orientations to party. The first efforts to do so made clear that reports of party identification were by no means identical with current party choices for President. Not only did the 1952 sample include many self-described "independents" who nevertheless had party choices at the moment; the sample included too an element representing millions of traditional Democratic identifiers who were prepared to vote for General Eisenhower. In other words, there was here a kind of discriminant validation: party identification was strongly associated with voting choice across the sample as a whole, but the association was sufficiently imperfect to sustain the belief that the conceptual distinction between enduring identifications and immediate behavioral choices was preserved in the empirical measures.

Nevertheless, survey responses which were confined to a single election could not dispose of the possibility that some people link themselves to a party in a purely nominal fashion, without these verbal responses having real motivational significance for current or future behavior. It seemed possible, for example, that large numbers of people might ordinarily call themselves Democrats because Franklin Roosevelt had accustomed them to doing so during the New Deal, but would as easily come to call themselves Republicans after several years' experience with an attractive Republican president in the White House.

The first contribution of longitudinal studies to resolving this sort of issue was to demonstrate how stable the distribution of party loyalties is. As reports of party identification began to be gathered from successive independent samples of the electorate, the proportions describing themselves as Republicans, Democrats, and Independents in fact differed so little from study to study that it was quite impossible to discern any genuine change from sampling fluctuation. This stability of party identification was the more noteworthy in view of the fluctuation of the vote from year to year. Indeed, the congressional vote, having swung strongly to the Republicans at the beginning of the Eisenhower period, in three successive elections m oved steadily back toward a division more in line with the distribution of party identification.

What successive samples implied as to the stability of party identification was confirmed by the individual panel data gathered from 1956 to 1960. Of course the small net change of party identification from year to year was bound to conceal some compensating streams of gross change; no large panel study has many turnover cells in which the frequencies vanish altogether. Nevertheless, by comparison with the rates of change of a variety of other individual-level measures of political attitude and behavior, the stability of party identification was most impressive, and this evidence strongly supported the view that our measure tapped a psychological orientation of great durability.

Clarifying the relationship between party loyalty and current voting choice, however,

---

[4] Paul F. Lazarsfeld, Bernard Berelson, and Hazel Gaudet, The People's Choice (New York: Duell Sloan and Pierce, 1944). The change of focus which the stability of party choice obliged these investigators to make is discussed in Peter H. Rossi, "Four Landmarks of Voting Research," Chapter 1 in American Voting Behavior, Eugene Burdick and Arthur J. Brodbeck, eds. (Glencoe, Illinois: The Free Press,1959)pp. 5-54.

required more than a demonstration of the stability of party identification. In three successive presidential elections--1952,1956, and 1960--the Republican candidate did better than could be expected on the basis of the distribution of party identification alone. In view of this it might still be claimed that millions of people who called themselves Democrats for reasons of family history or local tradition, especially in the South, were developing habitual Republican ties at the presidential level. A critical test of this hypothesis was made possible by the panel study from 1956 to 1960. If our measure of party identification were failing to detect an emerging group of presidential Republicans, the relationship between the set of Democrats supporting Eisenhower in 1956 and the set of Democrats supporting Nixon in 1960 ought to be a nesting one; that is, the group of traditional Democrats who voted Republican in at least one of these years ought to consist very largely of the presidential Republicans who voted against their nominal party both years, along with an additional element which voted Republican in the first of these elections, when the political tides were running a little more strongly to the Republican party. If the hypothesis as to the emergence of presidential Republicans were wrong, however, the forces producing deviation in the two elections would be independent, and the set of traditional Democrats voting for Eisenhower would overlap the set of Democrats voting for Nixon no more than we would expect if these two defections were statistically independent events.

In fact the truth lay much closer to the second of these models, and the two sets of defecting Democrats were found to be markedly non-overlapping. The inference to be drawn from this as to the motivational force of party identification was clear enough: despite the fact that in 1960 there would again be extensive Democrat defections, much the best prediction of what an Eisenhower Democrat would do in 1960 was one assigning him to his historic party. The best forecast of what an Eisenhower Democrat would do the next time around was to say that he would return to his traditional party loyalty, although another set of Democrats, especially Protestants who were obsessed by Kennedy's Catholicism, would vote Republican.

As repeated observations disposed of these issues of validity, the measure of party identification could be entered with increasing confidence into analytic models of the electorate's behavior. The "unbiassed" character of reported party loyalties has played a critical role in the effort to extract from the distribution of party identification a "normal" or "expected" division of the vote for the electorate as a whole, or for a given population grouping.[5] In fact, the "normal" division was found to depend on more than party identification alone: since the Republican Party has a disproportionate share of supporters who are better-educated, better-informed, and emotionally committed voters, Republicans *ceteris paribus* are less likely to be drawn away from their party by transient negative influences; hence, if the political issues and personalities of the moment favor both parties equally, we would nevertheless expect the Republicans to keep the support of a larger part of their identifiers at the polls (although they would lose the election owing to the Democratic preponderance of party identifiers).

Knowing how the vote would divide if it expressed only relatively enduring political orientations permits a surer description of transient political influences, and the capacity to separate long-term from short-term influences has in fact been a chief result of this work. The analysis of short-run forces on turnout and the party vote has provided a key to other aggregate properties of the electorate's behavior. For example, observing (1) that the support which a winning presidential candidate attracts from beyond his party is most likely to come from independents and persons of weak attachment to the other party, (2) that the presidential coattails will transfer some of this support to the congressional candidates of the same party, and (3) that independents and weak partisans are much more likely to drop out of the electorate at a mid-term congressional election gives a simple explanation of why it is that the president's party so often loses seats at an off-year election, an explanation which has nothing to do with increasing hostility to an administration's policies.[6]

Attitudinal components of the presidential vote. The immediate forces on the electorate's behavior are notoriously multivariate in character. At the very least, each presidential contest confronts the electorate with four principal actors--the two parties and their presidential candidates--which are objects of positive and negative popular feeling in varying degree. What is more, the grounds of favorable and unfavorable response involve a multiplicity of foreign and domestic issues, advantage or disadvantage of various social groups, the personal attributes of the presidential nominees, and so on. To assess the cognitive and affective content of these responses, the Center has begun each of its pre-election interviews with a very extended sequence of free-answer questions about the parties and candidates. Although the full qualitative variety of answers to these questions is carefully preserved in coding, the analytic use of this material makes some kind of data reduction mandatory. Accordingly, the partisan direction and frequency of responses have been used to place each sample respondent on several scales of attitude toward the actors of presidential politics. The simplest analysis has used four such scales, corresponding to the two

[5]This effort is discussed most fully by Philip Converse in "The Concept of a Normal Vote" Chapter 1 of Angus Campbell *et al*, *Elections and the Political Order* (New York: John Wiley & Sons, 1966), pp. 9 - 39.

[6]This explanation is developed by Angus Campbell in "Surge and Decline: A Study of Electoral Change," *Public Opinion Quarterly*, 24 (1960), pp. 397-418, and Chapter 3 of Campbell *et al.*, *Elections and the Political Order* (New York: John Wiley and Sons, 1966),pp. 40-62.

parties and two candidates, but it is at times useful to reorganize the material so as to form a somewhat larger number of scales, corresponding to the different grounds of evaluation of the parties and candidates.

Despite the exceeding variety of the attitudes which are involved in presidential voting, the elector must come in the end to a single preference. To resolve these final preferences--and the electorate's ultimate choice--into a set of attitudinal components we have combined the several dimensions according to a linear probability model whose weights could be estimated as the coefficients of a multiple regression equation. Multiplied by the displacement of the sample mean from the theoretical neutral point of a given dimension, the partial regression coefficient associated with the same dimension provided an estimate of how much the dimension had on average increased or lessened the likelihood of individual electors voting in this way. Alternatively, these quantities estimated the extent to which a given dimension had increased or lessened the winning party's majority.[7]

A model of this kind inevitably involves assumptions, and the extension of this research through time has allowed a more adequate test of these. Especially important was the test of the model's assumption that the individual's placement along the several dimensions of attitude is a sufficient explanation of his voting predisposition, in an immediate psychological sense, and that additional factors, including errors of measurement, have not biassed the probability estimates toward one party or the other. Formally speaking, this assumption implies that the estimating multiple regression equation will exhibit a constant term of 0.5. This was true in the Eisenhower elections, but the constant might have been 0.5 as a result of compensating errors. Hence, a good test of this assumption awaited elections in which the tides of politics would flow strongly in the Democratic direction. The test was not long in coming: from the second Eisenhower-Stevenson election to the Johnson-Goldwater contest the two-party division of the vote swung a remarkable 18 percentage points toward the Democrats. In none of the elections of this period did the model's estimate depart from the actual majority by more than two percent; indeed, the correlation of the estimated and actual figures exceeded .98.

The reality of issue beliefs. From the beginnings of survey studies of opinion, investigators have been aware of how frail are the means by which they seek to measure attitude formation in the mass public. Evidence of these frailties is distressingly plain in the changes of opinion that can be induced by subtle changes of question wording, and perhaps even more in the willingness of a portion of a sample to offer opinions on mythical or nonsense issues. Yet the measurement of issue attitudes is essential to the purposes of opinion research. Only if survey studies have a degree of success at it can they help answer some of the largest questions as to the place of the broad public in the political order. Certainly this is true of the questions of popular influence in government. The intercourse of public and political leaders on issues of public policy is a good deal what democratic theory is about.

The need to separate real from unreal opinions has strongly influenced the construction of the Center's political questionnaires. This need helped prompt the use of the free-answer questions to which I have alluded, questions which allow the respondent to talk about the parties and candidates in terms of his own choosing. In the design of more structured issue items these questionnaires went to unusual lengths to allow a respondent to reveal that he had no opinion on the subject rather than choose one or another response among a set of alternatives. This kind of permissiveness, however, did not remove the suspicion that the frequency of nominal responses was more than trifling. Several types of evidence, especially the exceedingly low interrelationship of issue items

---

[7]If a respondent's eventual behavior is scored 0 or 1 according to whether he votes Democratic or Republican, and the respondent's position on each of a set of I attitude dimensions is expressed in terms of sample standard deviations about a theoretical origin, the model expresses the probability of the respondent's voting Republican as the linear combination

$$P(R) = b_1X_1 + \ldots + b_IX_I \ .$$

Therefore, across the whole sample the average extent to which positive or negative attitude on the ith dimension may be said to have increased or lessened the probability of voting Republican depends on two quantities: (1) the coefficient b_i and (2) the displacement, $\bar{X}_i - X_i^o$, of the sample mean from the neutral point, X_i^o, of the dimension--that is, from the point where the sample mean would lie if attitude on the ith dimension were not more favorable to one party than the other. Hence the product

$$b_i\ (\bar{X}_i - X_i^o)$$

is an estimate of the contribution of the ith dimension to the winning majority, a contribution which would be nil either if b_i were to vanish or if $\bar{X}_i$ and X_i^o were to coincide. The difference

$$P_R - .5 = \sum_i b_i(\bar{X}_i - X_i^o)$$

is the model's estimate of the direction and extent to which the proportion P_R of the two-party vote cast for the Republican candidate will depart from fifty percent or, equivalently, of the direction and magnitude of the winning majority. Applications of the model to the Eisenhower elections are reported in Donald E. Stokes, Angus Campbell, and Warren E. Miller, "Components of Electoral Decision," American Political Science Review, 52, (1958), pp. 367-387; to electoral change over the past four presidential contests, in Donald E. Stokes, "Some Dynamic Elements of Contests for the Presidency," American Political Science Review, 60 (1966) pp. 19-28.

which might be expected to show higher intercorrelations, made a further analysis desirable.

An invaluable opportunity to examine these matters was presented by the panel study from 1956 to 1960. Pains were taken to ask many of the same issue questions in each of the three elections studied. The replies to certain of these questions confirmed the suspicion that nominal or random responses had frequently been given. The telltale indicator that this was true was the remarkable fact that on some issues the opinions held by our sample at the third period of time (in 1960) could be predicted as well from opinion held at the outset of the study (1956) as from opinions held midway (1958). This fact is exceedingly hostile to the view that the turnover of opinion in the sample reflected a change of genuine attitudes: if it had, the fit of 1958's opinions with those of 1960, or of 1956's opinions with those of 1958, would almost certainly have been closer than the fit of opinions held four years apart. In fact, the only plausible model of change which could account for the findings was one in which a part of the sample could be said to have real and stable opinions, the rest to have made a random selection among the opinion alternatives offered by the interviewer. What is more, the size of the sub-sample having authentic opinions in some cases was astonishingly small. For example, government provision of electricity and housing, questions which have attracted political debate at least since the time of the Roosevelt New Deal, apparently were matters of genuine attitude formation for something like one respondent in six, whereas almost five in six had volunteered an opinion in the initial interview. What my colleague, Philip Converse, has called "non-attitudes" proved to be a most obtrusive element of these data.[8]

THE CHANGING FOCUS OF RESEARCH

The longitudinal character of these studies has had a pervasive influence on the focus of research as well. Beyond the specific contribution to the development of explanatory models which I have tried to illustrate by several main examples, the extension of this work through time has had a number of critical side-effects. None of these is tightly determined by the lengthening time interval of the research, but in each case the extension through time has been a natural prelude to a new emphasis of these studies.

First of all, the fact that a series of elections has been encompassed within a single program of research inevitably has shifted attention from the individual voter to the electorate as a whole. The thought and action of the molecular citizen have not been lost sight of; the social and political variation afforded by a series of elections in fact has aided the refinement of models used to explain individual voting behavior. Rather, the extension of research over a number of elections has encouraged the development of related models which utilize what is known of individual behavior to deal with election outcomes as whole events. This is specifically true of the models of long- and short-term influences on the electorate to which I have referred.

Second, the extension of research to a number of elections has strengthened the sense of obligation to describe contemporary electoral history. This descriptive purpose in no sense conflicts with the theoretical purpose of developing explanatory models. On the contrary, the use of a correct model to explain an election outcome can be regarded as a form of description. In the giving of this sort of historical account, as so much else, nothing is so practical as a good theory.

Third, the interplay of theoretical and descriptive purpose in the historical present has deepened our interest in the past. This interest is partly a matter of looking for additional events to which current explanatory models apply. For example, the model of differential drop-out from the presidential electorate which will produce a loss of strength for the president's party in the following mid-term election can be used to account for an aspect of national party competition which has been present with remarkable consistency during the hundred-year life of our modern party system. Yet the interest in the past is also a matter of wanting to be clearer about the aspects of the present which need explaining; if, for example, the competition of parties since the Civil War gives unmistakable evidence of the presence of forces restoring the strength of the "weaker" party, we can pursue more diligently a search of contemporary data for clues as to the nature of these forces.[9] The interest in the past is a matter too of wanting to know the limits of the empirical domain to which contemporary models apply. Although the background of electoral behavior is much more varied over a twenty-year span of politics than it would be in a single or a very few elections, one era in one country will exhibit only limited differences in respect to things which are of great importance to voting. Having recourse to the past is a way of escaping these bonds, however difficult it may be to recapture historical materials equivalent to the data of contemporary research.

Fourth, the desire to analyze electoral behavior under broad changes of background factors, a desire strongly encouraged by the extension of research through time, has led naturally to a quest for comparative material. The politics of a single nation, even if traced over several historical periods, can hardly yield up more than

[8] The findings which I have touched, and the methods by which they are derived, are very fully discussed in Philip E. Converse, "The Nature of Belief Systems in Mass Publics," in Ideology and Discontent, David E. Apter, ed. (New York: The Free Press of Glencoe, 1964), pp. 206-61

[9] This sort of intercourse of survey and historical data in fact led me to review with Gudmund Iversen the evidence that powerful equilibrating forces must have operated in the American party system. See Donald E. Stokes and Gudmund R. Iversen, "On the Existence of Forces Restoring Party Competition," Public Opinion Quarterly, 26 (1962), pp. 159-171.

a limited variety in the political institutions and social structure which have such pervasive effects in electoral behavior. The yield can be increased by comparative analysis of electoral politics in nations whose social and institutional structure differs from our own. Accordingly, over a period of years the Survey Research Center has sought to carry forward analyses of this kind, and members of our group have collaborated with foreign scholars in projects for which the research site has variously been France, Norway, Great Britain, Canada and other nations.[10]

Finally, several of these developments have given new prominence to questions about the place of the electorate in the political system as a whole. The shift from micro- to macro-analysis and the search for social and institutional variation across time and national boundaries have enhanced our interest in the electorate's role within a wider political order. The party system has been a natural focus of this interest, and attention has been given the pervasive influence of the party milieu on voting, as well as the effects that voting behavior may have on the party system itself. Attention has also been given to institutional relationships that may link the electorate to other actors in the political system, especially the relationship of legislative representatives to their mass constituencies. This work has itself had a comparative aspect, and some of the design of a study of representation in the American Congress, undertaken in the 1950's, has been incorporated into a study of representation in the British Parliament.[11] Plans have been drawn to extend this comparative institutional analysis to other nations as well.

PATTERNS OF USE

The developments which I have touched characterize the work of the political research group at the Survey Research Center. I ought not to close, however, without mentioning the extent to which the Center's data have become a common resource of a much larger group of political scientists and scholars in related disciplines. As the Center's electoral series extended to more and more elections, an increasing number of requests were received for access to the data, requests which the Center sought to honor in a variety of ways.[12]

By the end of the 1950's a number of important publications had reported secondary analyses by scholars outside the Center's staff.[13] In time, however, the volume of requests, as well as the requests for training in data analysis, became so great that some new basis for sharing these data was needed. Accordingly, the Center joined with a number of other universities in forming an Inter-university Consortium for Political Research which would organize the archiving and use of data and training in their analysis. Over the three years of its life, the Consortium's membership has increased rapidly, and seventy-six North American and European institutions now are associated with the Survey Research Center in this way. Once the Consortium's archive had been created, it was expanded to include data collections other than the Center's electoral studies.[14] This archival work has been generously supported by the National Science Foundation, and a committee of the American Historical Association has helped guide a vast extension into antique American election returns, reported by county, back to the early years of the last century. The frequency of requests for data from member institutions has reached a very high level indeed: in the year from July 1965 to June 1966 several hundred distinct requests were processed by the staff.

The emergence of the Consortium as a data-sharing device clearly has widened the use of the Center's voting data as a resource for the study of American elections. Indeed, our desire to share these materials widely has led us to distribute to member universities the data gathered from each new election as soon as they are in machine-readable form, rather than reserving them until a primary analysis has been completed in Ann Arbor. This practice was followed in 1962 and again in 1964. It will be followed in the 1966 study now being prepared. Increasingly a whole profession has joined in the design and analysis of the Center's electoral studies.

---

[10] Several of the papers which have issued from these collaborations are collected in Angus Campbell _et al._, _Elections and the Political Order_. Op. cit.

[11] Selected findings of the American representation study appear in Warren E. Miller and Donald E. Stokes, "Constituency Influence in Congress," _American Political Science Review_, _57_ (1963), pp. 45-56.

[12] For example, two summer institutes for interested political scientists were held at the Center during the 1950's under the auspices of the Social Science Research Council.

[13] A partial list might include Morris Janowitz and Dwaine Marvick, _Competitive Pressure and Democratic Consent_ (Ann Arbor: Institute of Public Administration, 1956); Heinz Eulau, _Class and Party in the Eisenhower Years_ (New York: The Free Press of Glencoe, 1962); Robert E. Lane, "Political Personality and Electoral Choice," _American Political Science Review_, _49_ (1955) pp. 173-90; and Robert Agger, "Independents and Party Identifiers: Characteristics and Behavior in 1952," in _American Voting Behavior_, Eugene Burdick and Arthur J. Brodbeck, editors (Glencoe, Illinois: The Free Press, 1959), pp. 308-329.

[14] The list of additional studies would include the data from Gabriel Almond and Sidney Verba, _The Civic Culture_; Robert A. Dahl, _Who Governs_; Arthur S. Banks and Robert B. Textor, _A Cross-Polity Survey_; Bruce M. Russett _et al._, _World Handbook of Political and Social Indicators_; John C. Wahlke, Heinz Eulau, William Buchanan, and Leroy C. Ferguson, _The Legislative System_; Samuel A. Stouffer, _Communism, Conformity, and Civil Liberties_; U. S. Bureau of Census, _County and City Data Books_, _1952_, _1956_, _1962_.

ESTIMATES OF "ELIGIBLE" VOTERS IN SMALL AREAS: SOME FIRST APPROXIMATIONS*

Meyer Zitter and Donald E. Starsinic
U. S. Bureau of the Census

Figures available on the voting-age population and on votes cast indicate that even in national elections Americans turn out in relatively small numbers on election day. In 1964, for example, out of a total population of voting age of 114 million, only 71 million reportedly voted for president. In other words, less than two-thirds of the resident population of voting age voted for president. There is a tremendous State variability about the national average of voter participation, ranging from a low of 35 percent in Mississippi to almost 80 percent in Utah.

The population used as a denominator in determining the percentage voting is the total resident population of voting age, without regard to voter eligibility. This type of arithmetic results in relatively low voter participation rates. This paper represents an attempt to identify and to isolate the various categories of apparent nonvoters, to demonstrate a methodology of distributing the various components of nonvoters to States and smaller areas, and to illustrate the impact on State voter participation rates when adjusted base populations are used. The results may throw light on the general problem indicated in the Report of the President's Commission on Registration and Voter Participation (chaired by one of our speakers, Mr. Scammon), as to whether the low voter participation of the American electorate is a matter of disinterest, or more a product of restrictive legal and administrative procedures.

Nature And Scope Of Problem

In virtually every State in the United States, voter participation is dependent upon eligibility and two actions of the potential voter: (1) the voter must register at the place and during the time specified by the locality where he will cast his vote; and (2) the voter must cast his ballot on the day specified for the election.

In developing estimates of the components of the nonvoting population, we have concerned ourselves with two separate categories, each of which has a different degree of attachment to the voting process, and for which estimates of varying levels of accuracy can be derived. First, we have a group that is most likely to be ineligible to vote. The group includes aliens, who are clearly ineligible, persons failing to meet specific residence requirements of State, county, or precinct, and persons in specific kinds of institutions, such as correctional institutions, mental hospitals, and residential treatment centers, also usually ineligible.

The mobility of the American people is an important contributor to low voter participation because of length of residence requirements. About 21 million adults now move every year in the United States and are likely to have their voting eligibility affected because of residence rules. Of these, 3-1/2 million move across State lines and may become ineligible to vote because of State residence requirements. An additional 3-1/2 million move across county lines within the same State and are affected by county requirements on length of residence. The remaining 14 million local movers are affected to the extent that they cross local election district (precinct) lines. The majority of States (about 35) require one year of residence as a qualification for voting. About 14 States require only six months of residence as a qualification, but comparatively few of the most populous States are included in this group. To mitigate this length of delay in qualifying for voting, a number of States permit newcomers to vote for president and vice president if they were qualified as voters in their State of residence prior to their last move. A few States also permit the use of absentee ballots by persons who have moved from their States with insufficient opportunity to establish residence in their new States.

The second group this report deals with is made up of persons who are eligible to vote somewhere, but have a variety of obstacles in their paths. In many instances, their eligibility is in a place other than their current place of residence. This group includes Armed Forces (but not their dependents), college students away from home, and the population in such institutions as homes for the aged and dependent. Also included are persons who a priori appear eligible to vote by all criteria, but who are unexpectedly away from home on election day and thus cannot get to the polls to vote. This group includes persons called away on business or away on vacation (or in other travel status) and persons unexpectedly hospitalized.

Revised Estimates Of Voter Participation Rates

The bulk of the work of arriving at revised estimates of voter participation, in which the base resident population is modified to more closely approximate the population "exposed" to voting is found in the tables. They show each category that has been included in the estimates and the approximate number, by State. The

---

* The authors wish to gratefully acknowledge the assistance of Mrs. Mildred R. Stanback in compiling and processing basic source data used in deriving the estimates.

sources and methods of deriving the national totals and State distributions are also given. Table 1 shows revised estimates of voter participation rates for 1964; tables 2 and 3 indicate the various components of the ineligible and marginal voter categories and, in effect, illustrate the method of arriving at revised participation votes; and table 4 shows the percentage of nonvoters explained by our estimates.

At the national level, of the approximately 43 million persons of voting age who presumably did not vote in 1964 (or at least did not vote for president), about 14 million, or one-third, fall into one or the other of these categories. According to the estimates, there are 2.5 million aliens, 4.8 million persons failing residence requirements, and about 1 million in correctional and mental institutions; that is, about 8.3 million persons fall into the so-called "ineligible" group. In the second category, there are 1.6 million Armed Forces (station strength), 1/2 million other institutionalized, mainly in homes for the aged, and also 1.6 million students away at colleges. Of those otherwise eligible but away from home on election day, we estimate approximately 1.6 million as staying overnight in hotels and motels, and about 1/2 million in general (short-stay) hospitals. Although these national levels are highly approximate, they suggest the magnitude of the problem and provide overall controls in estimating State distribution.

On a State-by-State basis, there are substantial differences in the proportion of nonvoters that can be explained, that is, fall into one or another of our nonvoting classifications. Aliens, for example, are highly concentrated and are found in appreciable numbers in only a few States. Both California and New York are estimated to have in the neighborhood of one-half million aliens of voting age. Texas has about 200,000. Within the specific length of residence requirements, fast-growing States like California and Florida will have a relatively larger number of recent arrivals than less rapidly growing States with substantially less in-migration.

Although nationally it is estimated that about one-third of all nonvoters come under our criteria, there are ten States where more than half the nonvoters are accounted for here. In California, for example, two million of the almost four million nonvoters in 1964 fall into our classifications. New York, with one-third of the nonvoters accounted for, is estimated to have about 1.3 million persons out of its base voting-age population of some 11 million who cannot be expected to vote.

On an overall basis, voter participation rates increased by about 10 percent after deducting the nonvoting groups identified here; that is, for the nation as a whole for November 1964, voter participation increased from 62 percent to about 71 percent. A number of States make a relatively good showing in voter turnout. Nineteen States have voter participation rates in excess of 80 percent. On an unadjusted basis, no State had participation rates that high. Of course, there are many States where the adjustments suggested here have little impact on implied voter participation, since only a small portion of the nonvoters were estimated to fall into our categories. In many of the Southern States, for example, only about one out of four nonvoters can be accounted for here.

Sources Of Data And Adequacy Of Estimates

At this point, the utility of the estimates might be considered. Do they represent a set of usable numbers which provides relatively good guides to the State distribution of those ineligible to vote? The estimates are, of course, subject to unknown but probably a high degree of error. Not only is there a degree of uncertainty as to whether all persons assigned to a given category lacked voting rights in 1964, but there is also the possibility of overlap in the estimates of the population of each group. For example, persons failing residence requirements according to the estimates may also later be included as members of the Armed Forces.

Consideration of the sources provides the main guides into the acceptability of the estimates inasmuch as only fragmentary direct evidence on the potential error is available. Within the group labelled as "ineligible," the count of the number of aliens is relatively reliable, since the total number of aliens is reported annually by the Immigration and Naturalization Service. There is some uncertainty as to the number of voting age, but from census sources on "foreign born, in the United States in 1960, living abroad in 1955," we arrived at a rough percentage of the group that is of voting age. Thus the degree of error in this component is probably very small.

The residence requirements category is the largest component of ineligibility. To derive estimates for this group, we had available to us annual survey totals on interstate, inter-county, and intra-county movements. These were distributed to States on the basis of the 1955-60 distributions of the various categories for that period available from the 1960 Census. For each State, we allocated an appropriate proportion of the movers and migrants as being ineligible, based on the particular length of residence requirements for each area. No specific allowances were made for such probably nonvoting groups as drifters, hoboes, and persons of no fixed address.

Within this category (i.e., residence requirements) there is the problem of the 15 States that permit "newcomers" to register and vote. No attempt was made to determine how many people qualify for this particular category and actually go to the trouble of voting. In 1960 in California only 12,000 ballots were cast in this category, a number hardly large enough to have serious impact on our results. On the other hand, it is possible that many of these "new residents" are eligible for absentee ballots in their own States and choose that route for voting.

The estimates of persons in correctional or mental hospitals and institutions are basically from the 1960 Census, updated to take account of population growth since 1960. The estimates of this component are probably fairly firm. On the other hand, no allowances were made for former inmates of correctional institutions who may still be denied the right to vote.

We believe that the probability of error is largest within the group of "marginal voters." Here, for example, we do not know to what extent persons in this group cast absentee ballots. Nor do we know to what extent special arrangements may be made to vote for "shut-ins" such as those in homes for the aged or needy.

Practically all States permit persons who lived there before entering the Armed Forces to vote there. The military services apparently make strong efforts to circularize and send information to military personnel on absentee voting, but we do not know exactly how many take advantage of the opportunity. About 2.1 million servicemen of voting age (including those assigned overseas) would have been able to request an absentee ballot from their State of official residence in 1964. How many actually cast their votes and their distribution by State in 1964, however, are subject to conjecture, for the information available on the number of military ballots cast in the election is not sufficient to provide such detail.[1] What information we have, however, relates to 1960, and there appears to have been fairly light voting on the part of this group in that election year. In New York City, for example, somewhat less than 15,000 ballots were returned of a potential military vote of perhaps 75,000 to 80,000 (the estimated number of voting-age persons serving in the Armed Forces from New York City). Only 4,500 were counted in the city of Philadelphia; up to 10,000 were cast in Washington State (out of a potential military vote of 47,000); and about 4,000 in Rhode Island, of 15,000 possible votes.

Total votes cast by absentee ballot appear to have made up only a modest proportion of the population voting. From the limited information available, votes cast by absentee ballot run about five-percent of the total votes cast. In California in 1960, 243,000 absentee ballots were cast out of a total of 6-1/2 million ballots. In the State of Washington, 92,000 absentee ballots were cast out of a total of 1.3 million, and about 58,000 in Connecticut out of a total of 1.2 million votes cast. Because of the large number of persons identified as marginal voters, it appears that allowing for absentee ballots would have some, but not a very appreciable, impact on the overall level of nonvoters indicated in the tables.

The estimated number of persons away from home on election day, either away on business or in hospitals is also very rough. The overall totals on those away on business or vacation were derived from data published by the hotel-motel industry, citing 1963 Census of Business data. The totals were distributed to States in proportion to population. The number in hospitals was developed from data published by the American Medical Association and the National Center for Health Statistics. Perhaps many of those away from home on business may have anticipated their need to vote and arranged for absentee ballots. Despite the uncertainty of the level, it is obvious that a relatively large number of persons who may otherwise be eligible do not get to the polls in elections because they are away from home. In fact, the estimate of only two million in this category is probably too low.

Registration Statistics As A Data Source

The aforegoing represents a methodology and a framework for deriving approximate estimates of the population of voting age who have a clear field to the voting booth, or at least have no important obstacles to overcome on their way to the polls.

As election day approaches, it is the number of registered persons, of course, that actually determines the number of voting eligibles. In a fully automated and computerized society, one should expect the registration machinery to start whirling soon after the registration books are closed, and to generate tabular statistics more rapidly than the campaign speeches.

Unfortunately, such is not the case. In fact, there is a paucity of data readily available on registration statistics. Here is a set of administrative statistics that appears to have gone untapped over the years. Presumably, registration data could be compiled and summarized so as to present characteristics of the population registered to vote for all areas for which data are available. A highly sophisticated system of registration might even provide gross change data on registrants which may tell us something about changes in population composition and distribution.

---

[1] Recent information published by the Department of Defense indicates that overall voting by military personnel may be substantially higher than suggested here. See The Federal Voting Assistance Program, Fifth Report, prepared by the staff of the Federal Voting Assistance Program, Office of the Assistant Secretary of Defense (Manpower), Washington, D. C.

In order to learn more about registration statistics, it would be desirable for some group to undertake a survey of State and local sources of the nature and the availability of registration statistics. The survey should not be concerned with the legal requirements of the registration process, but rather, would emphasize the statistical viewpoint to determine the kind of information actually obtained during the registration process. Statistics indicating the extent to which such data are tabulated, summarized, or published by the State or local jurisdiction as well as the characteristics of information available would be of particular interest.

In the interim, a review of available State documents which describes the registration system of each State indicates the following:

Age is obtained in about 30 States; 16 States obtain race; 22 States indicate male or female; 26 States ask for occupation; most States call for State of birth, and a goodly number request marital status. Unfortunately, this summary tells us nothing about availability of such data; the subject needs further consideration.

Other

One final observation is in order. After the last general election in November 1964, the Census Bureau in its November 1964 Current Population Survey asked whether persons voted in the November election. This item is scheduled again for November 1966. We believe it would be highly desirable to obtain information about nonvoters or nonregistrants, so as to be able to separate the "ineligible" from the voluntary nonparticipants.

Table 1.--RESIDENT POPULATION OF VOTING AGE AND PERCENT VOTING, BY STATE: NOVEMBER 1, 1964

(Numbers in thousands)

Region, Division, and State	Resident population of voting age	Votes cast for president	Percent voting	Ineligibles or presumed nonvoters	Percent of "eligible" voters voting
UNITED STATES, TOTAL	113,795	70,642	62.1	14,015	70.8
Northeast	29,227	19,621	67.1	3,235	75.5
North Central	31,588	22,209	70.3	2,942	77.5
South	34,410	16,599	48.2	4,584	55.7
West	18,570	12,213	65.8	3,254	79.7
New England					
Maine	580	381	65.7	72	75.0
New Hampshire	398	288	72.4	54	83.8
Vermont	232	163	70.3	30	80.8
Massachusetts	3,267	2,345	71.8	490	84.4
Rhode Island	547	390	71.3	87	84.7
Connecticut	1,708	1,219	71.4	182	79.9
Middle Atlantic					
New York	11,280	7,166	63.5	1,348	72.2
New Jersey	4,131	2,848	68.9	404	76.4
Pennsylvania	7,085	4,823	68.1	568	74.0
East North Central					
Ohio	5,978	3,969	66.4	542	73.0
Indiana	2,831	2,092	73.9	226	80.3
Illinois	6,383	4,703	73.7	695	82.7
Michigan	4,673	3,203	68.6	407	75.1
Wisconsin	2,390	1,692	70.8	194	77.0
West North Central					
Minnesota	2,021	1,554	76.9	160	83.5
Iowa	1,636	1,185	72.4	122	78.2
Missouri	2,729	1,818	66.6	284	74.4
North Dakota	360	258	71.7	43	81.5
South Dakota	394	293	74.4	44	83.7
Nebraska	870	584	67.2	81	74.1
Kansas	1,324	858	64.8	143	72.7
South Atlantic					
Delaware	287	201	70.2	38	80.8
Maryland	2,003	1,116	55.7	271	64.4
District of Columbia ..	505	199	39.4	92	48.1
Virginia	2,538	1,042	41.1	414	49.1
West Virginia	1,064	792	74.4	80	80.5
North Carolina	2,751	1,425	51.8	274	57.5
South Carolina	1,357	525	38.7	177	44.4
Georgia	2,637	1,139	43.2	374	50.3
Florida	3,477	1,854	53.3	672	66.1
East South Central					
Kentucky	1,977	1,046	52.9	245	60.4
Tennessee	2,235	1,144	51.2	186	55.8
Alabama	1,923	690	35.9	207	40.2
Mississippi	1,231	409	33.2	209	40.0
West South Central					
Arkansas	1,123	560	49.9	118	55.7
Louisiana	1,901	896	47.2	225	53.5
Oklahoma	1,487	932	62.7	149	69.7
Texas	5,914	2,627	44.4	855	51.9

(continued)

Table 1 (continued)

Region, Division, and State	Resident population of voting age	Votes cast for president	Percent voting	Ineligibles or presumed nonvoters	Percent of "eligible" voters voting
Mountain					
Montana	396	279	70.4	50	80.5
Idaho	382	292	76.5	37	84.7
Wyoming	192	143	74.4	30	88.1
Colorado	1,122	777	69.3	189	83.3
New Mexico	517	329	63.6	99	78.7
Arizona	861	481	55.8	165	69.0
Utah	509	401	78.9	73	92.2
Nevada	249	135	54.3	31	61.9
Pacific					
Washington	1,751	1,258	71.8	279	85.5
Oregon	1,133	786	69.4	100	76.1
California	10,915	7,058	64.7	1,995	79.1
Alaska	139	67	48.3	62	87.2
Hawaii	402	207	51.5	145	80.6

Table 2.--ESTIMATES OF THE POPULATION INELIGIBLE TO VOTE, BY STATE: NOVEMBER 1, 1964

(Numbers in thousands)

Region, Division, and State	Total ineligibles	Percent of voting-age population	Aliens	Failing residence requirements: Total	State	County	Precinct	Correctional and mental institutions
UNITED STATES, TOTAL	8,282	7.3	2,470	4,820	3,142	890	789	992
Northeast	2,049	7.0	866	899	447	164	288	283
North Central	1,570	5.0	420	888	597	106	185	262
South	2,620	7.6	432	1,891	1,230	479	183	297
West	2,044	11.0	752	1,142	869	141	132	149
New England								
Maine	37	6.5	15	18	8	-	10	4
New Hampshire	32	8.1	8	21	8	3	10	3
Vermont	20	8.7	6	12	8	1	3	2
Massachusetts	318	9.7	102	183	67	-	116	34
Rhode Island	51	9.3	13	34	16	-	18	4
Connecticut	108	6.3	59	36	26	11	-	13
Middle Atlantic								
New York	927	8.2	456	338	154	136	48	132
New Jersey	238	5.8	129	76	63	13	-	33
Pennsylvania	316	4.5	78	180	98	-	83	58
East North Central								
Ohio	316	5.3	65	201	138	17	46	50
Indiana	109	3.8	20	65	38	13	14	23
Illinois	429	6.7	153	219	148	33	37	58
Michigan	214	4.6	102	72	41	-	30	40
Wisconsin	99	4.1	23	54	49	-	5	22

(continued)

Table 2 (continued)

Region, Division, and State	Total ineli-gibles	Percent of voting-age popu-lation	Aliens	Failing residence requirements				Correctional and mental institu-tions
				Total	State	County	Pre-cinct	
West North Central								
Minnesota	70	3.4	16	36	24	-	12	17
Iowa	47	2.9	7	30	18	10	3	9
Missouri	160	5.9	15	123	78	18	27	21
North Dakota	19	5.4	2	15	10	3	1	2
South Dakota	22	5.6	2	18	12	4	2	3
Nebraska	30	3.4	6	18	13	3	1	6
Kansas	56	4.2	8	37	27	4	6	11
South Atlantic								
Delaware	21	7.5	3	16	14	(Z)	1	2
Maryland	156	7.8	24	114	85	29	-	18
District of Columbia	51	10.2	13	29	29	-	-	9
Virginia	211	8.3	14	168	108	51	9	29
West Virginia	46	4.4	4	32	22	4	5	10
North Carolina	112	4.1	8	80	63	-	17	24
South Carolina	88	6.5	4	75	38	16	21	10
Georgia	168	6.4	9	128	79	50	-	31
Florida	497	14.3	132	341	303	38	-	24
East South Central								
Kentucky	118	6.0	5	97	49	26	22	16
Tennessee	92	4.1	5	70	59	12	-	17
Alabama	130	6.7	4	107	50	25	32	19
Mississippi	151	12.2	3	138	63	39	36	10
West South Central								
Arkansas	70	6.3	2	58	36	17	5	10
Louisiana	136	7.1	13	106	47	30	28	17
Oklahoma	58	3.9	6	40	26	9	5	12
Texas	514	8.7	185	291	159	132	-	39
Mountain								
Montana	26	6.5	4	19	16	1	2	2
Idaho	18	4.6	3	13	10	1	2	1
Wyoming	19	9.7	2	15	13	1	(Z)	2
Colorado	106	9.5	16	80	66	12	2	10
New Mexico	63	12.2	12	48	43	4	2	3
Arizona	119	13.8	33	80	76	1	4	5
Utah	42	8.2	9	30	20	5	5	3
Nevada	17	6.8	5	10	10	(Z)	(Z)	1
Pacific								
Washington	155	8.8	39	102	78	15	9	14
Oregon	48	4.2	16	23	23	-	-	9
California	1,340	12.3	575	669	468	101	99	96
Alaska	24	17.4	2	22	21	-	1	(Z)
Hawaii	68	16.9	36	30	24	-	7	2

Z = Less than 500.

Table 3.--ESTIMATES OF PRESUMED NONVOTERS, BY STATE: NOVEMBER 1, 1964

(Numbers in thousands)

Region, Division, and State	Total presumed nonvoters	Percent of voting-age population	Armed Forces	Homes for aged and needy	College students not living at home	Patients in short-term hospitals	Persons in travel status
UNITED STATES, TOTAL	5,733	5.0	1,611	508	1,560	460	1,594
Northeast	1,186	4.1	162	152	344	118	409
North Central	1,372	4.3	172	187	443	128	442
South	1,964	5.7	785	94	464	139	482
West	1,211	6.5	491	76	309	75	260
New England							
Maine	34	5.9	14	3	6	2	8
New Hampshire	22	5.4	5	4	5	2	6
Vermont	10	4.3	(Z)	2	4	1	3
Massachusetts	172	5.3	30	26	57	13	46
Rhode Island	36	6.5	16	4	6	2	8
Connecticut	74	4.3	10	10	22	7	24
Middle Atlantic							
New York	422	3.7	33	55	130	46	158
New Jersey	165	4.0	39	14	38	17	58
Pennsylvania	251	3.5	14	35	75	29	99
East North Central							
Ohio	226	3.8	16	31	72	24	84
Indiana	117	4.1	6	15	45	11	40
Illinois	266	4.2	38	33	79	26	89
Michigan	194	4.1	18	20	72	19	65
Wisconsin	95	4.0	4	16	32	10	33
West North Central							
Minnesota	90	4.4	4	17	32	8	28
Iowa	74	4.6	1	18	26	7	23
Missouri	125	4.6	25	16	34	11	38
North Dakota	24	6.7	8	2	7	1	5
South Dakota	22	5.5	5	2	7	2	6
Nebraska	52	5.9	15	7	13	4	12
Kansas	87	6.6	30	10	23	5	19
South Atlantic							
Delaware	16	5.6	8	1	3	1	4
Maryland	115	5.7	44	7	27	8	28
District of Columbia	41	8.1	11	4	17	2	7
Virginia	203	8.0	122	8	28	10	36
West Virginia	33	3.1	(Z)	3	10	4	15
North Carolina	162	5.9	73	7	32	11	39
South Carolina	88	6.5	48	2	14	5	19
Georgia	206	7.8	100	6	52	11	37
Florida	176	5.1	73	11	29	14	49
East South Central							
Kentucky	127	6.4	45	5	41	8	28
Tennessee	94	4.2	23	5	25	9	31
Alabama	77	4.0	18	3	22	8	27
Mississippi	58	4.7	17	3	16	5	17
West South Central							
Arkansas	48	4.2	13	3	12	5	16
Louisiana	90	4.7	28	3	24	8	27
Oklahoma	91	6.1	29	8	27	6	21
Texas	340	5.8	134	15	85	24	83

(continued)

Table 3 (continued)

Region, Division, and State	Total presumed nonvoters	Percent of voting-age population	Armed Forces	Homes for aged and needy	College students not living at home	Patients in short-term hospitals	Persons in travel status
Mountain							
Montana	24	6.0	8	2	7	2	6
Idaho	20	5.1	5	1	6	2	5
Wyoming	11	5.8	4	1	3	1	3
Colorado	82	7.3	31	5	26	5	16
New Mexico	36	7.0	17	1	9	2	7
Arizona	46	5.4	16	2	13	3	12
Utah	31	6.2	3	2	17	2	7
Nevada	14	5.4	7	(Z)	2	1	3
Pacific							
Washington	125	7.1	47	15	31	7	25
Oregon	52	4.6	5	8	19	5	16
California	655	6.0	251	38	170	44	153
Alaska	38	27.2	33	(Z)	2	1	2
Hawaii	77	19.1	64	1	4	2	6

Z = Less than 500.

Table 4.--PROPORTION OF NONVOTERS BY CATEGORY, BY STATE: NOVEMBER 1, 1964

Region, Division, and State	Total nonvoters (thousands)	Percent of nonvoters		
		Ineligibles or presumed nonvoters	Ineligibles	Presumed nonvoters
UNITED STATES, TOTAL	43,152	32.5	19.2	13.3
Northeast	9,605	33.7	21.3	12.3
North Central	9,379	31.4	16.7	14.6
South	17,810	25.7	14.7	11.0
West	6,357	51.2	32.2	19.0
New England				
Maine	199	36.2	18.9	17.3
New Hampshire	110	49.3	29.6	19.7
Vermont	69	43.9	29.4	14.6
Massachusetts	922	53.1	34.5	18.6
Rhode Island	157	55.1	32.4	22.8
Connecticut	489	37.2	22.1	15.1
Middle Atlantic				
New York	4,113	32.8	22.5	10.3
New Jersey	1,283	31.5	18.6	12.9
Pennsylvania	2,262	25.1	14.0	11.1
East North Central				
Ohio	2,009	27.0	15.7	11.3
Indiana	740	30.5	14.7	15.8
Illinois	1,680	41.4	25.5	15.9
Michigan	1,470	27.7	14.5	13.2
Wisconsin	698	27.8	14.1	13.6

(continued)

Table 4 (continued)

Region, Division, and State	Total nonvoters (thousands)	Percent of nonvoters: Ineligibles or presumed nonvoters	Ineligibles	Presumed nonvoters
West North Central				
Minnesota	467	34.2	14.9	19.3
Iowa	452	26.9	10.4	16.5
Missouri	911	31.2	17.5	13.7
North Dakota	102	42.7	19.0	23.7
South Dakota	101	43.5	22.0	21.5
Nebraska	285	28.5	10.4	18.1
Kansas	466	30.7	12.0	18.7
South Atlantic				
Delaware	85	43.9	25.1	18.9
Maryland	887	30.5	17.5	13.0
District of Columbia	306	30.1	16.8	13.3
Virginia	1,496	27.7	14.1	13.6
West Virginia	272	29.3	17.0	12.3
North Carolina	1,326	20.6	8.4	12.2
South Carolina	833	21.2	10.6	10.6
Georgia	1,497	25.0	11.2	13.7
Florida	1,622	41.5	30.6	10.8
East South Central				
Kentucky	931	26.3	12.7	13.6
Tennessee	1,091	17.0	8.5	8.6
Alabama	1,233	16.8	10.5	6.3
Mississippi	821	25.4	18.3	7.1
West South Central				
Arkansas	563	21.0	12.5	8.5
Louisiana	1,005	22.4	13.5	8.9
Oklahoma	555	26.8	10.5	16.3
Texas	3,287	26.0	15.6	10.4
Mountain				
Montana	117	42.4	22.0	20.4
Idaho	90	41.3	19.5	21.8
Wyoming	49	60.5	37.8	22.7
Colorado	345	54.7	30.9	23.8
New Mexico	188	52.7	33.6	19.1
Arizona	381	43.3	31.2	12.1
Utah	108	68.3	39.0	29.3
Nevada	114	26.8	15.0	11.9
Pacific				
Washington	493	56.7	31.4	25.3
Oregon	347	26.8	13.8	14.9
California	3,858	51.7	34.7	17.0
Alaska	72	86.3	33.6	52.7
Hawaii	195	74.4	34.9	39.5

SOURCES AND METHODS USED IN DEVELOPING DATA PRESENTED IN TABLES 1 TO 4

Votes Cast and Population of Voting Age

The population of voting age was obtained from Current Population Reports, Series P-25, No. 342. The number of votes cast for president was reported in Governmental Affairs Institute, America At The Polls, University of Pittsburgh Press, 1965.

Population "Ineligible" to Vote

Aliens.--The number of aliens who reported under the Alien Address Program is published by State of residence for 1964 in the Annual Report of the Immigration and Naturalization Service, 1965, Table 36. These data are not broken down by age. From the 1960 Census tables on foreign born population living in the United States in 1960 but abroad in 1955, it is estimated that 67 percent of this population was of voting age. Since the aliens registered have entered this country over a much broader period of time, it is to be expected that the proportion of voting age would be larger than indicated for the 5-year period. Accordingly, we have assumed that 75 percent of aliens reporting were of voting age in each State.

Movers failing residence requirements.--The number of interstate, intercounty and intracounty movers for the year ending April 1964 as reported from the Current Population Survey in Current Population Reports, Series P-20, No. 141, was used as a base for estimating movers who failed to meet residence requirements. These numbers relate to the civilian resident population plus those Armed Forces living with their families either on-post or off-post (approximately one-half of the resident Armed Forces). Each of the U. S. control totals was distributed by State by the pattern of interstate, intercounty, and intracounty movement, respectively, shown for the period 1955 to 1960 in table 100 of each State volume of the 1960 Census of Population. To these resultant estimates of movers were applied length of residence factors derived for each State from the World Almanac and Book of Facts, 1966, page 113. Factors for State residence requirements (e.g., .5 if six months residence is required) were applied to interstate movers, county residence requirements factors to intercounty movers, and precinct residence requirement factors to intracounty movers. It was arbitrarily assumed that one-half of all intracounty movers would have crossed precinct boundaries.

Whenever there was a precinct residence requirement, but none at the county level, precinct residence factors were applied to all intercounty movers as well as to one-half the intracounty movers.

Inmates of correctional and mental institutions.--The number of inmates of correctional and mental institutions in 1960 is given in tables 34 and 35, respectively, of the 1960 Census of Population, Inmates of Institutions, Series PC(2)-8A. These numbers were adjusted to a 1964 level by allowing for U. S. population increase since 1960.

Presumed Nonvoters

Armed Forces.--Armed Forces station strength for November 1, 1964, was obtained from Current Population Reports, Series P-25, No. 342.

Homes for the aged.--Inmates of homes for the aged and needy were obtained by State from table 37 of the 1960 Census report, Inmates of Institutions, adjusted to a 1964 level by allowing for U. S. population increase since 1960.

College students not living at home.--The total number of college students of voting age in 1960 are shown in table 101 of each State volume of the 1960 Census of Population. A ratio of the students of voting age to all students for each State was applied to the State distribution of college students not living at home contained in table 19 of the 1960 Census Report PC(2)-2B, Mobility for States and State Economic Areas. This provided an estimate of the 1960 distribution of college students of voting age not living at home. This distribution was adjusted to a 1964 level by using a ratio of total college fall enrollment, 1964 to 1959, as reported in the U. S. Office of Education's annual publication Opening (Fall) Enrollment in Higher Education.

Hospital patients.--Estimates of short-term patients in hospitals were developed by using the average daily census in short-term hospitals for 1964, as reported in the American Medical Association's journal, Hospitals: Guide Issue, 1965, Part II, page 450. This total was distributed by age on the basis of National Health Survey data on discharges from short-term hospitals during Fiscal Year 1964, as published in NCHS Series 10, No. 30, Vital and Health Statistics Reports: Data From the National Health Survey, June 1966. This control total was distributed by State on the basis of the 1964 population of voting age.

Persons in travel status.--A national estimate of the number of travelers on election day was developed on the basis of data available on hotel occupancy. From the Harris, Kerr, Forster & Company report, Trends in the Hotel-Motel Business, 1965, the average number of rooms occupied per day in 1964 was estimated from 1965 and 1963 data. This number was adjusted for number of guests per occupied rooms and for seasonal variation, and further adjusted to exclude residential hotels, making use of data published in pages 1 through 11 of this report. It was arbitrarily assumed that 10 percent of the persons occupying these rooms were under voting age. This control total was distributed by State on the basis of the 1964 population of voting age.

DISCUSSION

Richard W. Dodge - Bureau of the Census

These papers illustrate two approaches to the study and measurement of voting behavior. On the one hand, Professor Stokes has described the series of sample surveys on voting conducted by the Survey Research Center of the University of Michigan beginning with the 1948 election.

Meyer Zitter and Donald Starsinic, on the other hand, have a different purpose in view - that is, to provide state by state estimates of that portion of the nonvoting population who are unable to vote for legal and administrative reasons.

The Survey Research Center has emphasized in its nationwide surveys the psychological variables which explain political behavior, rather than such descriptive characteristics as age, income, and education. Although some would argue that there has been an undue emphasis on the psychological components in the Center's studies of political behavior, the insights offered, for example, on party identification, have enriched our understanding of American politics.

As Professor Stokes concedes however, students of politics also depend on the accurate and complete collection of basic election statistics by states and local governments. It is especially for persons who utilize these figures that the Zitter-Starsinic paper will be of interest.

The phenomenon of nonvoting has become a commonplace of American political commentary. The voting turnout in the United States is usually compared with the higher rates of other democratic nations. However, the would-be voter in this country must overcome many more administrative hurdles than do voters abroad. Principally, he must register, usually in person, before he can vote for the first time in his election district. He must also meet residence requirements which disfranchise many in a highly mobile society. Absentee voting laws in many states are unnecessarily restrictive. In spite of these and other obstacles, voting turnout in most nonsouthern states in Presidential years compares favorably with the performance of West European nations. It must be conceded, however, that this figure sags appreciably for mid-term elections.

Zitter and Starsinic have devised a way of estimating the approximately one-third of the resident population who have "legitimate" reasons for nonvoting. Subtracting these two groups from the voting age population, they arrive at a more realistic estimate of voting turnout. In general, I would agree with these approximations with a few reservations.

Census population estimates are based on place of residence. For members of the Armed Forces, this means where they are stationed rather than their state of voting residence. Although no reliable figures exist on the latter, it seems probable that subtracting these presumed nonvoters from the total of persons eligible by age may result in misleading figures if the majority of Armed Forces personnel register and vote in their home states.

Persons temporarily confined to their homes by illness are involuntary nonvoters by definition. This is another gray area where a good estimate is hard to come by, although Rossiter estimated that five million persons were unable to vote in the 1956 election because of illness.1/

The authors make a valid point concerning the availability of registration statistics. Although a few states and parts of states do not require formal registration, much would be gained by encouraging states to establish procedures for uniform collection and publication of these data. Sample surveys offer another means, seldom exploited, of gaining further enlightenment on the characteristics of the registered and the nonregistered.

I should like to conclude by describing briefly another study on the measurement of voting behavior in which the chairman and I participated. The Bureau of the Census became involved in this subject during the election of 1964 under a provision of Title VIII of the Civil Rights Act of 1964 which authorizes the Secretary of Commerce to "conduct a survey to compile registration and voting statistics in such geographic areas as may be recommended by the Commission on Civil Rights." In order to gain experience in this field, a single question was added to the November Current Population Survey to ascertain the number of age eligible persons who voted for President. The results, weighted by the usual CPS weighting procedures and adjusted to independent estimates of the civilian noninstitutional population, produced approximately six million more voters than the official figures, or a difference of seven percentage points.

We ascribed this discrepancy to four general factors without attempting to allocate the proportional share of each factor.2/ First, there is a tendency for respondents to overreport their voting, possibly increased by Census use of a household respondent to report for other family members.

1/ Clinton Rossiter, Parties and Politics in America, Ithaca, N. Y. 1960, p.31.
2/ For a more detailed discussion, see U. S. Bureau of the Census, Current Population Reports, Series P-20, No. 143, "Voter Participation in the National Election: November 1964," U. S. Government Printing Office, Wash. D.C. 1965, pp 4-5.

Second, the attribution of the characteristics of persons in interviewed households to households where no interview was obtained rests on the unlikely assumption that difficult to locate persons vote as frequently as the rest of the population.

Third, the Current Population Survey sample is known to be unsatisfactory in its coverage of certain groups such as nonwhite males age 21-24 who have a high level of nonvoting. In addition, the independent estimates of the voting age population are based on the 1960 Census and reflect undercounts known to exist in certain age groups.

Finally, there are two aspects of the election itself that have a bearing on the estimates - the invalidation of ballots and the fact that more people vote in an election than vote for any specific office, including that of President.

The data collected by the Bureau of the Census support previous findings in that nonvoters tend to be female, nonwhite, rural, of low income and education, the young and the elderly. The size of the Current Population Survey Sample (approximately 32,000 interviewed households) permitted an extensive cross-classification of variables, while the fact that the sample was based on households made possible an analysis of family voting. It has been proposed that the Census Bureau undertake another voting survey this fall (1966). Additional questions have been suggested on registration and reasons for nonregistration.3/

3/ Since the August meeting, Budget Bureau approval has been received for voting surveys in conjunction with the 1966 election. In addition to questions on voting and registration on the Current Population Survey, there will be two methodological tests - one obtaining information directly from the voter, rather than a household respondent, while the other will experiment with collecting voting and registration information by mail using two versions of question wording.

XV

CONTRIBUTED PAPERS I

Chairman, JOHN W. GRACZA, U. S. Bureau of Labor Statistics

A TEST FOR THE EFFECT OF TREATMENT OR TIME ON A BINOMIAL POPULATION

Haskel Benishay, State University of New York at Buffalo

I. Introduction

The influence exerted on the parameter of a binomial population by a change in underlying conditions during a given time interval is a common topic of research in economics, marketing, finance, opinion polling, medicine, and other fields. It is usually investigated either through a comparison of the means of two different samples, one taken at the beginning and another at the end of the period under consideration, or through the observation of the changes occurring during the period in the binomial attributes of one sample. In the latter approach, the individuals before and the individuals after the test period are the same and the binomial sample is a "matched" one, the "match" being in the sameness of the individuals over time. This paper provides a critical evaluation of the matched sample test currently available for the determination of the statistical significance of an influence exerted on a binomial population and develops an alternative test for the same end. The two tests are discussed and evaluated within the context of two probability models which are believed to cover a wide range of applications and of two major conditions of *a priori* information. The distinguishing characteristic of the two conditions is the existence or absence of *a priori* information about the extent of the changes which are expected to occur over time in the dichotomous characteristics of individuals even when the binomial parameter remains constant over time. It is suggested that the alternative method is more appropriate for the problem at hand for both probability models and for the case of *a priori* knowledge about change. In the absence of such knowledge a procedure based on the notions of the alternative test is proposed for situations described by both models.

II. The Problem

A certain influence or "treatment", like propaganda in a political setting, TV message in a drive to increase consumer demand, the administration of a drug in a medical experiment, or simply the passage of time, is presumed to be capable of increasing the number of "successes" in a binomial population. The presumed effect of such treatment is to be tested statistically on a matched random sample of n observations. Measurements on the binomial characteristics of the sample observations before and after the treatment are to be taken and evaluated in a test of significance.

In the specific context where the observations are individuals and the binomial characteristic is an answer of either yes or no to a particular question, the following procedure would be followed. Each individual is asked the question before and after an exposure to a treatment or an influence. The responses are recorded in four categories: (1) affirmative on both occasions; (2) negative on both occasions; (3) first "yes" and then "no"; (4) first "no" and then "yes". The problem is to determine by means of a test of significance whether attitudes have changed as a result of an intervening influence.

The customary test for this matched sample design compared the number of changes from one to zero against the number of changes from zero to one under the null hypothesis that the number of changes in each direction is one half of the total number of changes. The reasoning underlying this approach can be summarized approximately as follows: If there is no treatment effect in either the direction of increasing, or in the direction of decreasing, the number of ones, we would expect to observe only changes due to "chance". These "chance switches," negative changes from one to zero and positive changes from zero to one, would occur even in the complete absence of a treatment effect. Under the null hypothesis that the treatment exerts no effect at all and that the binomial population retains the same parameter in the absence of a treatment effect, the expectation is that the switches during the interval of time under consideration should be equally numerous in either direction. Or, put differently, the sum of the changes, both positive and negative, should be expected to equal zero.

Whatever it may have in its favor, the customary test is inappropriate for the problem at hand because it is a conditional test which does not take account of the total phenomenon under consideration. The conditional framework of the test is unsuitable because it tests the null hypotheses for a given number of total non-zero changes (sum of positive and negative changes) for any sample size. The conclusions of the test hold for the particular number of non-zero changes which occur not only in samples whose size equals that of the sample used in the test, but also in samples of any other size. As long as two samples of unequal size have the same number of non-zero changes, they are considered, in the framework of the traditional test, as equivalent preceding conditions. The test is defective because it neglects to incorporate an important and substantial amount of the total process under consideration. It does not at all take account of the non-changing individual in the process and thereby neglects altogether to consider the stable part of the process. It is incapable of providing a comprehensive answer to the truly relevant question: What is the statistical significance of a particular sample difference between the positive and negative changes for a given sample size, i.e., for the process as a whole including all the non-zero and zero changes?

The alternative test proposed and developed in this paper attempts to supplement and remedy the defects of the current test. As it turns out, the alternative test produces substantially different answers. For a given sample size it appears to reject the null hypothesis that no net change has occurred more often than the customary test.

In order to evaluate the traditional test and the proposed alternative in the context of clearly defined underlying processes the binomial process under consideration is described below in two alternative ways. Constructed are two probability models which purport to consist of the essential and relevant elements of the process as these elements appear in most empirical settings of actual problems. These models state explicitly the assumptions underlying the null hypothesis, assumptions which are made implicitly about the binomial process in the usual experimental context. Two sets of underlying assumptions characterize the null hypothesis:

1. The parameter of the binomial population remains stable over time. All zeros remain zeros and all ones remain ones. "Chance" in the form of observational error brings about not true but observed change in particular observations.

2. The parameter of the binomial population remains stable over time, although, due to "chance" conceived as the unsystematic effect of many variables, some ones become zeros and some zeros become ones.

III. Model I

Model of Errors in Measurement or Mirage Switches

In this model no individual in the binomial population really changes over time. Some individuals, however, appear to change because of the mirage effect of observational error.

Let P represent the probability of the existence of some dichotomous attribute (success, yes). Let the existence of this attribute be designated as usual by one. Let (1-P), to be referred to also as Q, represent the probability of its absence (failure, no) and let its absence be designated by zero.

The individuals in the population are assumed to be observed with an error. Specifically, some individuals with true values of one are observed as zeros. Some individuals with true values of zero are observed as ones. Let B represent the probability of observing or measuring an individual with an error and let the probability of observing its true value correctly be $G=1-B$.

The previous assumption provides a distinction between the true, one, zero, values and the observed values of individuals in a binomial population. It introduces, in principle and on the level of the population as a whole, the existence of observation or measurement errors. It creates a link between the need to deal with observed values subject to error and the fact that most discussions of problems associated with binomial populations omit the consideration of inevitable errors in the measurement of values.

Let t be a point in time and let t+z be a point in time some time later where z is the length of the period under consideration. Let the probability of a true one at t be P_t, at t+z be P_{t+z}, and let $P=P_t=P_{t+z}$. Let the probability of observing an individual at time t, with an error, be represented by B_t and at t+z, by B_{t+z} and let $B=B_t=B_{t+z}$.

Let the probability of observing a change of minus one, -1, i.e., a switch from one to zero, between t and t+z be P_{10}. In terms of previously defined probabilities P_{10} turns out to be:

(1.01)

$$PGB + QBG = BG$$

The term PGB represents the probability of being actually one and being observed correctly as one at t and being mistaken for zero at t+z. The term QBG represents the probability of being really zero and being observed erroneously as one at t and being observed correctly as zero at t+z.

Let the probability of observing a change of plus one, +1, i.e., a switch from zero to one between t and t+z, be P_{01}. In terms of previously defined probabilities P_{01} emerges as:

(1.02)

$$QGB + PBG = BG$$

The first term above, QGB, represents the probability of a true zero, first observed correctly, then observed erroneously; and the second term PBG is the probability of a true value of one observed first incorrectly, later with no error.

With similar reasoning P_{11}, the probability of observing a value of one both at t and t+z, i.e., a change of zero is:

(1.03)

$$PG^2 + QB^2$$

and P_{00}, the probability of observing a value of zero both at t and t+z, i.e., again, a change of zero, is:

(1.04)

$$QG^2 + PB^2$$

As becomes evident upon inspection, the probability BG is the variance of the error distribution. Thus the variance of the error distribution becomes the probability of observing a change of a plus one +1, between t and t+z, as well as the probability of observing a change of minus one, -1, between the same two points in time. Also, 2BG represents the probability of observing a change, i.e., negative and positive unity changes and (1-2BG) represents the probability of observing stability or a change of zero. The probability (1-2BG) is, of course, equal to the sum of (1.03) and (1.04) above.

To summarize, the observed change, a random variable henceforth referred to as D_1, can take on values -1, 0 and +1, with associated probabilities BG, (1-2BG) and BG respectively. And more generally, if BG is redefined as K, then associated probabilities of -1, 0 and +1, are K, 1-2K and K respectively. In a typical situation the probability of error is small and the bulk of the probability of the distribution of the changes is centered at zero as is shown in Figure I. The expectation of this distribution of changes, henceforth referred to as $E(D_1)$ will equal zero as can be easily demonstrated by a straightforward use of the definition of expectation:

(1.05)

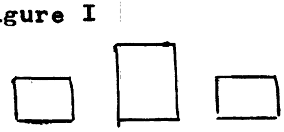
$$E(D_1)=(-1)BG+(0)(1-2BG)+(1)BG=0$$

Obtained in a similar fashion is the variance of the distribution of observed changes. The values of D_1, and their associated probabilities are employed in the definition of variance to obtain:

(1.06)

$$V(D_1) = (-1)^2BG + (1)^2BG = 2BG$$

Figure I



probabilities	BG	1-2BG	BG
probabilities	K	1-2K	K
D_1 values	-1	0	+1

Under the assumptions of Model I and particularly under the assumption that no change occurs in P between t and t+z, the expectation, E, of the distribution of the <u>sample sum</u>, S, of n independent observations randomly drawn from the distribution of changes is equal to zero since:

(1.07)

$$E(S) = \sum_{i=1}^{n} d_{1i} = nE(D_1) = 0.$$

The variance of the distribution of the sum, V(S), is equal to n times $V(D_1)$ which in turn equals n2BG.

(1.08)

$$V(S) = nV(D_1) = n2BG$$

Thus, in a Model I world, the null hypothesis implies an expectation of zero, variance of 2nBG or 2nK for the sample sum.

Under the alternative hypotheses associated with Model I, the parameter P can go either up or down in the following two clearly defined ways: (1) some true <u>zeros</u> becomes true <u>ones</u> and no change occurs in any true <u>ones</u>; or (2) some true <u>ones</u> become true <u>zeros</u> and no change occurs in the true <u>zeros</u>. A one right [left] tail test would imply for the alternative hypothesis the first way, (1), but not the second, (2) [the second way (2), but not the first, (1)]. A two tail test would make the alternative hypothesis either the first way, (1), or the second way, (2), of chaging P.

IV. Model II
Real Switches Model: No Error

In this model the parameter of the binomial population remains stable over time, but some true ones become true zeros and some true zeros become true ones due to a variety of "chance" influences not related to the particular factor whose effect is under investigation. No observational error is assumed to intervene in any way between the true values and their perception and recording.

Let P and Q be defined as they have been in Model I, such that P_t equals P_{t+z}. Let C_{01} be the probability that a true zero becomes a true one, C_{10} the probability that a true one becomes a true zero between t and t+z.

Given these assumptions, P at t+z can be viewed as:

(2.02)

$$P_{t+z} = P_t(1-C_{10}) + Q_tC_{01}$$

Since $P_{t+z} = P_t = P$,

(2.03)

$$P = P(1-C_{10}) + QC_{01}.$$

Expanded (2.03) indicates the nature of the addition to and subtraction from P.

(2.04)

$$P = P - PC_{10} + QC_{01}.$$

It becomes evident that as a consequence of preceding assumptions it is definitely, although implicitly, assumed that:

(2.05)

$$PC_{10} = QC_{01}$$

which in turn implies that:

(2.06)

$$P/Q = C_{01}/C_{10}$$

One more assumption is required to render Model II internally consistent. The probability of a switch away from an individual's own value must necessarily be constrained as follows:

(2.07)

$$C_{10} \overline{\overline{<}} Q/P$$

for $P \overline{\overline{>}} 1/2$. And similarly:

(2.08)

$$C_{01} \overline{\overline{<}} P/Q$$

for $Q \overline{\overline{>}} 1/2$. Without these constraints, it is impossible to maintain that the probability of positive changes is equal to the probability of negative changes without colliding with the model's logic as it now stands. The constraint simply acknowledges the fact that when more than half of the individuals are ones, it cannot be assumed that all ones become zeros, $C_{10} = 1$, and maintain the equality between negative and positive changes even if C_{01} is also assumed to equal unity.

The probabilities of observing

a switch are as follows:

from <u>one</u> to <u>zero</u>:

(2.09)

$$PC_{10}$$

from <u>zero</u> to <u>one</u>:

(2.10)

$$QC_{01}$$

The probability of remaining a <u>one</u> is:

(2.11)

$$P(1-C_{10})$$

and the probability of remaining a <u>zero</u> is:

(2.12)

$$Q(1-C_{01}).$$

Again, under the hypothesis that the treatment under consideration exerts no effect, we obtain under Model II a distribution of changes, D_2, whose values are -1, 0 and +1, and whose probabilities are PC_{10} [as in (2.09)], $1-PC_{10} - QC_{01}$, [the sum of (2.11) and (2.12)] and QC_{01} [as in (2.10)], respectively. As stated for D_1 in Model I (outlined above) the probabilities of the change, D_2, can be stated more generally as K for -1, (1-2K) for 0, and K for +1, in which case Figure I above may serve as a good description of the distribution of changes predicated on the assumptions of Model II. The null hypothesis under Model II for the sample sum, S, can be stated as it was under Model I: The expectation of the distribution of the sum statistic, S, is equal to zero. In a manner analogous to Model I, the alternative hypothesis can be: the expectation of the distribution of the sum of the changes is larger than zero (one tail), smaller than zero (one tail) and smaller or larger than zero (two tail). The difference is, of course, that any real change is a net change. It is the difference between positive true switches and negative true switches.

V. The Two Models and the Tests

It becomes evident that under the assumptions of both Model I and Model II we obtained distributions with common characteristics of observed changes. Under both models the possible values of the changes are -1, 0 and +1, and their associated probabilities can be expressed as K, (1-2K) and K respectively. We shall refer to this distribution generally as D. The difference between the models rests in the following: Under Model I, 2K must necessarily be equal to or smaller than 1/2 since by definition 2K equals 2BG and BG can be at most 1/4. In Model II, 2K is constrained in a different fashion. It can be as large as, or smaller than, twice the probability of <u>one</u> or <u>zero</u>, whichever is smaller. In most actual cases, however, and under both models, 2K is likely to be well below one-half (1/2) and probably closer to zero than to one-half.

The test of significance proposed in this paper as an alternative to the customary test, for situations depicted by either Model I or Model II, is based on a distribution of a statistic computed from a sample from the distribution of changes $D_1=D_2=D$. The statistic is the sum of n sample changes, d, "drawn" at random:

(3.01)

$$\sum_{i=1}^{n} d_i.$$

The expectation and variance of this statistic has been previously described as zero and $n2K$ respectively.

In terms of the distribution of the sample sum of changes, as a context for the comparison of the customary test and the one proposed as an alternative, it can be said that the customary test ignores totally the zero changes in the sample as if they have never existed, and treats the non-zero changes which appear in the sample, as if they were, <u>ex post</u>, a sample drawn from a distribution with only two possible values, -1 and +1, with associated probability of 1/2 for each.[1] The test proposed in this paper is different from the traditional test in that it takes account of the total process including the zero changes. The distribution for the test statistic proposed as an alternative in this paper can be worked out exactly for the case where previous knowledge of $2K$ is available. Where such <u>a priori</u> knowledge is not available it is proposed that confidence intervals be computed for the significance level.

VI. Advance Knowledge of Intrinsic Change

Advance <u>a priori</u> knowledge of the magnitude of K (or $2K$) in a Model I world may be based simply on actual knowledge of the probability of error B. If B is known independently, K may be obtained by means of the relationship between K and B.

(3.02)

$$K = B(1-B) = BY$$

[1] Treating the non-zero d outcomes of the sample as if they came (retroactively) from a population with -1, +1 values and probabilities of 0.5, 0.5 respectively is equivalent to what is done in the traditional test where the -1 values are redefined as 0 and the +1 values redefined as one.

In a Model II world, <u>a priori</u> knowledge of the intrinsic turnover is simply a direct knowledge of $2K$ or K.

If K is <u>known</u> <u>before</u> the start of the sampling process, then the test statistic, the sample sum, S, can be evaluated against its exact probability distribution which can be derived by analytical methods as is illustrated below. For large samples, the central limit theorem can be relied upon to make the normal (Gaussian) distribution an acceptable approximation for the distribution of S in which case the approximately normal distribution under consideration will be one with expectation of zero and a variance of $n2K$.

The probability distribution of the sample sum, S, can be derived by straightforward enumeration or by a method which is conditioned by the view that for a given sample size, the sample sum, S, is an outcome of a two stage process. The cumulative probability table for S, below, was computed by the second method.

The probability distribution of S can be derived by considering a draw from S as a two stage process. (a) A draw to determine the number, C_i, of non-zero outcomes in a sample size n where C_i ranges between 0 and n, i between 1 and $n+1$. The probabilities of C_i, $P(C_i)$, i.e., the probabilities of $C_1=0$, $C_2=1, \ldots C_{n+1}=n$ are usual binomial probabilities and are determined by $2K$ and n. (b) C_i draws from a distribution of +1, -1, with associated probabilities 1/2, 1/2 to determine S for a particular C_i. To compute the probabilities of values equal to or larger than S for sample size n, for all possible values of C_i, we compute (1) the probability of values equal to or larger than S for each C_i (2) we multiply the preceding probability by the probability of C_i, $P(C_i)$, and (3) sum this product for all S, for $i=1 \ldots n+1$ (for a given S, all terms from 0 to n). The probability of C_i is given by the binomial parameter $2K$, the probability of non-zero change and by the binomial expansion. The probability of values equal to or larger than S for C_i is given by the binomial parameter 1/2 and the binomial expansion.

There are (2n)+1 possible S values ranging from -n to +n in consecutive steps of unity. Since the probability distribution of S turns out to be symmetrical about zero, cumulative probabilities are presented only for values equal to or larger than X where X is zero or <u>positive</u> values of S, i.e., for values of S ranging from zero to n. A table is developed for sample size 1 to 10 and for probability of change 2K, equal to 0.2, [i.e., K=.1, (1-2K)=.8].

Upon comparison of the traditional test probabilities with those produced by our table, it becomes evident that the two methods produce different results. For example, take a case in which the number of non-zero changes equals two, C=2, and sample size is also 2. According to the traditional test the probability that S will be equal to or larger than 2 is the product (0.5)(0.5)=0.25. This result is obtained regardless of the values of 2K. According to the method proposed in this paper and given that 2K is equal to .2, the probability for the same outcome is 0.01 (see Table 1). For C of 2, for 2K of 0.2, but for a sample size of 10, the traditional test produces again the value 0.25 for the probability that S is equal to or larger than 2, the method proposed in this paper produces a probability of 0.13524132 for the same outcome (see Table 1).

VII. No Advance Knowledge of Intrinsic Change

When no knowledge of the magnitude of 2K is available before the sample is taken, it is proposed that the 2K be estimated from the data by the ratio of the observed number of non-zero changes to total observed changes. Once this estimate is available, the usual confidence intervals with Y percent confidence are computed with an upper limit and a lower one. These confidence limits are then used as if they were two known values of 2K to compute two levels of significance for the null hypothesis, a lower and an upper limit. It can then be said with Y percent confidence that the true level of significance falls between the upper and lower significance limits Y percent of the time, given, of course, that the null hypothesis of no net change is true.

TABLE 1

Cumulative Probability Table for Number of Positive (Negative) Changes Over Time in a Sample Size of n* from a Binomial Population. 2K equals 0.2.**

Sample Size	0	1	2	3	4	5
	6	7	8	9	10	
1	.90000000	.10000000				
2	.83000000	.17000000	.01000000			
3	.78000000	.22000000	.02500000	.00100000		
4	.74350000	.25650000	.04210000	.00330000	.00010000	
5	.71624000	.28376000	.05966000	.00686000	.00041000	.00001000
6	.69540200	.30459800	.07679000	.01149500	.00101500	.00004900
	.00000100					
7	.67910240	.32089760	.09304130	.01697650	.00196640	.00014080
	.00000570	.00000010				
8	.66606755	.33393245	.10822045	.02308197	.00328485	.00030985
	.00001865	.00000065	.00000001			
9	.65542519	.34457473	.12227779	.02961610	.00496706	.00057823
	.00004597	.00000238	.00000007	.00000000		
10	.64656984	.35343008	.13524132	.03641736	.00699308	.00096388
	.00009483	.00000651	.00000029	.00000000	.00000000	

* Sample size from 1 to 10

** The probability of positive and negative changes, 2K, is .02, The probability of +1 is 0.1. The probability of -1 is 0.1.

SEASON OF CONCEPTION: A NEGLECTED ANALYTIC VARIABLE
(A preliminary report on fetal losses)

Carl L. Erhardt, Frieda G. Nelson, and Jean Pakter, Department of Health, City of New York

With rare and limited exceptions, seasonal variation with respect to births and factors associated therewith has been examined on the basis of season of delivery. (1-9) It is our contention that seasonal analysis based on date of conception will yield more precise measurement of the effect on the fetus of time-limited exogenous events occurring during pregnancy.

<u>A Hypothetical Example</u>.

Assume an infant mortality rate of 25 per 1,000 live births. Based on the New York City total of 162,911 births in 1964, the standard error of such a rate is 0.39. To demonstrate a significant increase in the annual rate would require a level of 25.8 or an increment of 125 infant deaths. We now suppose that all babies are conceived and delivered equally throughout the year, or, on the annual total given, 12,531 each lunar month of 28 days.

Assuming again that the expected infant mortality among a group of 12,531 births is 25, the standard error of the monthly rate would be 1.39, with 95 percent confidence limits of 22.3 to 27.7. If, then, all the 125 additional deaths that are required to produce a significant increase in the <u>annual</u> rate occurred among a single monthly group, the rate for that month would obviously be affected to a highly significant degree. Actually the rate would be 35.0 compared to the 25.0 expected, or a 40 percent rise, well beyond the 27.7 upper limit of the stated confidence interval.

In this example, the increase of infant deaths is presumed to be the result of a single exogenous event (an epidemic, for example) that lasted four weeks. Its effect is exerted only on those fetuses that reached a gestational age of high susceptibility during the period when this exogenous event was operative. Therefore, the only fetuses that could be affected are those conceived during a specific lunar month. Thus, if we identify the live births deriving from that cohort of conceptions, the excess mortality is easily assessible.

But what happens if we look for this excess among groups of <u>delivered</u> infants? Not more than 46 percent of the affected fetuses will be delivered alive within the same lunar month. The remainder will be delivered over as many as six <u>other</u> months. Consequently, the effect of an event exerting stress on <u>one</u> cohort of conceptions will be distributed, to varying extents, among seven monthly groups of deliveries.

Table 1 has been constructed on the basis of the percentage distribution of live births in

Table 1. Distribution of live births for four-week intervals of delivery by four week intervals of conception*

Interval of delivery	All deliveries	Interval of conception								
		A	B	C	D	E	F	G	H	I
All conceptions	-	-	-	-	12,531	12,531	12,531	-	-	-
I	-	-	-	-	38	-	-	-	-	-
J	-	-	-	-	75	38	-	-	-	-
K	-	-	-	-	200	75	38	-	-	-
L	12,531	301	5338	5815	764	200	75	38	-	-
M	12,531	-	301	5338	5815	764	200	75	38	-
N	12,531	-	-	301	5338	5815	764	200	75	38
O	-	-	-	-	301	5338	5815	-	-	-
P	-	-	-	-	-	301	5338	-	-	-
Q	-	-	-	-	-	-	301	-	-	-

* Estimated from New York City 1964 data, assuming an equal distribution of totals by lunar month.

New York City in 1964. The columns portray the number of deliveries in specific lunar months that derive from conceptions in specific earlier months. The rows indicate the distribution of deliveries according to the month of their conception. Only the numbers pertinent to this illustrative example have been inserted.

The infant loss rate (whether computed vertically or horizontally) equals $\Sigma w_i r_i$, where the r_i's are gestation-specific rates and the w_i's are the proportion of the total live births delivered at that gestation. These w_i's are the same whether we deal with conception cohorts or delivery cohorts, as long as the number of conceptions and the distribution of live births resulting from those conceptions remain constant over time. Obviously, the r_i's for affected conception cohorts are uniformly increased, while for delivery groups only those r_i's referring to the affected conception cohorts contribute to an increase in the total rate.

Table 2. Estimated excess deaths* contributed to various groups of deliveries assuming the total excess of 125 deaths is limited to conceptions during a specified number of months.

Affected delivery groups	Estimated distribution of 125 excess infant deaths if increase limited to conceptions during		
	One month (D)	Two months (D,E)	Three months (D,E,F)
Total excess deaths	125.0	125.2	124.8
I	6.1	6.1	4.2
J	26.1	16.2	10.8
K	19.2	19.8	17.3
L	18.4	19.0	19.3
M	31.9	25.3	23.4
N	21.9	26.7	24.0
O	2.1	11.8	18.5
P	-	1.1	7.8
Q	-	-	0.7

* Estimated from New York City distribution of 1964 infant deaths by duration of gestation.

In Table 2 we show how 125 excess deaths in a single conception cohort (Col. D), two such cohorts (Cols. D and E) or three (Cols. D,E and F) would be distributed among subsequent groups of deliveries (Rows I through Q). The deaths are distributed by duration of gestation according to the 1964 New York City distribution. The figures for two and three cohorts assume the deleterious influence was spread over several months rather than limited to a single month.

Table 3. Infant mortality rates per 1,000 live births by month of delivery assuming an increase of 125 deaths limited to conceptions during a specified number of months.

Delivery months	Mortality rates if increase limited to		
	One month	Two months	Three months
I	25.5	25.5	25.4
J	27.1	26.3	25.9
K	26.6	26.6	26.4
L	26.5	26.5	26.6
M	27.6	27.0	26.9
N	26.7	27.2	26.9
O	25.2	26.0	26.5
P	-	25.1	25.6
Q	-	-	25.1

The resulting infant mortality rates for the delivery cohorts are shown in Table 3. It is clear from the upper confidence limit previously cited (27.7) that the infant mortality rate in not one of these delivery months reveals a rise sufficient to suggest that a time-limited exogenous event was responsible for the significant increase in the annual rate. Seasonal analysis of deliveries could therefore lead to inappropriate conclusions. On the other hand, whether the 125 excess deaths affect only one cohort of conceptions, or even if they are distributed among two or three such cohorts, the infant mortality rate among such cohorts would in each case be significantly raised to 35, 30 or 28.3 per 1,000 live births.

Source of Data.

The substantive data in this discussion derive from routine reports of live births and fetal deaths registered with the Department of Health of the City of New York. While we shall be investigating seasonality of infant losses, complications of pregnancy, congenital malformations, birth weight and duration of gestation according to season of conception, we concentrate in this paper on fetal losses.

In New York City, all fetal deaths are required to be reported regardless of period of gestation. It is recognized that underreporting exists particularly of very early losses where the woman herself may not have been sure of pregnancy and even losses of longer duration, especially where no medical attention was sought. Nevertheless, the sum of live births and fetal deaths reported represent the best estimate we have of total conceptions. From the annual patterns of both total reported conceptions and of live births, we conclude that underreporting is consistent by month over the three years covered in this paper, and that patterns based on conceptions throughout the three years are therefore factual representations of what happened even though the rates for early fetal deaths may be understated because of greater underestimation of numerators than of denominators.

It is also known that inaccuracies exist in reporting of the initial date of the last menstrual period (LMP), through carelessness, ignorance, or simple error. In a study of pregnancy losses made by the senior author some years ago, LMP dates reported by physicians at first prenatal visit were later compared with dates on corresponding certificates filed with the Department of Health. Errors in both directions were discovered but were found to be almost exactly compensating, with agreement as to date by far predominating. Hence, there is reason to believe that patterns will not be seriously disturbed by such errors. It must also be considered that we are here assuming that the date of LMP indicates the beginning of pregnancy, whereas in actuality conception usually does not occur until a couple of weeks after the LMP date. This factor has the effect of distributing the data about a half month earlier than conception actually occurs.

This study became possible because one item on the New York City certificates requests the initial date of the LMP. Its purpose on these documents is to permit calculation of the duration of gestation. On live birth certificates it is rarely unreported (1.6% of total live births in 1964) but more frequently is omitted on fetal death certificates (4.7% in 1964). Distribution by month of delivery of the cases where LMP date was not reported suggests there is no marked seasonal variation in failure to report this fact.

The month and year of LMP have been included on the statistical punch card prepared for all vital records. These cards were converted, beginning in 1960, to magnetic tapes as a space saving measure for future use even though EDP equipment had not yet been installed in the Department of Health. This study was made possible by the availability of these tapes, and this format also made possible editing that was not previously feasible. For example, the month and year of the LMP were checked against the month and year of delivery. For unreasonable combinations (such as LMP more than a year earlier than delivery or after delivery), the original records were checked and appropriate corrections made.

Fetal Losses.

One aspect of pregnancy problems is represented by casualties at various stages of pregnancy. If it could be determined that women whose pregnancies start at specific seasons were at higher risk than others of loss of the baby, special attention could be centered on women conceiving during those intervals. Moreover, the identification of unusually high losses among pregnancies starting during a specific period of time might lead to identification of environmental factors coincidental with the terms of these pregnancies that are implicated in the losses.

In the interest of time, we are not presenting the detailed tables here but this audience will want an approximation of the numbers involved. There are about 15,000 total reported conceptions monthly, more than 13,000 live births and nearly 2,000 fetal deaths. About one half of the fetal deaths are of less than 12 weeks gestation; another 30% occur at 12-19 weeks; the remainder (about 20%) is almost evenly split between 20-27 weeks and 28+, with slightly more in the latter group.

Total Fetal Mortality.

We have available to us thus far only three years of experience, 1962-1964, derived from events reported from 1962 through 1965. We have chosen to estimate seasonality by accumulating the experience of the three years. Because of aberrations in the observed rates, we are skeptical that the accumulations for this brief period provide a reliable estimate of the seasonality of fetal losses by period of gestation. Definitive patterns of seasonality will have to await the gathering of additional data. However, we do believe it is worth presenting the material analyzed thus far.

From Chart I it can be seen from the solid line based on the aggregate data of the three years that there appears to be a seasonal pattern with low fetal loss rates among conceptions occurring in the summer months and high rates among winter conceptions. However, for none of these three years individually is the aggregate typical. The pattern is followed in 1962 and in the latter half of 1964. From June of 1963 to July of 1964 the rates are uniformly and constantly higher than the monthly means for the three years.

Several possibilities exist. First, the 1962 data may well represent what ordinarily happens although it is also possible that the rates in that year are somewhat lower than might be usual. Second, there seems no doubt that fetal death rates were quite high among conceptions from approximately mid-1963 to mid-1964. Any seasonal pattern in these two years is obscured and yet, despite the variations in these two years, a seasonality appears to persist in the aggregate.

Fetal Deaths by Gestation.

It will be useful to examine separately the seasonal pattern of fetal losses by duration of gestation. Charts II to V present the data respectively for late fetal deaths (28 or more weeks gestation), for intermediate fetal deaths (20-27 weeks gestation) and for early fetal deaths, separately for those 12-19 weeks gestation and those less than 12 weeks gestation.

For each of these gestation intervals there is variation in losses by month of conception. The aggregate data for the three years suggests in each case that seasonality exists and that losses are more likely to be high among winter conceptions than for pregnancies beginning in the summer.

There are, of course, departures from this pattern for each gestation interval and it is our opinion that these departures obscure to some extent the seasonality pattern.

For example, among fetal deaths of 28 weeks or more gestation (Chart II), observed rates were unusually high among conceptions in the spring of 1963, low for those in the spring of 1962 and about average for those in the spring of 1964. These varying specific rates tend to cancel out each other and produce a fairly flat curve in the first half year in the aggregate.

Among fetal deaths at 20-27 weeks gestation (Chart III) a seasonal pattern becomes more evident. The noteworthy features here are the apparently higher than seasonal observed rates in the summer of 1963 and the unusually high rate in March of 1964.

Again, with reported fetal deaths of 12-19 weeks gestation (Chart IV) there appears to be a definite seasonal pattern. Here we seem to have relatively low rates in 1962 and peaks in July, November and December of 1963 and May of 1964.

Fetal deaths of less than 12 weeks gestation (Chart V) do not present as clear a picture. The extraordinarily high rates in early 1964 and the upward trend during 1963 are such that any seasonal pattern is obscured. The data for 1962 suggest the possibility of seasonality in early fetal losses. If this is so, then there appears to have been some factor or factors in 1963 and 1964 that disturbed this seasonality.

Comparison with Estimates.

The general theory on which we are examining variation by month of conception is that such

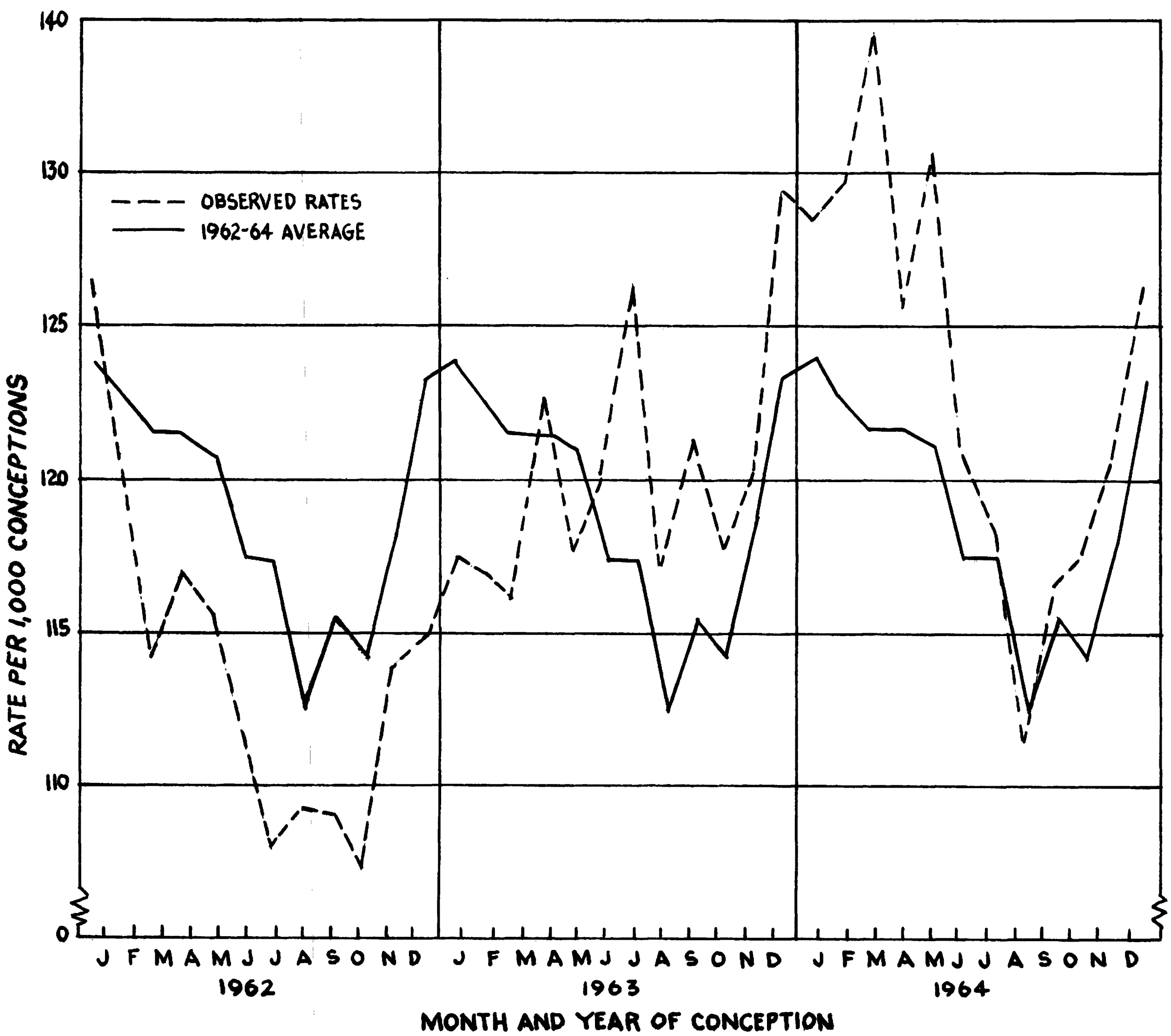


Chart I. All reported fetal deaths per 1,000 conceptions by month of conception: New York City, 1962-1964

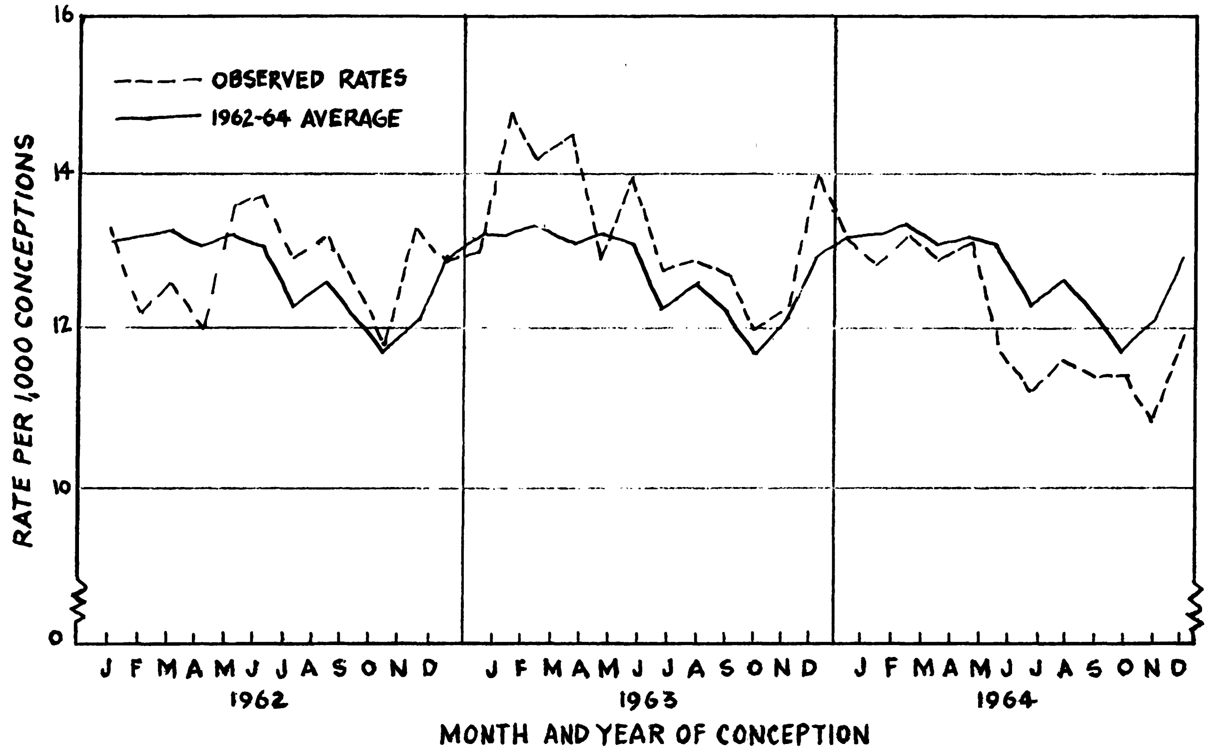


Chart II. Reported fetal deaths 28 and more weeks gestation per 1,000 conceptions by month of conception: New York City, 1962-1964

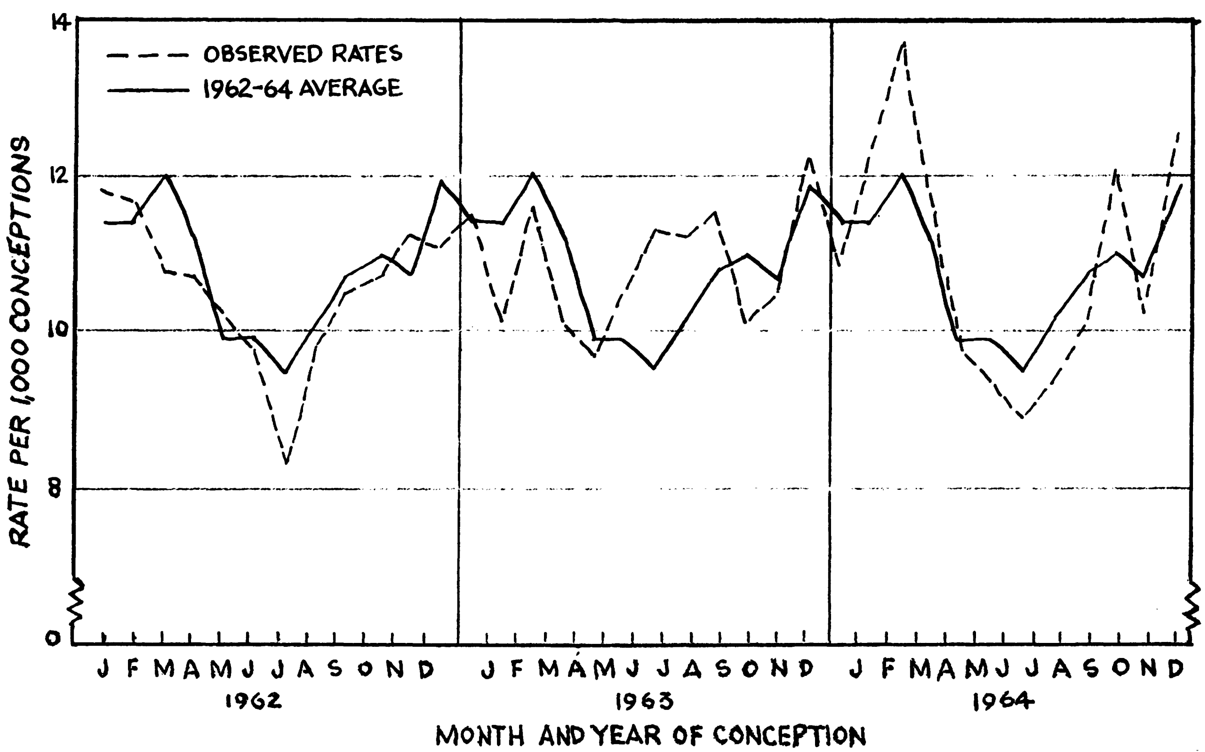


Chart III. Reported fetal deaths 20-27 weeks gestation per 1,000 conceptions by month of conception: New York City, 1962-1964

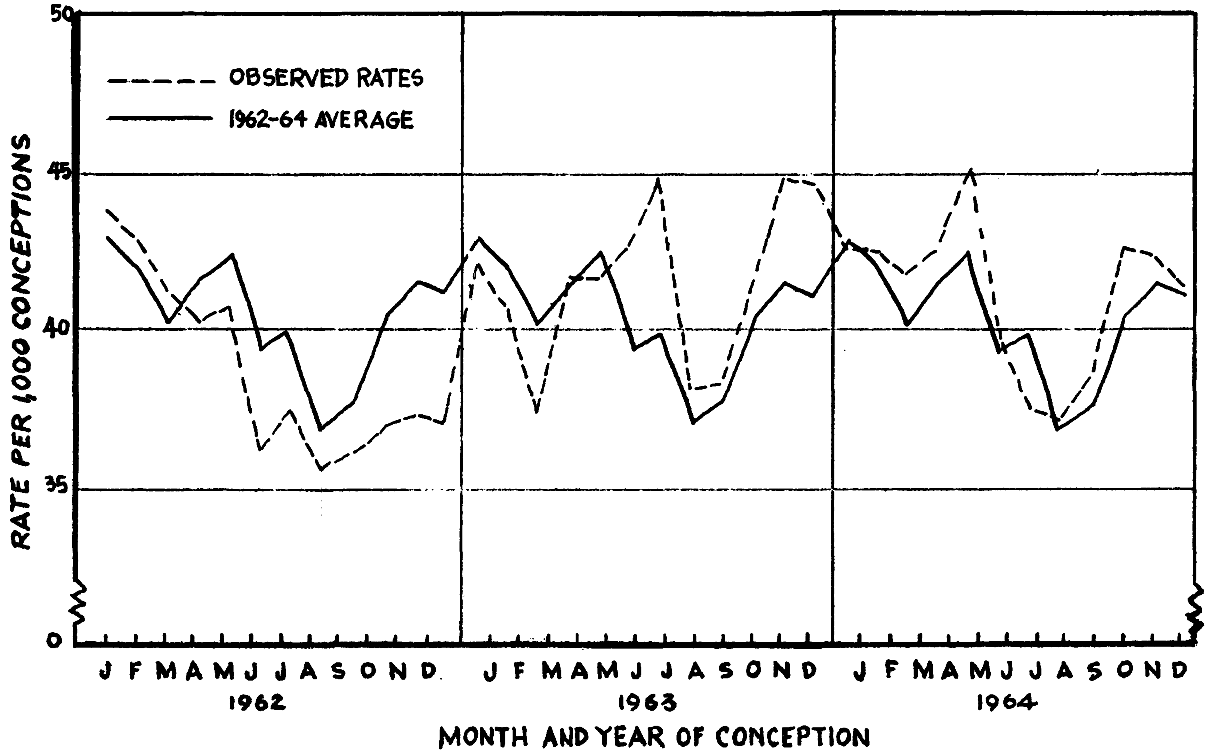


Chart IV. Reported fetal deaths 12-19 weeks gestation per 1,000 conceptions by month of conception: New York City, 1962-1964

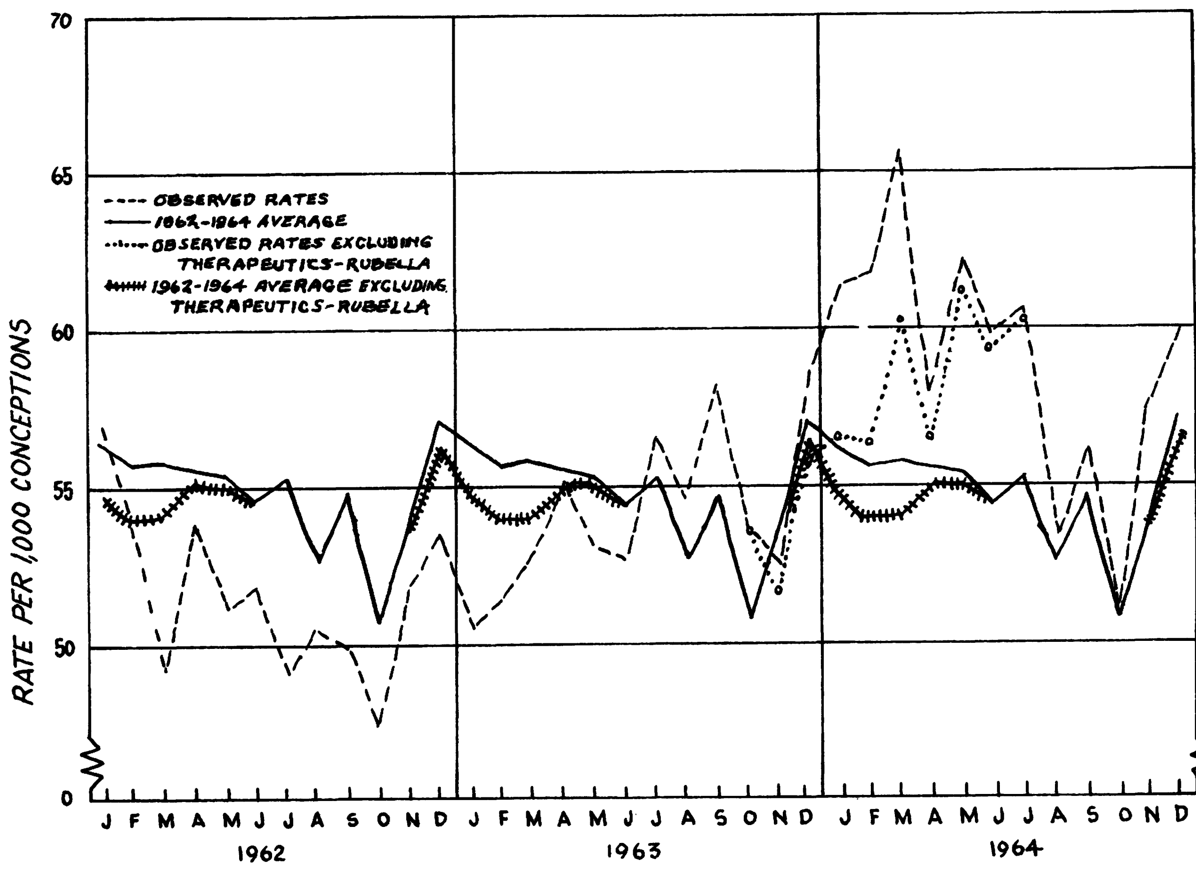


Chart V. Reported fetal deaths less than 12 weeks gestation per 1,000 conceptions by month of conception: New York City, 1962 - 1964

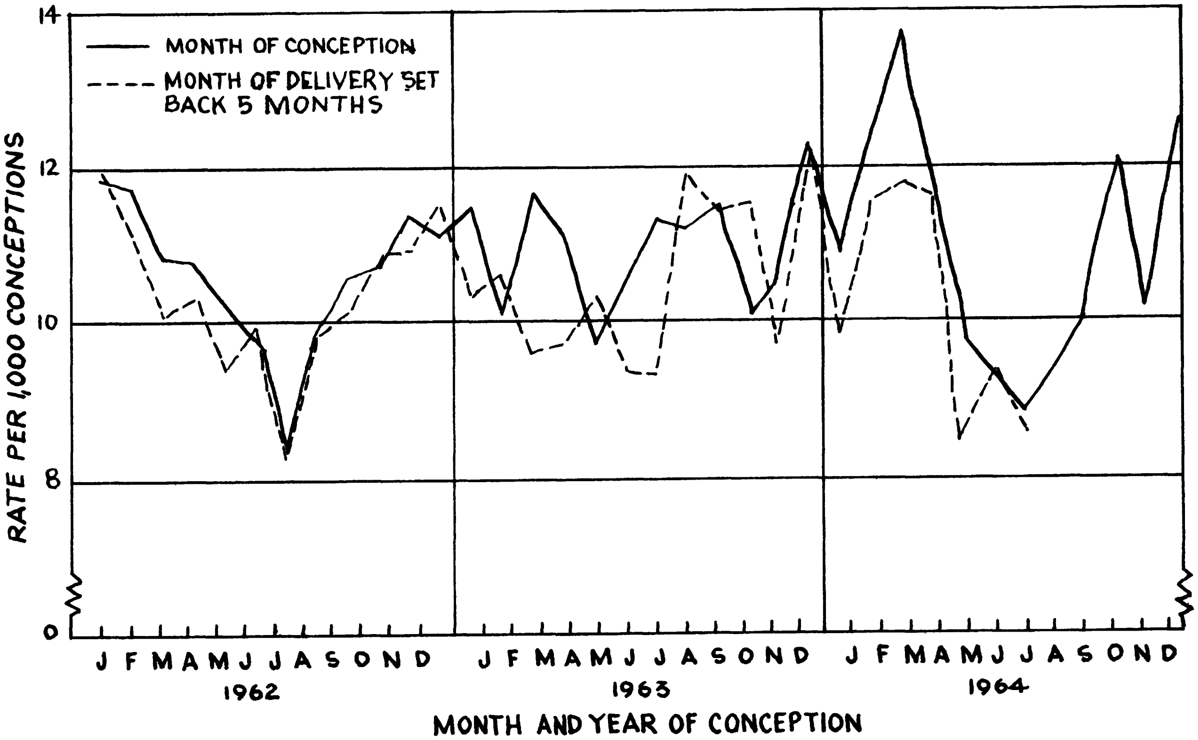


Chart VI. Reported fetal deaths 20-27 weeks gestation per 1,000 conceptions by month of conception and per 1,000 reported deliveries by month of delivery set back five months: New York City, 1962-1964

analysis will provide more precise measures than examination by season of delivery. We should then compare our observations with data routinely produced that might be used to estimate what happens among conception cohorts. Chart VI reproduces the observed data from Chart III for fetal deaths at 20-27 weeks gestation. Here we have also charted (dashed line) rates per 1,000 total deliveries reported as occurring in the same month as the fetal deaths of this gestation interval. The curve has then been shifted five months earlier to approximate the time of conception of the 20-27 week fetuses.

Limitation to fetal losses at a specific gestation interval in itself serves to minimize the effect of the estimation process. The close conformity of the two curves for 1962 suggests to us that during periods when no unusual extraneous factors are operating, the simple estimation process works quite well. However, from Spring, 1963, to April, 1964, when, as we have already commented, the course of the fetal loss rate seems to be aberrant, the two curves frequently are out of phase. Moreover, the sharp increase in March, 1964, is more clearly seen in the date of conception series.

Discussion.

Until further data become available we cannot say with certainty that there is a fundamental seasonal pattern of fetal losses by time of conception. Data in this paper suggest that such seasonality exists, but that, for at least a year of the three years under observation, any such pattern was suppressed by unusually high rates. Were there any events that might explain such aberrations?

Weekly mortality rates from influenza and pneumonia were unusually high in New York City, from late January to early March of 1963. In July of that year, a prolonged heat wave raised mortality rates in New York City to 13.4 and 13.7 per 1,000 population for the weeks ending June 28 and July 5, compared to an expected rate of about 10.2. An epidemic of German Measles built up during the Fall of 1963 with the peak number of case reports to the Department of Health reaching 5,720 in March and 5,731 in April of 1964. (Allowing for delay in reports, the epidemic peak in terms of onset of disease undoubtedly occurred somewhat earlier.)

In connection with the rubella outbreak, it is pertinent to mention that the number of early fetal losses counted in this paper for early 1964 includes therapeutic abortions. Starting with November, 1963, therapeutic abortions with reported indication that the mother had contracted German Measles increased substantially in New York City. Such abortions reached a peak with 78 reported among conceptions in each of the months of February and March, 1964. To indicate the effect of these therapeutic abortions, we recomputed the rates after their exclusion. The result is shown by the dotted line on Chart V for monthly observed rates and by the crossed line for the three-year aggregate. These reductions in the rates do not entirely remove the apparent excess for the period.

Any and all of these environmental events may have had an effect on fetal losses. Further work needs to be done to establish the facts epidemiologically and pathologically. It does not stretch credulity, however, that influenza virus, excessive heat or rubella virus may have been factors in producing higher than expected fetal loss rates among conceptions during 1963 and early 1964. We have not yet been able to investigate other possible influences, such as Widelock's series of infectious agents in the community (10) or even to examine carefully enough those we have cited.

While monthly analysis of fetal losses by date of occurrence would have indicated that fetal loss rates behaved strangely in 1963 and early 1964, we believe that analysis by date of conception points more directly at the time periods that require investigation for association with environmental conditions.

Summary

It is shown, by a hypothetical example, that a 40 percent increase in infant mortality among infants conceived during a one month period could not be detected as a significant increase in the annual rate nor in monthly rates computed in the usual way. It is concluded, therefore, that analysis by month of conception will yield more precise measures of the effect on the fetus of exogenous (environmental) factors during pregnancy.

In this paper are presented fetal loss rates by period of gestation according to month of conception for 1962, 1963 and 1964. The data suggest that fetal losses are low among summer conceptions and high among those in winter. Definitive seasonal patterns cannot be determined from these three years of observations, mainly because rates were almost continuously high throughout 1963 reaching a peak in early 1964.

Some conjectures are offered at this time that such events as high prevalence of influenza and pneumonia as evidenced by mortality in Spring, 1963; prolonged heat in July, 1963, and the German Measles epidemic from Fall 1963, to its peak about March, 1964, may have had a bearing on the relatively high loss rates among fetuses conceived during this period.

Acknowledgments: The imaginative assistance of Mrs. Catherine Laredo in all phases of this work is gratefully acknowledged. The research is supported by Research Grant H-106, Children's Bureau, U.S. Department of Health, Education and Welfare.

References

1. Macfarlane, W.V., and Spalding, D., "Seasonal Conception Rates in Australia". Med. J. Australia 47:121, Jan. 23, 1960
2. Tietze, K., "Zur Frage der Jahreszeitlichen Schwankungen der Geburts bezw.Konzeptionstermine". Therap. d. Gengenw. 102:955, Sept., 1963
3. Salber, Eva J. and Bradshaw, Evelyn S., "Seasonal Variation in Birth Weight". Brit. J. Soc. Med. 6:190, July, 1952
4. Mills, C.A. and Senior, Mrs. F.A., "Does Climate Affect the Human Conception Rate?" Arch. Int. Med. 46:921, 1930
5. Knobloch, Hilda and Pasamanick, Benjamin, "Seasonal Variation in the Births of the Mentally Deficient". Am. J. Public Health 48:1201, Sept., 1958
6. Pasamanick, Benjamin; Dinit, Simon and Knobloch, Hilda, "Socio-Economic and Seasonal Variations in Birth Rates". Milbank Memorial Fund Quarterly 38:248-254, July, 1960
7. Hewitt, David, "A Possible Seasonal Effect on Parturition". Am. J. Obs. and Gyn. 82:940-942, Oct., 1961
8. McKeown, Thomas, "Sources of Variation in the Incidence of Malformations". Congenital Malformations. J.B. Lippincott Co., Phila, and Montreal, 1961, pp. 45-52
9. Walker, James and Smith, Alwyn, "Anencephaly in Dundee, 1950-1959". Op. cit., pp. 257-263
10. Widelock, Daniel et al., "Surveillance of Infectious Disease by Serologic Methods". Am. J. Public Health. 55:578, April, 1965

SIMPLIFIED LEAST-SQUARES APPROACH FOR AN AGE OF COMPUTERS

Irving H. Siegel
The W. E. Upjohn Institute for Employment Research*

A Supermatrix System with Unprocessed Elements

The increasing availability of electronic data-processing services greatly enhances the practicability of an approach to least-squares adjustment developed some years ago by the present author. The straightforwardness of programming and of computer solution was demonstrated in tests conducted for the author by Daniel E. Cowgill in 1963-64 at the Computer Sciences Center of Research Analysis Corporation. The treatment of weighted and unweighted observations by the author's method and the design, in addition, of certain related adjustment procedures that do not employ the least-squares principle were outlined in a paper presented at the 1964 meeting of the American Statistical Association.1/ Below, reference is made again to the least-squares cases considered earlier, and the approach is shown to be also applicable to situations involving linear restrictions.

The author's concept should appeal to statisticians, who need only to organize the data in a certain manner that permits the delegation of literally all the mathematical operations to computer specialists. The method features instant design, without any arithmetic processing of the data whatsoever, of an algebraic equivalent of a set of conventional "normal equations." This equivalent system is actually a correct supermatrix equation constructed by mere arrangement of the original information "packages" with suitable "dunnage". The proper way to organize the data is easy for the statistician to learn and to teach his subordinates.

Data-processing specialists can assume the complete burden of programming and calculation. The traditional intermediate step of computing normal equations is avoidable altogether--or it can be incorporated, if desired, in the computing routine. The unprocessed data, even when voluminous, may be stored intact on tape; and, in addition, it is a simple matter to add, eliminate, or amend the observations, weights, and restraints.

The essence of the author's approach is to expand, rather than to compress, the original oblong matrix of the supposedly error-free observational data into an invertible square matrix. In the conventional approach, it will be recalled, the oblong is compressed by suitable operations into a number of "normal equations" equal to the number of unknown constants. Our alternative requires the construction instead of a design "supermatrix", which we shall call an S-matrix,2/ by the introduction of extra unprocessed information--e.g., conditions imposed on error terms, equal (i.e., unit) weights or unequal weights, and linear restraints. By such additions, the number of variables in the S-system as a whole is considerably enlarged. Explicit solutions are not needed for the extra variables, although they may as a rule be obtained routinely with modern computing equipment. The conditions imposed on the errors, incidentally, are restatements of the conditions expressed in the normal equations derived in the conventional approach.

Transposes of original data matrices play an important role in the formation of the design S-matrix.

---

* The author's views should not be ascribed to The W. E. Upjohn Institute for Employment Research.

1/ I. H. Siegel, "Least Squares with Less Effort", 1964 Business and Economic Statistics Section Proceedings of the American Statistical Association, pp. 284-85. See also two other pertinent papers by I. H. Siegel: "Least Squares 'Without Normal Equations'," Transactions of N.Y. Academy of Sciences, February 1962, pp. 362-371; and "Deferment of Computation in the Method of Least Squares", Mathematics of Computation, April 1965, pp. 329-331.

2/ At Research Analysis Corporation, "S-matrix" stood, not only for "supermatrix", but also for "swollen matrix" and, still better, for "Siegel matrix".

Specifically, the package of conditions imposed on the errors is a transpose of the matrix of the observed error-free coefficients. Hence, any puzzlement raised by the preceding paragraph with respect to the sources of additional needed information is groundless. Indeed, the method of least squares has, in addition to its other recognized merits, an unadvertised distinction of requiring "minimum information" for adjustment, inasmuch as it makes double use, through the transpose, of original observational data.

Our approach calls for the "stacking" of submatrix equations into the S-system. The oblong of error-free coefficients, its transpose, and a diagonal unit of weights are basic ingredients of all S-systems involved in least-squares adjustment. Another essential ingredient of the design S-matrix is the zero block--useful for padding and for spacing the other packages.

Organizing the Data: Unweighted, Weighted, Constrained Cases

The observational equations to be adjusted may be written in matrix form as:

$$Xb + e = Y.$$

Here X refers to n sets of observations on error-free coefficients of the k variables $b (n>k)$, Y is a column of n observed values, and e is a column of n unknown error terms to be purged from the elements of Y. This matrix equation becomes a submatrix equation in the S-system, comprising one tier.

Looking at the above matrix equation as though it were an ordinary linear equation in two unknowns, b and e, we may ask: What additional equations would yield a unique solution for these unknowns? The simplest of all possible companion equations is $X'e = 0$, where X' is the transpose of X. Using this equation as another tier, we have the S-system:

$$\begin{bmatrix} I & X \\ X' & 0 \end{bmatrix} \cdot \begin{bmatrix} e \\ b \end{bmatrix} = \begin{bmatrix} Y \\ 0 \end{bmatrix},$$

where I is a unit matrix and each 0 represents a zero block of appropriate dimensions. It may easily be verified that the two submatrix equations embedded in this S-system are equivalent to the unweighted least-squares matrix normal equation--i.e., equivalent to $X'Xb = X'Y$. It should also be observed that none of the elements in either tier has been processed arithmetically, yet the supermatrix system corresponds to the conventional normal equations. The system is ready for computer processing, being a solvable matrix equation rather than just a turgid array of information.

When the n observations are weighted, we have three tiers. We must introduce a diagonal n x n weight matrix and a set of n arbitrary unknown variables, λ, if all arithmetic preprocessing of the elements is to be avoided. The following S-system may readily be proved equivalent to the weighted matrix normal equation, $X'WXb = X'WY$:

$$\begin{bmatrix} 0 & I & X \\ I & W & 0 \\ X' & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda \\ e \\ b \end{bmatrix} = \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix}.$$

Constrained least-squares adjustment may also be handled by the stacking of submatrix equations containing no preprocessed elements. For example, the famous Gaussian "method of correlatives", commonly illustrated for angle or line measurements in texts treating conditioned observations,3/ is readily managed in a two-tier or three-tier construction. When three tiers are used, each represents an equation in one or more of three sets of variables--the unknown coefficients that we wish to determine, the residuals, and the "correlatives" (undetermined Lagrange multipliers). When the observation equations are written in non-matrix form, they are generally more numerous than the number of unknown constants (e.g., the angles of a triangle), while the conditions to be satisfied (e.g., the sum of the three angles equals 180°) are always fewer.

In modern treatments of constrained least squares, m linear restrictions of the form

$$Z = Rb$$

3/ See, for example, the classical discussions of Merriman and Chauvenet or the more recent works of Linnik and Arley-Buch.

are usually introduced, 4/ where Z is a known column vector of m elements, $n \geq k > m$, and R is a known matrix of order m x k. Arbitrary variables, $\emptyset$, are needed to express the conditions implicitly imposed on R', the transpose of R; they are m in number. The S-system becomes:

$$\begin{bmatrix} R' & X' & 0 \\ 0 & I & X \\ 0 & 0 & R \end{bmatrix} \cdot \begin{bmatrix} \emptyset \\ e \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ Y \\ Z \end{bmatrix}$$

None of the data packages contains processed elements. When the first two submatrix equations are combined, they become the more recognizable matrix normal equation, $X'Xb = X'Y + R'\emptyset$, which remains subject to the restriction stated in the third submatrix equation, $Z = Rb$.

4/ See, for example, A. S. Goldberger, Econometric Theory, Wiley, New York, 1964, pp. 256-257; and R. L. Plackett, Regression Analysis, Oxford, 1960, pp. 52-53.

MEASURING ECONOMIC GROWTH OF THE COMMUNITY OR SMALL AREA

Claude C. Haren, U. S. Department of Agriculture

Shifts in jobs and movements of people in the 1960's show signs of equaling or exceeding those of the preceding decade. Significantly, important differences also are appearing in the local orientation of the changes to metropolitan and nonmetropolitan, high and low income, and economically viable and nonviable areas. But, as the situation now stands, it will be necessary to wait for the results of the 1970 Census of Population for full verification, and even later for the kind of data coverage that will furnish adequate annual inventories of the transitions affecting communities or small areas. 1/

In the paper just distributed, the first section lists the steps that we are taking to meet the urgent need in the interim for data for local planning and development purposes. Then, aided by the accompanying charts and map, a brief review is added on what our research and experimentation have shown to be some of the major shifts underway.

The third section points out certain factors that make this form of direct measurement preferable to determinations of multiplier effects in evaluating contributions of development activities to changing employment and other patterns. Because of the comparative recency of much of the literature on limitations of the latter approach, the succeeding section examines various shortcomings in considerable detail.

In refocusing attention on direct mensuration, a comparison of local growth measures with our system of national economic indicators is followed by an evaluation of six suggested yardsticks of small area dynamics or viability. Inasmuch as fairly extensive restructuring of basic statistics is called for, the next section deals with special as well as conventional quality controls.

The mobility of today's youth and the limited statistics presently at hand on their migration merit addition of a separate section on measures for tracing their movements from home to jobs, school or military service.

The final or concluding section sums up how data supplied by direct measurement -- including information on the community impacts of specific projects or programs -- would assist in appraisals of the potential as well as present roles of area development activities.

Frequent additions have been made throughout the paper to cite experiences with today's impact and development models, underscore effects on area trends of the growing importance of the service industries, list advantages and weaknesses of individual growth indicators, and so on. With the reminder that these and other details are thus available to those interested, the few minutes remaining will be given over to highlights or main points of emphasis.

Research and Experimentation in Progress

Turning back the calendar briefly, migration during World War II and the initial postwar period resulted in widespread depopulation of farming, mining and other communities whose economies were resource oriented or based on obsolescent manufacturing facilities. Beginning in the 1950's, a succession of Federal, State and local programs were introduced to aid these and other economically disadvantaged areas.

Interest was increasing in the jobs and incomes created by various programs. At the same time, the national upturn in economic activity was gaining sufficient momentum to point to both an extension of new employment into additional localities, and a revitalization of the economies of the depressed areas.

Reports on employment and earnings covered under the Old Age and Survivors Insurance and State Unemployment Compensation programs supplied the most satisfactory initial source of data for measuring the changes taking place in these and other communities. With additional time, preliminary measures could be expanded to include yardsticks of the shifts occuring in total employment and incomes, along with changes in population, consumer purchasing power, and money and credit supplies.

An Economic Research Service publication now in preparation will contain further details on changes in employment and earnings in private nonfarm industries in the recent as compared with the earlier postwar period. A follow-up statistical bulletin will list labor force, population and other items for mid-decade, together with changes from 1959-60 benchmarks. Data also have been arranged for either periodic updating or special release by regions or subregions.

Changes Underway in the 1960's

The reduction shown by figure 1 in the metropolitan areas' share of annual increments in covered (OASI) employment in the 1962-64 period as compared with earlier years stemmed largely from the deceleration of former rates of expansion in many major growth centers. Also a factor was the initial sluggishness of the pickup in jobs in certain other metropolitan areas.

In the nonmetropolitan areas, gains were recorded not only in many communities with economic bases and populations of intermediate size,

1/ A county, metropolitan region, or other multicounty group. Excluded for present purposes are municipalities and other subcounty units for which secondary data generally are not currently published.

but in a relatively large number of rural and semirural communities, notably in the South and West. As in their metropolitan counterparts, metropolitan areas appear to be holding and possibly adding to their proportion of the continued buildup of employment in the past two years.

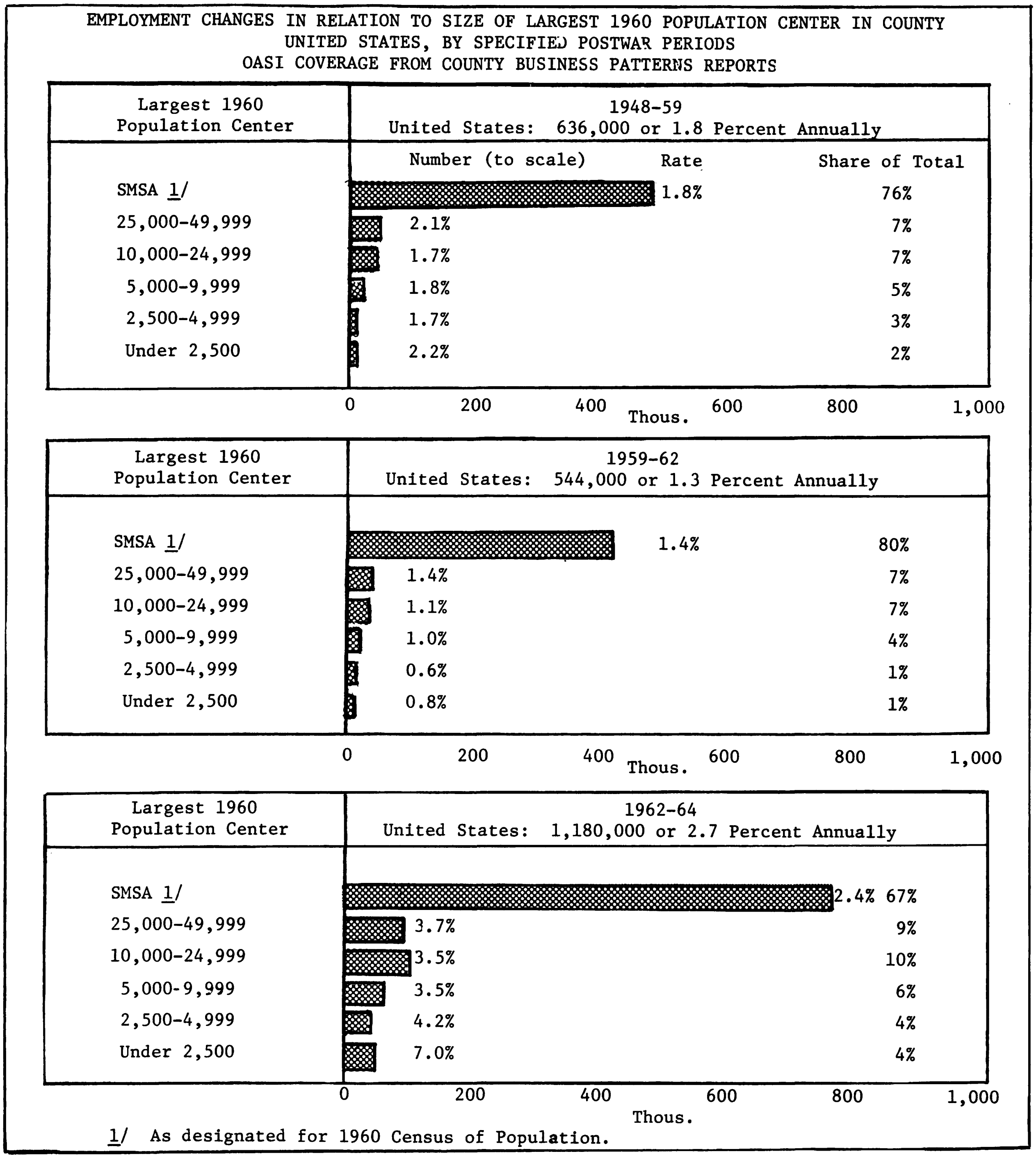


Figure 1

initial responses to the national upturn were mixed in such previously hard-hit coal mining and similar communities as those in Appalachia.

Information for full calendar year 1965 and in certain instances for first quarter 1966 are just beginning to become available. From the statistics presently at hand, however, the non-

Weighing Immediate Alternatives

In meeting requirements for local data, one approach would be to expand the number of jobs added by a specific program or facility by a factor that would account for the extra employment opportunities thus created. The other alternative would be to simply measure the aggregate changes taking place over the period, somewhat as we are doing, and then allocate the part attributable to particular activities.

As declines elsewhere in area economics frequently offset and even exceed gains from new sources of employment, sufficent data would have to be made available so the additions imputed by the multiplier process could be reconciled with the shifts actually occuring.

Because industrial mix as well as small size make many area economics especially subject to seasonal, cyclical and episodic forces, adjustments for trend analysis and projection purposes are as essential for local as for national data and probably more so. With area transitions also generally paralleling overall increases, gains in employment and incomes in the service sector not only have eclipsed additions to jobs in manufacturing and other

MEASURES OF ECONOMIC GROWTH
OF SMALL AREAS

PRIMARY INDICATORS

A. EMPLOYMENT (AND EARNINGS) IN PRIVATE NONFARM INDUSTRIES

B. TOTAL LABOR FORCE (THE EMPLOYED AND UNEMPLOYED)

C. POPULATION (AND HOUSEHOLDS)

D. AGGREGATE (AND PER CAPITA) INCOMES

E. CONSUMPTION EXPENDITURES: RETAIL SALES

F. MONEY AND CREDIT SUPPLIES: BANKING DEPOSITS (BANK AND S&L)

SECONDARY INDICATORS OR DATA QUALITY CONTROLS

A. VITAL RATES

B. SCHOOL ENROLLMENTS

C. MOTOR VEHICLE REGISTRATIONS

D. UTILITIES CONNECTIONS

E. RESIDENTIAL BUILDING AND DEMOLITION PERMITS

F. STATE INCOME TAX RECEIPTS

G. STATE SALES TAX COLLECTIONS

Figure 2

Also necessary would be enough data on commuting patterns to work and shopping centers so the share actually going to the immediate community could be ascertained.

Determinations of Multiplier Effects

Costs alone would fairly conclusively rule out the application to 3,000-odd counties of a multiplier approach that would minimize limitations of the sort mentioned. But, even for single-area projects, meeting these and other, more serious qualifications would necessitate major modifications of existing models.

goods-producing industries in many localities, but have been accompanied by sharp reductions in others. Together with the impact of general rather than specific increases in incomes, the added effect of the velocity of spending and respending actions in catalyzing area growth suggests that a restatement of the basic multiplier formula may also be required.

Direct Measurement

Measures of economic growth at small area levels serve the same broad purpose as our system of national economic indicators in pin-pointing changes in economic conditions. The difference is in scope, a reflection both of the comparatively limited local data available, and of what can be made to suffice -- against the tremendous backdrop of information at hand on movements in the economy at large -- in interpreting happenings locally.

The listing in figure 2 is restricted to the six sets of primary growth indicators that have proven particularly useful for yardstick purposes and that would involve no more than fairly inexpensive reprocessing of basic statistics. Included for expansion purposes are measures of earnings as well as employment for wages and salaries in private nonfarm industries, households as well as population, etc.

In choosing data on covered employment and earnings for the initial set of low-cost but reasonably reliable measures of local growth sought, one advantage recognized was the need for only minor modifications of published statistics. Another important consideration was the traditional role of shifts in the private nonfarm sector -- to which the data applied -- as bell-wethers of economic change at local as well as regional and national levels.

larly collected materials. As a result, estimates are more reliable than a high percentage of prevailing population projections, some illustrations of which are shown by table 1.

Changes in earnings and income furnish a better index of area viability than shifts in labor force or population inasmuch as increments in wages or income receipts as well as in the number of earners or recipients are included. Adjustments should be introduced to compensate for cost-of-living increases. Also helpful would be conversions to per capita values for purposes of relating local transitions to those at regional and national levels.

By including transfer payments and other nonwork receipts, shifts in income ordinarily would give a more complete picture of area viability than changes in earnings alone. But sharpness of comparisons can be seriously blurred by the intervention of more or less arbitrary deductions for certain taxes, or additions for such nonmoney items as value of food produced and consumed on farms, and rental value of farm and owner-occupied nonfarm dwellings. Adding still further to the advisability of restricting coverage to money income is the necessity for

Table 1.--Changes in population as determined from estimates by different methods, Ohio statewide and selected counties, 1960-65

Unit	Amount					Rate				
	1950-60	1950-65 Method				1950-60	1950-65 Method			
		One	Two	Three	Four		One	Two	Three	Four
	Thous.	Thous.	Thous.	Thous.	Thous.	Pct.	Pct.	Pct.	Pct.	Pct.
State:										
Four methods ---	2/	1676.2	1614.4	1250.4	1365.8	2/	17.4	16.7	12.9	14.1
Bureau Census --	1759.8	1074.0	1074.0	1074.0	1074.0	22.1	11.1	11.1	11.1	11.1
Difference ---	2/	602.2	542.4	176.4	291.8	2/	6.2	5.6	1.8	3.0
County:										
A - Franklin ---	179.6	191.6	202.6	133.6	123.6	35.7	28.2	29.8	19.7	18.2
B - Cuyahoga ---	258.4	220.0	155.8	190.0	286.6	18.6	13.4	9.5	11.5	17.5
C - Hamilton ---	140.2	143.2	145.2	98.0	109.0	19.4	16.7	16.9	11.4	12.7
D - Stark ------	57.2	49.6	50.0	40.4	40.4	20.2	14.6	14.8	11.9	11.9
E - Greene -----	35.8	37.2	35.8	28.0	17.0	60.7	39.7	38.2	29.9	18.1
F - Clermont ---	38.3	42.6	36.6	27.4	14.0	90.9	53.5	46.0	34.4	17.6
G - Darke ------	3.8	3.8	4.2	2.0	4.4	9.1	8.4	9.2	4.4	9.7
H - Athens -----	1.2	1.4	9.8	-0.6	3.2	2.5	3.0	20.8	-1.3	6.8
I - Perry ------	-1.1	-1.4	-1.4	-2.4	2.2	-3.9	-5.0	-5.0	-8.6	7.9

1/ Converted to 10 year equivalent for direct comparison with 1950-60 changes. 2/ Not Available.

Total labor force consists of all workers plus those seeking work, important additions to coverage in communities with large numbers of government employees, farm operators, farm workers, the self-employed in business and the professions, and the unemployed. By and large, statistics on the unemployed and those employed outside as well as in covered industries (OASI and State UI programs) are compiled from regu-

limiting correlations with consumer expenditures to the components of the latter that are for goods and hence show up in variations in retail sales.

Quality Controls

Such secondary indicators or quality controls as vital data, school enrollments and motor vehicle registrations can be applied in screening out some inconsistencies in population and other estimates. Because employment and other transitions tend to be identified with specific location, income and other characteristics, other disparities can be adjusted in the process of relating statistics on various changes to the county or area correlatives shown by figure 3.

Although the steps mentioned are helpful the key to effective restructuring rests with the processing of labor force, population, income and other data by subregions, labor market or trade areas, or some other intermediate or multicounty areal entity. At this as contrasted with the individual county level, differentials in employment and earnings between locations of work and points of residence tend to be at a minimum.

fluctuations in annual sales should be particularly helpful in this allocation process. Much information is available in addition on the manner and degree to which previous interrelationships between localities have been altered by plant additions or closures, highway improvements, and similar changes. Also variously at hand for adding to the consistency of the final allocations are statistics on residential building and demolition permits, and utility connections and discontinuances, together with the information supplied by special population censuses, and annual counts of occupied dwellings.

Tracing the Migration of Today's Youth

Present data can furnish only a fairly general notion of the work force and population changes resulting from the streams of migration between the local, metropolitan and regional

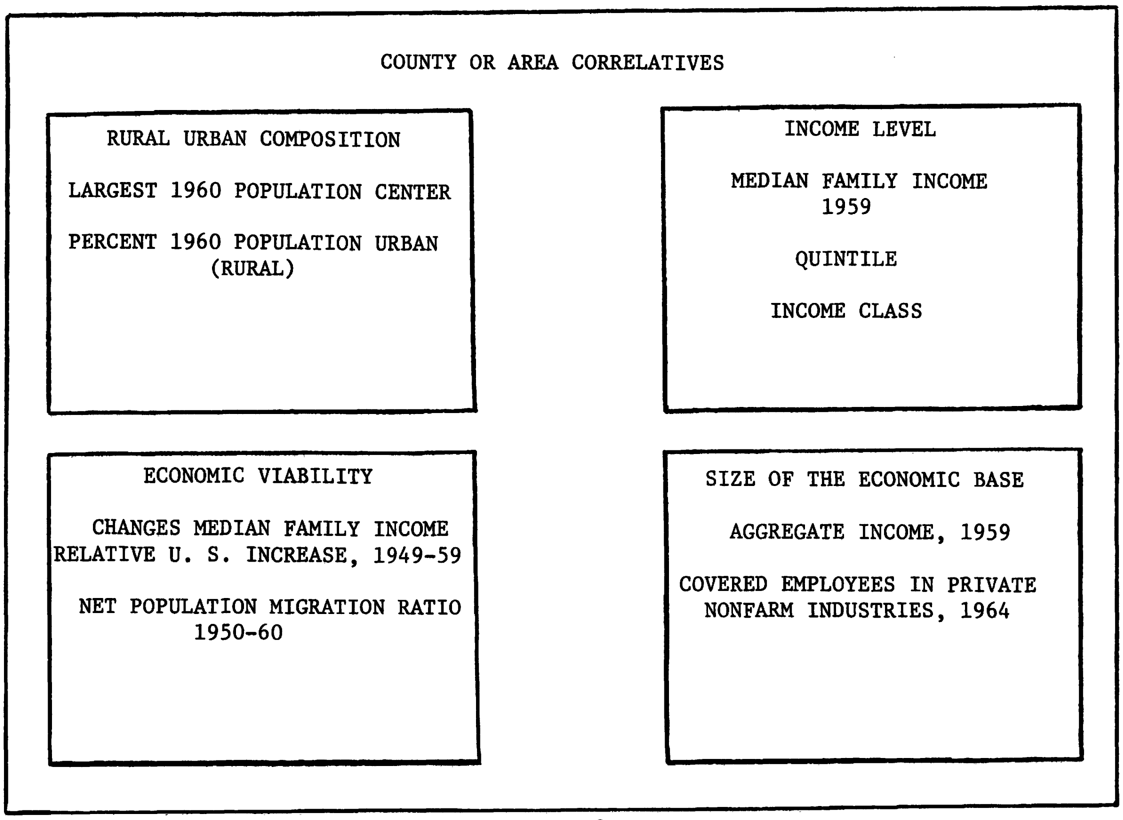


Figure 3

As with the impact of work commutation patterns, resolution of differentials between where people live and where they shop and bank also will usually be unnecessary at this level. With the opportunity thus presented for a thorough and relatively uncomplicated checkout of labor force participation and other ratios, the entire set of indicator values can be verified before having to proceed to final allocations to subareas.

Because of its sensitivity to changes in year-by-year incomes, the breakdown of sales to include those from types of establishments more locally oriented and less susceptible to

growth areas illustrated by figure 4. Even less specific information is provided on the changes involving youth -- the most mobile element in our population, and a group concerning whose movements very little actually is known.

Special classifications of jobholders and job seekers and the charting of local labor market, school and other destinations of high school graduating classes represent significent additions to the data needed. Information on places of residence furnished by college enrollees and military personnel can be applied to identify county or other areas of orgin. For those required to register, the records maintained by

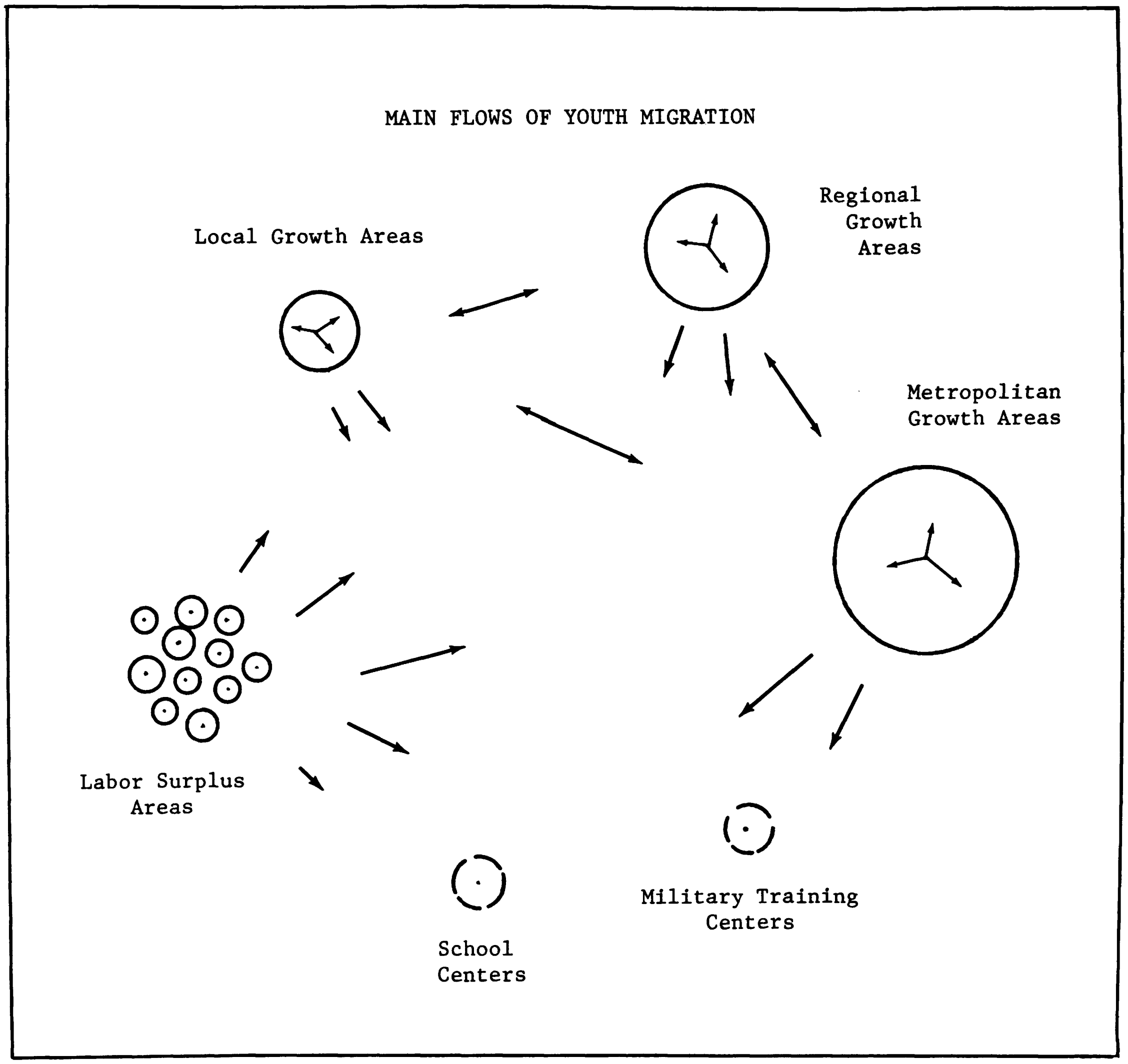


Figure 4

the State Selective Service Systems comprise an as yet almost untapped source of basic data on original and current places of residence, status, education, present occupation, etc.

Evaluating the Impacts of Community Development Activities

Regardless of whether feasible otherwise, direct measurement is not intended to, and cannot, supply a figure or set of figures on jobs created, income produced, and people added by various development activities, planned or unplanned. What can be readily furnished is a fairly detailed data sheet indicative of employment and other levels before and after an activity or a combination or series of activities is introduced.

Also supplied is a framework of reference for determining impacts of projected programs and facilities on communities with known economic and related characteristics. Provided in addition is a basis for examining variations between activities in the directness or indirectness and the immediacy or remoteness of their contributions to employment and income buildups, and for exploring their interrelationships with each other.

This points to both a need for comprehensive as opposed to superficial examination of what is happening in individual areas, and more importantly to the challanging opportunities presented for fully exploring the implications arising from the introduction of a variety of plans, programs, and facilities into equally varied community or small-area environments.

Suggested References

Andrews, R. B.
1963. Economic Planning for Small Areas. Land Econ. 39(2):143-155, 39(3):253-264

Biscoe, A. B., Jr. and Ray, D. G.
1965. Method Used for Estimating Personal Income by County in Florida. Bur. Business Research, U. Fla., var. p., Sept. (Working Manual).

Bureau of the Census, U. S. Dept. of Commerce
1963. Standard Metropolitan Statistical Areas in the United States as Defined October 18, 1963. Current Population Rpts., Tech. Studies Series P-23, No. 10, 16 pp., Dec. 5.

———— 1966. Americans at Mid-Decade. Current Population Rpts., Tech. Studies Series, P-23 No. 16, 31pp., Jan.

———— 1966. Inventory of State and Local Agencies Preparing Population Estimates Survey of 1965. Current Population Rpts., Population Estimates Series, P-25, No. 328, 25pp., Mar. 8.

Corry, O. C.
1964. Local Statistical Data, Needs and Adequacy from the Standpoint of State Government. American Statis. Assoc., Proceedings of 123rd Annual Meeting, Business and Econ. Statis, Sec., Cleveland, Ohio pp. 164-167, Dec. 27-30.

Council of State Governments.
1965. Administrative Officials Classified by Functions. Supplement II, The Book of the States, 90 pp., Chicago, Ill., July.

Fuchs, V. R.
1965. The Growing Importance of the Service Industries. Nat. Bur. Econ. Research occas. paper 96, Columbia U. Press, New York, 30 pp., Oct.

Haren, C. C.
1965. Regional and Local Differentials in Postwar Growth of Population Centers in Nonmetropolitan Areas: A Progress Report. Paper presented Southeastern Regional Science Association Annual Meeting, Charlottesville, Va. 6 pp., Apr. 23.

———— 1966. Responses in the South and Southeast to National Expansion in Economic Activity Since 1961. Paper presented Southeastern Regional Science Annual Meeting, U. Md., 22 pp., Apr. 1.

Hochwald, W.
1964. Local Statistical Data, Their Needs and Adequacies from the Viewpoint of Academic Research. Amer. Statis. Assoc., Proceedings of 123rd Annual Meeting, Business and Econ. Statis. Sec. Cleveland, Ohio, pp. 150-155, Dec. 27-30.

Hoover, E. M., and Chinitz, B.
1961. The Role of Accounts in the Economic Study of the Pittsburgh Metropolitan Region. In Design of Regional Accounts 281 pp., The John Hopkins Press, Baltimore.

Joint Economic Committee, 88th Congress, 2d Session, 1964. 1964 Supplement to Economic Indicators, Historic and Descriptive Background, 130 pp.

Leven, C. L.
1964. Regional and Interregional Accounts in Perspective. The Regional Science Assoc., Papers Vol. 13, pp. 127-144.

Meyer, J. R.
1963. Regional Economics: A Survey. The Amer. Econ. Rev. 53(1): 19-54.

New Hampshire Department of Employment Security
1965. Economic Changes in Each New Hampshire County 1957 to 1965. 120 pp., May.

North Carolina Employment Security Commission
1965. North Carolina Work Force Estimates by Labor Area. 226 pp., Oct.

Norton, J. H.
1965. Manpower Analysis for Metropolitan Areas: Critical Study of Currently Available Models and Data. Amer. Statis. Assoc., 1965 Proceedings of Business and Econ. Statis. Sec., Philadelphia, Pa., pp. 266-272, Sept. 8-11.

Park, Se-Hark
1965. The Economic Base Identification: An Appraisal. Land Econ. 41(4): 382-386.

Peach, W. N., et al
1965. Source Notes and Explanations to County Building Block Data for Regional Analysis. Research Foundation, Okla. State U., 138 pp., March.

Poole, R. W.
1964. Implications of Labor Characteristics and Commuting Patterns for Regional Analysis: A Case Study. Land Econ. 40(1): 110-116.

Thomas, M. D.
1964. The Export Base and Development Stages Theories of Regional Economic Growth: An Appraisal. Land Econ. 40(4): 422-432

Tiebout, C. M.
1962. The Community Economic Base Study. Supplementary Paper No. 16, Committee for Economic Development, New York, 86 pp., Dec.

———— 1965. Book Review: National Growth and Economic Change in the Upper Midwest by J.M. Henderson and A. O. Krueger: U. Minn. Press, Minneapolis. The Amer. Econ. Rev. 55(5): 1210-1212.

Wadsworth, H. A., and Conrad, J. M.
1965. Leakages Reducing Employment and Income Multipliers in Labor-Surplus Rural Areas. Jour. Farm Econ. 47(5): 1197-1202.

Weintraub, S. A.
1966. Keynesian Theory of Employment Growth and Income Distribution, Chilton Books, Phil., 147 pp., Feb.

THE STATISTICAL PROGRAM OF THE RAILROAD RETIREMENT BOARD

A. Benjamin, U.S. Railroad Retirement Board

The Railroad Retirement Board administers a retirement-survivor benefit program and a separate unemployment-sickness benefit program under the provisions of the Railroad Retirement and Railroad Unemployment Insurance Acts. The retirement and survivor benefit provisions are similar to those under the social security system but are generally more liberal. Unemployment benefits are higher than those under most State programs; only a few States have comparable sickness benefits. Detailed records are maintained on the operations of the benefit programs and on employment and payrolls in the railroad industry for statistical, actuarial, interagency financial transactions and for administrative purposes. These records are unique because they are the only ones covering a single industry under National laws.

Statistics are published on a universe basis with respect to (1) financial operations, (2) certain employment data, and (3) new awards and terminations of retirement and survivor benefits. Practically all other data are based on samples. Figures on the benefit programs and on employment and payrolls are presented in the Board's Annual Report, Monthly Review, and other publications. The financial data include the number and amount of benefits paid out, the receipts, expenditures and balances in the Railroad Retirement and Railroad Unemployment Insurance Accounts, etc.

Most of the statistical analyses and actuarial estimates are based on sample data. The samples may be called modified systematic samples since they are selected on the basis of specified combinations of digits of the social security account numbers. Since there is no known bias in this method of selection, the resulting samples are treated as random samples for the purpose of testing and inflation of the data. Sample studies are based on information developed from claims for benefits, reports of earnings and other administrative sources and assembled into statistical records which are kept up to date on a regular basis. Except on rare occasions, railroad employees are not asked to complete questionnaires or otherwise respond to inquiries to be used purely for statistical studies. Some of the major studies are discussed briefly below.

Service and Compensation

Beginning with 1937, the Board has published an annual census of all railroad employees who worked in the year. This volume has been used by both railroad labor and management in wage negotiations. The set of tables provides information on taxable earnings and months of service in the year by class of employer and by occupational groupings. Also, for employees of Class I railroads, the service and earnings are shown by individual occupation (the 128 Interstate Commerce Commission reporting divisions). The statistics on occupation are probably among the most comprehensive available in this field. The compilation also carries less detailed information on new entrants in the industry and other selected data including some historical information beginning with 1937.

4-Percent History Sample

A 4-percent sample is selected annually from the file representing the universe of employees who worked in the year. This file is used to update a permanent 4-percent history file of present or former railroad employees who worked in the industry at any time after 1936. The records contain certain cumulative information, such as total service and total taxable compensation earned after 1936, service months before 1937 and a code identifying each of the years that the employee worked in the railroad industry beginning with 1937. Current data include the employee's railroad occupation and his service months and compensation for each of the last 3 years of employment. (An outline of the major items contained in the record is attached.) One difference between this sample and a similar social security sample is that the Board's sample includes service before 1937 which is not relevant to the social security sample.

In addition to its use in the analysis of employment and payrolls, the 4-percent sample provides data on the basis of which cost factors for the triennial actuarial valuations are developed. Some of the more important of these are as follows:

1. Rates of withdrawal from, and of mortality in, active service, and exposures for developing rates of retirement and disability.

2. Numbers of active and inactive employees alive and not retired on the valuation date by year of entry, age, and length of service.

3. Age distribution of new entrants.

4. Past and future salary scales.

10-Percent of UI-SI Sample

Data on the railroad unemployment and sickness benefit program are developed largely from a 10-percent sample of beneficiaries maintained on a continuing basis. Comprehensive information on the unemployment and sickness experience of railroad workers has been made available in various studies on rates of sickness and unemployment which have contributed substantially to the knowledge in the field. In addition, comparisons with the various State plans and other studies have been published.

Analysis of the unemployment and sickness benefit program is based entirely on sample data. The size of the sample has varied--in recent years, it has been 10 percent. Because of the relatively small numbers involved, statistics on maternity benefits are on a 100-percent basis. The permanent beneficiary files contain such items of information as daily benefit rate, base-year compensation and service, total service, occupation, reason for unemployment, etc.

Financial Interchange

In 1951, the Railroad Retirement Act was amended to provide for a financial coordination between the railroad retirement and social security systems. The purpose of this legislation was to put the social security system in the same position financially in which it would have been had railroad employment been covered by that system beginning with 1937. This is carried out by joint annual determinations by the two agencies of the amounts to be transferred from one system to the other. Calculation of the amounts of these annual transfers involve two major steps: (1) determining the amount of taxes that would have been collected on railroad payrolls at social security rates, and (2) calculating the amount of benefits that would alternatively have been payable to railroad retirement beneficiaries if railroad service had been covered under the social security system. The resulting annual transfers involve hundreds of millions of dollars, so that the financial interchange is extremely important in the financing of the railroad retirement system.

The tax calculations are based on total taxable railroad payrolls adjusted on the basis of sample data to take into account the difference between the tax bases of the two systems. Under the social security system, there is an annual limit on earnings subject to tax, while railroad earnings are only taxed up to a monthly maximum. Also, under the social security system, if an employee worked for two employers, each is required to pay the taxes on earnings up to the annual limit. In the same situation under the railroad retirement system, each employer would pay taxes only on its share of the earnings up to the monthly maximum. Adjustments to take these differences into account are made on the basis of a special 1-percent sample of employees whose total (taxable and nontaxable) earnings are reported through special arrangements with railroads.

For the benefit computations, a 1-percent sample is used drawn from both retired and nonretired groups of employees in order to cover every possible situation in which benefits could have been paid by the social security system. At the present time, the sample consists of some 9,000 active cases. For the most part, these cases are processed individually on the basis of the actual record. In a few cases in which the actual record is not complete, actuarial techniques are used to estimate the amounts of benefits that would be payable under social security.

For purposes of testing and inflation of the sample results after the benefit calculations are completed, the sample cases are grouped into some 25 to 30 strata according to the age of the beneficiary, the composition of the beneficiary family, and the beginning date of the benefit, with the latter grouping reflecting the effective dates of changes in the tax base and benefit provisions in the Social Security Act.

Military Service Study

Under the Railroad Retirement Act, transfers are required to be made from general funds in the U.S. Treasury to the railroad retirement account to reimburse the account for the cost of crediting military service towards benefits. The amount to be transferred for creditable military service performed in the period from January 1937 through June 1963 consists of the railroad retirement taxes that would have been collected on imputed earnings of $160 a month. This amount (plus interest) was estimated on the basis of successive samples covering different periods. A sample of about one-fourth of 1 percent was selected from employees with railroad service through 1947; this was supplemented by 1-percent samples of new entrants in each of the subsequent years. These groups were refined by the elimination of employees who could not qualify for military service credit or for whom the probability of service in the armed forces was extremely low.

Questionnaires were sent to the remaining employees in the groups. If they had been or were still in military service, they were asked to supply information which could be used to verify their actual military service in the records of the Department of Defense. Appropriate employer and employee tax rates were applied to the creditable military service developed. The estimate was obtained by stratifying the sample results by period of military service and inflating the data for each stratum.

The additional cost to the retirement system of military service performed after June 1963 is based on the value of the additional benefits due to military service actually paid to beneficiaries. This cost is determined by actuarial techniques.

Retirement and Survivor Benefits

A complete statistical record is maintained of every benefit ever awarded by the Board. The file of benefits currently being paid is maintained and updated each month on a 100-percent basis. At the present time, this file includes records pertaining to about 1 million beneficiaries. Each record contains the amount of benefit, type of benefit, date of birth, railroad service and average earnings used in the computation of benefits, last railroad occupation, etc. Although most of the analytical tables on the retirement-survivor program are based on samples, the complete file is used for certain administrative operations.

The use of statistical records for administrative operations has greatly lightened the Board's workload over the years. This has been particularly true in connection with amendments involving changes in benefit rates for all or most beneficiaries since the Board's staff is small in relation to the number of beneficiaries on the rolls. The use of statistical records for this purpose began in 1946 when the Act was amended to provide a new system of survivor benefit. At that time, the statistical records provided the best source of information for carrying out the provisions of the new legislation. Since then the statistical records have been expanded to carry additional information designed to expedite administrative operations. This includes the identification of cases which must be checked for continuance of entitlement, and suspension and recertification actions, as well as the recomputation of benefits to reflect legislative changes.

April 1966

References

Robt. J. Myers, "Social Insurance and Allied Government Programs," (1965).

Benjamin Mandell, "The Continuous Work History Samples Under the Old Age and Survivors Insurance in the United States," International Social Security Association, (1957), "The Actuarial and Statistical Problems of Social Security," (1957), Vol. 2, 179-202.

Samuel A. Block, "The Railroad Social Insurance System," The Proceedings of the Conference of Actuaries in Public Practice, Vol. VI, (Feb. 1957).

U.S. Railroad Retirement Board and Department of Health, Education and Welfare, "RRB-SSA Financial Interchange," Summary and Documentation.

W. Matscheck and A. M. Niessen, "Sampling as a Solution to Certain Unusual Social Insurance Problems," First International Conference of Social Security Actuaries and Statisticians, (Nov. 1956).

Outline of 4-Percent Wage History Record

1. Social security account number.
2. Date of birth.
3. Sex.
4. Year last worked (or current year if employee is in active service).
5. Indication that employee has social security earnings at any time after 1936.
6. Year of original entry into railroad service (including years before 1937).
7. First year worked for a railroad after 1936.
8. Total prior service months (service before 1937).
9. Total subsequent service (service after 1936).
10. Indication of continuous service since year of entry.
11. Prior service average compensation.
12. Total subsequent compensation.
13. Number of calendar years in which employee worked after 1936.
14. Year of death or retirement.
15. Identification of death or retirement.
16. Indication that employee had more than one employer in current year.
17. Last railroad employer.
18. Occupation code (128 ICC reporting divisions)
19. Service months in current year or year last worked.
20. Taxable compensation in current year or year last worked.
21. Service months in first preceding year.
22. Taxable compensation in first preceding year.
23. Service months in second preceding year.
24. Service months in third preceding year.
25. Service months in fourth preceding year.
26. Codes indicating changes in occupation in last 4 years.
27. Identification of individual years in which service performed after 1936.
28. Creditable compensation in current year or year last worked under Railroad Unemployment Insurance Act.
29. Indication of military service.

XVI

CONTRIBUTED PAPERS II

Chairman, JOSEPH STEINBERG, Social Security Administration

SAMPLING ERRORS OF PREVALENCE ESTIMATES OF RARE DISEASES BASED ON SURVEYS OF MEDICAL SOURCES (Abstract)

Monroe G. Sirken, National Center for Health Statistics
and Z. W. Birnbaum, University of Washington

A prior publication derived the formulas for three unbiased estimates of the prevalence of a rare disease in the population, based on a stratified random sample of medical sources. Deriving unbiased estimates of the number of diagnosed cases based on patients reported by a sample of medical sources presented a special problem, since it is not uncommon for the patients to have been treated by more than one medical source. The three formulas differed in the extent that they utilized information about multiple reporting of patients collected in the sample survey.

The present paper derives the variance formula for each unbiased estimate of the prevalence of a rare disease that was presented in the earlier publication. The variances are derived under certain simplifying assumptions: (a) a medical source reports at most one patient; and (b) each patient is reported by the same number of sources. The paper compares the variances of the different estimates, and for each estimate notes the effect on the variance of the multiplicity of patient reporting.

EVALUATING RESPONSE ERROR IN THE REPORTING OF THE INCOME OF THE AGED: BENEFIT INCOME

Lawrence D. Haber, Social Security Administration

It is commonplace of survey research that if anything can possibly go wrong, it will. The collection of data in large scale interview surveys is filtered through a series of intermediaries; the phenomena studied are rarely observed directly. The ultimate source of information, the respondent, is for most purposes only another intermediary, with his own flaws and inadequacies in his ability to transmit accurate reports of the traits, characteristics, events or conditions about which the total stranger we have sent to his door is inquiring, generally for some reason which to the respondent is remote, abstract or irrelevant. He is asked to respond to questions out of context to their real life situation, concerning matters about which good friends, and sometimes even wives, are usually not expected to inquire. After completing the interview, or possibly that night, after putting the children to bed and washing the dishes, the interviewer reviews the questionnaire, to see that each space is appropriately filled in.

At survey headquarters, the one-hundred or so answers on the questionnaire are transferred to a record and manipulated to reduce them to the five, ten, twenty, or fifty dimensions which constitute the data of the survey. The data are then further manipulated to produce a set of tables and numbers which presumably have some bearing on the problem which prompted the study. Considering the multiple possibilities of error, it is reasonable and necessary to ask how good are the data we have collected.

Up to the point at which the interviewer knocked at the door, we had a precise estimate of the range of error expected from sampling, assuming an acceptable probability sampling design consistent with the objectives of the survey. At the respondent's door, however, errors attributable to the interviewer, the respondent or the interviewer-respondent interaction enter into the measurement. Regardless of the precision expected from the sample design, we have not accounted for the accuracy of the measurement until we have some notion of the kind and extent of the response errors entering into the estimates.

In this paper we will present data on the magnitude and type of response error and response bias in the reporting of social security benefit income and we will examine some of the characteristics associated with response bias. The data are based on a national area probability sample of all persons aged 62 and over, the 1963 Survey of the Aged, conducted by the Social Security Administration, in cooperation with the Bureau of the Census. Validation of the benefit income was obtained from a complete matching of the survey interviews with the benefit income data from Social Security record files.

The concern with response validity predates the sample survey by many years, as illustrated by the laws of perjury, rules of evidence and trial cross-examination procedures. Only in recent years has a systematic body of data and of theory on response error begun to develop.1/ Unlike sampling error, however, there is little existing theory as yet on which a priori judgments or estimates of response error may be based.

One of the difficulties in the study of response error is the limited availability of criterion data for assessment of the survey data. In many cases, the available criteria contain similar response errors, or a different but equally unknown set of estimation errors. In some areas of measurement, such as personality and attitudes, the existence of an external reality against which the data can be measured may itself be in question.

Most of the things with which we are concerned in survey research, however, do have an external reality or may be represented by events, acts, or conditions external to the reporting unit.

The 1963 Survey of the Aged provided an unusual opportunity to study response error in income reporting through the virtually complete matching of survey data from an area probability sample with a large record system, the Social Security benefit and earnings records. The survey data were collected in two stages and included information to identify respondents with the records and information on specific income sources, including benefit income and earnings. Interviews were completed with 7,500 survey units, consisting of a person aged 62 and over and his spouse, if any. Matching record information was located for all but 2 cases per 1,000 interviewed.2/ The benefit income reporting comparison is based on the total of 4,727 beneficiary units, representing 11 million couples and aged individuals. The sample is self-weighting, but has not been adjusted for population and non-interview factors in this analysis.

The matching operation was primarily undertaken to insure accurate reporting of beneficiary status and benefit income. The methodological analysis of benefit income reporting was a byproduct of the study rather than a major objective. Both benefit income amounts were kept in the data tape, but only the "true" or benefit record value was reported in the study, with the exception of a $60 range of error allowed to reduce the review and editing work load.3/

The accuracy of the respondent's income estimate was measured from the discrepancies between the amount recorded in the benefit record, the "true" value for the respondent, and the amount obtained in the interview. The mean benefit income reported is the survey estimate of the "true population value," without response error. The differences between the Social Security benefit record and the interview data represent response error, including both variable error and bias.

In order to examine some of the possible sources of response error, the data were analyzed by age, type of unit, education, income thirds, living arrangements, and amount of OASDI benefit income. Beneficiaries who did not correctly report their beneficiary status in the interview are included in the analysis.

Data from the record match

1. Response error and bias

Table 1 shows the distribution of reporting differences between the SSA benefit record and the interview for all beneficiary units, couples and the nonmarried. Almost half of the beneficiary income reports matched exactly with the benefit record and more than three-fourths of the beneficiaries were within a range of plus or minus $60. About one-fourth of the responses were distributed among over and underestimates of more than $60, but the proportion with underestimates was 8 percent higher. Most of the net differences in under and over reporting were large amounts in excess of $300.

The net income understatement or bias was $61, about 6 percent of the Mean Total Benefit Income (TBI) of $1,052, with a standard error of approximately $5. As table 2 indicates, almost all of the bias was attributable to the small proportion of units with large under reports: four-fifths of the bias was from errors in excess of $500. Response errors of less than $301 were almost entirely random variation and made little or no contribution to the bias.

Considering that more than half of the aged received less than $1,000 in benefits and only one in 20 received more than $2,000, errors in excess of $300 are quite sizable for this group and would seem to be more than simple inability to recall or work out an approximate total. Faulty reporting of beneficiary status may account for as much as a third of this; about two percent of the beneficiaries did not report their beneficiary status, but the variation in the proportion not reporting did not show any direct relationship to the proportion with large underreporting errors.

Other factors related to the particular conditions of benefit payments and to the questionnaire method may also have resulted in the omission of income amounts. Large back payments for delayed benefit awards, benefit terminations and suspensions may result in unusual or irregular payments. The questionnaire was set up primarily to obtain accurate totals on regular income and may thus have led to some understatement of irregular payments. Our data do not examine this possibility directly, but there is some support for this in the differences in the size and error rate of the bias by age groups.

2. Matching reports under $61

We were also interested in the effect on the income distribution and mean of the editing procedure allowing differences of up to $60 to be accepted as matching amounts. This was done in order to reduce the edit rejection rate and to eliminate extensive clerical checking for small differences.

As shown in tables 1 and 2, the $60 allowance was an effective procedure for reducing editing at little or no cost in bias. Reporting differences under $61 were randomly distributed between under and over reporting; 31 percent of the respondents had reporting errors under $61, but only 2 percent more under reported than over reported income. The net effect on the total benefit income was negligible, an average understatement of $1.

From an examination of the distribution, it would seem that our allowance was overly conservative and that we could have accepted all matching differences up to three hundred without affecting the mean Total Benefit Income (TBI) or introducing a serious distortion of the income distribution.

3. Sources of selective bias

The several variables selected for analysis were chosen on the basis of experience with or assumptions about their relationship to response validity. Differences between the Current Population Survey and Office of Business Economics estimates of the number of low-income families have been attributed, in part, to transfer income, including social security benefits.4/ Socio-economic factors such as education and income have frequently been found to be positively associated with accuracy of reporting.5/ Age, as an indicator of mental and physiological responsiveness should be negatively associated with validity, while living arrangements, as a measure of independent functioning, should have a positive relationship.

The results of this examination are summarized in table 3, for total beneficiary units. The relationship of these variables to response validity is examined further in tables 4 and 5, with marital status, sex and age held constant.

With the exception of living arrangements, none of the assumptions on the relationship between income reporting and characteristics were supported by the data on total beneficiaries. The younger beneficiaries have a greater net error than the beneficiaries aged 73 and over; the upper-income, larger benefit and higher education groups have greater net error rates than the lower income and education groups. The absolute differences, however, were relatively small and are not statistically significant for the Education and Income characteristics. Respondents with "no answer" on income, however, do have a significantly larger response bias.

The response error rates for beneficiary unit types, in table 4, are generally consistent with those for all beneficiary units, and provide little evidence of selective bias in the reporting of benefit income. Only in living arrangements and age are the percent bias relationships consistent and substantial.

In table 5, age appears to account for most of the bias variation in the benefit income groups, with the 62-64 year age group having biases two to three times that of the older age groups. This is also the age group most likely to have irregular payments because of recent benefit awards and interruptions in payments.

Respondents whose total income was unreported also tended to have among the highest rates of bias, even when age and marital status were controlled.

4. Effect of response error on the standard error

The survey estimate of the standard error of the mean, $41, includes sampling variability, but no response variability. Without the adjustment for bias, from the benefit record, the apparent standard error would have been approximately the same or slightly smaller than the standard error of the record data mean. The mean square error, or "true standard error," however, would have been almost twice that of the apparent standard error if there had been no correction for bias:

Survey estimate of standard error of mean corrected for bias.....	$\sigma_{\bar{x}}$	$41
Standard error of bias...........	$\sigma_{\bar{e}}$	5
True standard error of reported mean (mean square error).......	$\sqrt{MSE}$	74*

$(MSE = \sigma_{\bar{x}} + \sigma_{\bar{e}} + 2r\,\sigma_{\bar{x}}\,\sigma_{\bar{e}} + (Bias)^2)$

* Assuming the coefficient of correlation, $r \leq \pm .5$

DISCUSSION

On the basis of these data, we can say with reasonable confidence that the reporting of Social Security benefit income is relatively accurate in surveys of the aged. With few exceptions, respondents reported benefit income with great accuracy. The bias, or net error was small and was confined to understatements by 5 to 7 percent of the reporting units, rather than to constant error or consistent under-reporting. The bias, moreover, was relatively consistent among education and income groups and was selectively distributed to a marked degree only by age.

The extent of understatement of benefit income was also consistent with the findings of a variety of other income studies. The 1950 Census underestimated the National Income accounts by 9 percent.6/ As Guthrie reported last year, the 1960 Consumer Expenditures study underestimated income by 6 percent and the Michigan Consumer Finance Surveys ranged from 3 percent to 13 percent in income underestimates during 1947-1955.7/

Aside from the immediate purposes of the Aged Survey, the data also have some general implications for methods of controlling response bias. The data strongly suggest that income reporting is not a problem of the sensitivity of the instrument. There was little tendency towards "yard-stick" error or consistent individual under-reporting of benefit income. The bias was due almost entirely to a small proportion of cases with gross under-statements. It was also associated with incomplete reporting in other financial areas. A good part of this would be accounted for by people who did not report beneficiary status, probably because of suspended payments and other changes in their beneficiary status. Others may have omitted lump-sum and retroactive payments.

As Ferber concluded from his study of time deposits, it is doubtful that methods designed to increase the accuracy of response overall would be effective for those income reports. Most respondents are already providing adequate answers. The deviant case is the one which presents the problem.

One possibility would be the use of questions directed specifically towards recall of unusual or one-time payments. A good part of the net error came from those respondents who did not answer other income questions, however, and it is doubtful that much more could be obtained from them through the refinement of questioning techniques.

Adjustment procedures which take into account not only the demographic characteristics of non-respondents, but also non-response as an attribute may provide more effective means of accounting for response bias. This will, of course, require more knowledge of the special attributes of non-responders and more validation research. Non-responders on total income, for example, had a mean benefit income equal to the high and middle income thirds. Adjustments for non-response which take this into account as an attribute should, therefore, produce better estimates.

The extent to which these findings may be generalized to other questions and other population can, of course, only be determined by further study. Our research plans include a variety of validation studies related to Social Security record data. We are, for example, preparing a similar analysis for earnings from the 1963 Aged Survey. Data on earnings and benefit income reporting will be available for the under age 65 disabled population, from the 1966 Survey of Disabled Adults. Comparison data on the reliability of diagnostic information will also be developed. The proposed Longitudinal Survey Retirement should provide data on the temporal nature of response error.

The developing computer technology and access to large-scale data systems are extending the possibilities for validation studies and much more research may now be done on response error. It is to be hoped that the growing body of research will lead to a more standard treatment of response estimation procedures and that in the future these procedures will be made available in at least as much detail as is now published for sampling procedures.

References

1/ See, for example, William G. Madow, "On Some Aspects of Response Error Measurement," 1965 Proceedings of the Social Statistics Section, American Statistical Association, Washington, D.C. pp. 182-192, 1965; John B. Lansing, Gerald P. Ginsburg and Kaisa Braaten, An Investigation of Response Error, Studies in Consumer Savings No. 2, 1961, pp. 188-204; and William G. Cochran, Sampling Techniques, 2nd ed., John Wiley and Sons, New York, 1963, pp. 374-389.

2/ The procedures for record matching are described by Lawrence D. Haber, "Methodological Analyses in the 1963 Survey of the Aged," paper presented at the Gerontological Society meeting, October 1964, Minneapolis, Minnesota; the sample design is summarized in "Technical Note on Source and Reliability of the Estimates for the 1963 Survey of the Aged," Social Security Bulletin, July 1964, pp. 26-28.

3/ The benefit income reported in the study also included minor editing adjustments to resolve differences arising from exclusion of children's benefits and changes in marital status.

4/ Selma F. Goldsmith, "Low-Income Families and Measures of Income Inequality," Review of Social Economy, 20:1 (March 1962), pp. 1-19; and "The Relation of Census Income Distribution Statistics to Other Income Data," An Appraisal of the 1950 Census Income Data, vol. 13, Studies in Income and Wealth, National Bureau of Economic Research, Princeton, 1958, pp. 70-83.

5/ See, for example, Robert Ferber, "The Reliability of Consumer Surveys of Financial Holdings: Time Deposits," J. American Statistical Association, 60:309 (March 1965), pp. 148-163; and Lansing, op.cit., pp.180-181.

6/ Herman P. Miller, Income of the American People, John Wiley & Sons, New York, 1955, p. 15.

7/ Harold W. Guthrie, "Some Methodological Issues in Validation Studies," 1965 Proceeding, op. cit., pp. 193-196.

Table 1.-- Distribution of response errors for OASDI benefit income by type of beneficiary unit: Percent of beneficiary units aged 62 and over

Reporting difference (SSA - Interview)	Beneficiary units: Total	Married couples	Nonmarried: Men	Nonmarried: Women
Sample N...........	4,727	2,137	713	1,877
Percent............	100	100	100	100
Under-reporting				
$500 or more...........	6	8	4	6
301-500................	3	3	2	3
101-300................	4	5	3	4
61-100................	2	2	2	2
Matching report (±$60)....	77	73	82	79
Exact..................	46	39	74	44
±1-60..................	31	34	8	35
Over-reporting				
61-100..................	2	2	2	2
101-300.................	4	6	4	2
301-500.................	1	1	1	1
501 or more.............	1	1	1	1
Percent of units Net response error (NRE) ($61 or more).........	8	8	3	9
Mean (dollars)				
Total benefit income (TBI).................	$1,052	$1,351	$911	$764
Net under-reported..	$61	$82	$29	$50
Percent of TBI......	5.8	6.1	3.2	6.6

Table 2.--Percent distribution of response errors: Units and aggregate dollars

Reporting difference (SSA-Interview)	Total beneficiary units 62 and over: Percent of units: Total (Under plus over-reporting)	Percent of units: Net under-reporting (Under minus over-reporting)	Net under-reported dollars: Percent of mean TBI	Net under-reported dollars: Percent of bias
Base..........	4,727	4,727	$1,052	$61
Percent.......	100	100	100	100
Size of difference				
$501+	7	5	4.8	82
301-500......	4	2	0.8	14
101-300......	8	1	0.2	2
61-100.......	4	*	0.0	*
1-60.........	31	2	0.1	2
None.........	46	---	---	---

* Less than one-half of one percent.

Table 3.--Mean Total Benefit Income and Response Bias: Total Beneficiary Units

Selected Characteristics	Number of Units	Mean Total Benefit Income	Net Under-report	
			Mean Bias	Percent of TBI
Total (62 & over)	4,727	$1,052	$61	5.8
Age				
62-64	449	834	99	11.9
65-72	2,309	1,095	71	6.5
73 & over	1,969	1,050	41	3.9
Living Arrangements				
In household-no relatives	2,969	1,099	47	4.3
In household-with relatives	1,702	1,980	82	8.4
Institutionalized	56	717	194	27.0
Education				
Less than 9 years	2,884	1,049	58	5.6
9-11 years	556	1,060	47	4.5
12 years or more	979	1,106	73	6.6
NA	308	892	79	8.8
OASDI Benefit Income				
$0-499	755	416	2	0.5
500-999	1,752	765	35	4.6
1,000-1,499	1,364	1,220	68	5.5
1,500-1,999	568	1,769	96	5.4
2,000-2,499	244	2,133	163	7.7
2,500 & over	44	2,885	916	31.7
Income Terciles a/ (65 & over)	4,278	$1,075	$57	5.3
Lowest	1,072	898	29	3.2
Middle	1,401	1,135	46	4.0
Highest	1,323	1,132	54	4.7
NA	480	1,136	165	14.5

a/ Terciles were defined for each array of unit types separately and represent the position of the unit among couples, non-married men and non-married women.

Table 4 --Mean total benefit income and response bias by type of beneficiary unit

Specified characteristics	Number of units			Mean total benefit income(TBI)			Mean net under-reported			Percent of TBI under-reported		
	Married couples	Nonmarried		Married couples	Nonmarried		Married couples	Nonmarried		Married couples	Nonmarried	
		Men	Women		Men	Women		Men	Women		Men	Women
Total (62 and over).........	2,137	713	1,877	$1,351	$911	$764	$82	$29	$50	6.1	3.2	6.6
Age												
62-64.....................	204	42	203	991	**	696	166	**	42	16.8	**	6.1
65-72.....................	1,114	296	899	1,357	978	809	95	44	51	7.0	4.5	6.4
73 and over...............	819	375	775	1,433	878	729	43	15	51	3.0	1.7	7.0
Living arrangements												
In household-no relatives.	1,559	401	1,009	1,355	921	775	61	17	37	4.5	1.8	4.7
In household-with relatives	575	292	835	1,346	909	753	137	39	60	10.2	4.3	7.9
In institution............	3	20	33	**	**	**	**	**	**	**	**	**
Education												
Less than 9th grade.......	1,383	465	1,036	1,333	894	739	73	21	56	5.5	2.3	7.5
9-11th grade..............	248	71	237	1,343	980	788	70	36	27	5.2	3.7	3.4
12th grade or more........	417	104	458	1,453	991	816	126	20	37	8.6	2.0	4.5
NA........................	89	73	146	1,193	842	734	54	85	91	4.5	10.1	12.4
OASDI Benefit income												
$0-499.....................	154	129	472	351	431	433	-31	17	8	-9.0	3.9	2.0
500-999....................	467	290	995	764	784	760	38	17	39	5.0	2.2	5.1
1,000-1,499................	676	286	402	1,264	1,230	1,139	56	34	112	4.4	2.7	9.8
1,500-1,999................	557	6	5	1,771	**	**	86	**	**	4.9	**	**
2,000-2,499................	240	1	3	2,133	**	**	157	**	**	7.3	**	**
2,500 or more..............	43	1	---	**	**	---	**	**	---	**	**	---
Income terciles (Total 65 and over).............	1,933	671	1,674	1,390	922	772	73	28	51	5.3	3.0	6.6
Lowest.....................	568	168	336	1,165	678	555	28	22	34	2.4	3.2	6.1
Middle.....................	615	241	545	1,528	938	780	68	20	32	4.5	2.2	4.1
Highest....................	509	213	603	1,465	1,088	866	75	24	46	5.2	2.2	5.3
NA.........................	241	49	190	1,406	**	837	186	**	153	13.2	**	18.3

** Mean and percents not computed for bases of less than 50 cases.

Table 5.--Mean total benefit income and response bias by selected characteristics and age of head of beneficiary unit

Specified characteristics	Number of units			Mean total benefit income(TBI)			Mean net under-reported			Percent of TBI under-reported		
	Age of head			Age of head			Age of head			Age of head		
	62-64	65-72	73 and over	62-64	65-72	73 and over	62-64	65-72	73 and over	62-64	65-72	73 and over
Total (62 and over)........	449	2,306	1,969	$834	$1,095	$1,050	$99	$71	$41	11.9	6.5	3.9
Living arrangements												
In household-no relatives.	274	1,460	1,235	800	1,133	1,125	65	56	31	8.1	5.0	2.8
In household-with relatives	173	834	695	887	1,035	937	154	94	50	17.4	9.1	5.3
In institution...........	2	15	39	**	**	**	**	**	**	**	**	**
Education												
Less than 9th grade.......	268	1,415	1,201	836	1,103	1,032	90	71	36	10.8	6.4	3.5
9-11th grade..............	68	302	186	877	1,096	1,069	109	51	19	12.4	4.6	1.8
12th grade or more........	90	474	415	835	1,100	1,171	157	69	59	18.7	6.2	5.1
NA........................	23	118	167	**	983	858	**	143	51	**	14.5	6.0
OASDI Benefit income												
$0-499....................	137	314	304	324	402	472	-23	7	8	-7.1	1.8	1.6
500-999...................	166	801	785	749	767	767	70	29	34	9.3	3.8	4.4
1,000-1,499...............	110	748	506	1,180	1,223	1,225	113	79	41	9.6	6.4	3.4
1,500-1,999...............	25	276	267	**	1,771	1,766	**	118	46	**	6.7	2.6
2,000-2,499...............	6	137	101	**	2,132	2,130	**	189	125	**	8.9	5.9
2,500 or more.............	5	33	6	**	**	**	**	**	**	**	**	**
Income terciles												
Lowest....................	---	449	623	---	911	888	---	49	14	---	5.3	1.6
Middle....................	---	756	645	---	1,181	1,082	---	51	40	---	4.3	3.7
Highest...................	---	840	485	---	1,106	1,176	---	75	16	---	6.8	1.4
NA........................	---	264	216	---	1,130	1,142	---	158	173	---	14.0	15.1

** Mean and percent not computed for bases of less than 50 cases.

RECENT RESEARCH IN REINTERVIEW PROCEDURES

Barbara A. Powell
U.S. Bureau of the Census

1. Introduction.

Reinterview procedures are used extensively by the Bureau of the Census for evaluating the quality of censuses and surveys. In particular after both the 1950 Census and the 1960 Census, a large-scale reinterview program was conducted to evaluate the quality of census results. There are two kinds of errors which can affect census data -- coverage errors and content errors. Coverage errors result from persons having been missed or having been counted more than once. Content errors result from the assignment of persons to incorrect classifications in the census tabulations on characteristics of persons who were counted. The reinterview programs conducted after the 1950 and 1960 Censuses were designed, in part, to measure content errors.

However, the kind of reinterview procedure used can itself affect the measurement of the quality of the original results. The purpose of this paper is to evaluate the different kinds of reinterview procedures used in measuring the quality of the 1960 Census and to identify a "best" procedure for use with a census or a nonrecurrent survey. In evaluating the reinterview procedures, two main problems are discussed:

a. The effect of the time lag between the census or survey and the reinterview -- A question frequently asked is whether the time lag between the census and reinterview has a deteriorating effect on the reinterview data. In order to get a partial answer to this question we tried to estimate the effect of having the reinterview three months rather than six months after the Census.

b. The effect of the reinterviewers having access to the original responses -- In many reinterview situations, the reinterview results and the original results are "reconciled." The reinterview responses are compared with the original responses for identical persons and where differences exist, an effort is made, with the help of the respondent, to decide upon the proper response. The reconciliation process may take place at the same time as the reinterview or it may take place at a later time. In the case where the reconciliation is to be done immediately following the reinterview, the reinterviewer is given the original results and told to conduct the reinterview without looking at the original responses. He is then required to compare the responses and to reconcile any differences. He is instructed not to change either the census or the reinterview response, but to enter the reconciled response in a separate place. However, it is thought that the accessibility of the original responses has an effect on the reinterview data. We tried to estimate the effect on the reinterview data and on the measurement of the quality of the census results.

In reference to the first problem, the time lag, our data show that for most characteristics -- age, school enrollment, and number of children -- the additional three months' time lag had no identifiable effect on the data. However, for mobility and income items, a reinterview closer in time to the original interview produced better results.

In reference to the second problem, the accessibility of the original responses, our data show that for characteristics such as age, mobility, type and level of school, and number of children, the situation where the reinterviewer had access to the original results had no identifiable effect on the reinterview data. For school enrollment, educational attainment, and income items, this kind of reinterview procedure had a decided effect on the reinterview data.

In summary, for the purpose of measuring the quality of a census or a nonrecurrent survey, the best kind of reinterview procedure to use is one in which the reinterview is close in time to the census or survey and one in which the reinterviewers are not given access to the original responses.

2. The Model.

The mathematical model underlying this study was developed by Hansen, Hurwitz and Bershad [2]. In this model, the term "survey" is used for either a census or a survey. The survey is regarded as being repeatable, under the same general conditions, in such a way that repetitions relate to the same point in time and such that the results of any one trial are not influenced by any earlier trial. A single census, then, is viewed as a random sample of one trial from among such a set of repetitions, even though, in practice, independent repetitions of the census may be impossible. A reinterview is also viewed as a sample from among this set of repetitions.

With these assumptions, x_{jtG} is defined as a random variable whose value is as follows:

$x_{jtG} = 1$ if the sample person, j, has the characteristic of interest in trial t in a survey conducted under general conditions, G.
$= 0$ otherwise.

Assuming an equal probability selection method, an estimate of the proportion of the population having this characteristic is:

$$p_{tG} = \frac{1}{n_t} \sum_j^{n_t} x_{jtG} \qquad (1)$$

where n_t is the number of persons in the sample in trial t, and G specifies the general conditions under which the survey takes place. The general conditions are composed of several factors -- the kind of questionnaire used, the training and instructions to interviewers, the method of payment of the interviewers, the time of the year during which the interviews were conducted, the sponsor of the survey, and other related items.

a. The gross difference rate and simple response variance[1].

Let $E\, p_{tG} = P_G$ be the expected value over all samples of persons and all trials. Now think of repeated measurements on one person in the population, say the j-th person. This conditional expected value is:

$$E\, x_{jtG} = P_{jG} \qquad (2)$$

The response deviation for a given sample person is:

$$d_{jtG} = x_{jtG} - P_{jG}, \qquad (3)$$

the difference between the measurement for the j-th person on the t-th trial and the expected value for that person over all trials.

Using this notation, Hansen, Hurwitz and Bershad defined the response variance as:

$$\sigma^2_{\bar{d}G} = \frac{1}{n} E\, d^2_{jtG} + \frac{(n-1)}{n} E(d_{jtG} d_{ktG}) \qquad (4)$$

Now, $\sigma^2_{dG} = E(d_{jtG} - E\, d_{jtG})^2$ by definition. From equation (3), it is obvious that $E\, d_{jtG} = 0$. So

$$\sigma^2_{dG} = E\, d^2_{jtG}. \qquad (5)$$

Also, $\rho_{dG} = \dfrac{E(d_{jtG} d_{ktG})}{\sigma^2_{dG}}$. With these definitions, equation (4) becomes:

$$\sigma^2_{\bar{d}G} = \frac{\sigma^2_{dG}}{n} + \frac{(n-1)}{n} \rho_{dG} \sigma^2_{dG} \qquad (6)$$

[1] The discussion of the gross-difference rate and simple response variance is based on The Estimation and Interpretation of Gross Differences and the Simple Response Variance, by Hansen, Hurwitz and Pritzker.

In the first term is the simple response variance which reflects the basic trial-to-trial variability in response. In the second term is the correlated response variance which reflects the correlation among the response deviations within a trial. In this study, we are interested only in the simple response variance.

From equation (5),

$$\sigma^2_{dG} = E\, d^2_{jtG} = E(x_{jtG} - P_{jG})^2 = E\, x^2_{jtG} + E\, P^2_{jG} - 2\, E P_{jG} x_{jtG}. \qquad (7)$$

Since x_{jtG} is a zero-one variate, $x^2_{jtG} = x_{jtG}$. Then the expected value over trials for a fixed person is

$$\begin{aligned} \sigma^2_{dG} &= E\, x_{jtG} + E\, P^2_{jG} - 2\, E P_{jG} x_{jtG} \\ &= E\, P_{jG} + E\, P^2_{jG} - 2\, E P^2_{jG} \\ &= E(P_{jG} - P^2_{jG}) \end{aligned} \qquad (8)$$

Now, when these values are averaged over all persons in the population, the simple response variance is:

$$\sigma^2_{dG} = \frac{1}{N} \sum_j^N P_{jG}(1 - P_{jG}). \qquad (9)$$

We are now interested in getting an estimate of σ^2_{dG}. For each person included in a reinterview study, we have the original census measurement (x_{jtG}) as well as the reinterview measurement ($x_{jt'G'}$). For each person, we can get the difference, $x_{jtG} - x_{jt'G'}$. Then, let

$$g = \frac{1}{n} \sum_j^n (x_{jtG} - x_{jt'G'})^2 \qquad (10)$$

$$E(g) = \frac{n}{n} E(x_{jtG} - x_{jt'G'})^2 \qquad (11)$$

$$= E(x^2_{jtG}) + E(x^2_{jt'G'}) - 2\, E(x_{jtG} x_{jt'G'}) \qquad (12)$$

Since x_{jtG} and $x_{jt'G'}$ are zero-one variates, $x^2_{jtG} = x_{jtG}$ and $x^2_{jt'G'} = x_{jt'G'}$.

Then, $E(g) = E x_{jtG} + E x_{jt'G'} - 2 E x_{jtG} x_{jt'G'}$ (13)

If the survey conditions, G and G', are the same, and if the two trials, t and t', are independent, then equation (13) becomes:

$$E(g) = 2\, E\, P_{jG} - 2\, E(P^2_{jG}) \qquad (14)$$

$$= 2\left[\frac{1}{N} \sum_j^N P_{jG} - \frac{1}{N} \sum_j^N P^2_{jG}\right]$$

$$E(g) = \frac{2 \sum_{j}^{N} P_{jG}(1 - P_{jG})}{N} \qquad (15)$$

Thus, $E(g) = 2\,\sigma^2_{dG}$.

However, as pointed out after equation (13), $E(g) = 2\,\sigma^2_{dG}$ is true only when G and G' are the identical survey conditions and the census and reinterview measurements on identical persons are independent. Hansen, Hurwitz and Pritzker [3] point out that g/2 will be a poor estimator of σ^2_{dG} whenever there is a large positive correlation among the response deviations on the census and the reinterview. The reason for this is as follows. From equation (11) we have

$$E(g) = E(x_{jtG} - x_{jt'G'})^2$$

$$= E[(x_{jtG} - P_{jG}) - (x_{jt'G'} - P_{jG'}) + (P_{jG} - P_{jG'})]^2 \qquad (16)$$

$$= E(x_{jtG} - P_{jG})^2 + E(x_{jt'G'} - P_{jG'})^2 - 2E(x_{jtG} - P_{jG})(x_{jt'G'} - P_{jG'}) + E(P_{jG} - P_{jG'})^2 \qquad (17)$$

and the remaining two cross-product terms vanish. Then,

$$E(g) = \sigma^2_{dG} + \sigma^2_{dG'} - 2\rho_{dG,dG'}\sigma_{dG}\sigma_{dG'} + E(P_{jG} - P_{jG'})^2 \qquad (18)$$

where

$$\rho_{dG,dG'} = \frac{E(x_{jtG} - P_{jG})(x_{jt'G'} - P_{jG'})}{\sigma_{dG}\sigma_{dG'}}$$

Now, if the census conditions, G, and the reinterview conditions, G', are identical $P_{jG} = P_{jG'}$, and the last term will drop out. Also $\sigma^2_{dG} = \sigma^2_{dG'}$, so that

$$E(g) = 2\sigma^2_{dG} - 2\rho_{dG,dG'}\sigma^2_{dG}$$

$$= 2\sigma^2_{dG}(1 - \rho_{dG,dG'}) \qquad (19)$$

Therefore, where the reinterview measurements are not independent of the census measurements, g/2 is an understatement of the simple response variance by the amount of the between-trial covariance of response deviations. Estimates of this covariance are presented in Section 4.

Disregarding this covariance term, let us see how σ^2_{dG} can be estimated from the data available. The diagram below shows the results of the comparison of the census data with the reinterview data.

Diagram A.--COMPARISON OF RESULTS OF CENSUS AND REINTERVIEW FOR IDENTICAL PERSONS

Reinterview results	Census results		
	$x_{jtG} = 1$	$x_{jtG} = 0$	Total
$x_{jt'G'} = 1$	a	b	a+b
$x_{jt'G'} = 0$	c	d	c+d
Total	a+c	b+d	n

From equation (10) we have

$$g = \frac{1}{n}\sum_{j}^{n} x^2_{jtG} + \frac{1}{n}\sum_{j}^{n} x^2_{jt'G'} - \frac{2}{n}\sum_{j}^{n} x_{jtG}x_{jt'G'}$$

$$= \frac{1}{n}\sum_{j}^{n} x_{jtG} + \frac{1}{n}\sum_{j}^{n} x_{jt'G'} - \frac{2}{n}\sum_{j}^{n} x_{jtG}x_{jt'G'} \qquad (20)$$

Substituting the appropriate values shown in Diagram A for the quantities in equation (20) we have:

$$g = \frac{a+c}{n} + \frac{a+b}{n} - \frac{2a}{n}$$

$$= \frac{b+c}{n} \qquad (21)$$

This is the equation for the gross-difference rate.

b. Net-difference rate.

In evaluating a census statistic, we are interested in the square of the bias as well as the variance of that statistic. If we view the reinterview as providing a standard measurement, then the bias of a census statistic is the expected value of the census measurement minus the expected value of the reinterview measurement. (The reinterview may provide better measurements for items such as income, where the respondent is asked many detailed questions. He may tend to report things that he did not think of when answering the census question.) The bias then is $E\,p_{tG} - E\,p_{t'G'}$. The estimate of bias is $p_{tG} - p_{t'G'}$. However,

$$p_{tG} - p_{t'G'} = \frac{1}{n}\sum_{j}^{n} x_{jtG} - \frac{1}{n}\sum_{j}^{n} x_{jt'G'} \qquad (22)$$

Using the notation of Diagram A, equation (22) becomes

$$\frac{a+c}{n} - \frac{a+b}{n} = \frac{c-b}{n} \qquad (23)$$

This term, which is an estimate of the bias of the census statistic, is referred to as the net-difference rate.

Using the gross- and net-difference rates as estimators of the simple response variance and bias, respectively, we shall try to evaluate different kinds of reinterview procedures.

3. The Study Design.

The 1960 Census provided an opportunity to evaluate different kinds of reinterview procedures. Each person selected in one of the reinterview samples was a person who had been enumerated in the census. In fact, the person had been included in the 25 percent of the population who had been asked to give census information on migration, education, number of children, labor force, and income. So the original interview was the census interview taken under prevailing census conditions.

Because the reinterview samples were much smaller than the census, they could be handled on a more intensive basis. First, the reinterviewers were hired on a more selective basis than the census interviewers. Second, the training of these reinterviewers was carefully done by a few Washington personnel. Third, the reinterview questionnaire was a very detailed one. From a combination of superior interviewers with intensive training on a detailed questionnaire, we hoped to get answers which could be regarded as being of higher quality than those obtained in the census.

In order to evaluate the different types of reinterview procedures, three independent, multistage probability samples of the 1960 Census enumerated population were selected.

Sample I was a selection of 396 1960 Census Enumeration Districts (EDs) from a selection of 148 primary sampling units (PSUs).[2] Within each ED, a cluster of housing units was selected and all the persons within the housing units were included in the reinterview sample. The sample consisted of about 4,900 persons in 1,450 housing units.

A second sample of 1,003 EDs in 268 PSUs was selected. This sample was split into two parts, one housing unit being designated as Sample II and the next housing unit in the sample being designated as Sample III. Each sample was composed of about 5,450 persons in 1,650 housing units.

Following the sample selection, persons to do the reinterviewing were hired. Interviewers from the Current Population Survey (CPS) were given first priority. These people were part of a permanent staff of highly trained and closely supervised interviewers. If a CPS interviewer was not available, a census interviewer who had been recommended as doing a particularly good job in the census was hired. The reinterviewers were given intensive training on the reinterview questionnaire by Bureau of the Census personnel. This was in contrast to the training of the census interviewers which was several times removed from the original Washington training personnel.

In addition to the differences in the kinds of interviewers hired, the training given, and the type of questionnaire, there were also other differences between the census and the reinterview situations. For example, information at the reinterview was obtained from the best respondent in the household (usually the person himself), whenever possible, instead of just any responsible member of the household, as in the census. Also, interviewers were supervised more closely than in the census, and their pay was on an hourly basis rather than on a piece-rate basis.

In July, 1960, the reinterviews for Sample I were conducted. The reinterviewers had no knowledge of the census responses for the sample persons. After the field work was completed, a comparison of the census and reinterview responses for identical persons was made by regular census personnel. Where differences were found, a census subject-matter specialist reviewed the case. It was that specialist who decided whether a reconciliation of the census and reinterview answers was to be attempted by another interview with the sample person. However, even if there was a large discrepancy, the case was not always reconciled. The reconciliation was to be done in October, at the same time as the reinterviews for the other two samples. If the reconciliation case was in an area distant from the areas where the other samples were located, the reconciliation was not attempted.

In October, 1960, the reinterviews for Samples II and III were conducted. In Sample II, the reinterviewers were supplied with the census data for the sample persons. They were instructed to complete the reinterview questionnaire, then to look at the census responses and try to reconcile any differences between the census and reinterview answers. This was the "on-the-spot" reconciliation process.

In Sample III, the reinterviewers were not supplied with the census responses for the sample persons. They were instructed to complete the reinterview and leave the household. No reconciliation of census and reinterview data was ever attempted.

From the three samples, we have available five sets of data as shown below.

Samples	Unreconciled responses	Reconciled responses
Sample I	D_{11} (July)	D_{12} (October)
Sample II	D_{21} (October)	D_{22} (October)
Sample III	D_3 (October)	---

2 The 3,103 counties and independent cities in the United States were combined into 1,891 PSUs, each PSU being one or more contiguous counties. Three hundred and thirty-three of these PSUs were included in the Current Population Survey at the time of the 1960 Census. It was from these 333 that the sample of 148 PSUs was selected.

D_{11} refers to data from Sample I, before reconciliation; D_{12} to Sample I after reconciliation; D_{21} to Sample II before reconciliation, and so forth. The data fall into three categories: (1) gross-difference rates, (2) net-difference rates, and (3) standard errors of differences between gross-difference rates or between net-difference rates.

By making appropriate comparisons among the five sets of data, we have evaluated the reinterview procedures. However, there are limitations to the data. Some of these are as follows:

a. There was a difference among the samples in the number of non-interview cases. The percentage of non-interview cases by sample is shown below:

Sample I	6.2%
Sample II	11.4%
Sample III	10.1%

The non-interview rate for Sample I looks considerably lower than for Samples II and III. The non-interview rate for Sample I was originally very high. All households for which there were no responses in July were included with the enumeration of Sample II in October. Therefore, some Sample I cases were given two chances for responding.

b. Some differences among the samples arose in processing the data. Samples I and II were processed at the same time. All census, original reinterview, and final, reconciled, reinterview answers were coded to special FOSDIC[3] data sheets, which were, in turn, converted to magnetic tapes. A series of detailed computer edits were performed on the data in order to insure the quality of the transcription and coding process. Specially trained clerks checked the original documents for cases failing edit. A correction process was instituted for the cases needing correction after edit.

Two years later, data for Sample III were transcribed and coded. Instructions were altered to take into account the lack of reconciliation in Sample III. The same computer edits were performed on the coded data and, where possible, the same kind of edit correction process was carried out. However, the two-year gap, revised instructions, and different clerks may have caused a change in the results for some items.

All characteristics were studied carefully for possible processing differences. One item which was known to exhibit differences due to processing was omitted from the analysis which follows. With the exception of that item, the remaining characteristics did not show any differences which were known to have been caused by processing.

c. Because the data are based on samples, comparisons among them are limited by the sample size. Perhaps some differences which exist among the sets of data are not apparent because of the sample size.

[3] FOSDIC stands for Film Optical Sensing Device for Input to Computers.

d. Where there was no reconciliation of census and reinterview data, we assumed independence between the reinterview and census interview. The assumption is not correct due to the "conditioning" effect of repeated interviewing of the same household. (See [6] for some results on conditioning effects on collection of expenditures data.) However, the conditioning effect is probably small in comparison with the reconciliation effect.

4. Estimation of the Between-Trial Covariance.

As mentioned in Section 2, the estimates of the simple response variance provided by g/2 are poor whenever there is a high correlation among the response deviations on the census and on the reinterview. With the five sets of data available it is possible to get an estimate of an upper-bound for this between-trial covariance under different reinterview procedures.

Suppose we are given that G' is an improved procedure over G. In general, G' is an improved procedure over G when:

$$\text{for} \quad P_{jG} > .5 \qquad P_{jG'} > P_{jG}$$
$$\text{for} \quad P_{jG} < .5 \qquad P_{jG'} < P_{jG} \tag{24}$$

Suppose, that G reflected the census conditions and G' the conditions of the reinterview.

Let g_{11} be the gross-difference rate estimated from the comparison of census responses with the responses from Sample I before reconciliation. Let g_{22} be the gross-difference rate estimated from the comparison of census responses with the reconciled Sample II responses. We shall view the reconciled reinterview as an improved procedure. Using equation (11) we have

$$\begin{aligned} E(g_{22}) &= E(x_{jtG} - x_{jt'G'})^2 \\ &= E[(x_{jtG}-P_G)-(x_{jt'G'}-P_{G'})+(P_G-P_{G'})]^2 \\ &= E(x_{jtG}-P_G)^2 + E(x_{jt'G'}-P_{G'})^2 + (P_G-P_{G'})^2 \\ &\quad - 2E(x_{jtG} - P_G)(x_{jt'G'} - P_{G'}) \end{aligned} \tag{25}$$

We have seen that

$E(x_{jtG}-P_G)^2 = P_G(1-P_G)$ and $E(x_{jt'G'}-P_{G'})^2 = P_{G'}(1-P_{G'})$. Then, adding and subtracting P_{jG} and $P_{jG'}$ in the last term of equation (25):

$$\begin{aligned} E(g_{22}) &= P_G(1-P_G) + P_{G'}(1-P_{G'}) + (P_G-P_{G'})^2 \\ &\quad - 2E(x_{jtG}-P_{jG})(x_{jt'G'}-P_{jG'}) \\ &\quad - 2E(P_{jG}-P_G)(P_{jG'}-P_{G'}) \end{aligned} \tag{26}$$

Since

$$\rho_{dG,dG'}\sigma_{dG}\sigma_{dG'} = E(x_{jtG}-P_{jG})(x_{jt'G'}-P_{jG'}),$$

the between-trial covariance of response deviations,

$$\rho_{dG,dG'}\sigma_{dG}\sigma_{dG'} = \tfrac{1}{2}[P_G(1-P_G)+P_{G'}(1-P_{G'})+(P_G-P_{G'})^2 - E(g_{22})]-E(P_{jG}-P_G)(P_{jG'}-P_{G'}) \quad (27)$$

Hansen, Hurwitz and Pritzker [3] show that

$$E(P_{jG}-P_G)(P_{jG'}-P_{G'}) > \sigma^2_{P_{jG}}+(.5-P_G)(P_{G'}-P_G) \quad (28)$$

where $\sigma^2_{P_{jG}} = E(P_{jG} - P_G)^2$

is the sampling variance. Since the total variance is the sum of response variance and the sampling variance

$$\sigma^2_{P_{jG}} = P_G(1 - P_G) - \sigma^2_{dG}. \quad (29)$$

Then,

$$\rho_{dG,dG'}\sigma_{dG}\sigma_{dG'} < \tfrac{1}{2}[P_G(1-P_G)+P_{G'}(1-P_{G'})+(P_G-P_{G'})^2 - E(g_{22})] - P_G(1-P_G) + \sigma^2_{dG} - (.5-P_G)(P_{G'}-P_G) \quad (30)$$

When all terms are multiplied out, equation (30) becomes:

$$\rho_{dG,dG'}\sigma_{dG}\sigma_{dG'} < \sigma^2_{dG} - \tfrac{1}{2}E(g_{22}) \quad (31)$$

From the data of Sample I before reconciliation, a good estimate of σ^2_{dG} can be made. This estimate is:

$$\tfrac{1}{2}E(g_{11}) = \sigma^2_{dG}. \quad (32)$$

Therefore, an upper bound for the between-trial covariance among response deviations can be estimated by:

$$\rho_{dG,dG'}\sigma_{dG}\sigma_{dG'} < \tfrac{1}{2}E(g_{11} - g_{22}) \quad (33)$$

We are able to get several estimates of this between-trial covariance since we have five sets of data. The sets can be ordered, in an arbitrary way, by the degree of dependence between the original interview and the reinterview. The assumption is made that the two sets of data before reconciliation are less dependent than the two sets of data after reconciliation and that the data from an "on-the-spot" reconciliation process are the most dependent. Diagram B shows the ordering.

Diagram B.
ORDERING OF SETS OF DATA BY ANTICIPATED DECREASING DEGREE OF DEPENDENCE OF REINTERVIEW ON CENSUS INTERVIEW

Set of data	Survey conditions	Gross-difference rate
Sample II: After reconciliation	G_{22}	g_{22}
Sample I: After reconciliation	G_{12}	g_{12}
Sample II: Before reconciliation	G_{21}	g_{21}
Sample I: Before reconciliation	G_{11}	g_{11}
Sample III:	G_3	g_3

G_{11} and G_3 were both survey conditions in which the reconciliation process did not affect the reinterview results. $g_{11}/2$ was selected to provide the estimate of σ^2_{dG} in the estimation of the upper bound for the between-trial covariance.

G_{21} was a set of survey conditions in which the reinterviewer had the census data with him during the reinterview. The estimate of the between-trial covariance provided by $(g_{11}-g_{21})/2$ will show the degree of dependence due to the reinterviewers having access to the census data. Similarly, $(g_{11}-g_{12})/2$ will show the effect of an independent reconciliation process; and $(g_{11}-g_{12})/2$ will show the effect of an "on-the-spot" reconciliation process.

Let us look at the results of this kind of comparison. Table I which follows shows this kind of comparison for two characteristics -- educational attainment and other income, females. The upper bounds for the covariance appear in columns (6), (7) and (8). The ratios shown in columns (9), (10) and (11) are estimates of $\rho_{dG,dG'}$ if the simple response variances under all the varying reinterview conditions are the same. We would expect the estimates from Sample II after reconciliation to exhibit the largest covariance estimates. This holds true for eight of the 15 educational attainment items and ten of the 15 income items.

Another thing apparent from this limited comparison is that, in some instances, Sample II before reconciliation exhibits more dependence (a higher covariance estimate) than does Sample I after reconciliation. For eight of the 15 education items, and 11 of the 15 income items, the covariance from Sample II before reconciliation is at least as large or larger than the covariance from Sample I after reconciliation. This is an indication that the reinterviewers having access to the census data, even before reconciling differences, provides a dependent situation. This corresponds to the findings in the Current Population Survey. Practically speaking, if g/2 is used as an estimate of the simple response variance, where g is estimated from a reinterview situation where the reinterviewers have access to the original responses, the value will be underestimated.

It may be of more interest to compare estimates of the upper bound for the between-trial covariance for an item as a whole, rather than for each category within an item. In order to get a gross-difference rate per item, the gross-difference rates over all categories within an item were averaged. Table 2 which follows shows the average gross-difference rates, estimates of the upper bounds for the between-trial covariances, and ratios of the estimates of the upper bounds to the estimate of the simple response variance for an item.

Table 1.--COMPARISON OF ESTIMATES OF THE UPPER BOUNDS OF THE BETWEEN-TRIAL COVARIANCE AMONG RESPONSE DEVIATIONS FOR DIFFERENT DEGREES OF DEPENDENCE BETWEEN CENSUS AND REINTERVIEW

Characteristics	Estimated gross-difference rates: Sample I: Unreconciled g_{11} (1)	Sample I: Reconciled g_{12} (2)	Sample II: Unreconciled g_{21} (3)	Sample II: Reconciled g_{22} (4)	Estimated simple response variance $g_{11}/2$ (5)	Estimated upper bounds of covariance: $\frac{(g_{11}-g_{12})}{2}$ (6)	$\frac{(g_{11}-g_{21})}{2}$ (7)	$\frac{(g_{11}-g_{22})}{2}$ (8)	Ratio of estimators: (6)÷(5) (9)	(7)÷(5) (10)	(8)÷(5) (11)
Educational Attainment:											
No school..............	.0132	.0094	.0120	.0104	.0066	.0019	.0006	.0014	.29	.09	.21
Elementary 1-2 years....	.0156	.0118	.0128	.0118	.0078	.0019	.0014	.0019	.24	.18	.24
3-4 years....	.0378	.0306	.0346	.0320	.0189	.0036	.0016	.0029	.19	.08	.15
5-6 years....	.0558	.0456	.0424	.0392	.0279	.0051	.0067	.0083	.18	.24	.30
7 years....	.0574	.0498	.0520	.0462	.0287	.0038	.0027	.0056	.13	.09	.20
8 years....	.1080	.0940	.0742	.0746	.0540	.0070	.0169	.0167	.13	.31	.31
High School 1 year.....	.0712	.0542	.0450	.0444	.0356	.0085	.0131	.0134	.24	.37	.38
2 years....	.0662	.0504	.0554	.0506	.0331	.0079	.0054	.0078	.24	.16	.24
3 years....	.0460	.0388	.0500	.0480	.0230	.0036	-.0020	-.0010	.16	-.09	-.04
4 years....	.0804	.0674	.0796	.0764	.0402	.0065	.0004	.0020	.16	.01	.05
College 1 year.....	.0346	.0270	.0236	.0230	.0173	.0038	.0055	.0058	.22	.32	.34
2 years....	.0354	.0294	.0180	.0190	.0177	.0030	.0087	.0082	.17	.49	.46
3 years....	.0202	.0160	.0130	.0116	.0101	.0021	.0036	.0043	.21	.36	.43
4 years....	.0216	.0204	.0152	.0140	.0108	.0006	.0032	.0038	.06	.30	.35
5 years or more....	.0154	.0122	.0104	.0059	.0077	.0016	.0025	.0047	.21	.32	.61
Other Income, Females:											
No income..............	.1604	.1482	.1314	.1152	.0802	.0061	.0145	.0226	.08	.18	.28
$1 to $499 or loss......	.1382	.1238	.1014	.0894	.0691	.0072	.0184	.0244	.10	.27	.35
$500 to $999............	.0724	.0674	.0544	.0470	.0362	.0025	.0090	.0127	.07	.25	.35
$1,000 to $1,499........	.0284	.0272	.0332	.0282	.0142	.0006	-.0024	.0001	.04	-.17	.01
$1,500 to $1,999........	.0224	.0192	.0162	.0146	.0112	.0016	.0031	.0039	.14	.28	.35
$2,000 to $2,499........	.0136	.0132	.0116	.0104	.0068	.0002	.0010	.0016	.03	.15	.24
$2,500 to $2,999........	.0092	.0092	.0034	.0026	.0046	.0000	.0029	.0033	.00	.63	.72
$3,000 to $3,499........	.0026	.0026	.0026	.0024	.0013	.0000	.0000	.0001	.00	.00	.08
$3,500 to $3,999........	.0024	.0024	.0012	.0010	.0012	.0000	.0006	.0007	.00	.50	.58
$4,000 to $4,499........	.0010	.0004	.0000	.0006	.0005	.0003	.0005	.0002	.60	1.00	.40
$4,500 to $4,999........	.0026	.0020	.0004	.0004	.0013	.0003	.0011	.0011	.23	.85	.85
$5,000 to $5,999........	.0006	.0006	.0012	.0012	.0003	.0000	-.0003	-.0003	.00	-1.00	-1.00
$6,000 to $6,999........	.0006	.0006	.0000	.0000	.0003	.0000	.0003	.0003	.00	1.00	1.00
$7,000 to $9,999........	.0000	.0000	.0012	.0012	.0000	.0000	-.0006	-.0006	.00	undefined	undefined
$10,000 and over........	.0000	.0000	.0006	.0006	.0000	.0000	-.0003	-.0003	.00	undefined	undefined

Table 2.--ESTIMATES OF THE AVERAGE GROSS-DIFFERENCE RATES AND OF UPPER BOUNDS FOR BETWEEN-TRIAL COVARIANCE BY ITEM FOR DIFFERENT DEGREES OF DEPENDENCE BETWEEN CENSUS AND REINTERVIEW

Characteristics	Estimated average gross-difference rates				Estimated simple response variance	Estimated upper bounds of covariance			Ratio of estimators		
	Sample I		Sample II								
	Unreconciled	Reconciled	Unreconciled	Reconciled							
	g_{11}	g_{12}	g_{21}	g_{22}	$g_{11}/2$	$\frac{(g_{11}-g_{12})}{2}$	$\frac{(g_{11}-g_{21})}{2}$	$\frac{(g_{11}-g_{22})}{2}$	(6)÷(5)	(7)÷(5)	(8)÷(5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Sex.........................	.0121	.0081	.0092	.0096	.0060	.0020	.0014	.0012	.33	.23	.20
Color.......................	.0082	.0051	.0041	.0037	.0041	.0016	.0020	.0022	.39	.49	.54
Male age....................	.0049	.0043	.0059	.0055	.0024	.0003	-.0005	-.0003	.12	-.21	-.12
Female age..................	.0050	.0046	.0055	.0052	.0025	.0002	-.0002	-.0001	.08	-.08	-.04
White age...................	.0046	.0043	.0049	.0047	.0023	.0002	-.0002	.0000	.09	-.09	.00
Nonwhite age................	.0077	.0063	.0123	.0110	.0038	.0007	-.0023	-.0016	.18	-.60	-.42
White male age..............	.0045	.0040	.0049	.0049	.0022	.0002	-.0002	-.0002	.09	-.09	-.09
White female age............	.0047	.0045	.0047	.0045	.0024	.0001	.0000	.0001	.04	.00	.04
Nonwhite male age...........	.0065	.0059	.0124	.0109	.0032	.0003	-.0030	-.0022	.09	-.94	-.69
Nonwhite female age.........	.0081	.0059	.0123	.0112	.0040	.0011	-.0021	-.0016	.28	-.52	-.40
1955 residence..............	.0248	.0220	.0308	.0264	.0124	.0014	-.0030	-.0008	.11	-.24	-.06
Type and level school.......	.0176	.0164	.0222	.0130	.0088	.0006	-.0023	.0023	.07	-.26	.26
School enrollment...........	.0246	.0214	.0394	.0145	.0123	.0016	-.0074	.0050	.13	-.20	.41
Educational attainment......	.0453	.0371	.0359	.0338	.0226	.0041	.0047	.0058	.18	.21	.26
Number children.............	.0153	.0124	.0132	.0126	.0076	.0014	.0010	.0014	.18	.13	.18
Total income, all...........	.0512	.0499	.0453	.0425	.0256	.0006	.0030	.0044	.02	.12	.17
Total income, males.........	.0549	.0568	.0515	.0496	.0274	-.0010	.0017	.0026	-.04	.06	.09
Total income, females.......	.0449	.0427	.0387	.0359	.0224	.0011	.0031	.0045	.05	.14	.20
Self-employed income, males..	.0169	.0160	.0154	.0136	.0084	.0004	.0008	.0016	.05	.10	.19
Self-employed income, females	.0039	.0035	.0040	.0031	.0020	.0002	.0000	.0004	.10	.00	.20
Other income, males.........	.0383	.0355	.0312	.0277	.0192	.0014	.0036	.0053	.07	.19	.28
Other income, females.......	.0303	.0278	.0239	.0210	.0152	.0012	.0032	.0046	.08	.21	.30

Some things are immediately apparent from Table 2.

a. The independent reconciliation process is an improved procedure over a process in which there is no reconciliation at all. This is illustrated by the figures in column (6) which show positive estimates of the between-trial covariance for all but one item.

b. Sample II age classifications behaved in a peculiar way, especially for nonwhite groups. This was probably due to some kind of processing difference.

c. For 13 of the 22 items, data from Sample II after reconciliation produced the highest covariance estimates. If the age items are omitted, data from Sample II after reconciliation produced the highest estimates for 12 of the 14 remaining items.

d. Again, omitting the age items, the estimates of covariance are higher for eight of the 14 remaining items for the Sample II data before reconciliation than the Sample I data after reconciliation.

It seems clear from the data above that the way to get a good estimate of the simple response variance is to use unreconciled data from a procedure where the reinterviewers had no knowledge of the original responses.

5. Effect of Time Lag Between Census and Reinterview.

It was thought that a reinterview that occurred several months after the original interview might not provide data as accurate as could be obtained by having the reinterview closer to the original interview. In this part of the study we were able to compare data based on a match of reinterview and census data for identical persons when the reinterview was three months after the census interview and when the reinterview was six months after the census interview.

As explained in Section 3, the persons in Sample I were reinterviewed in July, 1960 -- about three months after the census interview; the persons in Sample III were reinterviewed in October, 1960 -- about six months after the census interview. In neither of these samples did the reinterviewer have access to the census data. We can compare g_{11} and g_3 (gross-difference rates) and b_{11} and b_3 (net-difference rates) to measure the effect of the time lag.

Ideally we would like to be able to identify a reinterview procedure which is "better" than other reinterview procedures. However, determining what "better" is presents difficulties. We would prefer a reinterview procedure that had a smaller simple response variance than any other procedure. Let us return to some of the estimators of response variance in Section 2 for help in identifying what a "better" procedure is. The expected value of the gross-difference rate from equation (17) was:

$$E(g) = E(x_{jtG}-P_{jG})^2+E(x_{jt'G'}-P_{jG'})^2+E(P_{jG}-P_{jG'})^2 - 2E(x_{jtG}-P_{jG})(x_{jt'G'}-P_{jG'}) \quad (17)$$

Adding and subtracting the same term and rearranging the terms gives

$$\begin{aligned} E(g) = {} & E(x_{jtG}-P_{jG})^2 + E(x_{jt'G'}-P_{jG'})^2 \\ & - 2E(x_{jtG}-P_{jG})(x_{jt'G'}-P_{jG'}) \\ & + [E(P_{jG}-P_{jG'})]^2 + E(P_{jG}-P_{jG'})^2 \\ & - [E(P_{jG}-P_{jG'})]^2 \end{aligned} \quad (34)$$

The third term from the end is equal to the square of the bias $(B)^2$. The last two terms together are equal to the variance of $(P_{jG}-P_{jG'})$.

Equation (34) may be expressed as:

$$E(g) = \sigma^2_{dG} + \sigma^2_{dG'} - 2\rho_{dG,dG'}\sigma_{dG}\sigma_{dG'} + (B)^2 + Var(P_{jG} - P_{jG'}) \quad (35)$$

Then, making the assumption that $Var(P_{jG}-P_{jG'})$ is very small, we can express the expected value of the gross-difference rate from Sample I as:

$$E(g_{11}) = \sigma^2_{dG}+\sigma^2_{dG_{11}} -2\rho_{dG,dG_{11}}\sigma_{dG}\sigma_{dG_{11}} +B^2_{11} \quad (36)$$

The expected value of the gross-difference rate from Sample III can be expressed in a similar fashion. However, for both these samples, the between-trial covariance is relatively small, since the reinterviewer did not have access to the census data. So, for the purposes of this section, we will drop the covariance term. Then we have:

$$\begin{aligned} E(g_{11}) &= \sigma^2_{dG} + \sigma^2_{dG_{11}} + B^2_{11} \\ E(g_3) &= \sigma^2_{dG} + \sigma^2_{dG_3} + B^2_3 \end{aligned} \quad (37)$$

We have estimates of g_{11}, g_3, B_{11} and B_3 from the samples. Using those estimates in the following equations:

$$\begin{aligned} E(g_{11}) - B^2_{11} &= \sigma^2_{dG} + \sigma^2_{dG_{11}} \\ E(g_3) - B^2_3 &= \sigma^2_{dG} + \sigma^2_{dG_3} \end{aligned} \quad (38)$$

we see that if $E(g_{11}) - B^2_{11}$ is greater than $E(g_3) - B^2_3$ this implies that the simple response variance of Sample I is greater than the simple response variance of Sample III.

We see, then, that if the gross-difference rates for the two samples are the same for a given item but one net-difference rate is much larger than the other, the procedure which produces the larger net-difference rate is the "better" procedure. Similarly, if the net-difference rates for a given item are about the same for both procedures, the procedure producing the smaller gross-difference rate is the "better" procedure.

In order to compare the gross- and net-difference rates of Sample I with those of Sample III the variances of the gross and net differences were computed. It has been shown [1] that an estimate of the sampling variance of the gross-difference rate is:

$$\mathrm{Var}(g) = g/n - g^2/n \qquad (39)$$

The assumptions for this estimate to be valid are: (1) simple random sampling, (2) independence of the two measurements on the elements, and (3) uncorrelated response deviations. These conditions are not fully met, so the estimates of variances will be underestimates. Since g^2/n is small in comparison with g/n, the last term was not used in computing the variance. The sampling variance of the net-difference rate is also g/n. Then, since Samples I and III are independent, the sampling variance of the difference between the gross-difference rates (or the net-difference rates) is the sum of the estimated variances for each of the samples.

The results of the comparison of gross- and net-difference rates for Samples I and III showed that for most items, we could detect no difference between Samples I and III. However, for some items, some very interesting differences were found.

First, gross-difference rates for all items for both samples were compared and then the net-difference rates were compared. The type of problem that arose in making the comparisons is shown in Table 3 which follows. In this table the gross-difference rates for "nonwhite male age" were compared. Consider the category "50 to 54 years." The difference between the gross-difference rates, as shown in column (5), was -.0004. This seems like a very small difference and we might immediately decide the two reinterview procedures did not differ very much for that item. However, the standard error of that difference, in column (6), is .0099. With a standard error that large, it is impossible for us to make any definitive statement about the difference between the gross-difference rates.

Now, look at the cateogy "40 to 44 years." Here the difference between gross-difference rates is -.0166. This looks like a very large difference and we might conclude that the two reinterview procedures are producing very different results for that category. However, a glance at the standard error of the difference, .0103, shows that even though the estimated difference is large, the standard error of the difference is so large that again we can't make any definitive statment about the difference between the gross-difference rates.

If two reinterview procedures actually produced gross-difference rates as different as those shown in Table 3, we would like to be able to identify the procedure producing lower gross-difference rates. However, in every case, we must say that the sample size is not large enough to permit us to identify differences between the gross-difference rates. The same kind of situation occurred for all age items studied. Since we were also not able to differentiate between the net-difference rates for age items, we have no reason to prefer one procedure over the other, at least for age items.

Table 3.--COMPARISON OF GROSS-DIFFERENCE RATES FOR SAMPLES I AND III FOR NONWHITE MALE AGE

Category	g_{11} [1] (1)	$s_{g_{11}}$ (2)	g_3 [2] (3)	s_{g_3} (4)	$g_{11}-g_3$ (5)	$s_{g_{11}-g_3}$ (6)
0 to 4 years.............	.0162	.0079	.0084	.0060	.0078	.0099
5 to 9 years.............	.0114	.0066	.0084	.0060	.0030	.0089
10 to 14 years............	.0076	.0054	.0000	.0000	.0076	.0054
15 to 19 years............	.0124	.0068	.0000	.0000	.0124	.0068
20 to 24 years............	.0000	.0000	.0082	.0060	-.0082	.0060
25 to 29 years............	.0090	.0059	.0166	.0085	-.0076	.0103
30 to 34 years............	.0042	.0040	.0124	.0073	-.0082	.0084
35 to 39 years............	.0042	.0040	.0082	.0060	-.0040	.0072
40 to 44 years............	.0042	.0040	.0208	.0095	-.0166	.0103
45 to 49 years............	.0152	.0076	.0168	.0085	-.0016	.0114
50 to 54 years............	.0118	.0067	.0122	.0073	-.0004	.0099
55 to 59 years............	.0098	.0069	.0082	.0060	.0016	.0085
60 to 64 years............	.0056	.0046	.0042	.0043	.0014	.0063
65 to 69 years............	.0056	.0046	.0000	.0000	.0056	.0046
70 to 74 years............	.0000	.0000	.0040	.0042	-.0040	.0042
75 to 79 years............	.0000	.0000	.0040	.0042	-.0040	.0042
80 to 84 years............	.0000	.0000	.0042	.0043	-.0042	.0043
85 years or over..........	.0000	.0000	.0042	.0043	-.0042	.0043

1 Gross-difference rate from Sample I where reinterview was in July, 1960.

2 Gross-difference rate from Sample III where reinterview was in October, 1960.

The next item studied was "1955 residence." For this item there were very large differences between the gross-difference rates for the two samples for every category but one. These estimated differences were also large in comparison with their standard errors as is shown below.

Table 4.--COMPARISON OF GROSS-DIFFERENCE RATES FOR SAMPLES I AND III FOR 1955 RESIDENCE

Category	g_{11}[1] (1)	g_3[2] (2)	$g_{11}-g_3$ (3)	$s_{g_{11}-g_3}$[3] (4)
Same house........	.0434	.0628	-.0194	.0090
Different house, same county.....	.0506	.0772	-.0266	.0099
Different county, same State......	.0126	.0274	-.0148	.0056
Different State...	.0136	.0266	-.0130	.0056
Abroad............	.0038	.0036	.0002	.0023

1 Gross-difference rate from Sample I where reinterview was in July, 1960.

2 Gross-difference rate from Sample III where reinterview was in October, 1960.

3 $s_{g_{11}-g_3}$ estimated by $\sqrt{\frac{g_{11}}{n_{11}} + \frac{g_3}{n_3}}$ was multiplied by 1.8 to account for sampling households rather than persons.

The differences between the net-difference rates for this item were not large in comparison to their standard errors. For this item, $\sigma^2_{dG_3}$ is larger than $\sigma^2_{dG_{11}}$. Therefore we see that by having the reinterview six months rather than three months after the census interview, the simple response variance was increased.

For two of the education items (educational attainment and school enrollment) we are again in the position of not being able to detect differences between the gross-difference rates or the net-difference rates. For the item on "type and level of school" Sample III gross-difference rates are higher. Let

$$\phi_{11} = g_{11} - b^2_{11}$$

$$\phi_3 = g_3 - b^2_3$$

Table 5 below shows ϕ_{11} and ϕ_3 for this item.

Table 5.--COMPARISON OF ϕ_{11} AND ϕ_3 FOR TYPE AND LEVEL OF SCHOOL

Category	ϕ_{11}	ϕ_3
Public elementary............	.0322	.0552
Private elementary...........	.0263	.0355
Public high school...........	.0228	.0374
Private high school..........	.0068	.0059
Public college...............	.0138	.0167
Private college..............	.0064	.0069

The values of ϕ were not affected much by the bias terms, so are based mostly on the gross-difference rates. Notice that ϕ_3 is larger than ϕ_{11} for all but one category. Thus, the reinterview later in time produced larger simple response variances.

Turning now to the income items, we find some very interesting differences. Most of the differences occur in the "no income" or "$1-$499 or loss" categories. There were no differences found in the self-employment income tables. Table 6 shows the items for which there was a large difference between the gross-difference rates or the net-difference rates.

Table 6.--COMPARISON OF GROSS- AND NET-DIFFERENCE RATES FOR SPECIFIC INCOME CATEGORIES

Item and category	g_{11} (1)	g_3 (2)	b_1 (3)	b_3 (4)	ϕ_{11} (5)	ϕ_3 (6)	$s_{g_{11}-g_3}$ or $s_{b_{11}-b_3}$ (7)
Total Income, All:							
No income..................	.1182	.1210	.0574	.0045	.1149	.1210	.0085
$1 to $499 or loss.........	.1012	.1142	-.0311	-.0021	.1002	.1142	.0081
Total Income, Male:							
No income..................	.0814	.0668	.0272	.0002	.0807	.0668	.0097
$5,000 to $5,999...........	.0386	.0874	-.0017	-.0093	.0386	.0873	.0089
Total Income, Female:							
No income..................	.1530	.1678	.0874	.0108	.1454	.1677	.0137
$1 to $499 or loss.........	.1408	.1552	-.0484	.0000	.1385	.1552	.0131
Other Income, Male:							
No income..................	.1812	.1876	.1239	.0651	.1658	.1838	.0156
$1 to $499 or loss.........	.1770	.1760	-.0990	-.0462	.1672	.1739	.0152
Other Income, Female:							
No income..................	.1604	.1532	.1010	.0492	.1502	.1508	.0137

Look at the first two categories. The gross-difference rates are not too different in comparison with their standard errors, but the net-difference rates are very different. The estimate of the bias from Sample I is very large. This pattern holds for all but one of the items shown in the table.

Notice that all but one of the items are "no income" or "$1 to $499 or loss." For each of these items, having the reinterview closer in time to the original interview permitted the detection of large biases in these items. Therefore, having the reinterview closer in time seems to be important in identifying persons with small amounts of income that are not reported in the census.

In summary, then, for most items we were not able to detect differences between the two reinterview procedures. In the cases in which the sample size was large enough to detect differences between procedures, the Sample I procedure, before reconciliation, seemed to be a "better" procedure than the Sample III procedure. So at least for mobility, type and level of school, and income items, we prefer a reinterview procedure which specifies that the reinterview be closer in time to the original interview. For other items, we do not yet have enough evidence to prefer one of these procedures over the other.

6. Effect of a Dependent Reinterview Situation.

There is evidence from the Current Population Survey (CPS) that an interviewing situation in which the reinterviewers are provided with the original responses does not insure independence of the reinterview data. A sample of the households included in the CPS every month is selected for reinterview. For 20 percent of the reinterview sample, the original CPS data are not supplied to the reinterviewer; for the remaining 80 percent, the original CPS data are supplied to the reinterviewer. A comparison of the results of these two samples shows that the gross-difference rates for the 80 percent sample were about one-half of the gross-difference rates for the 20 percent sample. (See Technical Paper No. 6, The Current Population Survey Reinterview Program.)

In the conference on Evaluation of the 1960 Censuses held on February 6-7, 1959, it was stated that an independent reinterview gives a more valid gross-difference rate and also a better indication of the bias. For this reason, it was planned for the evaluation program to provide a comparison between a completely independent reinterview and this potentially dependent reinterview. Sample II, for which the reinterviews were conducted at the same time of the year as the Sample III reinterviews, would provide such a comparison since, in Sample II, the reinterviewers were provided with the original census data. The comparisons which follow are based on the data from Sample II before reconciliation and Sample III. The comparisons are made between g_{21} and g_3 and between b_{21} and b_3.

We would like to be able to make the same kind of comparisons with these data as we did with Samples I and III data. However, the dependent reinterview situation somewhat complicates the matter. From equation (36) we see

$$E(g_{21}) = \sigma^2_{dG} + \sigma^2_{dG_{21}} - 2\rho_{dG,dG_{21}}\sigma_{dG}\sigma_{dG_{21}} + B^2_{21}.$$

Therefore, we have a comparison of the following quantities:

$$\begin{cases} E(g_{21}) - B^2_{21} = \sigma^2_{dG} + \sigma^2_{dG_{21}} - 2\rho_{dG,dG_{21}}\sigma_{dG}\sigma_{dG_{21}} \\ E(g_3) - B^2_3 = \sigma^2_{dG} + \sigma^2_{dG_3}. \end{cases}$$

We see that if $E(g_{21}) - B^2_{21}$ is greater than $E(g_3) - B^2_3$, then the simple response variance from Sample II is greater than that from Sample III. However, if $E(g_3) - B^2_3$ is greater than $E(g_{21}) - B^2_{21}$, we don't know if $\sigma^2_{dG_{21}}$ is smaller than $\sigma^2_{dG_3}$ unless we can estimate the size of the covariance term. Therefore, for the comparisons which follow, if we find differences of the latter kind, we will estimate the covariance term by $(g_{11} - g_{21})/2$ where g_{11} is the gross-difference rate from Sample I.

For the age items, there were no large differences between either the gross-or the net-difference rates for Samples I and II. The same thing occurred for the "1955 residence" item and "type and level of school." However the "school enrollment" item showed rather large differences between Samples I and II. Table 7 summarizes these differences.

Table 7.--COMPARISON OF GROSS- AND NET-DIFFERENCE RATES FOR SAMPLES II AND III FOR SCHOOL ENROLLMENT

Category	g_3 (1)	g_{21} (2)	b_3 (3)	b_{21} (4)	ϕ_3 (5)	ϕ_{21} (6)	$s_{g_3-g_{21}}$ or $s_{b_3-b_{21}}$ (7)
Kindergarten or elementary 1	.0158	.0288	-.0099	.0180	.0157	.0285	.0067
2	.0332	.0576	-.0014	.0062	.0332	.0576	.0097
3	.0386	.0670	.0059	.0013	.0386	.0670	.0099
4	.0432	.0690	.0032	-.0035	.0432	.0690	.0101
5	.0408	.0580	-.0021	-.0055	.0408	.0580	.0097
6	.0380	.0550	-.0130	.0093	.0378	.0549	.0093
7	.0370	.0560	.0101	-.0119	.0369	.0559	.0089
8	.0266	.0344	.0042	-.0060	.0266	.0344	.0081
High school 1	.0164	.0488	.0086	.0093	.0163	.0487	.0091
2	.0276	.0590	-.0176	-.0082	.0273	.0589	.0086
3	.0306	.0484	.0071	.0071	.0305	.0483	.0082
4	.0248	.0400	-.0019	-.0051	.0248	.0400	.0078
College.... 1	.0198	.0230	.0014	-.0098	.0198	.0229	.0059
2	.0072	.0064	.0072	-.0003	.0071	.0064	.0038
3	.0042	.0102	-.0042	-.0008	.0042	.0102	.0046
4	.0092	.0070	-.0025	.0000	.0092	.0070	.0034
Five or more	.0048	.0018	.0048	.0000	.0048	.0018	.0022

The categories for which there were large differences between the samples in either the gross- or net-difference rates were underlined. Note that for all categories below the college level, ϕ_{21} was greater than ϕ_3. For some reason, this dependent procedure produced larger simple response variances.

For the educational attainment item the differences went in the other direction and the covariance term was estimated. We still compute ϕ_{21} and ϕ_3 but then add $2\rho_{dG,dG_{21}}\sigma_{dG}\sigma_{dG_{21}}$ to ϕ_{21} before comparing it with ϕ_3. If ϕ_{21} plus the covariance term is still smaller than ϕ_3, we assume that the simple response variance of Sample II is smaller than that of Sample III. Table 8 shows these comparisons. The categories for which the gross or net differences for the two samples were very large in comparison with the standard errors of the differences are underlined. However, for 14 of the 15 items ϕ_{21} was smaller than ϕ_3.

Table 8.--COMPARISON OF GROSS- AND NET-DIFFERENCE RATES FOR SAMPLES II AND III FOR EDUCATIONAL ATTAINMENT

Category	ϕ_{21} (1)	$\frac{g_{11}-g_{21}}{2}$ (2)	$\phi_{21}+2\left(\frac{g_{11}-g_{21}}{2}\right)$ (3)	ϕ_3 (4)
No school....	.0120	.0006	.0132	.0124
Elementary...				
1-2	.0128	.0014	.0156	.0197
3-4	.0346	.0016	.0378	.0432
5-6	.0424	.0067	.0558	.0718
7	.0520	.0027	.0574	.0708
8	.0742	.0169	.1080	.1043
High school..				
1	.0450	.0131	.0712	.0571
2	.0554	.0087	.0728	.0714
3	.0500	[1]	[1]	.0478
4	.0796	.0004	.0804	.0806
College......				
1	.0236	.0055	.0346	.0256
2	.0180	.0087	.0354	.0362
3	.0130	.0036	.0202	.0146
4	.0152	.0032	.0216	.0210
Five or more	.0104	.0025	.0154	.0132

[1] Estimate of covariance was negative.

Now the values in column (3) are almost the same as the gross-difference rates estimated from Sample I for those underlined. When we compared the gross-differences from Samples I and III in the previous section we found no significant differences. Therefore, after accounting for the covariance term we see that the simple response variances from Samples II and III are not really different. However, the gross-difference rate is substantially reduced in Sample II and should not be used as an estimate of the simple response variance.

Turning to the income items, we found a few categories for which there seemed to be large differences between the gross-difference rates or net-difference rates. For these items, even after accounting for the covariance term, as shown in Table 9, the simple response variances were smaller for Sample II, except for the "no income" category for "total income, males."

Table 9.--COMPARISON OF GROSS- AND NET-DIFFERENCE RATES FOR SOME INCOME CATEGORIES

Item and category	ϕ_{21} (1)	$\frac{g_{11}-g_{21}}{2}$ (2)	$\phi_{21}+2(\frac{g_{11}-g_{21}}{2})$ (3)	ϕ_3 (4)
Total Income, All:				
No income..................	.1069	.0046	.1161	.1210
$1 to $499 or loss..........	.0813	.0099	.1013	.1142
Total Income, Males:				
No income..................	.0800	.0000	.0800	.0668
Total Income, Females:				
$1 to $499 or loss..........	.1120	.0142	.1404	.1552
Self-employment Income, Males:				
$1 to $499 or loss..........	.0202	.0054	.0310	.0438

We see that in a dependent reinterview procedure, for most items there was no reduction in the simple response variance. However, there were significant reductions in the estimated gross-difference rate due to a large positive between-trial covariance. If the gross-difference rates from this procedure were used for estimating the simple response variance, the estimates would be too low.

REFERENCES

[1] Bershad, Max A. (1964). Comparison of the Variances of the Net and Gross Error Rates. Unpublished memorandum.

[2] Hansen, Morris H., Hurwitz, William N., and Bershad, Max A. (1961). Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute, vol.38, No. 2, pp. 359-374.

[3] Hansen, Morris H., Hurwitz, William N., and Pritzker, Leon (1964). The Estimation and Interpretation of Gross Differences and the Simple Response Variance. Contributions to Statistics Presented to Professor P.C. Mahalanobis on the Occasion of His 70th Birthday. Pergamon Press, Calcutta, India.

[4] U.S. Bureau of the Census (1963). The Current Population Survey Reinterview Program, Some Notes and Discussion. Technical Paper No. 6, Washington, D.C.

[5] U.S. Bureau of the Census (1964). Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by Reinterviews. Series ER60, No. 4. Washington, D.C.

[6] U.S. Bureau of the Census (1965). Response Errors in Collection of Expenditures Data by Household Interviews: An Experimental Study. Technical Paper No. 11. Washington, D.C.

CHANGING THE PROBABILITIES OF SELECTION WHEN TWO UNITS ARE SELECTED WITH PPS WITHOUT REPLACEMENT

Ivan P. Fellegi
Dominion Bureau of Statistics, Ottawa

Introduction

Let the population consist of N units and assume that a "measure of size", m_i (i=1, 2, ..., N), is associated with each unit. A general problem of selecting n units with probability proportional to size (pps) without replacement is to devise sampling schemes such that π_i, the probability that the i-th unit is in the sample is proportional to m_i, i.e.

$$\pi_i = nm_i/\Sigma m_i \tag{1}$$

Setting
$$p_i = \frac{m_i}{\Sigma m_i} \tag{2}$$

(1) can be rewritten as $\pi_i = np_i$ (3)

The present author proposed a scheme [2] of successively selecting n units without replacement in such a way that the probability δ_i^k of selecting the i-th unit at the k-th selection equals p_i:

$$\delta_i^k = p_i \qquad i=1,2,\ldots,N;\ k=1,2,\ldots,n \tag{4}$$

Since the events of selecting unit i at any two draws are mutually exclusive, it follows that

$$\pi_i = \sum_{k=1}^{n} \delta_i^k = np_i$$

This general method of selection has some particularly useful features for n=2. A method has been outlined in [2] (and a FORTRAN programme is available) to solve the following system of equations for q_i(i=1,2,...,N) and a:

$$aq_i(1-q_i) = 2p_i \qquad i=1,2,\ldots,N \tag{5}$$

$$\sum_{i=1}^{N} q_i = 1 \tag{6}$$

$$a = 2/(1 - \sum_{i=1}^{N} q_i^2) \tag{7}$$

The selection of two units is then carried out* by selecting the first unit with probabilities $\{p_i;\ i=1,2,\ldots,N\}$ and having selected unit i at the first draw, selecting the second unit from among the remaining units with probabilities $\{q_j/(1-q_i);\ j=1,2,\ldots,N;\ j\neq i\}$. The probability of selecting units i and j in that order is, therefore, equal to

$$p_i \frac{q_j}{1-q_i} \tag{8}$$

while the probability of selecting the same units in reverse order is equal to

$$p_j \frac{q_i}{1-q_j} \tag{9}$$

It follows from (5) that the expressions under (8) and (9) are both equal to

$$\frac{1}{2} aq_iq_j = p_i \frac{q_j}{1-q_i} = p_j \frac{q_i}{1-q_j} \tag{10}$$

and hence π_{ij}, the probability of selecting units i and j in either order is equal to

$$\pi_{ij} = aq_iq_j \tag{11}$$

Consequently, given that units i and j are in the sample, the conditional probability of selecting them in either order equals 1/2. Formula (4) follows from formulae (5) to (11).

The problem posed in the present paper is the following. Suppose a sample of two units had been selected as described above, with probabilities proportional to the measures of size $\{m_i\}$. The two units might, for example, be two primary sampling units (areas) and the measures of size might be census counts. Some years later a new census is taken, resulting in new measures of size M_i. It is desirable to select a sample of two psu's with the following properties: I) The probabilities of selection are proportional to the new measures of size. II) Since, however, the original sample of two psu's often represents a capital investment (listings of households might have been prepared, enumerators trained, etc.), it is desirable to maximize the overlap between the old and new sample. More precisely, if X is a random variable (depending on the sampling scheme for the selection of the new sample), such that

X = 0 if the new sample coincides with the old
= 1 if the new and old samples have one unit in common

---

* Brewer [1] proved that equations (5) to (7) always have a unique solution if only $p_i < \frac{1}{2}$ (i=1,2,...,N).

= 2 if the new and old samples are disjoint

then it is desirable to have a sampling scheme for the selection of the new sample for which

$$E(X) = \text{expected number of rejections}$$

is minimized. III) it is also desirable to have the joint probabilities of selecting two units in the new sample satisfy equations (4) to (11) in terms of P_i, Q_i, A and Π_{ij} (defined in terms of the new size measures in analogy to p_i, q_i, a and π_{ij}) so that the process of revising the size measures in the future can be repeated again (actually property III implies I).

The procedure outlined below achieves I) and III) and achieves II) approximately.

The procedure

Keyfitz [3] provided a procedure to achieve the objectives outlined above when the sample consists of one unit. In this simple case sampling with and without replacement become indistinguishable and property I) automatically implies III). Since, however, the present procedure follows the methods of Keyfitz, it is instructive to recapitulate them here briefly.

A sample of one unit had been selected with probabilities $\{p_j\}$. The desired new probabilities are $\{P_j\}$. The procedure of changing the probabilities of selection is summarized as follows: first the original unit is subjected to a test of retention; it is retained with probability

$$R_j$$

rejected with probability

$$1-R_j$$

In case of rejection, the second step consists of a subsequent selection. The new probability of selecting unit j is therefore

$$P_j = p_j R_j + S_j$$

S_j being the probability of selecting unit j at the second step. The objective is to maximize R_j (or minimize S_j) within the equation above. Clearly if

$$R_j < 1 \qquad \text{and} \qquad S_j > 0$$

the procedure is not optimal, since R_j could be further increased, S_j further decreased. In this case the probability of rejection is positive

$$1-R_j > 0$$

Hence a procedure is not optimal if it permits the rejection of a unit followed by its subsequent re-selection. The converse is also true: a procedure is optimal if the set of units with a positive probability of rejection and the set of units with a positive probability of subsequent selection are disjoint. The Keyfitz procedure is optimal in this sense. The set of units with positive probability of rejection is the set of decreasing units for which

$$p_j > P_j$$

None of these units can be selected subsequently. The set of units with positive probability of subsequent selection is the set of units which increased, i.e. for which

$$p_j \leqq P_j$$

Formally

$$I = \{i : p_i \leqq P_i\} \quad \text{and} \quad D = \{d : p_d > P_d\} \qquad (12)$$

The Keyfitz procedure consists of retaining the originally selected unit if it is in I. If the originally selected unit (d) is in D, then it is retained with probability $R_d = P_d/p_d$, rejected with probability $1-P_d/p_d$. Finally, if a unit in D is rejected, then a unit in I is selected with probabilities

$$S_i = \frac{P_i - p_i}{\sum_{i\varepsilon I} (P_i - p_i)} \qquad i\varepsilon I \qquad (13)$$

It is easy to show that the probability of unit j being in the sample after the adjustment is equal to $P_j (j=1,2,\ldots,N)$ and that the procedure is optimal in the sense of property II) above (it minimizes the expected number of rejections). These properties are based on the observation that since

$$\sum_{j=1}^{N} p_j = \sum_{j=1}^{N} P_j = 1$$

therefore

$$\sum_{d\varepsilon D} (p_d - P_d) = \sum_{i\varepsilon I} (P_i - p_i) \qquad (14)$$

The present procedure starts with the observation that the sampling procedure by which the original two units were selected is symmetric [2] i.e. the conditional probability of the two units having been selected in either order is equal to 1/2. Consequently if one does not know the order in which the units had been selected, one can "recreate" the order at random, e.g. by selecting a random number r ($0 < r \leqq 1$) and assuming that if $0 < r \leqq 0.5$ than the order in which units i and j had been selected was higher subscript first, lower subscript second; if $0.5 < r \leqq 1$ then the order in which the units had been selected was the reverse.

After this step we may assume that we have an ordered sample of two units (i first, j second) whose joint probability of selection was

$$ {}_i\pi_j = p_i \frac{q_j}{1-q_i} = \frac{1}{2} aq_iq_j \tag{15} $$

Next we observe that the (unconditional) probability of unit i having been selected first was equal to p_i. Since (4) is to be maintained (in terms of the new measures) after the adjustment, the probability of unit i to be selected as the first should be equal to P_i. The Keyfitz procedure can be applied to this end without modification: the two sets I and D, defined under (12), are formed and the Keyfitz rules for retention, rejection and reselection are applied.

After the adjustment of the probability of selection of the first unit the probability of selection of the second unit is adjusted. Here the Keyfitz principle has to be modified slightly. To illustrate this point, consider the sets

$$ I_i = \{j:\ j\neq i \text{ and } \frac{q_j}{1-q_i} \leqq \frac{Q_j}{1-Q_i}\} \qquad i=1,2,\ldots,N $$

$$ D_i = \{j:\ j\neq i \text{ and } \frac{q_j}{1-q_i} > \frac{Q_j}{1-Q_i}\} $$

where $q_j/(1-q_i)$ and $Q_j/(1-Q_i)$ are the conditional probabilities of selecting unit j after i was selected.

Suppose in (15) $i\varepsilon D$ and $j\varepsilon D_i$. According to the Keyfitz principle unit i is retained with probability

$$ \frac{P_i}{p_i} $$

and unit j is retained with probability

$$ \frac{Q_j}{1-Q_i} \Big/ \frac{q_j}{1-q_i} $$

This adjustment yields a probability of selection for units i and j in that order

$$ {}_i\Pi_j = {}_i\pi_j \frac{P_i}{p_i} \cdot \frac{Q_j}{1-Q_i} \Big/ \frac{q_j}{1-q_i} $$

$$ = P_i \frac{Q_j}{1-Q_i} = \frac{1}{2} AQ_iQ_j \tag{16} $$

provided that neither unit has a chance of being re-selected once rejected (which is part of the Keyfitz principle). Formula (16) is the desirable equivalent of (15) after the adjustment. This reasonably simple procedure, however, has to be modified (as far as the adjustment of the probability of selection of the second unit is concerned!) if the first unit happens to be rejected. The reason for this is the fact that as long as the first unit is retained the Keyfitz procedure can be applied at the second step to the resulting conditional distribution (the condition being the selection of unit i at the first step). If, however, the first unit is rejected and another one is selected (e.g. i') then one has to compare

$$ \frac{q_j}{1-q_i} $$

with

$$ \frac{Q_j}{1-Q_{i'}}, $$

i.e. one is compelled to consider two different conditional distributions.

The following modification of the Keyfitz principle yields the desirable result:

Theorem 1.

Suppose that a sample of two units had been selected without replacement with probabilities proportional to the measures of size $\{m_i\}$ satisfying formulae (1) to (11) (using the procedure described in [2]). Let $\{M_i\}$ be the new measures of size and let $\{P_i\}$, $\{Q_i\}$ and A be computed from formulae (2), (5), (6) and (7). Then the application of the rules below will result in a sample having the properties (1) to (11) in terms $\{M_i\}$, $\{P_i\}$, $\{Q_i\}$, and A.

Define the subsets of the set of integers between 1 and N as follows

$$ I = \{c:\ p_c \leqq P_c\} \tag{17} $$

$$ D = \{c:\ p_c > P_c\} \tag{18} $$

$$ I_j = \{c:\ c\neq j \text{ and } \frac{q_c}{1-q_j} \leqq \frac{Q_c}{1-Q_j}\} \quad \text{for } j=1,2,\ldots,N \tag{19} $$

$$ D_j = \{c:\ c\neq j \text{ and } \frac{q_c}{1-q_j} > \frac{Q_c}{1-Q_j}\} \quad \text{for } j=1,2,\ldots,N \tag{20} $$

$$ I_{id} = \{c:\ c\neq i,\ c\neq d \text{ and } \frac{q_c}{1-q_i-q_d} \leqq \frac{Q_c}{1-Q_i}\} \quad \text{for } i\varepsilon I \text{ and } d\varepsilon D \tag{21} $$

$$ D_{id} = \{c:\ c\neq i,\ c\neq d \text{ and } \frac{q_c}{1-q_i-q_d} > \frac{Q_c}{1-Q_i}\} \quad \text{for } i\varepsilon I \text{ and } d\varepsilon D \tag{22} $$

Denote

$$S = \sum_{c\epsilon I} (P_c - p_c) = \sum_{c\epsilon D} (p_c - P_c) \quad (23)$$

$$S_j = \sum_{c\epsilon I_j} \left(\frac{Q_c}{1-Q_j} - \frac{q_c}{1-q_j}\right)$$

$$= \sum_{c\epsilon D_j} \left(\frac{q_c}{1-q_j} - \frac{Q_c}{1-Q_j}\right) \quad (24)$$

$$S_{id} = \sum_{c\epsilon I_{id}} \left(\frac{Q_c}{1-Q_i} - \frac{q_c}{1-q_i-q_d}\right) \quad (25)$$

Rule 1: Given the original sample of two units determine the order in which they will be subjected to a test of rejection by selecting a random number r ($0 < r \leq 1$). If $0 < r \leq 0.5$ the unit with the lower subscript will be adjusted first (called "the first unit"); if $0.5 < r \leq 1$ then the unit with the higher subscript will be the "first unit".

Rule 2: Apply to the first unit the Keyfitz rule, i.e. a) if unit c is the first and $c\epsilon I$ then retain it; b) if $c\epsilon D$ retain it with probability P_c/p_c, reject with probability $1-P_c/p_c$ and in this case select a unit from I with probability $(P_c-p_c)/S$.

Rule 3: If the first unit was retained (with subscript j) and the second unit had subscript k:

(a) retain the second unit if $k\epsilon I_j$

(b) if $k\epsilon D_j$ retain it with probability

$$\frac{Q_k}{1-Q_j} \Big/ \frac{q_k}{1-q_j} = \frac{Q_k(1-q_j)}{q_k(1-Q_j)}$$

reject it otherwise; if it is rejected, select another one from I_j with probabilities

$$\frac{1}{S_j}\left(\frac{Q_c}{1-Q_j} - \frac{q_c}{1-q_j}\right); \qquad c\epsilon I_j$$

Rule 4: If the first unit (d) was in D and if it was rejected and replaced by $i\epsilon I$ and if the second unit had subscript k

(a) retain the second unit if $k\epsilon I_{id}$

(b) if $k\epsilon D_{id}$, retain it with probability

$$\frac{Q_k}{1-Q_i} \Big/ \frac{q_k}{1-q_i-q_d} = \frac{Q_k(1-q_i-q_d)}{q_k(1-Q_i)}$$

reject it otherwise; if it is rejected, select d with probability

$$\frac{Q_d}{Q_d+(1-Q_i)\, S_{id}}$$

and if d is not selected, select a unit from I_{id} with probabilities

$$\frac{1}{S_{id}}\left(\frac{Q_c}{1-Q_i} - \frac{q_c}{1-q_i-q_d}\right), \qquad c\epsilon I_{id}$$

(c) if k=i, select from among all the units, excepting i, with probabilities

$$\frac{Q_c}{1-Q_i} \qquad c=1,2,\ldots,N; \quad c\neq i$$

Two lemmas are required for the proof of theorem 1:

Lemma 1. Let S_{id} be defined by formula (25). Then

$$S_{id} = \sum_{c\epsilon I_{id}} \left(\frac{Q_c}{1-Q_i} - \frac{q_c}{1-q_i-q_d}\right)$$

$$= \sum_{c\epsilon D_{id}} \left(\frac{q_c}{1-q_i-q_d} - \frac{Q_c}{1-Q_i}\right) - \frac{Q_d}{1-Q_i} \quad (26)$$

Lemma 2. Suppose that in the original sample the first unit (as obtained from Rule 1) was unit $d\epsilon D$, the second unit was $k(\neq d)$ and that the application of Rule 2 resulted in the replacement of d by $i\epsilon I$. Denote the conditional probability of obtaining (through retention or through rejection and a new selection) unit β as the second unit of the new sample by

$$P_\beta(d\to i,\ k)$$

Then

$$\sum_{\substack{k=1\\k\neq d}}^{N} q_k P_\beta(d\to i,\ k) = (1-q_d)\,\frac{Q_\beta}{1-Q_i} \quad (27)$$

for any $d\epsilon D$, $i\epsilon I$ and any $\beta(\neq i)$.

The proof of lemma 1 follows immediately from (6) (i.e. the analogous result for Q_i) and from the observation that the union of D_{id} and I_{id} is the set of integers from 1 to N excepting i and d. The proof of lemma 2 will be presented after the proof of theorem 1.

Proof of theorem 1:

Denoting the first and second units of the new sample by α and β respectively, there are four possibilities:

(i) $\alpha\epsilon I$, $\beta\epsilon I_\alpha$

(ii) $\alpha\epsilon I$, $\beta\epsilon D_\alpha$

(iii) $\alpha\epsilon D$, $\beta\epsilon I_\alpha$

(iv) $\alpha\epsilon D$, $\beta\epsilon D_\alpha$

It will be shown that in each case the probability of obtaining units α and β in that order in the new sample is equal to

$$\frac{1}{2} AQ_\alpha Q_\beta \tag{28}$$

This will complete the proof of the theorem in that it will then follow from (5) (in terms of A, P_i, Q_i) that the probability of obtaining in the new sample unit α as the first unit is equal to

$$\sum_{\beta(\neq\alpha)} \frac{1}{2} AQ_\alpha Q_\beta = \frac{1}{2} AQ_\alpha(1-Q_\alpha) = P_\alpha$$

$$\alpha=1, 2, \ldots, N$$

Also the probability of obtaining unit β as the second unit is equal to

$$\sum_{\alpha(\neq\beta)} \frac{1}{2} AQ_\alpha Q_\beta = \frac{1}{2} A(1-Q_\beta)\, Q_\beta = P_\beta$$

$$\beta=1, 2, \ldots, N$$

and from the symmetry of (28) the probability of obtaining α and β in either order is equal to

$$\Pi_{\alpha\beta} = AQ_\alpha Q_\beta$$

In the following proofs we assume that rule 1 had already been applied and that consequently the original sample had already been ordered.

Proof of case (i): $\alpha\varepsilon I$, $\beta\varepsilon I_\alpha$

Such a sample can arise in the following three ways: the original sample was α and β in that order in which case they are both retained (rules 2 and 3a); the original sample was α and $k\varepsilon D_\alpha$, α was retained and $\beta\varepsilon I_\alpha$ selected (rules 2 and 3b); the first unit was $d\varepsilon D$, it was rejected, $\alpha\varepsilon I$ selected (rule 2) then β was obtained for the second unit as in lemma 2. The three probabilities corresponding to these three events are as follows:

$$\frac{1}{2} aq_\alpha q_\beta +$$

$$+ \sum_{k\varepsilon D_\alpha} \frac{1}{2} aq_\alpha q_k \left[1- \frac{Q_k(1-q_\alpha)}{q_k(1-Q_\alpha)}\right] \frac{1}{S_\alpha} \left(\frac{Q_\beta}{1-Q_\alpha} - \frac{q_\beta}{1-q_\alpha}\right) +$$

$$+ \sum_{d\varepsilon D} \sum_{k(\neq d)} \frac{1}{2} aq_d q_k \left(1- \frac{P_d}{p_d}\right) \frac{1}{S}(P_\alpha-p_\alpha)P_\beta(d\to\alpha, k) =$$

$$= \frac{1}{2} aq_\alpha q_\beta +$$

$$+ \frac{1}{2} aq_\alpha(1-q_\alpha) \frac{1}{S_\alpha} \left(\frac{Q_\beta}{1-Q_\alpha} - \frac{q_\beta}{1-q_\alpha}\right) \sum_{k\varepsilon D_\alpha} \left(\frac{q_k}{1-q_\alpha} - \frac{Q_k}{1-Q_\alpha}\right)+$$

$$+ \frac{1}{S} (P_\alpha-p_\alpha) \sum_{d\varepsilon D} \frac{1}{2} aq_d\left(1 - \frac{P_d}{p_d}\right) (1-q_d) \frac{Q_\beta}{1-Q_\alpha}$$

the last term following from lemma 2. Applying (5), (10), (23) and (24) we obtain

$$\frac{1}{2} aq_\alpha q_\beta +$$

$$+ p_\alpha \frac{Q_\beta}{1-Q_\alpha} - \frac{1}{2} aq_\alpha q_\beta$$

$$+ (P_\alpha-p_\alpha) \frac{Q_\beta}{1-Q_\alpha} = P_\alpha \frac{Q_\beta}{1-Q_\alpha} = \frac{1}{2} AQ_\alpha Q_\beta$$

Proof of case (ii): $\alpha\varepsilon I$, $\beta\varepsilon D_\alpha$

Such a sample can arise in two ways: the original sample was α and β and β was retained (α is certainly retained by rule 2); the first unit was $d\varepsilon D$, it was rejected, $\alpha\varepsilon I$ selected (rule 2) then β was obtained for the second unit as in lemma 2. The probabilities corresponding to these events are as follow:

$$\frac{1}{2} aq_\alpha q_\beta \frac{Q_\beta(1-q_\alpha)}{q_\beta(1-Q_\alpha)} +$$

$$+ \sum_{d\varepsilon D} \sum_{k(\neq d)} \frac{1}{2} aq_d q_k \left(1- \frac{P_d}{p_d}\right) \frac{1}{S}(P_\alpha-p_\alpha)\, P_\beta\, (d\to\alpha, k)$$

$$= \frac{1}{2} aq_\alpha(1-q_\alpha) \frac{Q_\beta}{1-Q_\alpha} +$$

$$+ (P_\alpha-p_\alpha) \frac{Q_\beta}{1-Q_\alpha}$$

the second term on the right hand side following from lemma 2 in a way identical to the manipulation of the last term in the proof of (i). Applying (5) to the first term we obtain

$$p_\alpha \frac{Q_\beta}{1-Q_\alpha} + (P_\alpha-p_\alpha) \frac{Q_\beta}{1-Q_\alpha} = P_\alpha \frac{Q_\beta}{1-Q_\alpha} = \frac{1}{2} AQ_\alpha Q_\beta$$

Proof of case (iii): $\alpha\varepsilon D$, $\beta\varepsilon I_\alpha$

Such a sample can arise in two ways: the original sample was α and β and α was retained (rule 2; β is then automatically retained by rule 3a); the original sample was α and $k\varepsilon D_\alpha$, α retained (rule 2) k rejected and $\beta\varepsilon I_\alpha$ selected (rule 3b). The probabilities corresponding to these events are as follow:

$$\frac{1}{2} aq_\alpha q_\beta \frac{P_\alpha}{p_\alpha} +$$

$$+ \sum_{k\varepsilon D_\alpha} \frac{1}{2} aq_\alpha q_k \frac{P_\alpha}{p_\alpha} \left(1 - \frac{Q_k(1-q_\alpha)}{q_k(1-Q_\alpha)}\right) \frac{1}{S_j}$$

$$\left(\frac{Q_\beta}{1-Q_\alpha} - \frac{q_\beta}{1-q_\alpha}\right) =$$

$$= \frac{1}{2} aq_\alpha q_\beta \frac{P_\alpha}{p_\alpha} +$$

$$+ \frac{1}{2} aq_\alpha(1-q_\alpha) \frac{P_\alpha}{p_\alpha} \frac{1}{S_j} \left(\frac{Q_\beta}{1-Q_\alpha} - \frac{q_\beta}{1-q_\alpha}\right)$$

$$\sum_{k\epsilon D_\alpha} \left(\frac{q_k}{1-q_\alpha} - \frac{Q_k}{1-Q_\alpha}\right)$$

$$= P_\alpha \frac{q_\beta}{1-q_\alpha} +$$

$$+ P_\alpha \left(\frac{Q_\beta}{1-Q_\alpha} - \frac{q_\beta}{1-q_\alpha}\right)$$

the first term following from (5), the second from (5) and (24). We obtain

$$P_\alpha \frac{Q_\beta}{1-Q_\alpha} = \frac{1}{2} AQ_\alpha Q_\beta$$

Proof of case (iv): $\alpha\epsilon D$, $\beta\epsilon D_\alpha$

Such a sample can only arise in one way: the original sample was α and β, α was retained (rule 2) and β was retained (rule 3b). The probability of this event is

$$\frac{1}{2} aq_\alpha q_\beta \frac{P_\alpha}{p_\alpha} \frac{Q_\beta(1-q_\alpha)}{q_\beta(1-Q_\alpha)} = P_\alpha \frac{Q_\beta}{1-Q_\alpha} = \frac{1}{2} AQ_\alpha Q_\beta$$

after two applications of (5).

This completes the proof of theorem 1.

Proof of lemma 2:

The original sample consisted of units $d\epsilon D$ and $k(\neq d)$. The application of rule 2 resulted in the rejection of d and its replacement by $i\epsilon I$. Lemma 2 is to be proved for all $\beta(\neq i)$, i.e. for the cases when $\beta\epsilon D_{id}$, $\beta\epsilon I_{id}$ and $\beta=d$.

(a) If $\beta\epsilon D_{id}$:

According to rule 4a) and rule 4b) if $k\epsilon I_{id}$ or if $k\epsilon D_{id}$ but $k\neq\beta$

$$P_\beta\ (d\to i,\ k) = 0$$

According to rule 4b) if $k=\beta\epsilon D_{id}$

$$P_\beta\ (d\to i,\ \beta) = \frac{Q_\beta(1-q_i-q_d)}{q_\beta(1-Q_i)}$$

According to rule 4c)

$$P_\beta\ (d\to i,\ i) = \frac{Q_\beta}{1-Q_i}$$

Consequently

$$\sum_{\substack{k=1\\(k\neq d)}}^{N} q_k P_\beta(d\to i,\ k) = q_\beta \frac{Q_\beta(1-q_i-q_d)}{q_\beta(1-Q_i)} + q_i \frac{Q_\beta}{1-Q_i}$$

$$= (1-q_d) \frac{Q_\beta}{1-Q_i}$$

(b) If $\beta\epsilon I_{id}$

According to rule 4a) if $k\epsilon I_{id}$ but $k\neq\beta$ then

$$P_\beta(d\to i,\ k) = 0$$

and

$$P_\beta(d\to i,\ \beta) = 1$$

According to rule 4c)

$$P_\beta(d\to i,\ i) = \frac{Q_\beta}{1-Q_i}$$

According to 4b) if $k\epsilon D_{id}$

$$P_\beta(d\to i,\ k) = \left(1- \frac{Q_k(1-q_i-q_d)}{q_k(1-Q_i)}\right)\left(1- \frac{Q_d}{Q_d+(1-Q_i)S_{id}}\right)$$

$$\frac{1}{S_{id}} \left(\frac{Q_\beta}{1-Q_i} - \frac{q_\beta}{1-q_i-q_d}\right)$$

Hence

$$\sum_{\substack{k=1\\(k\neq d)}}^{N} q_k\ P_\beta(d\to i,\ k) = q_\beta + q_i \frac{Q_\beta}{1-Q_i} +$$

$$+ \left(1- \frac{Q_d}{Q_d+(1-Q_i)S_{id}}\right) \frac{1}{S_{id}} \left(\frac{Q_\beta}{1-Q_i} - \frac{q_\beta}{1-q_i-q_d}\right)$$

$$(1-q_i-q_d) \sum_{k\epsilon D_{id}} \left(\frac{q_k}{1-q_i-q_d} - \frac{Q_k}{1-Q_i}\right)$$

Applying lemma 1 to the last term we obtain

$$q_\beta + q_i \frac{Q_\beta}{1-Q_i} + \frac{(1-Q_i)S_{id}}{Q_d+(1-Q_i)S_{id}} \frac{1}{S_{id}} \left(\frac{Q_\beta}{1-Q_i} - \frac{q_\beta}{1-q_i-q_d}\right)$$

$$(1-q_i-q_d)\ \left(S_{id} + \frac{Q_d}{1-Q_i}\right)$$

$$= (1-q_d) \frac{Q_\beta}{1-Q_i}$$

(c) If $\beta=d$

According to rule 4b) if $k\epsilon I_{i\beta}$

$$P_\beta\ (\beta\to i,\ k) = 0$$

According to rule 4c)

$$P_\beta\ (\beta\to i,\ i) = \frac{Q_\beta}{1-Q_i}$$

According to rule 4b) if $k\epsilon D_{i\beta}$

$$P_\beta(\beta\to i,\ k) = (1-\frac{Q_k(1-q_i-q_\beta)}{q_k(1-Q_i)})\ \frac{Q_\beta}{Q_\beta+(1-Q_i)S_{i\beta}}$$

Consequently

$$\sum_{\substack{k=1\\(k\neq\beta)}}^{N} q_k P_\beta(\beta\to i,k)=q_i\ \frac{Q_\beta}{1-Q_i} + \frac{Q_\beta}{Q_\beta+(1-Q_i)S_{i\beta}}\ (1-q_i-q_\beta)$$

$$\sum_{k\epsilon D_{i\beta}} (\frac{q_k}{1-q_i-q_\beta} - \frac{Q_k}{1-Q_i})$$

Applying lemma 1 to the last term, we obtain

$$q_i\ \frac{Q_\beta}{1-Q_i} + \frac{Q_\beta}{Q_\beta+(1-Q_i)S_{i\beta}}\ (1-q_i-q_\beta)\ (S_{i\beta} + \frac{Q_\beta}{1-Q_i})$$

$$= (1-q_\beta)\ \frac{Q_\beta}{1-Q_i}$$

This completes the proof of Lemma 2.

The expected number of rejections

It has been proved in the previous section that the procedure as described by the rule of theorem 1 satisfies objectives I) and III) as set out in the introduction. With respect to the objective of minimizing the expected rejections the following may be said. The procedure would be optimal if it rejected originally selected units only to the extent necessary, i.e. if it rejected a unit only if its probability of selection for the new sample had to be diminished, and if in this case the original probability of selection times the probability of retention were equal the desired new probability of rejection. A procedure is not optimum if a rejected unit can be re-selected since this means that the rejection procedure renders its probability for the new sample too small (it is rejected with an unnecessarily large probability).

The present procedure is such that a first unit (or a second unit), once rejected, can never be re-selected as a first unit (or a second unit). There is, however, one (and only one) condition, under which a unit rejected as a first unit can be re-selected as a second unit. This is embodied in rules 4b) and 4c) when a first unit $d\epsilon D$ is rejected, $i\epsilon I$ is selected, the second unit k is in D_{id} or k=i, it is rejected and d is re-selected as a second unit. It is not too difficult to show that if one did not permit the re-selection of d as a second unit at all, then the procedure would not yield the required probabilities for the new sample. Whether the present procedure is actually optimal, however, is not known. Neither is it known whether there is an optimal procedure at all (subject to objectives I) and III)). Yet the probability of the event described above which may result in a departure from optimality is sufficiently small to render the statement concerning the "near optimality" of the procedure more or less justified.

The actual formula for the expected number of rejections can be derived without difficulty.

Theorem 2.

The expected number of rejections when applying the rules of theorem 1 is given by

$$\sum_{i\epsilon I} p_i S_i + \sum_{d\epsilon D} P_d S_d + S + \sum_{i\epsilon I}\ \sum_{d\epsilon D} (P_i-p_i)(p_d-P_d)\ \frac{S_{id}}{S}$$

$$- \sum_{d\epsilon D}\ \sum_{i\epsilon I} \frac{1}{2}\ aq_d q_i\ (1-\frac{P_d}{p_d})\ \frac{P_i-p_i}{S}\ \frac{Q_d}{1-Q_i}$$

$$- \sum_{d\epsilon D}\ \sum_{i\epsilon I} \frac{1}{2}\ aq_d q_i\ (1-\frac{P_d}{p_d})\ \frac{P_i-p_i}{S}\ S_{id} \qquad (29)$$

and an upper bound of (29) is given by

$$\sum_{i=1}^{N} P_i S_i + S \qquad (30)$$

Proof:

The event "at least one rejection" is composed of the following mutually exclusive events:

(i) the first unit is rejected

(ii) the first unit is retained and the second unit rejected.

It is easy to show that the probability of event (i), as a result of applying rule 2, is given by

$$S = \sum_{i\epsilon I} (P_i-p_i) = \sum_{d\epsilon D} (p_d-P_d) \qquad (31)$$

The probability of event (ii) is given by (rule 3)

$$\sum_{i\epsilon I}\ \sum_{k\epsilon D_i} \frac{1}{2}aq_i q_k(1-\frac{Q_k(1-q_i)}{q_k(1-Q_i)}) + \sum_{d\epsilon D}\ \sum_{k\epsilon D_i} \frac{1}{2}aq_d q_k\ \frac{P_d}{p_d}\ (1-\frac{Q_k(1-q_d)}{q_k(1-Q_d)})$$

$$= \sum_{i\epsilon I} p_i \sum_{k\epsilon D_i} (\frac{q_k}{1-q_i} - \frac{Q_k}{1-Q_i}) + \sum_{d\epsilon D} P_d \sum_{k\epsilon D_i} (\frac{q_k}{1-q_d} - \frac{Q_k}{1-Q_d})$$

$$= \sum_{i\epsilon I} p_i S_i + \sum_{d\epsilon D} P_d S_d \qquad (32)$$

the middle line following from (5).

(31) plus (32) represent the probability of at least one rejection in the course of selecting the new sample. It is possible to arrive at the new sample via rejections and obtain the old sample. This happens according to rule 4c) with the probability ($d\epsilon D$ and $i\epsilon I$ constituted the first sample in that order, d was rejected and i selected for the first unit in the new sample and d was then selected for the second unit in the new sample):

$$\sum_{d\epsilon D} \sum_{i\epsilon I} \frac{1}{2} a q_d q_i \left(1-\frac{P_d}{p_d}\right) \frac{P_i-p_i}{S} \frac{Q_d}{1-Q_i} \qquad (33)$$

Consequently (31) plus (32) minus (33) represents the probability that the new sample is not identical with the old. Adding to this the probability of the event that the new sample will be entirely different from the old we shall obtain the expected number of rejections. The probability of the latter event (rule 4b) is given by

$$\sum_{d\epsilon D} \sum_{i\epsilon I} \sum_{k\epsilon D_{id}} \frac{1}{2} a q_d q_k \left(1-\frac{P_d}{p_d}\right) \frac{P_i-p_i}{S}$$

$$\left(1-\frac{Q_k(1-q_i-q_d)}{q_k\ (1-Q_i)}\right) \left(1-\frac{Q_d}{Q_d+(1-Q_i)S_{id}}\right)$$

$$= \sum_{d\epsilon D} \sum_{i\epsilon I} \frac{1}{2} a q_d \left(1-\frac{P_d}{p_d}\right) \frac{P_i-p_i}{S} (1-q_i-q_d) S_{id}$$

$$= \sum_{d\epsilon D} \sum_{i\epsilon I} (p_d-P_d)(P_i-p_i) \frac{S_{id}}{S} - \sum_{d\epsilon D} \sum_{i\epsilon I} \frac{1}{2} a q_d q_i$$

$$\left(1-\frac{P_d}{p_d}\right) \frac{P_i-p_i}{S} S_{id} \qquad (34)$$

the second line following from lemma 1.

Since the expected number of rejections is given by

(31)+(32)-(33)+(34)

this completes the proof of (29).

In order to prove (30), we note that I_{id} is a subset of I_i, consequently

$$S_{id} \leq S_i$$

hence

$$\sum_{i\epsilon I} \sum_{d\epsilon D} (P_i-p_i)(p_d-P_d) \frac{S_{id}}{S} \leq$$

$$\leq \sum_{i\epsilon I} \sum_{d\epsilon D} (P_i-p_i) \frac{p_d-P_d}{S} S_i$$

$$= \sum_{i\epsilon I} (P_i-p_i) S_i \qquad (35)$$

Omitting the last two terms in (29) and substituting the right hand side of (35) for the fourth term in (29) proves (30).

An alternative procedure

An alternative procedure, also satisfying objectives I) and III) of the introduction is a direct application of the Keyfitz procedure to the entire sample (instead of the individual units). The procedure can be briefly described as follows:

Let π_{ij} be the probabilities of the original samples, Π_{ij} the desired new probabilities given by (11) in terms of A, Q_i, Q_j. Define the sets of distinct samples (ordered pairs of integers) as follows:

$$K = \{(i,j):\ i<j \text{ and } \pi_{ij} \leq \Pi_{ij}\}$$

$$L = \{(i,j):\ i<j \text{ and } \pi_{ij} > \Pi_{ij}\}$$

Let

$$T = \sum_{(i,j)\epsilon K} (\Pi_{ij}-\pi_{ij}) = \sum_{(i,j)\epsilon L} (\pi_{ij}-\Pi_{ij})$$

If the units of the old sample are in K, retain them. If they are in L retain them with probability Π_{ij}/π_{ij}. If the old sample is rejected select a new sample from K with probabilities

$$\frac{\Pi_{ij}-\pi_{ij}}{T}$$

Keyfitz's proof concerning his procedure applies here without change to show that this procedure satisfies objective III) and hence also I). It also follows from the Keyfitz proof that this procedure maximizes the probability of maintaining the entire old sample (i.e. it minimizes the probability of at least one rejection). In case of a rejection, however, the entire old sample is rejected. The procedure is not optimal, since it permits the re-selection of a rejected unit (if a sample is rejected and a new sample is selected, one of the units of the new sample may be identical to one of the units of the old).

Numerical examples show the relative performance of the two procedures: the alternative procedure results in 25 to 40 per cent (in one example 100 per cent) more rejections than the procedure of theorem 1. It is also easier to apply (although the rules appear to be more complicated) since under that procedure in any given concrete case at most 2N comparisons need to be made between old and new probabilities while the alternative procedure requires N(N-1)/2 such comparisons.

Numerical examples

In examples 1 to 4 the same "original measures of size" are used.

In example 1 the "new measures of size" are obtained by reversing the ordering of the original measures of size: the unit with the smallest original size measure becomes the largest, etc. This represents a drastic change in the original size distribution. In example 2 the new size measures are obtained by interchanging consecutive pairs of the original size measures. This represents "intermittent moderate growth". In example 3 the fourth and fifth original measures of size are interchanged to obtain the new measures of size. This represents "some small growth and some small decline among the larger units". In example 4 the new measures are identical to the old except that the largest measure of size was doubled. This represents "large growth of a few units".

Examples 5 to 8 are analogous to examples 1 to 4 except that the population is larger (N=20) and there is a substantially higher variation in the original size measures.

The expected number of rejections, as in (29), and its upper bound, as in (30), is shown. Also the expected number of rejections under the alternative procedure of the previous section is shown.

Examples 1 to 4

i	m_i	M_i Ex. 1	Ex. 2	Ex. 3	Ex. 4
1	10	22	14	10	10
2	14	19	10	14	14
3	17	18	18	17	17
4	18	17	17	19	18
5	19	14	22	18	19
6	22	10	19	22	44
Expected No. of rejections		0.3759	0.1671	0.0239	0.3143
Upper bound of Expected No. of rejections		0.4000	0.1752	0.0245	0.3237
Alternative Expected No. of rejections		0.4998	0.1999	0.0342	0.4455

Examples 5 to 8

i	m_i	M_i Ex. 5	Ex. 6	Ex. 7	Ex. 8
1	5	95	10	5	5
2	10	90	5	10	10
3	15	85	20	15	15
4	20	80	15	20	20
5	25	75	30	25	25
6	30	70	25	30	30
7	35	65	40	35	35
8	40	60	35	40	40
9	45	55	50	45	45
10	50	50	45	50	50
11	50	50	55	50	50
12	55	45	50	55	55
13	60	40	65	60	60
14	65	35	60	65	65
15	70	30	75	70	70
16	75	25	70	80	170
17	80	20	85	75	80
18	85	15	80	85	135
19	90	10	95	90	90
20	95	5	90	95	200
Expected No. of rejections		0.9039	0.1012	0.0108	0.3066
Upper bound of Expected No. of rejections		0.9194	0.1027	0.0109	0.3109
Alternative Expected No. of rejections		1.2170	0.1252	0.0193	0.4051

In seven of the eight examples the alternative procedure yields between 20 to 45 per cent more rejections than the procedure of theorem 1. In example 7 the alternative procedure yields almost 75 per cent more rejections. In every case the upper bound for the expected number of rejections provided satisfactory approximations.

References

[1] Brewer, K.R.W.: "A note on Fellegi's method of sampling without replacement with probability proportional to size". To be published in the Journal of the American Statistical Association.

[2] Fellegi, I.P.: "Sampling with varying probabilities without replacement: rotating and non-rotating samples". Journal of the American Statistical Association, 58 (1963), 183-201.

[3] Keyfitz, N.: "Sampling with probabilities proportional to size: adjustment for changes in probabilities". Journal of the American Statistical Association, 46 (1951), 105-109.

A MONTE CARLO STUDY OF SOME RATIO ESTIMATORS

J. N. K. Rao and LeNelle D. Beegle
Texas A & M University

Summary

Small-sample efficiencies of eight ratio estimators of the population ratio $\bar{Y}/\bar{X}$ are investigated by Monte Carlo methods, assuming a linear regression of y on x and x normally distributed. Efficiencies of three variance estimators are also investigated. Sample criteria for which expected values exist are used. From this study, Mickey's unbiased ratio estimator and the approximately unbiased ratio estimator obtained by Quenouille's method appear promising.

1. Introduction

Ratio estimators are often employed in sample surveys for estimating the population mean $\bar{Y}$ of a characteristic of interest 'y' or the population ratio $\bar{Y}/\bar{X}$ utilizing a supplementary variate 'x' that is positively correlated with 'y'. It is well-known that the classical ratio estimator is biased and often, in practice, the bias may be negligible compared to standard error and can be neglected. However, the bias may become considerable in surveys with many strata and small or moderate size samples within strata if it is considered appropriate to use "seperate ratio estimators". In these situations, the use of unbiased or approximately unbiased (i.e., estimators with a smaller bias than the classical ratio estimator) ratio estimators may be of great advantage. Therefore, in recent years, considerable attention has been given to the development of unbiased and approximately unbiased ratio estimators.

In this paper we shall, without loss of generality, confine ourselves to estimation of ratios (assuming the population mean $\bar{X}$ is known). Further, to simplify the discussion we shall confine ourselves to simple random sampling and assume the population size N is infinite.

If a simple random sample of n pairs (y_i, x_i) is drawn, the classical ratio estimator of $R=\bar{Y}/\bar{X}$ is given by

$$r = \frac{\bar{y}}{\bar{x}} \qquad (1)$$

where $\bar{y}$ and $\bar{x}$ are the sample means of y_i and x_i respectively. The usual estimator of the variance of r is:

$$v(r) = \frac{1}{n\bar{X}^2}\left(s_y^2 - 2rs_{xy} + r^2 s_x^2\right) \qquad (2)$$

where s_x^2 and s_y^2 are the sample mean squares of x_i and y_i respectively and s_{xy} is the sample mean product of x_i and y_i. It is well known that the bias of v(r) is of order 1/n. Kokan (1963) has investigated the large-sample stabilities of v(r) and the unbiased variance estimator:

$$v\left(\frac{\bar{y}}{\bar{X}}\right) = \frac{s_y^2}{\bar{X}^2}, \qquad (3)$$

where $\bar{y}/\bar{X}$ is the estimator of R not using the sample x-information. He has shown that the coefficient of variation of v(r) is always larger than that of $v(\bar{y}/\bar{X})$ for a bivariate normal distribution, and this property also holds for a bivariate log normal distribution for certain ranges of the parameters.

Hartley and Ross (1954) were the first to give an exact upper bound for the bias of r and the unbiased ratio estimator:

$$t_1 = \bar{r} + \frac{n}{(n-1)\bar{X}}(\bar{y} - \bar{r}\,\bar{x}) \qquad (4)$$

where $\bar{r}$ is the sample mean of $r_i = y_i/x_i$. Goodman and Hartley (1958) have shown that the variance of t_1 will often be larger than the variance of r, for large n.

We now consider two ratio estimators based on dividing the sample at random into g groups, each of size m, where n = mg. Following Mickey (1959), an unbiased estimator of R is given by

$$t_2 = \bar{r}_g + \frac{g}{\bar{X}}(\bar{y} - \bar{r}_g\bar{x}) \qquad (5)$$

where $\bar{r}_g = \sum_1^g r_j'/g$ and r_j' is the classical ratio estimator computed from the sample after omitting the j-th group, i.e., $r_j' = (n\bar{y}-m\bar{y}_j)/(n\bar{x}-m\bar{x}_j)$ where $\bar{y}_j$ and $\bar{x}_j$ are the sample means computed from the j-th group. It may be noted that t_2 reduces to t_1 for the important case of n=2.

Quenouille (1956) proposed a method of reducing estimation bias from order 1/n to $1/n^2$, based on random division of the sample into groups. Durbin (1959) applied this method to ratio estimators and has shown that the estimator:

$$t_3 = gr - (g-1)\bar{r}_g = \frac{1}{g}\sum_1^g r_{Qj} \qquad (6)$$

where $r_{Qj} = gr - (g-1)r_j'$ (called pseudo-values by by Tukey), has bias of order n^{-2} at most. Durbin has further shown that if the regression of y on x is linear (i.e., $y = \alpha + \beta x + e$ where e is the error term) and x normally distributed (model 1, say), t_3 with g = 2 has a smaller asymptotic variance than r. J. Rao (1965) has shown, for the above model, that both asymptotic bias and asymptotic variance of t_3 are decreasing functions of g, so that g=n would be the optimum choice for large or moderately large n. Durbin (1959) has also considered the case where the regression of y on x is linear, but x has a gamma distribution (model 2, say). He has shown that, although the variance t_3 with g=2 compared to the variance of r is slightly increased, the reduction in bias is such that the mean square error of t_3 is reduced. Recently, J. Rao and Webster (1966) have shown, assuming model 2, that both bias and variance t_3 are decreasing functions of g, so that g=n would be the optimum choice. They have also shown that t_3 has a smaller variance than r for g > 2. It may be noted that the results for model 2 are exact for any sample size, n.

Following Tukey (1958) we can use the simple estimator:

$$v(t_3) = g^{-1}(g-1)^{-1}\sum_1^g(r_{Qj} - t_3)^2 \qquad (7)$$

as the variance estimator of t_3, since the g estimators r_{Qj} may be treated as approximately independent and $t_3 = \Sigma\, r_{Qj}/g$. Lauh and Williams (1963) have made a Monte Carlo study of the stabilities of v(r) and $v(t_3)$ with g=n, for small samples (n=2 to 9), using model 1 and model 2 (with exponential distribution for x) and assuming α=0 (i.e., regression through the origin). It was shown that the Monte Carlo variances of v(r) and $v(t_3)$ are about the same for model 1, whereas, for model 2, the variance of $v(t_3)$ is considerably smaller than that of v(r). [3]Incidentally, Lauh and Williams have used $v(t_3)$ as the estimator of the variance of r rather than of the variance of t_3.

Tin (1965) investigated the large-sample bias, variance and approach to normality of the following estimators: r, t_3 with g=2, Beale's estimator

$$t_4 = r\,\frac{\left(1 + \frac{1}{n}\frac{s_{xy}}{\bar{x}\,\bar{y}}\right)}{\left(1 + \frac{1}{n}\frac{s_x^2}{\bar{x}^2}\right)} \qquad (8)$$

and the modified estimator

$$t_5 = r\left[1 + \frac{1}{n}\left(\frac{s_{xy}}{\bar{x}\,\bar{y}} - \frac{s_x^2}{\bar{x}^2}\right)\right] \qquad (9)$$

His comparison shows that t_5 is slightly better than t_4 which in turn is better than t_3 with g=2. Tin has also made a Monte Carlo study for large and moderately large samples (n=50, 200, and 1000), using model 1 and the results are in agreement with his mathematical results. J. Rao and Webster (1966) made an exact comparison of t_5 and t_3 assuming model 2. Their comparison shows that the precisions of t_5 and t_3 with g=n are about the same.

Tin (1965) has also considered the estimator:

$$t_6 = \frac{g}{g-1}\,r - \frac{1}{g(g-1)}\sum_1^g \frac{\bar{y}_j}{\bar{x}_j}\,. \qquad (10)$$

It may be noted that t_6 is identical to t_3 when g=2. Also, t_6 is based on the group means $\bar{y}_j$ and $\bar{x}_j$, whereas, t_3 is based on the complementary means $(n\bar{y} - m\bar{y}_j)/(n-m)$ and $(n\bar{x} - m\bar{x}_j)/(n-m)$. Hartley (see Pascual, 1961 and Sastry, 1965) has earlier proposed t_6 with g=n and Murthy and Nanjamma (1959) have used t_6 when g independent and interpenetrating sub-samples each of size m are available.

Pascual (1961) proposed the following estimator obtained by estimating the bias of r approximately:

$$t_7 = r + \frac{1}{(n-1)\bar{X}}\,(\bar{y} - \bar{r}\,\bar{x})\ . \qquad (11)$$

It may be noted that the investigations by Pascual (1961) and Sastry (1965) regarding t_6 (with g=n) and t_7 are not very satisfactory, since they assume the higher order population moments of $\delta x_i = (x_i - \bar{X})/\bar{X}$ and $\delta y_i = (y_i - \bar{Y})/\bar{Y}$ are negligible and $|\delta x_i| < 1$, whereas, to develop asymptotic theories for r, t_2, t_3, t_4 and t_5 we need only assume that n is large or moderately large and $|\bar{x}-\bar{X}|/\bar{X}<1$ or $|\bar{x}_j'-\bar{X}|/\bar{X}<1$ where $\bar{x}_j' = (n\bar{x}-m\bar{x}_j)/(n-m)$. Similarly, for the estimator t_6 ($g \neq n$) an asymptotic theory would not be satisfactory if m is small.

It is clear that efficiency comparisons in small or moderate size samples would be more valuable since these are the cases in which freedom from bias may be important. Therefore, in the present paper, we make a Monte Carlo study of the efficiencies of the eight ratio estimators r, t_1, ..., t_7 and the three variance estimators $v(\bar{y}/\bar{X})$, v(r) and $v(t_3)$ for small and moderate size samples, using model 1. It may be noted that, unlike under model 1, exact analytical comparisons of the estimators can be made under model 2 for any sample size -- some exact results, under model 2, have already been given by J. Rao and Webster

(1966) and further work is in progress.

We have used Lauh and Williams' model as well as that of Tin for our study. The model of Lauh and Williams is ideal for ratio estimators since it is assumed that the regression is through the origin and the coefficient of variation of x, C_x, is small. On the other hand, Tin's model is not so favorable since the regression is not through the origin and C_x is not small. It would be interesting, therefore, to study the performances of the estimators under both the models.

2. Monte Carlo Study

In Lauh and Williams' model, x_i is N(10, 4) and y_i is defined as $5(x_i + e_i)$ where e_i is N(0, 1) and independent of x_i. Therefore, the correlation, ρ, between y_i and x_i is 0.89 and $C_x = 0.2$. In Tin's model, x_i and y_i have a bivariate normal distribution with $\overline{X} = 5$, $\sigma_x^2 = 45$, $\overline{Y} = 15$, $\sigma_y^2 = 500$ and $\rho = 0.4$, 0.6 or 0.8. To carry out the experiment on the computer, the model $y_i = \alpha + \beta x_i + e_i$ was used, where $\alpha=\overline{Y}-\beta\overline{X}$, $\beta = \rho\sigma_y/\sigma_x$, and e_i is $N[0,\sigma_y^2(1-\rho^2)]$ and distributed independently of x_i.

Using the IBM 7094 pairs of random numbers (u_{1i}, u_{2i}) were generated from a rectangular distribution with mean 1/2 and range 1 and were transfromed into standard normal variates w_{1i} and by the transformation:

$$w_{1i} = (-2 \log_e u_{1i})^{1/2} \sin 2\pi u_{2i}$$
$$w_{2i} = (-2 \log_e u_{1i})^{1/2} \cos 2\pi u_{2i}$$

(see Box and Muller, 1958). Then the pairs (x_i, y_i) were computed as follows: For Lauh and Williams' model, $x_i = \overline{X} + w_{1i}\sigma_x$ and $y_i = 5(x_i + w_{2i})$; for Tin's model, $x_i = \overline{X} + w_{1i}\sigma_x$ and $y_i = \alpha + \beta x_i + e_i$ where $e_i = w_{2i}\sigma_y(1-\rho^2)^{1/2}$. For each selected n, 1000 samples of n pairs (x_i,y_i) were generated and the eight ratio estimators and the three variance estimators were computed from each sample. Thus we have 1000 values of each estimator for each selected n.

The ratio estimators all have Cauchy distributions when the distribution of (x_i,y_i) is bivariate normal so that the population moments do not exist. It may be meaningless, therefore, to use the variance (or mean square error) of the 1000 values as a sample criterion in comparing the estimators. However, the variance would be a satisfactory criterion for Lauh and Williams' model since $\overline{X}$ is so large compared with σ_x that the range of x is effectively positive. On the other hand, for Tin's model the probability of getting a negative x-value is not negligible and the criterion of variance can lead to meaningless results, particularly for small n. Tin's results based on the variance may, however, be fairly satisfactory since he considered only large or moderately large n(n = 50, 200 and 1000). Therefore, we have used, for Tin's model, the following criteria for which expected values exist, and employed the variance criterion for Lauh and Williams' model:

(1) Concentration. Proportion of the value of an estimator in some *a priori* neighborhood around the population ratio R. In this study, this interval is chosen as (R - 0.1, R + 0.1). An estimator T is more efficient than another estimator S if its concentration is larger than that of S.

(2) Interquartile range. Distance between the upper and lower quartile points. Thus it is a range which contains one-half of the 1000 values of an estimator. An estimator T is more efficient than S if its interquartile range is smaller than that of S.

2.1. Results for Lauh and Williams' Model

The variances of the eight ratio estimators r, t_1, ..., t_7 (obtained from 1000 samples) are given in Table 1 for n=4, 6, 8 and 12. It may be noted that all the estimators are unbiased under this model. It is clear, from Table 1, that there are very little differences in the variances of the eight estimators, even for small n, and hence it does not matter which ratio estimator is used. We may, however, still make the following observations: (1) the optimum number of groups, g, for the estimators t_2, t_3 and t_6 is n, (2) for n>4, there are virtually no differences in the variances of t_1, t_2 (g=n), t_3(g=n), t_4, t_5, t_6(g=n) and t_7; for n=4, t_1 has a slightly larger variance, (3) the variances of t_1, ..., t_7 are slightly smaller than that of r for n<12.

Turning to the three variance estimators $v(\overline{y}/\overline{X})$, v(r) and $v(t_3)$, Table 2 gives the coefficients of variation of these estimators for n=4, 6, 8 and 12 -- the criterion of coefficient of variation is more appropriate here since the variance estimators do not have the same mean.

The following tentative conclusions may be drawn from Table 2: (1) Coefficient of variation of $v(t_3)$ decreases considerably as g increases. (2) For any n, the coefficients of variation of v(r) and $v(t_3)$ with g=n are essentially equal. (3) Coefficient of variation of v(r) is slightly larger than that of $v(\overline{y}/\overline{X})$ (for n>4) -- this is in agreement with Kokan's (1963) asymptotic result that the increase in the coefficient of variation of v(r) over that of $v(\overline{y}/\overline{X})$ would be small if C_x is small. We may conclude that the variance estimators v(r) and $v(t_3)$ with g=n are quite stable, for any n, compared to $v(\overline{y}/\overline{X})$, if the regression is approximately through the origin and C_x is small.

Table 1. Variances of the eight ratio estimators r, $t_1, \ldots, t_7$ (obtained from 1000 samples) for selected values of n (Lauh and Williams' model)

Estimator \ n		4	6	8	12
Classical: r		0.0665	0.0398	0.0322	0.0219
Hartley-Ross: t_1		0.0661	0.0394	0.0318	0.0218
Mickey: t_2	g = 2		0.0397	0.0321	0.0218
	$g = \frac{n}{2}$		0.0397	0.0319	0.0218
	g = n	0.0659	0.0394	0.0318	0.0218
Quenouille: t_3	g = 2		0.0397	0.0321	0.0218
	$g = \frac{n}{2}$		0.0397	0.0318	0.0218
	g = n	0.0658	0.0394	0.0318	0.0218
Beale: t_4		0.0659	0.0394	0.0318	0.0218
Modified: t_5		0.0659	0.0394	0.0318	0.0218
Tin: t_6	g = 2		0.0397	0.0321	0.0218
	$g = \frac{n}{2}$		0.0397	0.0318	0.0218
Hartley:	g = n	0.0659	0.0394	0.0317	0.0218
Pascual: t_7		0.0659	0.0394	0.0317	0.0218

Table 2

Coefficients of Variation of the Variance Estimators $v(\bar{y}/\bar{X})$, v(r) and $v(t_3)$ for Selected Values of n (Lauh and Williams' Model)

Variance Estimator \ n		4	6	8	12
$v(\bar{y}/\bar{X})$		0.85	0.61	0.52	0.42
v(r)		0.86	0.67	0.55	0.43
$v(t_3)$:	g = 2		1.45	1.38	1.34
	$g = \frac{n}{2}$		1.03	0.81	0.64
	g = n	0.87	0.68	0.54	0.44

2.2 Results for Tin's Model

The interquartile ranges (for n=4, 6, 10, 20 and 50) and the concentrations (for n=10, 20, and 50) of the eight ratio estimators (obtained from 1000 samples) are given in Table 3 for ρ=0.6, and in Table 4 for ρ=0.8. We have also computed the concentrations for n=4 and 6, but the values are erratic and, hence, unreliable -- this may be probably due to the narrowness of the interval (R-0.1, R+0.1). The variances are also given in Table 3 for ρ=0.6 and n=10 to show that the criterion of variance may lead to meaningless results for small samples. The criterion of variance may, however, become satifactory with large or moderately large n, excepting for those estimators based on the individual ratios y_i/x_i or the group ratios $\bar{y}_j/\bar{x}_j$ for small m. Therefore, we have included the variances of r, t_2, t_3, t_4, t_5 and t_6 (g=2) for n=50 --- our values of the variance of r(0.313 with ρ=0.6 and 0.150 with ρ=0.8) are fairly close to those obtained from the usual asymptotic variance formula (0.293 with ρ=0.6 and 0.148 with ρ=0.8), whereas Tin's values (0.376 with ρ=0.6 and 0.238 with ρ=0.8) are markedly different.

The following tentative conclusions may be drawn from Tables 3 and 4: (1) Unlike under Lauh and Williams' model, the differences in the efficiences of the estimators are quite significant. (2) The interquartile ranges (for all n) and the variances (for n=50) of t_2 and t_3 appear to decrease as g increases, but, for n > 10, the values for g = n/2 and g=n are essentially equal. The concentrations of t_2 and t_3 seem to increase as g increases, excepting that in one case (n=10, ρ=0.6) the concentration of t_2 is 7.1% for g=2, 7.3% for g = n/2 and 7.0% for g=n. In any case, the combined evidence of the three criteria seems to indicate that the optimum value of g for t_2 and t_3 may be taken as n. Moreover, with g=n there is no random splitting involved and the ratio $\bar{r}_g$ is simply given by $n^{-1} \sum^{n} (n\bar{y} - y_j)/(n\bar{x} - x_j)$. (3) Efficiency of t_6^1 is maximum at g=2 unlike under Lauh and Williams' model. This may be because t_6 is biased under Tin's model and as Tin pointed out, the bias increases as g increases. (4) All three criteria indicate that the efficiencies of t_2(g=n), t_3(g=n), t_4 and t_5 are about the same for n $\geq$ 10; for n=4 and 6, t_4 appears to be more efficient than the others. All the above four estimators are more efficient than r, expecting that in one case (n=50, ρ=0.8), the concentrations are essentially equal. (5) r is more efficient than t_6(g=2) for n=4 and 6; for n $\geq$ 10 the efficiencies are about the same, excepting that in one case (n=10, ρ=0.6) the concentration of t_6(g=2) is somewhat low. (6) r is more efficient than Hartley's estimator t_6 (g=n), Pascual's estimator t_7 and the Hartley-Ross unbiased estimator t_1. The efficiencies of t_6(g=n), t_7 and t_1 are about the same for n=4 and 6; for n $\geq$ 10, t_1 is less efficient than t_6(g=n) and t_7.

Turning to the variance estimators, the interquartile ranges (for n=10, 20 and 50) and coefficients of variation (for n=50) of $v(\bar{y}/\bar{X})$, v(r) and $v(t_3)$ are given in Table 5. We have not computed the concentrations around the variances because, as pointed out earlier, the Monte Carlo variances of r and t_3 would be meaningless for small samples. First, considering the criterion of coefficient of variation (for n=50) it is clear from Table 5 that the coefficient of variation of $v(t_3)$ decreases considerably as g increases, so that the optimum value of g is n. The coefficients of variation of $v(t_3)$ (with g=n) and v(r) are essentially equal, but both are considerably larger than the coefficient of variation of $v(\bar{y}/\bar{X})$-- this is in agreement with Kokan's (1963) result because C_x is not small. Turning to the criterion of interquartile range, it may be noted that, in comparing the variance estimators, it would be more appropriate to take the interquartile ranges given in Table 5 relative to the interquartile ranges of the corresponding estimators of R. Now since the interquartile range of t_3 (for any n) is smaller than that of r which in turn is considerably smaller (for ρ=0.8) than that of $\bar{y}/\bar{X}$, it follows from Table 5 that v(r) is slightly more efficient than $v(t_3)$ with g=n (particularly for n=10), but $v(\bar{y}/\bar{X})$ is considerably more efficient than v(r) and $v(t_3)$ with g=n.

3. Concluding Remarks

The approximately unbiased ratio estimators cannot be expected to help when very small samples are taken within strata. In such situations, Mickey's unbiased estimator (with g=n) may be promising since it behaves well under ideal conditions and, under non-ideal conditions, it is considerably more efficient than the Hartley-Ross unbiased estimator and also slightly more efficient than the classical estimator. However, for the important case of n=2 in each stratum, Mickey's estimator is identical to the Hartley-Ross estimator so that no improvement can be achieved. If an approximately unbiased estimator serves the purpose, then Quenouille's estimator (with g=n), Beale's estimator and the modified estimator look favorable, and it does not matter which one is used (although Beale's estimator looks slightly more efficient for n=4 and 6). However, Quenouille's estimator has the added advantage of having a simple variance estimator (Tukey's variance estimator). The approximately unbiased estimators of Hartley and Pascual appear to be unsatisfactory compared to the above approximately unbiased estimators.

Since both Tukey's variance estimator and the usual variance estimator of the classical ratio estimator are considerably less efficient, under non-ideal conditions, than the variance estimator

TABLE 3

Interquartile ranges (I.R.), concentrations (C) and variances (V) of the eight ratio estimators $r, t_1, .., t_7$ (obtained from 1000 samples) for selected values of n and $\rho = 0.6$ (Tin's model)

Estimator		n = 4	n = 6	n = 10			n = 20		n = 50		
		I.R.	I.R.	I.R.	% C	V	I.R.	% C	I.R.	% C	V
Classical: r		2.8	2.3	1.8	6.5	51	1.07	9.1	0.72	17.1	0.31
Hartley-Ross: t_1		4.0	3.2	2.5	4.5	1988	1.63	6.8	1.05	12.8	
Mickey: t_2	$g = 2$	3.7	2.5	1.9	7.1	90	1.09	9.0	0.71	16.3	0.31
	$g = \frac{n}{2}$	3.7	2.3	1.5	7.3	466	0.96	10.7	0.68	18.5	0.29
	$g = n$	2.6	1.7	1.4	7.0	637	0.96	10.7	0.68	18.9	0.28
Quenouille: t_3	$g = 2$	3.5	2.6	1.9	5.0	404	1.06	9.1	0.71	16.4	0.31
	$g = \frac{n}{2}$	3.5	2.4	1.5	6.1	6407	0.93	9.5	0.67	18.6	0.28
	$g = n$	2.7	1.9	1.4	7.7	11508	0.93	11.4	0.67	19.2	0.28
Beale: t_4		1.7	1.6	1.4	6.8	1	0.96	10.9	0.68	19.2	0.28
Modified: t_5		2.1	1.7	1.4	6.7	66544792	0.94	11.6	0.68	19.1	0.28
	$g = 2$	3.5	2.6	1.9	5.0	404	1.06	9.1	0.71	16.4	0.31
Tin: t_6	$g = \frac{n}{2}$	3.5	3.1	2.3	5.0	2494	1.29	8.5	0.78	14.6	
Hartley:	$g = n$	4.0	3.0	2.2	4.9	119	1.24	7.6	0.77	15.3	
Pascual: t_7		4.0	3.1	2.2	5.7	166	1.26	7.2	0.77	14.9	

of the estimator which does not use the supplementary information, caution is needed in the indiscriminate use of ratio estimators.

It may be noted that the main reason for using Mickey's unbiased estimator or an approximately unbiased estimator over the classical estimator is to eliminate or reduce bias in situations where freedom from bias is important; however, it is gratifying that these estimators may, in fact, lead to small or moderate gains in efficiency over the classical estimator.

The conclusions from this study are not necessarily applicable to all populations, since we have considered only two particular models; however, these models reflect many situations that are encountered in practice. Also, our study is of necessity empirical in nature and not mathematical. Clearly therefore, further work, mathematical as well as empirical, using other models and actual data is needed. To this end, we are at the present time investigating the following problems:

(1) Asymptotic results for Mickey's estimator along the lines of J. Rao (1965) using Durbin's model 1.
(2) Exact mathematical results for the eight ratio estimators and the three variance estimators considered in this paper, using Durbin's model 2.
(3) Mathematical and Monte Carlo results when x has a log normal distribution.
(4) Empirical results using several sets of real data.
(5) Results for other models.

References

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates, Ann. Math. Statist., 29, 610-1.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios, Biometrika, 46, 477-80.

Goodman, L. A. and Hartley, H. O. (1958). The precision of unbiased ratio-type estimators, J. Amer. Statist. Assoc., 53-491-508.

Hartley, H. O. and Ross, A. (1954). Unbiased ratio estimators. Nature, 174, 270-1.

TABLE 4

Interquartile ranges (I.R.), concentrations (C) and variances (V) of the eight ratio estimators $r, t_1, .., t_7$ (obtained from 1000 samples) for selected values of n and $\rho = 0.8$ (Tin's model)

Estimator		n = 4	n = 6	n = 10		n = 20		n = 50		
		I.R.	I.R.	I.R.	% C	I.R.	% C	I.R.	% C	V
Classical: r		1.9	1.6	1.3	7.7	0.81	15.4	0.50	23	0.15
Hartley-Ross: t_1		2.8	2.2	1.9	5.6	1.20	9.2	0.73	15	
Mickey: t_2	g = 2	2.3	1.8	1.3	8.0	0.80	15.4	0.51	22	0.15
	$g = \frac{n}{2}$	2.3	1.6	1.1	9.1	0.72	16.0	0.49	23	0.14
	g = n	1.9	1.2	1.0	10.5	0.72	16.0	0.49	23	0.14
Quenouille: t_3	g = 2	2.5	1.9	1.3	8.3	0.81	14.8	0.51	22	0.15
	$g = \frac{n}{2}$	2.5	1.7	1.1	8.5	0.70	16.0	0.49	23	0.14
	g = n	2.0	1.4	1.0	10.0	0.70	16.2	0.48	23	0.14
Beale: t_4		1.3	1.2	1.0	9.9	0.71	16.2	0.48	23	0.14
Modified: t_5		1.7	1.3	1.0	10.3	0.69	16.0	0.48	23	0.14
Tin: t_6	g = 2	2.5	1.9	1.3	8.3	0.81	14.8	0.51	22	0.15
	$g = \frac{n}{2}$	2.5	2.1	1.5	7.7	0.93	11.1	0.56	22	
Hartley:	g = n	2.9	2.1	1.5	6.1	0.88	12.7	0.55	20	
Pascual: t_7		2.8	2.2	1.6	7.1	0.89	13.3	0.55	20	

Table 5

Interquartile ranges and coefficients of variation of the variance estimators $v(\bar{y}/\bar{X})$, $v(r)$ and $v(t_3)$ for selected values of n and ρ (Tin's model)

Variance Estimator		$\rho = 0.6$				$\rho = 0.8$			
		Interquartile range			Coeff. of Variation	Interquartile range			Coeff. of Variation
		n=10	n=20	n=50	n=50	n=10	n=20	n=50	n=50
$v(\bar{y}/\bar{X})$		1.3	0.43	0.11	0.20	1.21	0.42	0.10	0.20
$v(r)$		3.1	0.97	0.20	0.61	1.60	0.45	0.09	0.51
$v(t_3)$:	g=2	4.2	1.55	0.47	2.21	2.26	0.81	0.22	2.22
	$g=\frac{n}{2}$	4.4	1.08	0.22	0.66	2.24	0.52	0.10	0.58
	g=n	3.7	0.97	0.20	0.61	1.88	0.46	0.10	0.51

Kokan, A. R. (1963). A note on the stability of the estimates of standard errors of the ordinary mean estimate and the ratio estimate, Cal. Statist. Assoc. Bull., 12, 149-58.

Lauh, E. and Williams, W. H. (1963). Some small sample results for the variance of a ratio, Proc. Amer. Statist. Assoc., (Social Statistics Section), 273-83.

Mickey, M. R. (1959). Some finite population unbiased ratio and regression estimators, J. Amer. Statist. Assoc., 54, 594-612.

Murthy, M. N. and Nanjamma, N. S. (1959). Almost unbiased ratio estimators based on interpenetrating subsamples, Sankhya, 21, 299-314.

Pascual, J. N. (1961). Unbiased ratio estimators in stratified sampling, J. Amer. Statis. Assoc., 56, 70-87.

Quenouille, M. H. (1956). Notes on bias in estimation, Biometrika, 43, 353-60.

Rao, J. N. K. (1965). A note on the estimation of ratios by Quenouille's method, Biometrika, 52, 647-9.

Rao, J. N. K. and Webster, J. T. (1966). On two methods of bias reduction in the estimation of ratios, Biometrika, 53 (in press).

Sastry, K. V. R. (1965). Unbiased ratio estimators, J. Ind. Soc. Agric. Statist., 17, 19-29.

Tin, M. (1965). Comparison of some ratio estimators, J. Amer. Statist. Assoc., 60, 294-307.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples, Ann. Math. Statist. (Abstract), 29, 614.

This rearch was supported by the National Science Foundation, under Grant GP-6400.

XVII

CONTRIBUTED PAPERS III

Chairman, MORTON S. RAFF, U. S. Bureau of Labor Statistics

A MARKOVIAN ANALYSIS OF MIGRATION DIFFERENTIALS[1]

Andrei Rogers, University of California

1. Introduction

Over the past twenty years, quantitative models of internal migration have received considerable attention in the social sciences, particularly in the areas of sociology and demography. A vast amount of data have been collected, and numerous mathematical models have been proposed to account for apparent empirical regularities. These indicate that migration is a clearly patterned nonrandom phenomenon which is subject to scientific explanation and, therefore, perhaps ultimately may be forecast with a reasonable degree of accuracy.

Internal migration may be approached from two different points of view: from the point of view of migration streams and from the point of view of migration differentials. These are not mutually exclusive conceptualizations, but each concentrates on a particular aspect of migration. Migration stream analysis focuses on the volume and direction of place-to-place movements. The analysis of migration differentials selects as its principal subject of inquiry the differences in the characteristics of migrants and nonmigrants and the differences between migrant sub-groups. Whereas the analysis of streams is concerned primarily with the effect that variations in environmental conditions at origins and destinations have on volumes of flow, the study of differentials is concerned with the traits of migrants in various age-sex-income-race classifications. Thus the problem shifts from that of accounting for changes in flow patterns to explaining in what respects migrants differ from the general population. In short, differential migration is concerned with the study of those migrant categories which have a disproportionately greater or smaller percentage of migrants than is found in the population as a whole.

The definitive work on migration differentials continues to be that of Dorothy S. Thomas, whose exhaustive findings on this topic were published almost thirty years ago.[2] Since that time several significant analyses of migration differentials have appeared. Bogue and Hagood, by cross-classifying stream characteristics, simultaneously consider the joint effects of income, age, occupation, employment, marital status and education on migration.[3] Beshers and Nishiura suggest a theory of internal migration differentials.[4] The principal hypotheses which consistently reappear in these and other studies are:

1. Young adults are the most mobile segment of the population.
2. Males tend to be more migratory than females.
3. Unemployed persons are more likely to move than employed persons.
4. Whites move more than non-whites.
5. Professionals are among the most mobile elements of the population.

Paralleling the growing interest in quantitative analysis of migration phenomena has been the emergence of Markov chain theory as a methodological tool for analyzing social, industrial and geographic mobility. Markov chains have been used to examine intergenerational mobility,[5] to study the movement of workers between industries,[6] and to project future population totals for Census Divisions in the United States.[7] By and large, however, the empirical results have been disappointing. What at first appeared as a powerful new technique for temporal analysis has been found to be generally inapplicable in much of sociological and demographic research. Fundamentally, the discouraging results stem from the restrictive assumption of unchanging movement probabilities. Such an assumption, of course, is unrealistic in light of our knowledge concerning mobility in general and interregional migration in particular. Transition probabilities vary over time as well as over space. Moreover they are dependent on differential socioeconomic, demographic and political situations at origins and destinations. Thus one may justifiably conclude that Markov chain analysis may be more useful in analyses of past migration flows and of very little practical use in efforts to forecast future place-to-place movements. However, though of limited utility in temporal analysis, it appears that Markovian concepts do provide useful indices for purposes of differential analysis. Thus despite its limited success in accounting for interregional migration streams, Markov chain theory does supply useful insights concerning the observed differential behavior of a population of migrant cohorts at a given point in time.

This paper describes an investigation of migration differentials in California. The data are the U.S. Census reported flows for the 1955-1960 time period and supplementary estimates provided by a recent study completed for the California State Development Plan.[8] The method of analysis utilizes the Markovian concepts of transition matrices, mean first passage times and equilibrium distributions.

2. Markovian Analysis of Migration Differentials

Consider an interregional system of m regions and a population composed of n cohorts. Define a cohort as a group of persons who behave independently but according to an identical migration structure. That is, assume that a member of cohort r behaves independently of all other members and according to an m by m transition matrix P_r. Then we may estimate each element of P_r by means of observed proportions taken over a cohort class, i.e.,

$$_{r}p_{ij} = \frac{_{r}k_{ij}}{\sum_{j=1}^{m} {_{r}k_{ij}}}, \quad (r = 1, 2, \ldots.., n) \quad (i, j = 1, 2, \ldots.., m)$$

where k_{ij} = the number of people, who during a specified time period, moved from region i to region j.

With cohort-specific data on migration propensities, we may begin to study the changes of state that a single individual is likely to undergo in light of the transition structure of his cohort class. More specifically, for each cohort, we may identify current movement characteristics and thereby establish a series of intra-cohort contrasts. Three properties of transition structures serve as particularly useful indices: the cohort's transition matrix, the associated mean first passage time matrix and the equilibrium vector.

Transition Matrices

Cohort-specific transition matrices provide a great deal of information about the mobility of migrant classes. In particular, their diagonal elements provide an immediate dimension along which we may contrast the degree of overall mobility of different migrant groups. For example, consider a hypothetical system of only two regions, A and B, and a population divided into two broad cohort classes, white and non-white. Let us suppose that if an individual, in the white cohort class, is in region A there is a 50 per cent chance that he will move to region B during the unit time interval. If the person is currently in region B, however, with probability 1/4 he will move to A during the same time period. Assume, further, that for the non-white cohort class the corresponding probabilities are 1/4 and 1/5. In matrix form we have then:

$$P_W = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \end{array} \qquad P_{NW} = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 3/4 & 1/4 \\ 1/5 & 4/5 \end{pmatrix} \end{array}.$$

Immediately we observe that the diagonal elements of P_{NW} are greater than the corresponding entries in P_W. From this we may infer that non-whites are less mobile than whites. If our interregional system contained more than two regions, we would, in addition, be in a position to compare the relative "attraction" of alternate destinations for different migrant cohorts.

Mean First Passage Times

Frequently it is desirable to study the length of time that it takes an average individual to move from state i to state j for the first time. The distribution describing this random variable is called the first passage time distribution. Its mean is commonly referred to as the mean first passage time.

Turning to our two-region example, consider the probability that an individual currently in region A will move to region B, <u>for the first time</u>, in n time periods. Denote this probability by $g_{AB}^{(n)}$ and begin with n equal to 1. Then,

$$g_{AB}^{(1)} = p_{AB}$$

$$g_{AB}^{(2)} = p_{AA} \cdot p_{AB}$$

and by substitution

$$g_{AB}^{(2)} = p_{AA} g_{AB}^{(1)}$$

The above equations merely state than an individual's probability of going from A to B, <u>for the first time</u>, in one time period is p_{AB} (by definition), and the probability of doing this in two steps is the product of the probability of remaining in A during the first time period and the probability of moving to B during the second time period.

Extending the argument to the general case, for this two-region example, we have:

$$\begin{aligned} g_{AB}^{(n)} &= p_{AA} \cdot g_{AB}^{(n-1)} \\ &= p_{AA} \cdot p_{AA} \cdot g_{AB}^{(n-2)} \\ &= \ldots \\ &= (p_{AA})^{n-1} \cdot p_{AB} \end{aligned}$$

This function is called the first passage time distribution. Since $p_{AA} = 1 - p_{AB}$, we have

$$g_{AB}^{(n)} = p_{AB}(1 - p_{AB})^{n-1}$$

which is the geometric distribution with a mean

$$m_{AB} = \frac{1}{p_{AB}}$$

This statistic is defined as the mean first passage time and represents the average number of time periods required for a person in region A to visit region B for the first time. The matrix M, consisting of entries m_{ij}, is defined as the mean first passage time matrix.

Returning to our numerical example, we find for the white cohort:

$$g_{AB}^{(n)} = (1/2)\ (1/2)^{n-1}$$

and

$$m_{AB} = \frac{1}{1/2} = 2 \, .$$

In general, the mean first passage times of a Markov chain may be found by recursively applying the following equation:[9]

$$m_{ij} = p_{ij} + \sum_{k \neq i} p_{ik} m_{ki} \, .$$

Kemeny and Snell, however, offer a more convenient matrix formulation:

$$M = (I - Z + EZ_{dg})D$$

where

- D = a diagonal matrix with elements $d_{ii} = 1/a_i$;
- E = a matrix with all elements equal to 1;
- I = the identity matrix;
- Z = the fundamental matrix;
- Z_{dg} = the Z matrix with all off-diagonal entries set equal to 0.

The fundamental matrix, Z, is defined by the following equation:

$$Z = (I - (P-A))^{-1} \, ,$$

where

- P = the matrix of transition probabilities;
- A = a matrix with each row identically equal to the equilibrium vector $\underline{a}$.

The computation of the matrix of mean first passage times may be illustrated by returning to our example:

$$Z_W = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \left[\begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} - \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix} \right] \right\}^{-1}$$

$$= \begin{pmatrix} 5/6 & 1/6 \\ 1/12 & 11/12 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 11/9 & -2/9 \\ -1/9 & 10/9 \end{pmatrix} ,$$

and

$$M_W = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 11/9 & -2/9 \\ -1/9 & 10/9 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 11/9 & 0 \\ 0 & 10/9 \end{pmatrix} \right] \begin{pmatrix} 3 & 0 \\ 0 & 1\,1/2 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 2 \\ 4 & 1\,1/2 \end{pmatrix} .$$

As a check, notice that m_{12} is again equal to 2.

Repeating the above computation for the non-white cohort, we have:

$$M_{NW} = \begin{pmatrix} 2\,1/4 & 4 \\ 5 & 1\,4/9 \end{pmatrix} .$$

Mean first passage times provide a measure of a particular kind of contiguity--one based on interchange probabilities rather than distance. Thus they may be viewed as indices of aspatial interregional distance. Let us define this aspatial measure of proximity as "migrant distance."

With reference to our two-region, two-cohort example, we may make both an intra-cohort observation and an inter-cohort contrast:

(1) "Migrant distance" from region A to region B, for both cohorts, is "shorter" than the distance from B to A. This asymmetry suggests that, on the basis of actual migrant exchange, B is "closer" to the population at A than A is to the population at B.

(2) White "migrant distance" from region A to region B is "shorter" than non-white "migrant distance" between the same two regions.

Equilibrium or Limiting State Probabilities

The transition matrix P provides a great deal of information about the Markov process described above. For example, it allows us to derive the probability that an individual currently residing in region A will be in region B after 2 years. This "event" can occur only in one of two mutually exclusive and collectively exhaustive ways:

(1) the individual remains in A during the first year and migrates to B during the

second year;

(2) the individual migrates to B during the first year and remains in B during the second year.

Therefore, for the white cohort,

$$p_{AB}^{(2)} = p_{AA}p_{AB} + p_{AB}p_{BB}$$
$$= (1/2)(1/2) + (1/2)(3/4)$$
$$= 5/8 .$$

With analogous arguments we find:

$$p_{AA}^{(2)} = (1/2)(1/2) + (1/2)(1/4) = 3/8,$$

$$p_{BA}^{(2)} = (1/4)(1/2) + (3/4)(1/4) = 5/16,$$

$$p_{BB}^{(2)} = (1/4)(1/2) + (3/4)(3/4) = 11/16.$$

These numbers can be presented in a matrix:

$$P_W^{(2)} = \begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{pmatrix} \end{array}.$$

The matrix $P^{(2)}$ describes movement between two periods of time. Similarly $P^{(n)}$ describes movement during n time periods. It should now become apparent that the transition matrix P, in a Markov model, completely determines the character of the migration process. Therefore, it is possible to use this short term data to compare the movement patterns of different classes of individuals, to project these into the future, and to assess what are the intrinsic distributional consequences of a particular movement structure.

The essential feature of representing Markov processes by transition matrices stems from the ease with which nth order transition probabilities may be derived by matrix multiplication. In particular, the multiplication of the transition probability matrix P by itself, n number of times, yields the nth-order transition probabilities. For example, it can be shown that:

$$P^{(2)} = P \cdot P = P^2$$

and, in general,

$$P^{(n)} = P^n \quad .$$

This can be demonstrated by our example:

$$P_W = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} = \begin{pmatrix} .50 & .50 \\ .25 & .75 \end{pmatrix}$$

$$P_W^2 = \begin{pmatrix} .38 & .62 \\ .31 & .69 \end{pmatrix}$$

$$P_W^3 = \begin{pmatrix} .34 & .66 \\ .33 & .67 \end{pmatrix}$$

$$P_W^5 = \begin{pmatrix} .33 & .67 \\ .33 & .67 \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix} .$$

Similarly, for the non-white cohort:

$$P_{NW}^8 = \begin{pmatrix} 4/9 & 5/9 \\ 4/9 & 5/9 \end{pmatrix} .$$

An interesting and very important feature of a class of Markov processes, defined as "ergodic" chains, is illustrated by the above matrices. It will be noted that initially the transition probabilities are different for each of the two states. That is, a migrant's destination is heavily influenced by his place of origin. However, after n powers of the transition matrix are calculated, it becomes apparent that the effect of the starting state diminishes. For example, for the white cohort this occurs when n is equal to 5. For this and larger values of n, the rows of the transition matrix are identical. This means that as n increases, $p_{ij}^{(n)}$, the probability of migrating from i to j in n years, approaches a limit p_j which is independent of i. At this point the system is said to be in "equilibrium" or to have reached a "steady state."

Comparing the equilibrium vectors of the white and non-white cohorts in our example suggests something of the long-term implications of current behavior. It is an abstract index, to be sure, since "death" is not included as a possible end-state. Nevertheless, the steady state vector may be viewed as a kind of "speedometer" which describes the ultimate consequences of the current movement pattern if it remains unchanged. Instead of assuming that the driver doesn't die and that his car continues at exactly the same speed, we assume that the migrant doesn't die and that the transition probabilities remain constant.

In our example, we note that on the basis of current trends it appears that the white cohort is favoring region B as a destination. A similar observation may be made with respect to the non-white cohort.

3. Migration Differentials in California: Some Empirical Results

According to the Census of 1960, over 2.1 million persons migrated to California between 1955 and 1960 while slightly under a million departed, thus producing a net increase of some 1.2 million people over the five-year period.[10] Origins and destinations for these migrants, by 19 State Economic Areas, have been published[11] and total age- and color-specific intrastate flows and transition matrices have been estimated.[12] For ease of exposition, however, we shall structure the discussion around selected matrices of a smaller order. In particular, we shall focus on the reduced versions which are exhibited in Tables 1 through 6.[13]

Transition Matrices

Several interesting findings are suggested by the transition matrices. These are by no means surprising and, indeed, merely support relatively well-established demographic hypotheses.

First, it is clear that the transition probabilities have not remained constant over time. The population has become much more mobile both at the interstate and the intrastate levels. Second, there are significant differences between the characteristics of white and non-white flow patterns. Non-white probabilities are considerably higher than white probabilities in urban to urban transfers and much lower in urban to suburban-rural movements. Finally, considerable differences appear to exist between the migration structures of various age groups. The most mobile age groups are the 15 to 19 and 20 to 24 age groups; the least mobile are the post-65-year age groups.

Temporal Differentials: The transition matrix for California SEA's has changed considerably over time. This is immediately apparent from even the most cursory examination of Table 1. In every instance the diagonal element of the 1935-1940 matrix is larger than the corresponding diagonal element of the 1955-1960 matrix. This points to the greater mobility of today's population. For example, for the 1935-1940 cohort, the probability that a member of the San Francisco-Oakland population moves out of that SMSA is less than .09. The corresponding figure for the 1955-1960 time interval is almost .15. The change for other SEA's is less striking, but is significant nevertheless.

Color Differentials: Two major points should be noted concerning the white and non-white transition matrices presented in Tables 2 and 6. First, the data clearly show that, on the whole, whites are more mobile than non-whites. Every diagonal element of the non-white matrix in Table 2 is larger than the corresponding element of the white matrix. For example, the probability that an individual of the non-white cohort in the Los Angeles-Long Beach SMSA moves out of that subregion during the 1955-1960 period is less than 0.6. The corresponding figure for whites is exactly twice that number.

The second finding concerns rural to urban transfers. Non-white movements are primarily urban to urban migrations. Non-white probabilities are relatively higher than white probabilities in SMSA to SMSA movements, but are much lower in SMSA to non-SMSA transfers.

Age Differentials: Tables 3, 4, and 5 highlight the age-specific mobility pattern which emerges out of an analysis of the transition matrices of the 17 age cohorts in California. Although considerable differences exist between individual SEA's, the overall distribution is unmistakable. The probability of leaving an SEA is highest for the 15- to 19- and 20- to 24-year age groups and lowest for the post-65-year age groups. The distribution is unimodal and resembles the Gamma distribution. The high values are distributed around .40 with the low values approaching zero. The maximum is attained by the South Central Coast SEA. Here the probability that an individual in the 15- to 19-year age group moves out of this SEA is almost .44.

Mean First Passage Times

Tables 7, 8, and 9 present mean first passage time matrices for six of the eight transition matrices appearing in Tables 1, 2, 4, and 5. The actual values of these "migrant distances" are quite meaningless; however, when considered in relative terms, they suggest several interesting findings concerning spatial and aspatial contiguities among California's major SMSA's.

A quick glance at the 1935-1940 and 1955-1960 mean first passage time matrices reveals changes both in intra- and inter-matrix levels. On the whole, it is clear that migrant distances declined over the twenty-year period--a reflection of increased geographical mobility. Other changes, however, are equally noteworthy. Perhaps the most noticeable is the shortening of migrant distances in relation to the distance between the Los Angeles-Long Beach and San Francisco-Oakland SMSA's. For example, whereas during the 1935-1940 period the migrant distance from Los Angeles to San Jose was over four times that of the migrant distance between Los Angeles and San Francisco, in 1955-1960 this ratio declined to two to one.

Differences both within and between the white and non-white mean first passage time matrices are quite apparent. Particularly striking are the non-white migrant distances to the San Jose SMSA. The non-white migrant distance between the San Francisco Oakland and the San Jose SMSA's, for example, is nine times the reverse distance and thirteen times the distance between the San Francisco and Los Angeles SMSA's. This probably is a reflection of the racial discrimination in San Jose's housing market.

The mean first passage time matrices for the 20- to 24- and 65- to 69-year age groups differ considerably in absolute values but are very similar in relative terms. This is an indication that, although the former age group is much more mobile than the latter age group, their movement patterns are quite similar. For example, in both matrices the distance from San Jose to Sacramento is three times that of the reverse distance.

Finally, it is interesting to note the total absence of any significant correlation between interregional highway-mileage distances (Table 10) and interregional migrant distances as measured by mean first passage times. Table 11 presents the correlations between each of the mean first passage time matrices in Tables 7, 8, and 9 and the interregional distances shown in Table 10. Clearly the spatial and aspatial measures of interregional distances are totally unrelated.

Equilibrium Distributions

The migration differentials revealed by the transition matrices in Tables 1, 2, 3, 4, 5, and 6 are readily recognizable. Differences in the propensity to move are immediately apparent. Not so obvious perhaps, are the implied distributional consequences of the various transition structures. For example, a comparison of the equilibrium vectors of the 1935-1940 and the 1955-1960 transition matrices suggests that California's share of the national population is going to taper off at a lower level than indicated by pre-World War II trends. This is not immediately apparent from a consideration of the transition matrices alone.

At more disaggregated levels, the equilibrium solutions present a detailed, quantitative picture of the spatial implications of current mobility trends. Moreover, they provide indications of temporal changes and of differentials between migrant sub-classes.

Temporal Differentials: The temporal changes in the values of the equilibrium vectors for California's population have little meaning other than as an index of the direction of changes in regional preferences over time. Perhaps the most significant finding in Table 1 is the decline in the equilibrium probabilities of the San Francisco-Oakland and Los Angeles-Long Beach SMSA's. This, however, is not an unexpected trend, especially when viewed against the increasing equilibrium probabilities of the San Jose, Sacramento and San Diego SMSA's.

Color Differentials: The most striking finding arising out of the equilibrium vectors in Table 2 is the overwhelming expected concentration of non-whites in the Los Angeles and San Francisco regions. Of the projected non-white share for California, well over half are expected to settle in the Los Angeles-Long Beach SMSA and about a fifth should locate in the San Francisco-Oakland SMSA. This is in marked contrast to the white equilibrium vector. The latter exhibits a relatively more uniform distribution, though it too shows a significant concentration in the Los Angeles subregion.

Age Differentials: Despite considerable differences between age-specific transition matrices, the equilibrium vectors of the six age groups analyzed in Tables 3, 4, and 5 are, on the whole, quite similar. The major difference appears in the California-Rest of the U.S. probability allocation. Thus, for example, whereas for the 20- to 24-year age group this division is .193-.807, for the 65- to 69-year age group the corresponding split is .138-.862. Among the five SMSA's, however, the vector does not vary substantially between age groups.

4. Conclusion

This paper has borrowed concepts from Markov chain theory to identify and analyze migration differentials. Transition matrices were used to establish the movement propensities of each migrant cohort. Mean first passage times defined aspatial measures of interregional "migrant distance." Finally, equilibrium distributions pointed to the distributional tendencies of different classes of migrants.

The basic Markovian model is conceptually simple and rests on very strict assumptions concerning human behavior. Because of this, it is an analytic system which shows only limited promise as a tool for long-term forecasting of interregional flows. However, as a technique for analyzing differential behavior during an observed period, it appears to provide insights which are not readily obtainable by other means.

5. Footnotes

1 This study began in 1964 as one of several Phase II California State Development Plan Studies conducted by the Center for Planning and Development Research at the University of California. The research was prepared under contract to the State Office of Planning and was financed in part through an urban planning grant from the Housing and Home Finance Agency, under the provisions of Section 701 of the Housing Act of 1954, as amended. The author is indebted for this financial assistance.

2 Dorothy S. Thomas, Research Memorandum on Migration Differentials (New York: Social Science Research Council, 1938), Bulletin 43.

3 Donald J. Bogue and Margaret Jarmon Hagood, Differential Migration in the Corn and Cotton Belts (Oxford, Ohio: Scripps Foundation, 1953).

4 James M. Beshers and Eleanor N. Nishuira, "A Theory of Internal Migration Differentials," Social Forces (1961), 39, pp. 214-218.

5 S. J. Prais, "Measuring Social Mobility," Journal of the Royal Statistical Society, Series A (1955), pp. 56-66; John G. Kemeny and J. Laurie Snell, Finite Markov Chains (Princeton,

N.J.: D. Van Nostrand Co., 1960), pp. 191-200.

6 Isadore Blumen, Marvin Kogan and Philip J. McCarthy, The Industrial Mobility of Labor as a Probability Process, Vol. VI, Cornell Studies of Industrial and Labor Relations (Ithaca, New York: The New York State School of Industrial and Labor Relations, Cornell University, 1955).

7 James D. Tarver and William R. Gurley, "A Stochastic Analysis of Geographic Mobility and Population Projections of the Census Divisions in the United States," Demography, II (1965), pp. 134-139. Also see: Robert McGinnis and John E. Pilger, "On a Model for Temporal Analysis," paper presented at the 58th Annual Meeting of the American Sociological Association, Los Angeles, August 29, 1963.

8 Andrei Rogers, An Analysis of Interregional Migration in California, Center for Planning and Development Research, University of California, Berkeley, California, December 1965.

9 For a derivation of this equation, see: John G. Kemeny and J. Laurie Snell, op. cit., pp. 78-80. Their notation has been retained in order to reduce possible confusion.

10 Financial and Population Research Section, California Migration: 1955-1960, California Department of Finance, Sacramento, 1964, p. 1.

11 U.S. Census of Population, 1960, Mobility for States and State Economic Areas, U.S. Bureau of the Census, Department of Commerce, 1963.

12 These were developed in the study: Andrei Rogers, Projected Population Growth in California Regions: 1960-1980, Center for Planning and Development Research, University of California, December 1965. For estimating procedures see pp. 15-17 of that study.

13 Transition matrices for the 1935-1940 time period were derived from interregional flow data reported in: Donald J. Bogue, Henry S. Shryock, Jr., and Siegfried A. Hoermann, Subregional Migration in the United States, 1935-40, Volume I: Streams of Migration (Oxford, Ohio: Scripps Foundation. Miami University, 1957).

TABLE 1. TRANSITION MATRICES AND EQUILIBRIUM DISTRIBUTIONS FOR CALIFORNIA: BY TIME PERIOD.*

A. 1935-1940 Total Flows

	A	B	C	F	G	CAL.	U.S.
A	.9139	.0067	.0049	.0615	.0022	.0293	.0265
B	.0575	.8529	.0056	.0121	.0034	.0459	.0226
C	.0379	.0030	.8434	.0125	.0019	.0741	.0272
F	.0096	.0012	.0013	.9215	.0058	.0242	.0364
G	.0147	.0014	.0043	.0498	.8371	.0208	.0719
CAL.	.0280	.0059	.0086	.0031	.0044	.8912	.0288
U.S.	.0009	.0001	.0001	.0033	.0004	.0020	.9932

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0387	.0050	.0053	.0688	.0065	.0479	.8279)

B. 1955-1960 Total Flows

	A	B	C	F	G	CAL.	U.S.
A	.8543	.0203	.0070	.0172	.0053	.0363	.0596
B	.0460	.8271	.0053	.0155	.0043	.0465	.0553
C	.0247	.0061	.8165	.0142	.0034	.0667	.0684
F	.0076	.0043	.0030	.8907	.0078	.0324	.0542
G	.0120	.0046	.0019	.0371	.7923	.0255	.1266
CAL.	.0209	.0099	.0109	.0327	.0078	.8538	.0640
U.S.	.0017	.0006	.0004	.0056	.0016	.0028	.9873

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0253	.0107	.0070	.0667	.0116	.0456	.8331)

*A = S.F. - Oakland
B = San Jose
C = Sacramento
F = Los Angeles
G = San Diego
Cal. = Rest of California
U.S. = Rest of the U.S.

TABLE 2. TRANSITION MATRICES AND EQUILIBRIUM DISTRIBUTIONS FOR CALIFORNIA: BY COLOR

A. 1955-1960 White Flows

	A	B	C	F	G	CAL.	U.S.
A	.8465	.0221	.0073	.0172	.0056	.0388	.0625
B	.0453	.8269	.0053	.0154	.0044	.0465	.0562
C	.0247	.0063	.8118	.0141	.0035	.0689	.0707
F	.0077	.0046	.0032	.8863	.0082	.0339	.0561
G	.0117	.0048	.0019	.0367	.7897	.0260	.1292
CAL.	.0208	.0102	.0110	.0326	.0080	.8526	.0648
U.S.	.0018	.0006	.0004	.0058	.0017	.0030	.9867

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0249	.0112	.0071	.0661	.0121	.0477	.8309)

B. 1955-1960 Non-white Flows

	A	B	C	F	G	CAL.	U.S.
A	.9174	.0059	.0044	.0171	.0026	.0162	.0364
B	.0660	.8341	.0052	.0190	.0027	.0439	.0291
C	.0245	.0031	.8792	.0156	.0016	.0376	.0384
F	.0062	.0009	.0011	.9437	.0030	.0143	.0308
G	.0166	.0011	.0012	.0444	.8425	.0182	.0760
CAL.	.0234	.0048	.0095	.0356	.0058	.8728	.0481
U.S.	.0016	.0001	.0002	.0045	.0007	.0011	.9918

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0371	.0033	.0060	.1023	.0073	.0273	.8167)

TABLE 3. TRANSITION MATRICES AND EQUILIBRIUM DISTRIBUTIONS FOR CALIFORNIA: BY AGE GROUP

A. 1955-1960 Flows for Age Group #2: 5 to 9 years

	A	B	C	F	G	CAL.	U.S.
A	.8458	.0215	.0074	.0182	.0056	.0384	.0631
B	.0469	.8238	.0054	.0158	.0044	.0473	.0564
C	.0256	.0063	.8101	.0147	.0035	.0690	.0708
F	.0081	.0046	.0032	.8834	.0083	.0346	.0578
G	.0116	.0045	.0018	.0358	.7997	.0245	.1221
CAL.	.0215	.0101	.0113	.0338	.0081	.8497	.0655
U.S.	.0016	.0005	.0004	.0050	.0014	.0025	.9886

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0223	.0095	.0064	.0573	.0109	.0408	.8527)

B. 1955-1960 Flows for Age Group #4: 15 to 19 years

	A	B	C	F	G	CAL.	U.S.
A	.6952	.0424	.0146	.0359	.0111	.0761	.1247
B	.0740	.7221	.0086	.0250	.0069	.0745	.0889
C	.0433	.0107	.6782	.0249	.0059	.1170	.1200
F	.0142	.0081	.0057	.7952	.0146	.0607	.1015
G	.0223	.0086	.0035	.0690	.6136	.0475	.2355
CAL.	.0334	.0167	.0168	.0518	.0123	.7668	.1022
U.S.	.0033	.0011	.0008	.0106	.0030	.0052	.9760

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0228	.0125	.0075	.0666	.0116	.0533	.8255)

TABLE 4. TRANSITION MATRICES AND EQUILIBRIUM DISTRIBUTIONS FOR CALIFORNIA: BY AGE GROUP

A. 1955-1960 Flows for Age Group #5: 20 to 24 years

	A	B	C	F	G	CAL.	U.S.
A	.7288	.0377	.0130	.0320	.0099	.0676	.1110
B	.0710	.7332	.0082	.0240	.0067	.0715	.0854
C	.0445	.0110	.6693	.0256	.0061	.1202	.1233
F	.0138	.0079	.0056	.8006	.0142	.0590	.0989
G	.0204	.0079	.0032	.0630	.6468	.0435	.2152
CAL.	.0324	.0161	.0164	.0531	.0271	.7668	.1025
U.S.	.0036	.0012	.0009	.0116	.0033	.0057	.9737

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0272	.0138	.0078	.0735	.0137	.0570	.8070)

B. 1955-1960 Flows for Age Group #8: 35 to 39 years

	A	B	C	F	G	CAL.	U.S.
A	.8825	.0163	.0056	.0138	.0043	.0294	.0481
B	.0391	.8531	.0045	.0132	.0037	.0394	.0470
C	.0215	.0053	.8403	.0124	.0029	.0581	.0595
F	.0063	.0036	.0025	.9097	.0064	.0267	.0448
G	.0087	.0033	.0013	.0268	.8497	.0186	.0916
CAL.	.0186	.0086	.0099	.0291	.0069	.8701	.0568
U.S.	.0016	.0005	.0004	.0050	.0014	.0025	.9886

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0283	.0109	.0074	.0713	.0140	.0455	.8226)

TABLE 5. TRANSITION MATRICES AND EQUILIBRIUM DISTRIBUTIONS FOR CALIFORNIA: BY AGE GROUP

A. 1955-1960 Flows for Age Group #11: 50 to 54 years

	A	B	C	F	G	CAL.	U.S.
A	.9293	.0098	.0034	.0083	.0026	.0177	.0289
B	.0256	.9038	.0030	.0086	.0024	.0258	.0308
C	.0130	.0032	.9034	.0075	.0018	.0351	.0360
F	.0043	.0025	.0017	.9378	.0044	.0185	.0308
G	.0053	.0020	.0008	.0163	.9090	.0111	.0555
CAL.	.0137	.0062	.0074	.0205	.0050	.9062	.0410
U.S.	.0008	.0003	.0002	.0026	.0007	.0013	.9941

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0252	.0094	.0067	.0557	.0123	.0341	.8566)

B. 1955-1960 Flows for Age Group #14: 65 to 69 years

	A	B	C	F	G	CAL.	U.S.
A	.9383	.0086	.0030	.0073	.0023	.0152	.0253
B	.0228	.9145	.0027	.0077	.0021	.0228	.0274
C	.0108	.0026	.9195	.0063	.0015	.0292	.0301
F	.0038	.0022	.0015	.9456	.0039	.0160	.0270
G	.0046	.0018	.0007	.0142	.9206	.0097	.0484
CAL.	.0116	.0052	.0060	.0175	.0042	.9206	.0349
U.S.	.0007	.0002	.0002	.0021	.0006	.0011	.9951

Equilibrium Vector:

	A	B	C	F	G	CAL.	U.S.
a =	(.0245	.0083	.0069	.0525	.0119	.0336	.8622)

TABLE 6. TRANSITION MATRICES AND EQUILIBRIUM DISTRIBUTIONS FOR CALIFORNIA: BY SMSA AND NON-SMSA FLOWS

A. Total Flows 1955-1960

	S.	N.S.	U.S.
SMSA	.9167	.0211	.0622
NON-SMSA	.1120	.8218	.0662
U.S.	.0114	.0013	.9873

B. White Flows 1955-1960

	S.	N.S.	U.S.
SMSA	.9135	.0221	.0644
NON-SMSA	.1121	.8214	.0665
U.S.	.0119	.0014	.9867

C. Non-white Flows 1955-1960

	S.	N.S.	U.S.
SMSA	.9556	.0085	.0359
NON-SMSA	.1101	.8326	.0573
U.S.	.0077	.0005	.9918

Equilibrium Vector:

	S.	N.S.	U.S.
A. a =	.1451	.0232	.8317
B. a =	.1459	.0246	.8295
C. a =	.1695	.0111	.8194

D. Flows for 15-19 Age Group 1955-1960

	S.	N.S.	U.S.
SMSA	.8422	.0391	.1187
NON-SMSA	.1788	.7150	.1062
U.S.	.0215	.0025	.9760

E. Flows for 35-39 Age Group 1955-1960

	S.	N.S.	U.S.
SMSA	.9319	.0175	.0506
NON-SMSA	.0995	.8416	.0589
U.S.	.0102	.0012	.9886

F. Flows for 50-54 Age Group 1955-1960

	S.	N.S.	U.S.
SMSA	.9554	.0114	.0332
NON-SMSA	.0708	.8873	.0419
U.S.	.0053	.0006	.9941

Equilibrium Vector:

	S.	N.S.	U.S.
D. a =	.1436	.0270	.8294
E. a =	.1571	.0236	.8193
F. a =	.1293	.0176	.8530

TABLE 7. MEAN FIRST PASSAGE TIMES: BY TIME PERIOD

A. 1935-1940 Total Flows

	A	B	C	F	G	CAL.	U.S.
A	25.8	1304.2	1174.6	145.0	915.1	146.5	33.3
B	179.0	200.0	1161.7	145.5	915.0	127.2	34.1
C	215.8	1339.2	188.7	144.4	913.5	106.1	33.0
F	291.9	1400.9	1239.8	14.5	889.8	171.0	29.5
G	301.7	1416.0	1237.1	136.2	153.8	192.1	23.4
CAL.	238.1	1329.2	1154.7	128.6	898.8	20.9	32.2
U.S.	382.1	1492.7	1333.5	216.4	990.3	264.1	1.2

B. 1955-1960 Total Flows

	A	B	C	F	G	CAL.	U.S.
A	39.5	485.2	749.8	119.3	398.0	118.1	16.3
B	204.9	93.5	753.8	119.4	399.3	110.6	16.6
C	238.0	530.2	142.9	120.8	402.2	102.1	15.4
F	264.1	541.2	771.3	15.0	389.0	121.3	16.8
G	271.5	554.3	791.3	116.1	86.2	143.1	11.2
CAL.	243.4	521.9	737.6	109.8	392.4	21.9	15.8
U.S.	300.1	583.8	817.2	143.0	421.7	174.5	1.2

TABLE 8. MEAN FIRST PASSAGE TIMES: BY COLOR

A. 1955-1960 White Flows

	A	B	C	F	G	CAL.	U.S.
A	40.2	460.5	720.6	116.4	376.6	111.2	15.7
B	199.4	89.3	725.0	116.4	378.0	105.1	16.2
C	230.8	505.9	140.8	117.8	380.8	96.6	15.0
F	255.2	515.8	740.7	15.1	367.9	114.5	16.3
G	262.7	529.1	760.8	113.2	82.6	135.6	10.9
CAL.	235.6	497.5	709.0	107.0	371.3	21.0	15.4
U.S.	289.1	557.6	785.4	138.4	399.0	165.1	1.2

B. 1955-1960 Non-white Flows

	A	B	C	F	G	CAL.	U.S.
A	27.0	1762.5	1335.1	135.9	846.2	248.0	27.1
B	190.1	303.0	1319.7	130.2	841.2	204.0	27.9
C	257.2	1810.4	166.7	136.0	851.1	207.8	26.3
F	309.6	1866.5	1388.2	9.8	838.8	252.4	29.1
G	303.5	1876.0	1400.6	123.7	137.0	270.4	20.7
CAL.	272.0	1808.0	1305.8	121.8	830.2	36.6	25.0
U.S.	363.0	1937.6	1455.0	180.3	890.0	339.3	1.2

TABLE 9. MEAN FIRST PASSAGE TIMES: BY AGE GROUP

A. 1955-1960 Flows for Age Group #5: 20 to 24 years

	A	B	C	F	G	CAL.	U.S.
A	36.8	243.3	373.2	59.3	198.3	58.7	9.2
B	103.1	72.5	375.9	59.9	199.5	55.6	10.0
C	119.1	265.8	128.2	60.1	200.5	50.8	8.9
F	132.0	271.3	383.7	13.6	193.9	60.4	9.6
G	135.3	277.7	393.3	57.6	73.0	71.0	6.5
CAL.	122.5	261.9	368.8	54.8	195.8	17.5	9.4
U.S.	148.5	291.7	404.8	70.0	209.1	85.9	1.2

B. 1955-1960 Flows for Age Group #14: 65 to 69 years

	A	B	C	F	G	CAL.	U.S.
A	40.8	1267.0	1733.8	307.3	1011.1	298.7	35.3
B	497.2	120.5	1740.3	305.3	1013.4	277.8	34.2
C	580.2	1385.6	145.0	310.5	1022.2	257.7	32.8
F	641.0	1408.7	1781.9	19.0	986.1	303.6	34.2
G	664.6	1448.9	1833.5	301.8	84.0	363.0	25.7
CAL.	589.2	1361.9	1702.0	280.9	995.2	29.8	31.4
U.S.	736.6	1529.5	1893.7	373.0	1078.2	443.2	1.2

TABLE 10. INTERREGIONAL DISTANCES*

	A	B	C	F	G	CAL.	U.S.
A	--	48	89	403	522	--	--
B	48	--	125	366	485	--	--
C	89	125	--	383	502	--	--
F	403	366	383	--	120	--	--
G	522	485	502	120	--	--	--
CAL.	--	--	--	--	--	--	--
U.S.	--	--	--	--	--	--	--

* County seat to county seat highway mileages.

TABLE 11. CORRELATIONS BETWEEN INTERREGIONAL MEAN FIRST PASSAGE TIMES AND INTERREGIONAL DISTANCES*

	R
Temporal:	
1935-1940 matrix	.024
1955-1960 matrix	-.012
Color:	
White	-.015
Non-white	-.047
Age:	
20- to 24-year age group	-.014
65- to 69-year age group	-.005

* Computed on the basis of twenty observations.

AN EMPIRICAL STUDY OF ALTERNATIVE ESTIMATORS FOR μ AND σ^2

Donald T. Searls, Westat Research, Inc.

In two previous papers [1], [2] estimators for μ were developed which tended to minimize the effect of large true observations occurring in small samples. Proofs that these estimators could have smaller mean-squared errors than $\bar{y}$ were presented.

This paper will consider analagous estimators for σ^2, however since the proofs become extremely cumbersome the gains will be demonstrated with an empirical sampling study.

The first estimator considered is the one where observations larger than a predetermined cutoff point, t, are discarded. This procedure leads to the estimators $\bar{y}_1$, and s_1^2.

$$\bar{y}_1 = \frac{\sum_{j=1}^{r} y_j}{r} \qquad (y_j < t) \quad (r > 0)$$

$$= t \; ; \qquad (r = 0)$$

$$s_1^2 = \frac{\sum_{j=1}^{r} (y_j - \bar{y}_1)^2}{r - 1} , \qquad (y_j < t) \quad (r \geq 2)$$

$$= 0 , \qquad (r < 2)$$

where r is the number of observations less than t.

The next estimators are formed by substituting the value of t for those observations greater than t. For the estimator of σ^2 only one t is used.

$$\bar{y}_t = \frac{\sum_{j=1}^{r} y_j + (n - r)t}{n} \qquad (0 \leq r \leq n) \quad (y_j \leq t)$$

Let $T = \sum_{j=1}^{r} y_j + t$,

$$S = \sum_{j=1}^{r} y_j^2 + t^2 .$$

then

$$s_t^2 = \frac{(r + 1)\, S - T^2}{r(r + 1)} , \qquad (r \geq 1)$$

$$= 0 . \qquad (r = 0)$$

If one is unwilling to discard observations an optimum weight for the observations [2] and squared deviations can be derived.

$$\bar{y}_w = [1/(n + v^2)] \sum_{j=1}^{n} y_j$$

where $\quad v^2 = \sigma^2/\mu^2$.

A similar development for estimating σ^2 gives:

$$s_w^2 = W \sum_{j=1}^{n} (y_j - \bar{y})^2$$

$$\text{MSE}\,(s_w^2) = W^2\,(n-1)^2\,(\mu_4-\sigma^4)/n + \sigma^4[1-(n-1)W]^2 .$$

Differentiating with respect to W and solving gives approximately

$$W = 1/\,[(n-1) + \beta_2] .$$

Terms with small contributions as n increases have been deleted.

Hence

$$s_w^2 = \frac{\sum_{j=1}^{n} (y_j - \bar{y})^2}{(n - 1) + \beta_2}$$

where $\beta_2 = \mu_4/\sigma^4$.

Since $\beta_2 = 3$ for the normal distribution,

$$s_w^2 = \frac{\sum_{j=1}^{n} (y_j - \bar{y})^2}{n + 2} \quad \text{for this situation.}$$

Positively skewed populations were used to demonstrate the effectiveness of the alternatives. See table 1 for the tabulation of the 1000 observations obtained from 200 samples of size 5 from each population.

Table 1. Number of observations by intervals and other characteristics for populations used.

	Distribution	
Interval	I	II
0-1	667	134
1-2	187	562
2-3	68	171
3-4	41	75
4-5	19	38
5-6	9	9
6-7	4	5
7-8	4	2
8-9	1	2
9+	0	2
Total	1000	1000

Parameter	I	II
μ	.959	1.866
σ^2	1.556	1.328
CV	1.301	.618

The 1000 observations were considered as the complete populations for comparison purposes. Distribution I has a coefficient of variation greater than one (1.301) and distribution II has a coefficient of variation less than one (.618) .

Tables 2 and 3 present the mean-squared errors for different cutoff points and sample sizes for the two distributions. Results for samples of sizes 10 and 20 were inferred from the results for samples of size 5 by adding together the bias and the variance with the appropriate division.

Figures 1 and 2 demonstrate graphically the gains achieved with the alternatives s_1^2 and s_t^2 for samples of size 20. Similar patterns are shown in the tables for $\bar{y}_1$ and $\bar{y}_t$.

It is evident that s^2_t and $\bar{y}_t$ achieve gains over a wider region than do s_1^2 and $\bar{y}_1$. Also, lower mean-squared errors are achieved by s_t^2 and $\bar{y}_t$. However s_1^2 and $\bar{y}_1$ are superior in the region which would normally be of the most interest and they are simpler to work with since they are the estimators formed by ignoring the large observations and using only those remaining. For these reasons it would appear to be preferable to use s_1^2 and $\bar{y}_1$ when sampling from distributions of the types used in this study. Of course if no information is available beforehand as to what might constitute a "large" observation the choosing of the cutoff point becomes very difficult. The results indicate that gains are achieved for a rather large region, however if a mistake is made and the point is chosen too small, rather disastrous consequences are obtained. In practice the sampler frequently has some very general ideas concerning minimum and maximum possible values and these plus information concerning the general shape of the distribution will generally provide sufficient information to intelligently pick a t value.

Figure 3 presents the distribution of s_t^2 (t=4) and s^2 for the 200 samples of size 5 from distribution II.

REFERENCES

[1] Searls, Donald T., "Some Alternative Estimators for a Population Mean," 1964 Social Statistics Proceedings of the American Statistical Association.

[2] Searls, Donald T., " The Utilization of a Known Coefficient of Variation in the Estimation Procedure, " Journal of the American Statistical Association, 59(1964), 1225-1226.

Table 2. Mean-squared errors for alternative estimators of μ and σ^2 for distribution I.

n	Cutoff point t	$\bar{y}$	$\bar{y}_1$	$\bar{y}_t$	s^2	$s^2{}_1$	$s^2{}_t$
		Mean squared errors					
5	9	.319	.319	.319	4.363	4.363	4.363
	7	.319	.273	.309	4.363	2.519	3.524
	6	.319	.264	.296	4.363	1.880	2.799
	5	.319	.233	.274	4.363	1.348	1.879
	4	.319	.216	.244	4.363	1.141	1.174
	3	.319	.209	.200	4.363	1.265	.864
	2	.319	.244	.156	4.363	1.633	1.178
	1	.319	.465	.222	4.363	2.201	1.930
	0	.319	.920	.920	4.363	2.421	2.421
10	9	.160	.160	.160	2.181	2.181	2.181
	7	.160	.137	.155	2.181	1.287	1.764
	6	.160	.134	.148	2.181	1.003	1.410
	5	.160	.121	.138	2.181	.819	.971
	4	.160	.122	.123	2.181	.879	.683
	3	.160	.143	.105	2.181	1.182	.691
	2	.160	.207	.100	2.181	1.609	1.142
	1	.160	.450	.204	2.181	2.198	1.927
	0	.160	.920	.920	2.181	2.421	2.421
20	9	.080	.080	.080	1.091	1.091	1.091
	7	.080	.069	.077	1.091	.670	.883
	6	.080	.069	.074	1.091	.565	.711
	5	.080	.066	.069	1.091	.554	.517
	4	.080	.076	.063	1.091	.748	.438
	3	.080	.110	.058	1.091	1.141	.605
	2	.080	.189	.072	1.091	1.597	1.124
	1	.080	.442	.196	1.091	2.197	1.925
	0	.080	.920	.920	1.091	2.421	2.421

Table 3. Mean-squared errors for alternative estimators of μ and σ^2 for distribution II.

n	Cutoff point t	Mean squared errors $\bar{y}$	$\bar{y}_1$	$\bar{y}_t$	s^2	s^2_1	s^2_t
5	11	.240	.240	.240	3.941	3.941	3.941
	9	.240	.221	.236	3.941	2.319	3.332
	8	.240	.212	.231	3.941	1.718	2.824
	7	.240	.205	.220	3.941	1.278	2.037
	6	.240	.197	.209	3.941	.954	1.449
	5	.240	.173	.194	3.941	.803	.954
	4	.240	.164	.163	3.941	.839	.649
	3	.240	.201	.129	3.941	1.106	.755
	0	.240	3.483	3.483	3.941	1.764	1.764
10	11	.120	.120	.120	1.971	1.971	1.971
	9	.120	.110	.118	1.971	1.168	1.666
	8	.120	.106	.115	1.971	.882	1.414
	7	.120	.104	.110	1.971	.681	1.026
	6	.120	.101	.105	1.971	.557	.743
	5	.120	.091	.097	1.971	.526	.523
	4	.120	.103	.084	1.971	.721	.451
	3	.120	.164	.077	1.971	1.070	.706
	0	.120	3.483	3.483	1.971	1.764	1.764
20	11	.060	.060	.060	.985	.985	.985
	9	.060	.055	.059	.985	.593	.834
	8	.060	.054	.058	.985	.464	.708
	7	.060	.053	.055	.985	.383	.520
	6	.060	.052	.053	.985	.359	.390
	5	.060	.050	.049	.985	.386	.307
	4	.060	.072	.044	.985	.662	.352
	3	.060	.141	.051	.985	1.051	.681
	0	.060	3.483	3.483	.985	1.764	1.764

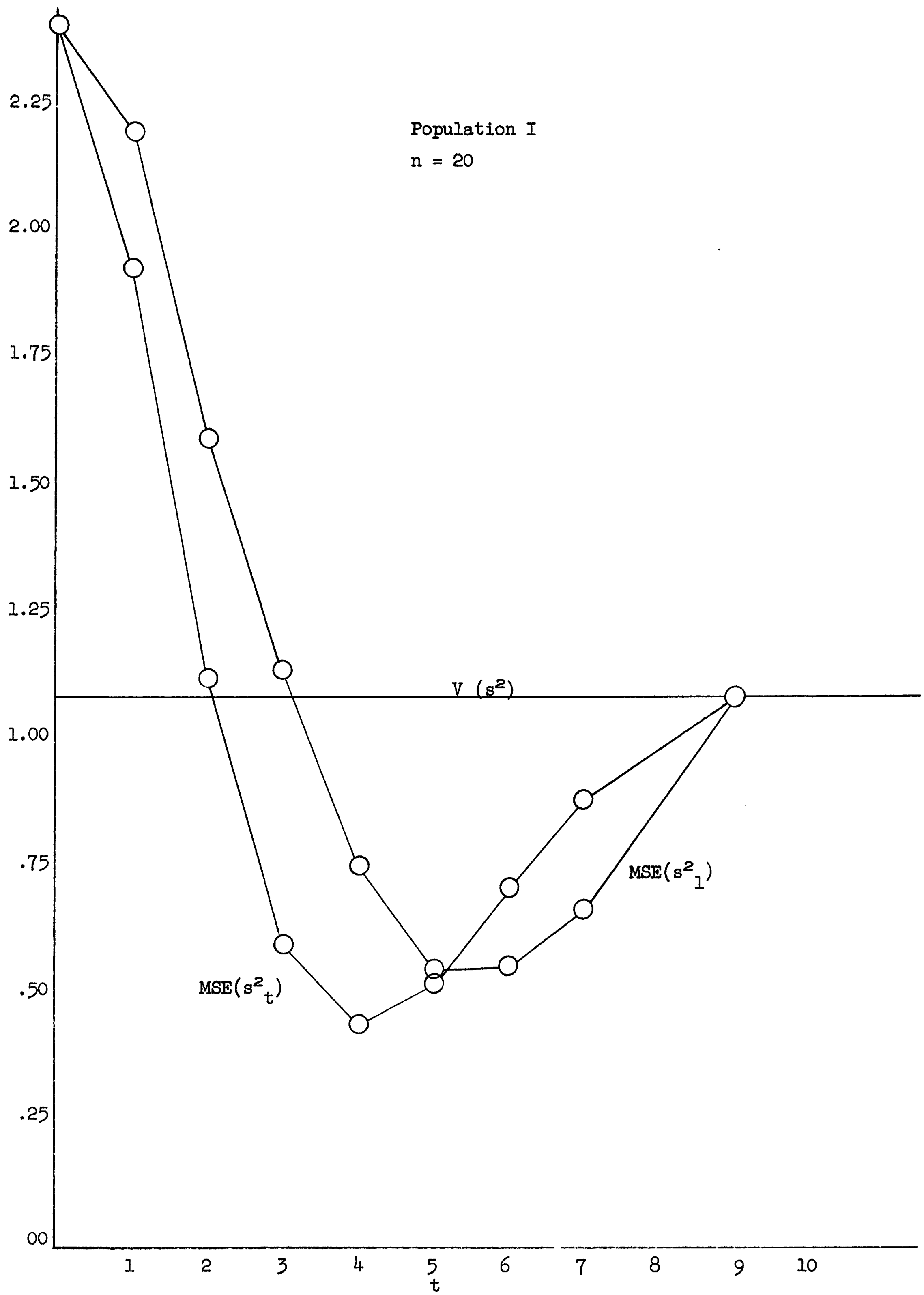


Figure 1. Mean-squared errors versus cutoff point - population I.

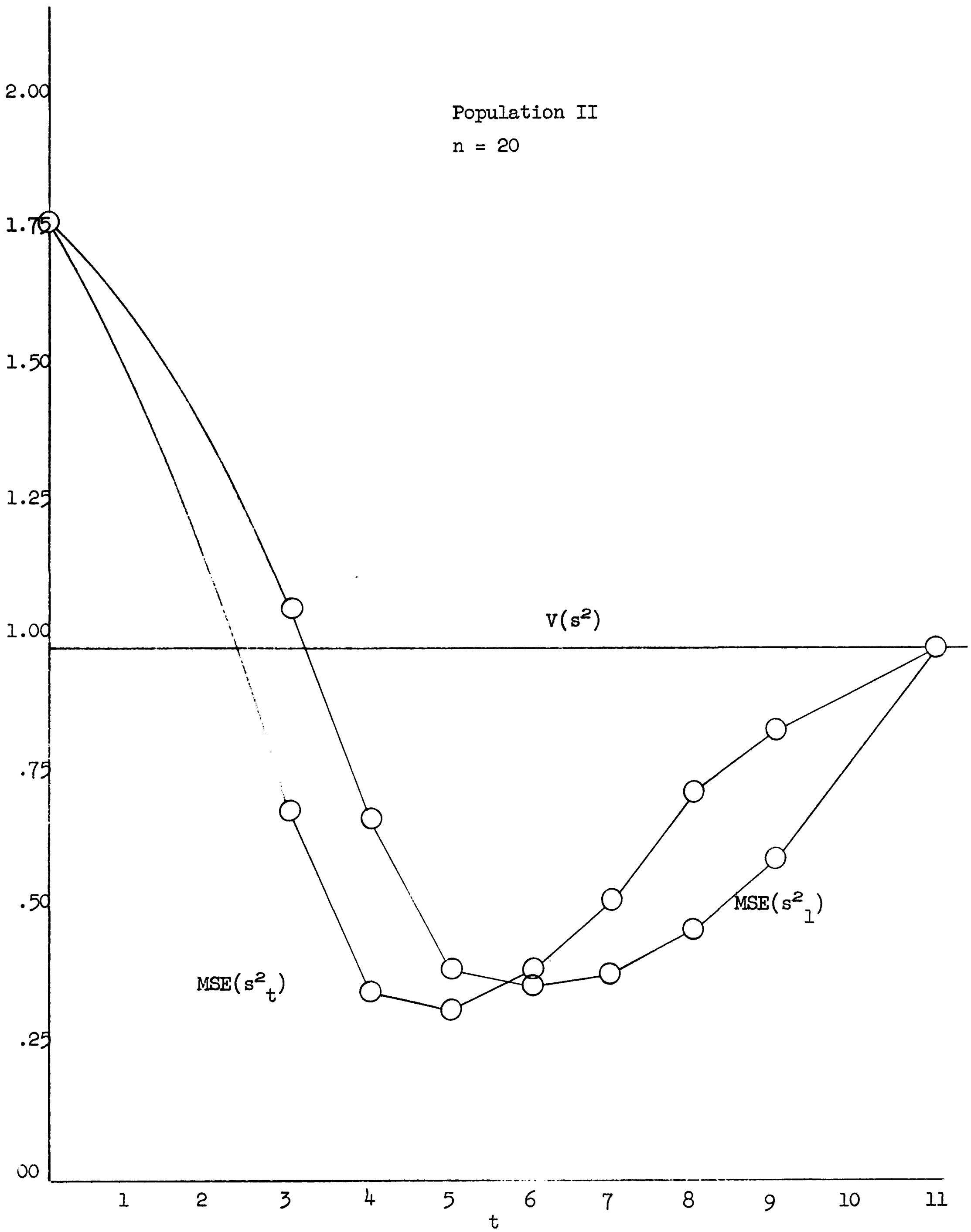


Figure 2. Mean-squared errors versus cutoff point - population II.

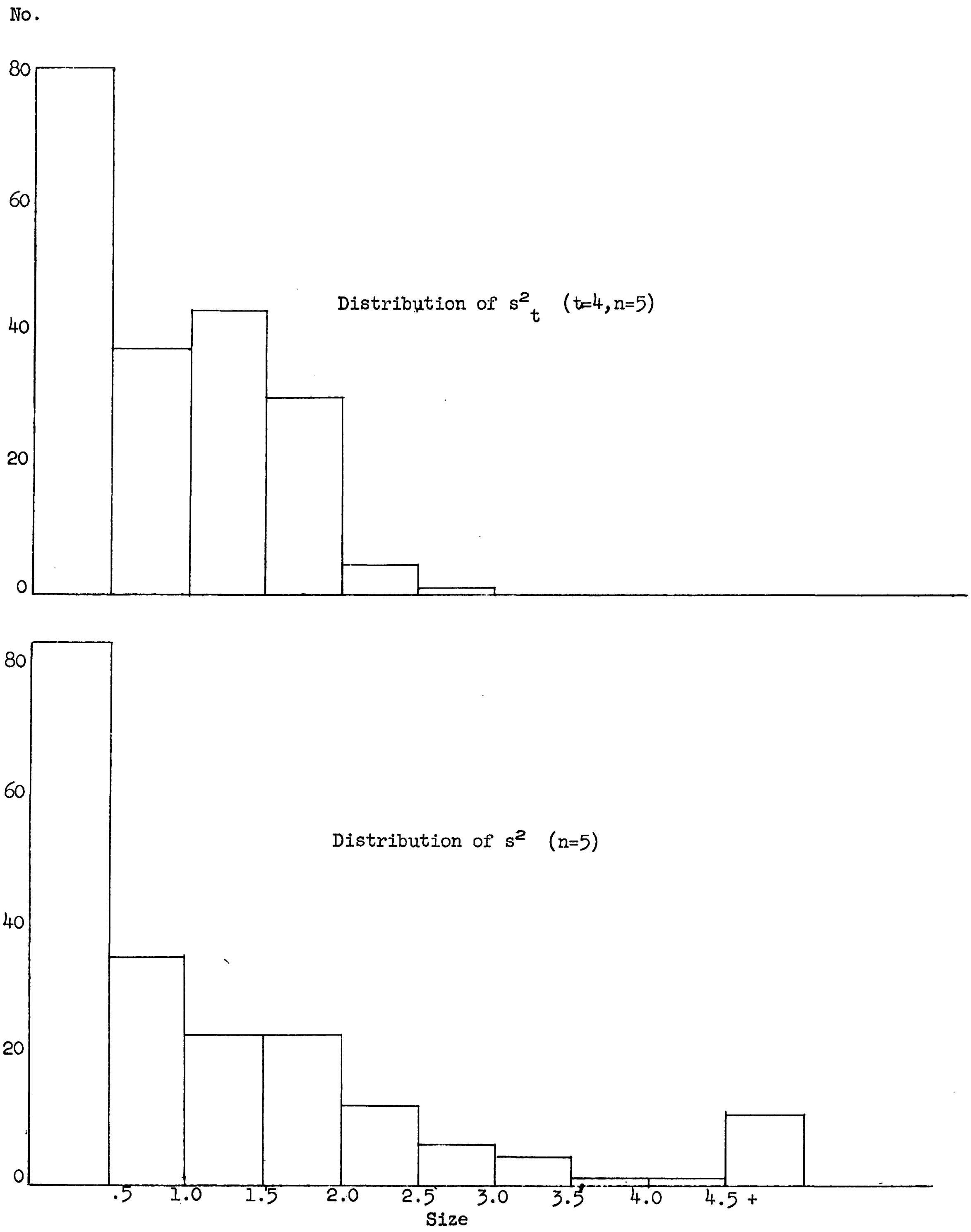


Figure 3. Comparison of distributions of s_t^2 and s^2. (Distribution II)

LEMUEL SHATTUCK: AMERICAN PIONEER IN POPULATION STATISTICS

Frank H. Godley, National Center for Health Statistics

Data collection procedures, formats, and content will be revised in 1968 by the two major arms providing population statistics for the United States. The U.S. Bureau of the Census will hold a dress rehearsal of the 1970 Decennial Census of Population in April 1968, while the National Division of Vital Statistics will have instituted its new Standard Certificates of Birth, Death, Marriage, and Divorce by January 1 of that year.

This paper is intended to gain an historical perspective on the collection of population statistics in the United States. Specifically, investigation is of the contributions of Lemuel Shattuck, a co-founder of the American Statistical Association in 1939.

In his publicizing Circular of April 1840, Shattuck listed nineteen subjects in which the newly formed American Statistical Association was interested. Number two on this list was "Population.--The Census of different periods; the Births, Marriages, and Deaths, specifying the diseases, sex, age, and months of the year, when each death took place; Boards of Health, prevalence of Epidemics, and other diseases" (1). Shattuck himself was to become the most significant contributor, as we shall see, to population statistics in America.

As one of the chief planners of the 1850 Census of the United States, Lemuel Shattuck introduced not only new questions, but also new analytical concepts. The first of these, and the most revolutionary, was that of making the individual, rather than the family, the unit of analysis. This innovation he first applied to his 1845 Census of Boston.

The consequences of shifting from the analytically cumbersome family unit to the individual unit were profound. One consequence is, of course, that much more information (age, sex, education, etc.) can be collected for the individual than for the household. The census schedule needs only to provide a single line for each member of the family, so that records for families or households may be kept together (2).

The purposes of such a census of population fall in three classes, according to Shattuck: political, public health, and social-scientific (3). In addition to the political purpose of allocating legislative Representatives and raising revenue (as stated in the United States Constitution), census enumerations, Shattuck recognized, are indispensable in assessing the morbidity and mortality of a community. This is the public health purpose, which introduced the concept of relating the number of cases of disease or death to the size of the population at risk. A third purpose of a census was social-scientific. Census data on the age composition, for example, can be used to forecast the population of school age and thereby to plan school building.

In order to achieve such social-scientific and public health purposes, Shattuck listed ten classes of facts that should be included in every census. They are age, sex, race, nativity, residence in cities, occupation, marital status, education, economic status, and housing facilities. Age, believed Shattuck, was the most crucial in public health studies. He stressed the necessity for exactness in the reporting of age. To him, the principal value of a population by age was its comparability with annual decedents by age. That is, age-specific mortality was to be a powerful tool for comparing the sanitary conditions of different communities or of the same community with its past. While most of the facts recommended by Shattuck were intended for public health purposes, they have gradually acquired utility to the social scientist.

One of Shattuck's recommendations was to tabulate the occupations of males aged 15 years or older. Occupation in the 19th Century revealed significant differentials in mortality, but these have largely disappeared in modern times. More important is the current use of occupation as an index or component of socio-economic status.

Education, in those embryonic days of public health, was measured by the crude standard of literacy. The proportion literate, thought Shattuck, should influence the sanitation of a community. His proposed question on the census was "Can you read or write?" In modern times, literacy is often taken as the minimum level of educational attainment and is used in cross-cultural comparisons. Within cultures or countries, however, the number of years of school completed has become the preferred index of educational attainment. In conjunction with occupation or by itself, years of school is used as an indicator of social class and in studies of social mobility.

Suspecting that property owners enjoyed better health conditions than tenants, Shattuck recommended a census question on the ownership of real estate. This indicator of economic status was phrased "means of subsistence and comfort."

Housing facilities were the last of Shattuck's ten facts essential to a census. The recommended questions were the number of persons per family and per house. It is probable that the overcrowding from rapid urbanization and centralization of the population in the Nineteenth Century impressed upon Shattuck the sordid health conditions arising therefrom.

A principle can be drawn from these early census recommendations. It is that census questions must have a basis either in some current, unanswered questions of a practical nature, or in some body of knowledge of a theoretical nature. In the Nineteenth Century, the most serious and pressing questions were posed by the high and wide-spread levels of acute contagious disease. Medicine provided the theoretical context for posing these questions.

Today, in the mid-Twentieth Century, the most pressing problems are perhaps somewhat different: they raise questions of the functioning of the society and man, the social animal, rather than man, the biological animal. Behavioral sciences, more often, provide the theoretical framework for answering these questions. Just as the census officials have responsibility for recognizing this, we as social scientists have an obligation of posing the practical and theoretical problems of the day in terms of data needs. This is what Lemuel Shattuck did one century ago.

The significance of Lemuel Shattuck's census proposals were also largely to aid the vital statistics branch of population statistics. By first removing vital statistics from an unreliable collection mechanism, the census, and then by substituting an effective registration system, Shattuck provided for a continuous file of potentially accurate birth, death, and marriage records.

Secondly, recognizing the need to relate vital statistics to the exposed population, those at risk of the event, he revolutionized the census by substituting the individual for the family as the unit of analysis. This created a base of detailed, cross-classified populations for comparing the incidence of vital events as well as that of disease in different communities or time periods. As a result, detailed studies of demographic processes were possible. Vital events could be properly related to the individuals exposed to specified risks of birth and death. It became possible to calculate person-years of exposure, as Shattuck demonstrated in his reports on the 1845 Census of Boston.

The greatest problem of early vital registration statistics was their uneven coverage of different areas of the United States. The problem of regional variation in the quality and quantity of vital statistics is a common one faced by all growing registration systems. A few highly developed registration centers, often in the large cities, report vital events relatively accurately and completely, while in the less developed areas, bias and nonreporting of vital events is high.

These considerations force a choice between two alternative registration systems: in one, geographic coverage is complete, but error and nonreporting are frequent; the alternative provides high-quality vital statistics, but entails partial geographic coverage.

In the Nineteenth Century United States, the latter alternative was embodied by the registration systems of a few major cities such as New York, Boston, and Baltimore. The alternative to this, however, was carried out by the Census of Population, using a survey method to count vital events and measure vital rates for the Nation as a whole. Shattuck disapproved of the Census method, arguing that it was highly unreliable and yielded virtually useless vital data.

It was unreliable to tally births and deaths from a census, he argued, because the time lag between the occurrence of the vital event and its recording was simply too long. In the interim, people forget. By census time they tend to misstate not only the date of the occurrence of the birth or death, but also the characteristics of the persons concerned. Moreover, the family units and their associated households may be lost by migration, marriage, divorce, and death.

These disadvantages also apply to the sample survey methods used by some of the newer nations to estimate vital events. Again there is a time lag between the date of a vital event and that of the interview. Moreover, a survey at one point or period in time may not necessarily represent the calendar year. There is a seasonal variation of births, deaths, and marriages. Unless overlapping or continuous surveys are undertaken, the respondents are forced to tax their often unreliable memories about the dates and particulars of events occurring in other seasons of the same year and sometimes in previous years. Some additional reasons have been advanced in modern times. A. Mitra, the Registrar-General of India, contends that the sample survey method can never replace universal vital registration because sample data for local areas contain too large a sampling error, especially when cross-classified, and that these data are of no use for legal identification purposes (4).

But notwithstanding the vehement protestations of Lemuel Shattuck, the United States Census served as the principal means of collecting National statistics on births, deaths, and marriages from 1850 (when, ironically, Shattuck himself designed the questions and tables) until 1910. Consequently, reliable vital statistics were not collected in the United States until the Twentieth Century, or nearly 50 years after Shattuck's proposals.

Despite all of Shattuck's highly desirable and advanced plans, considerable organized opposition was voiced. In his <u>Report to the Sanitary Commission of Massachusetts, 1850</u>, Shattuck listed some of the more persistent arguments against asking detailed questions of every person in the census:

1. The invasion of personal and family privacy would be intolerable.

2. It constitutes an invasion of "Divine

Providence," who alone should decide when and of what humans should die; public health shouldn't intervene.

3. Local communities are potentially threatened by government possession of such ancillary information as is proposed for the census.

4. The census would be "too statistical; you can prove anything by figures."

Alongside each of these objections and the others, Shattuck published rebuttals in his Report (5). Since most of his proposals were incorporated into the 1850 Census of the United States - excepting that of no birth and death questions - and in view of a complete nationwide system of vital statistics today, the efforts of this pioneer in American demography appear to have been successful, the obstacles overcome.

REFERENCES

(1) American Statistician, "125th Anniversary Issue," Volume 19, Number 2 (April 1965), p. 31.

(2) National Office of Vital Statistics, Vital Statistics of the United States, 1950, Volume I (Washington, 1954), pp. 5-6.

(3) Lemuel Shattuck, et. al., Report of the Sanitary Commission of Massachusetts, 1850, (reprinted, Cambridge, 1948), passim.

(4) India Ministry of Home Affairs, The Registrar-General, Vital Statistics of India for 1961 (New Delhi, 1964), p. 1.

(5) Lemuel Shattuck, et. al., op. cit., pp. 283-292.

POVERTY IN THE SOUTHWEST: A COMPARATIVE STUDY OF MEXICAN-AMERICAN, NONWHITE AND ANGLO FAMILIES*

Frank G. Mittelbach and Grace F. Marshall,
University of California, Los Angeles

Much of the recent research on poverty in the United States focuses on the differentials between whites and nonwhites in our population. For many purposes this distinction will suffice, since it highlights the association between deprivation and skin-color. But, in the five southwestern states of Arizona, California, Colorado, New Mexico and Texas, the white-nonwhite dichotomy may hide as much as it reveals. Within this vast area, comprising 21 percent of the land area of the United States and 16 percent of its population, is concentrated the overwhelming bulk of the people of Mexican descent in this country.

"Mexican-Americans," a generic term that we shall use to include people of Mexican, Spanish, or mixed Indian descent, are classified as white. Members of this minority group are usually physically or otherwise identifiable, however, and their experiences with the rest of American society partially parallel those of other immigrant groups of relatively recent vintage, on the one hand, and those of nonwhites, and particularly Negroes, on the other. By every yardstick available, poverty among Mexican-Americans is a serious problem both from the standpoint of the ethnic community and for society at large.

The Position of Mexican-Americans in the United States

Mexican-Americans comprise the second-largest disadvantaged minority group in the United States, the largest, of course, being Negroes. Of an estimated 3.8 million Mexican-Americans in the United States in 1960, approximately 87 percent resided in the Southwest. Only 45 percent of the total were of foreign stock, (that is, either born in Mexico or of Mexican parentage); the remainder were the descendants of still earlier immigrants. The social and economic conditions of Mexican-Americans are reflected in the tabulations of the U.S. Bureau of the Census relating to white persons of Spanish surname.

Contrary to widespread impression, Mexican-Americans are highly urbanized. In the Southwest, 79 percent of the Spanish-surname persons lived in urban areas in 1960, compared to 80 percent of nonwhites and 81 percent of the Anglos (white minus white persons of Spanish surname). Clearly, the image of Mexican-Americans as being primarily a rural population is out of date. They vary little in this characteristic from the population at large.[1] Most Mexican-Americans are engaged in non-agricultural pursuits and reside in other than rural areas.

Within the cities, Mexican-Americans are highly segregated residentially. As Moore and Mittelbach have shown, this residential segregation is widespread but varies greatly in degree from city to city.[2] Moreover, residential segregation of Mexican-Americans from the dominant population does not assume the dimensions of that between Negroes and Anglos. The pervasive geographical distance of Mexican-Americans from Anglos, (and parenthetically also from Negroes) is merely one indication of general social distance.

Other indicators show a similar picture. For example, one presumably important distinction is that between first, second, and third-or-more generation immigrants. It might seem reasonable to assume that socio-economic position tends to improve the further people are removed from immigrant status. No long series are available to subject this proposition to careful scrutiny. Median income data for Spanish-surname persons by nativity and parentage, however, indicate that there is no such straightforward relationship. While natives of native parentage fare better than do the foreign born, natives of foreign or mixed parentage enjoy still higher incomes than do either of the others. This holds for males in all five southwestern states, and for the rural as well as the urban population.** (Table 1)

In another study, Fogel highlights the above income differentials more sharply by holding age, as well as sex and area, constant.[3] In 1959 the income of Spanish-surname urban males was higher for the children of immigrants than for those who were themselves foreign born; however, income of third-or-more generation Mexican-American males was lower than that of the second-generation immigrant, although it exceeded that of the foreign born. These differences were not attributable to lower educational attainment among natives of native parentage than for the children of immigrants. In fact, the opposite was found to be true.

---

*This paper is derived largely from our monograph, The Burden of Poverty, (Mexican-American Study Project, Graduate School of Business Administration, University of California, Los Angeles, 1966), which is part of a comprehensive study of the Mexican-American population in the United States. We gratefully acknowledge the comments and assistance of Leo Grebler, Joan Moore and Ralph Guzman. The authors, however, take full responsibility for all data and findings.

**The same conclusion appears to apply to the income of females as well; however, so many of the female incomes were in the undifferentiated class of less than $1,000 that exact comparisons are impossible.

Table 1

Median Income by Nativity and Parentage for
White Males of Spanish Surname in Five Southwest States, 1960
Urban and Rural

Nativity Class	Total	Urban	Rural Nonfarm	Rural Farm
All Classes	$2,804	$3,197	$1,871	$1,531
Native of Native Parentage	2,689	3,071	1,890	1,495
Native of Foreign or Mixed Parentage	3,345	3,650	2,152	1,892
Native of Mexican or Mixed Parentage	3,114	3,426	1,971	1,648
Foreign Born (Total)	2,307	2,742	1,610	1,423
Foreign Born (in Mexico)	2,158	2,602	1,564	1,374

Source: 1960 U.S. Census of Population, PC(2) 1B, Table 6.

Table 2

Number and Percent of Poor Families in
Various Population Groups in the Southwest, 1960

Population Group	All Families (1)	Poor Families[a] (2)	Percent of Poor in Each Group (3)	Poor in Each Group as Percent Of All Poor (4)
Total	7,356,866	1,451,655	19.7%	100.0%
White	6,766,367	1,205,729	17.8	83.1
Anglo	6,068,340	962,826	15.9	66.4
Spanish Surname	698,027	242,903	34.8	16.7
Nonwhite	590,299	245,926	41.7	17.0

[a]Families with annual income under $3,000 in 1959.

Source: 1960 U.S. Census of Population, PC(2) 1B Table 5; Vol. I, State Volumes, Table 65.

It is beyond the scope of the present paper to describe or analyze comprehensively the socioeconomic status of the population of Mexican descent in the United States. A great many factors should enter into such a discussion and we can only sketch some of the outlines. Unquestionably, more recent immigrants, as well as the descendants of earlier settlers in areas of the Southwest which at one time were part of Mexico, frequently have found themselves in a hostile social, economic and political environment. A great deal of discrimination was experienced in the labor market and elsewhere, although this may have diminished somewhat in recent years. The relative recency of large-scale immigration from Mexico into the United States, which reached its peak in the 1920's after immigration from other parts of the world had begun to decline, also may partially explain present conditions of this population.[4] The proximity to Mexico, the continued stream of immigration from that country, the uncertainty of the intentions of immigrants from Mexico to remain here, plus other factors have contributed to the maintenance of cultural boundaries. In turn, these boundaries were further strengthened by the attitudes of the host society. The chain of causality is, indeed, complex.

With this brief introduction to the subpopulations with which we shall be concerned, we turn to our examination of poverty. Using data from the 1960 Censuses of Population and Housing, we shall analyze poverty among Mexican-Americans and present comparisons with other major population groups.

The Incidence of Poverty Among Families

For present purposes, we have adopted the widely-used statistical "poverty line" of $3,000 family income per year. The concept and definitions of poverty are the subject of many recent writings. Our choice of definition was largely determined by availability of data. An alternative measure of poverty was tried but it failed to produce any major changes in the number and percent of poor families in the Southwest.[5]

The highly-condensed data in Table 2 tell a great deal. Nearly 243,000 Spanish-surname families in the Southwest, or about 35 percent of all such families, were in the poverty group in 1960. The number of poor nonwhite families was slightly larger (almost 246,000); however, fewer nonwhite families than Spanish-surname families resided in the Southwest. Consequently, the incidence of poverty among nonwhites was still greater than among families of Spanish surname (almost 42 percent for nonwhites.) The incidence was far smaller for "Anglos," the group representing the dominant society. The relative frequency of poverty among Spanish-surname families was more than twice the Anglo rate. In the case of nonwhites the frequency of poverty was two and one-half times that of the Anglos.

Numerically, poor Anglo families by far exceeded the poor of both minority groups combined. As the last column of Table 2 shows, two thirds of all poor families in the Southwest were Anglo, 16.7 percent were in the Spanish-surname group, and 17.0 percent were nonwhite. In other words, for every poor family in both minorities combined there were nearly two Anglo families, a reminder that below-minimum income affects many people not in ethnic or racial minorities.

The initial data show also the distortion that results from the frequent failure to distinguish between Anglo and Spanish-surname families within the "white" class. Although the Spanish-surname families represent only 10.3 percent of all white families in the Southwest, irrespective of income, they account for 20.1 percent of the poor white families. Consequently, analyses of poverty in the Southwest that do not consider Spanish-surname persons separately from Anglos hide the existence of a minority which, by the income criterion used here, is almost as disadvantaged as the nonwhite minority. In terms of social action, this failure to differentiate is bound to ignore problems that are more specific to the Spanish-surname population and less prevalent among Anglos. Further, this failure to separate the Spanish-surname group from the rest of the white population in effect reduces the income gap between Anglos and nonwhites in the Southwest.

Selected Characteristics of the Poor

This section analyzes some of the special characteristics of the poor in the subpopulations. National studies have shown that it is the aged, broken families, farmers and farm workers, the minorities, and the unemployed or underemployed who are most likely to be poor.[6] These are frequently overlapping categories, and the risk of being poor rises if a family falls within more than one of these groups. When the higher incidence of poverty among the minorities is recognized and minority status held constant, the question remains, "Who are the poor in the Spanish-surname population?" It will be shown that selected types of Spanish-surname families are particularly afflicted by low incomes, as is true also for Anglos and the nonwhites. Here again, the answers are derived from the 1960 Census, and particularly the One-In-One-Thousand sample.[7] The situation may differ in small detail at the present time; however, the structural relationships with which we are concerned here do not change significantly over a number of years.

Table 3 (last column) shows that the incidence of poverty among Spanish-surname, nonwhite and, in fact, all families is especially high when the head of the family is employed in farming, is 65 years old and over, or is a female, this last item being an indication of a broken family. About one-third of the Spanish-surname families generally are poor, but when one considers the subgroupings it becomes evident that a Mexican-American family is more than twice as likely to be poor if the head is in farming or is a female (69 and 68 percent are poor, respectively), and almost as unfortunate if the head is

Table 3

Selected Characteristics of Poor and All Families for Three Populations in the Southwest, 1960
(Families may be counted more than once)

Characteristics	Total	Percent of Total	Poor	Percent of Poor	Percent Poor of Total
	(1)	(2)	(3)	(4)	(5)
Spanish-Surname Families					
Number of Families	702,000	100.0	241,000	100.0	34.3
Characteristics of Head					
Occupation Farmer or Farm Worker	81,000	11.5	56,000	23.2	69.1
65 Years old & over	63,000	9.0	38,000	15.8	60.3
Female Head	90,000	12.8	61,000	25.3	67.8
Head 0-4 Years of Education Completed	223,000	31.8	127,000	52.7	57.0
All Families					
Number of Families	7,356,885	100.0	1,451,655	100.0	19.7
Characteristics of Head					
Employed Farmers and Farm Workers	383,023	5.2	167,364	11.5	43.7
65 Years old & over	865,651	11.7	415,431	28.6	48.2
Female Head	660,013	9.0	327,724	22.6	49.7
Nonwhite Families					
Number of Families	590,514	100.0	245,926	100.0	41.7
Characteristics of Head					
Employed Farmers and Farm Workers	37,676	6.4	22,245	9.0	59.0
65 Years old & over	58,717	9.9	41,287	16.8	70.3
Female Head	106,871	18.1	77,864	31.7	72.9

Source: 1960 U.S. Census of Population, Vol. I, State Volumes, Tables 65, 110, 139 and 145; Spanish Surname, which is estimated, is from the One-in-One-Thousand Sample (See Reference 7).

over 65 (60 percent). Similar patterns are found for nonwhites. Over-all, 42 percent of the nonwhite families are poor, but over 70 percent are poor if the family head is a female or is elderly. The disadvantage of employment in farming is not quite so high for nonwhites (59 percent poor) as for Mexican-Americans. Generally, then, among Spanish-surname and nonwhite families with the three characteristics examined, poverty is so common that it is the rule, not the exception.

Farm employment, old age, and broken families account also for a large proportion of poverty among all families, irrespective of ethnicity. Nearly one-fifth of all families in the Southwest are poor, but, among families whose head is a farmer or farm worker, 44 percent are in this category. For families headed by an aged person, the incidence of poverty is 48 percent, and for those with a female head, 50 percent. The relatively high proportions of poverty-stricken families of these types in the general population reflect the statistical weight of minority families, among whom the three characteristics are unusually numerous.

A note of caution is called for at this point. The immediate causes of poverty among farm, elderly and broken families are unquestionably related to these circumstances. But, further reflection suggests a more complex chain of causality. For example, the fact that families with elderly persons as heads are prone to be poor results from the low earning capacity of the breadwinner. Many factors may be associated with this inability to earn sufficient income, including physical impairments, discrimination against older persons in the labor market, and technological obsolescence of skills. Moreover, in many such families the head is over 65 and is a woman whose responsibilities in the home, or lack of marketable skills, prevent her from earning a sufficient amount to take the family out of the poverty group. On a longer perspective, it will often be found that the present poverty status in families with an elderly head results from experiences or personal difficulties in the past. Lack of opportunity or ability to prepare for voluntary or involuntary retirement or, for that matter, the breadwinner's death, will have contributed to the present plight of the family. Among minority families, the breadwinners' experiences in the labor market during their younger days will have prevented the accumulation of savings for the day when they are no longer able to work or compete with younger job applicants. Moreover, industries and firms employing large numbers of minority persons often have no retirement and pension programs. Consequently, at old age they are heavily dependent upon social security and welfare programs to meet their needs, and even these funds may be lacking or given sparingly. In any case, social security and welfare support cannot (and are not designed to) provide enough income to take a family out of the poverty group.

Another illustration is the prevalence of poverty among broken families. Divorce, desertion, or early death of the male head will in many cases place the remaining family in the poverty group even if the original unit was better off. In the instance of divorce, two households will be formed. As a unit, the family may not have been poor, but the separate parts of the unit may both be classified as poor. This is not merely a matter of definition. Certain expenses, such as rent, will now be duplicated. In this case it may not seem unreasonable to attribute poverty status to the breaking up of the family. However, the breaking up of the family may be as much a result of poverty as the cause. The stresses within the family in the face of low income often lead to divorce or desertion. If so, the causality runs in the other direction -- from poverty to the breaking up of the family.

The incidence of broken families among some minority groups is unusually high. To the extent that female family headship is an indication of this phenomenon, it will be noted in Table 3 that 13 percent of the Spanish-surname families in the Southwest are of this type irrespective of income size. The comparable figures are 18 percent for nonwhites and 9 percent for all families (including the two minority groups). The problem is less severe in the Spanish-surname population than among nonwhites, and especially Negroes; but it can be inferred that the incidence of broken families among Mexican-Americans is far greater than among Anglos. It is also far greater than the commonly-accepted notion of the traditional strength of the family in this segment of the population would suggest. This subject will be pursued in other parts of the Mexican-American Study Project. In the present context, the relatively high incidence of broken families in the Spanish-surname group warrants emphasis because of its bearing on the incidence of poverty.

It would be erroneous, however, to conclude that the relatively large number of broken families in selected minority groups results from inherent instabilities. Considerable evidence supports the thesis that the experience of disadvantaged minority groups in American society tends to undermine the self-confidence, and indeed the identity, of minority-group men. In turn, the stability of the family is threatened. Though this generalization has been derived primarily from studies of American Negroes, it is quite likely that it also applies to Mexican-Americans. A history of discrimination, comparatively high unemployment, low income, and alienation from both the dominant community (and sometimes their own group) has characterized the Mexican-American minority as well as the Negro, and all of these and other disabilities are likely to weaken the fibers which hold families together. This is a problem, then, of both the larger society and the minority.[8]

Similar caveats on causality apply to other family or personal characteristics associated with a high incidence of poverty. For example, poverty is often correlated with low educational attainment, but the long-range factors which have contributed to this condition are often deeply imbedded in history and the whole structure of our society.

The data on the Spanish-surname population in Table 3 make it quite apparent that many of the families who are poverty-stricken have more than one of the characteristics discussed so far, and others as well. The percentages in the Spanish-surname portion of the table add to more than 100. This reflects the substantive fact that poor farmers also often have a low level of educational attainment, etc.

An attempt has been made to pin down these multiple characteristics of the poor more closely. Special tabulations were prepared for the Spanish-surname poor, in which each of the families was counted only once. The data are presented in Table 4. The procedures for preparing the estimates were as follows: First, the record was searched to determine if the head of the family had the occupation of farmer or farm worker. If so, the family was recorded in this group no matter what the other characteristics were. If not a farmer or farm worker, the records were searched further to establish if the head was 65 years old or over. Again, if the answer was positive the family was classified under "head 65 years and over" and nowhere else. The search continued in this manner until all families had been classified under one of the six specified categories in the order indicated in Table 4. If they did not fit any of these classifications, the families were placed in the "other" group. This procedure was applied to both poor and non-poor families as well as to persons not in families. While the ordering is subjective, it is not without rationale.[9]

In the aggregate, the overwhelming majority of the Spanish-surname poor families, a little under 83 percent, had a head with one or more of the six specified characteristics which have been identified as strongly associated with poverty status. In comparison, only 39 percent of the non-poor families fell under one or more of these six headings.

Moreover, the new data make it doubly clear that poverty among Spanish-surname families is associated with several characteristics rather than any one. Perhaps the best illustration is the low education of the heads of poor families. As shown in Table 3, 127,000 families who are poor have a head whose educational attainment is four years of schooling or less, more than those in any other category. Generally, persons with such a low level of education are considered to be functionally illiterate, a term applied by the United States Army in World War II to "persons incapable of understanding the kinds of written instructions needed to carry out basic military tasks." One would expect and, in fact, does find a relatively high incidence of poverty among families of this type. They represent about 53 percent of all Spanish-surname families who are poor, and 57 percent of the families headed by persons with such low educational preparation are in the poverty group. Table 4 provides a new perspective. After classifying poor families first by reference to the other characteristics, only 14 percent of the poor have heads who are functionally illiterate. In other words, low education among the poor is associated with farm occupation, family headship of 65 years and over, female family headship, and so on.

The foregoing analysis has served to demonstrate more specifically the multiple characteristics of minority families and individuals that are related to their poverty, in addition to minority status. In this fashion, it has extended the search for the causes of economic deprivation. But it stops short of establishing the order of links in the chain of events and circumstances that produce poverty. Where does it all begin, and what are the processes through which an initial shortcoming becomes transformed into more or less permanently inadequate income?

To take the case of poverty associated with low educational attainment, what are the reasons for poor schooling -- inability of schools to adapt their system (and their teachers) to the needs of Mexican-American children, or a low value placed on schooling at home, or the pressing need to get a job at an early age because of the parents' poverty, or the youngsters' expectations of a low payoff from education as they observe discrimination in labor markets? And are the formal schooling requirements of employers really necessary for the performance of certain jobs which, if Mexican-Americans could obtain them, would be sufficient to take members of this minority group out of the poverty class? Or, have these requirements sometimes been imposed as a convenient device to screen out applicants of certain ethnic minorities?

Similar questions could be raised about the other characteristics. Explanations will vary from group to group and even from person to person within each group, and this is not the point of the questions. The real point is rather whether major institutions in our society operate in such fashion that they produce, aggravate, and perpetuate poverty among minorities.

In conclusion, better understanding of the multiple determinants of poverty is important mainly because it will help suggest approaches to effective public action. One of these is to alleviate the problem through traditional welfare and some of the more recently-developed anti-poverty programs -- essentially a redistribution of income. Another is illustrated by public aids for better schooling and manpower training - or retraining - programs to help persons and deprived groups overcome specific handicaps and thus increase their earning capacity and their productive contribution to the economy.

Also, there remains the challenge of general economic growth sufficiently rapid and sustained to move a maximum number of poor families out of the poverty class. It seems that the period of high-level economic activity during World War II, and the associated large demand for labor, came closer to reaching this objective for Mexican-Americans as well as others than did any period

Table 4

Estimate of Characteristics of
Poor and Non-Poor Spanish-Surname Families and Individuals: 1960
(Families and Individuals Counted Only Once)[a]

THE POOR

	Number of families[b]	Members of families	Head's Children under 18	Persons[c] not in families
Total Number	241,000	1,092,000	527,000	105,000
Total Percent	100.0%	100.0%	100.0%	100.0%
Head farmer or farm laborer	23.2	28.7	29.6	31.4
Head 65 years and over	14.1	9.1	2.3	16.2
Female family head	20.3	17.8	18.6	22.9
Head employed less than 13 weeks[d]	4.6	5.1	6.5	3.8
Head under 25 years	6.6	4.8	3.6	15.2
Head 0-4 years education	13.7	15.0	17.3	3.8
Other	17.4	19.6	22.2	6.7

THE NON-POOR

	Number of families	Members of families	Head's Children under 18	Persons not in families
Total Number	461,000	2,233,000	1,007,000	75,000
Total Percent	100.0%	100.0%	100.0%	100.0%
Head farmer or farm laborer	5.4	6.6	6.9	12.0
Head 65 years and over	5.2	4.1	0.6	9.3
Female family head	4.6	4.1	3.7	18.7
Head employed less than 13 weeks[d]	4.3	4.9	5.4	8.0
Head under 25 years	6.9	5.2	4.8	10.7
Head 0-4 years education	12.6	14.7	14.9	12.0
Other	61.0	60.4	63.9	29.3

---

[a] See text for further explanation.

[b] Poor families are defined as having income of less than \$3,000 in 1959. Number of families equals number of family heads.

[c] Individuals were classified as poor if their income in 1959 was under \$1,500. Inmates were excluded.

[d] Includes unemployed.

Source: From the One-in-One-Thousand Sample, 1960 Censuses of Population and Housing (See Reference 7).

thereafter, no matter how prosperous. It would be a stupendous failure of our economic system if the same goal could not be accomplished without war. The long-run solution calls for larger and more widely-shared wealth -- shared by people in different income classes and regardless of minority status.

References

1. This generalization is valid for the Southwest as a whole; but, there are interesting variations between the five states. See Joan W. Moore and Frank G. Mittelbach, Residential Segregation in the Urban Southwest, (Los Angeles: Mexican-American Study Project, Graduate School of Business Administration, University of California, 1966) Advance Report 4, pp. 5-10.

2. Ibid., pp. 15-22.

3. Walter Fogel, Education and Income of Mexican-Americans in the Southwest, (Los Angeles: Mexican-American Study Project, Graduate School of Business Administration, University of California, 1965) Advance Report 1, pp. 5-7.

4. Leo Grebler, Mexican Immigration to the United States: The Record and Its Implications, (Los Angeles: Mexican-American Study Project, Graduate School of Business Administration, University of California, 1966) Advance Report 2, Chap. II.

5. When families were classified as poor if their incomes were less than half the Southwest median income, the incidence of poverty in each of the five states was the same as when the $3,000 line was used, because half the 1959 median family income was $2,996, rather precisely at the line already adopted. This methodology was adapted from Victor R. Fuchs, "Toward a Theory of Poverty," in The Concept of Poverty, published by the Chamber of Commerce of the United States, Washington, D. C., 1965.

6. See, for example, Herman P. Miller, Rich Man, Poor Man, (New York: Thomas Y. Crowell Co., 1964) pp. 58-80.

7. For a detailed discussion see: U.S. Department of Commerce, Bureau of the Census, U.S. Censuses of Population and Housing, 1/1,000, 1/10,000, Two National Samples of the Population of the United States, Description and Technical Documentation (Washington, 1963) and supplements thereto. Users of these data are obliged to include the following notation: "Certain data used in this publication were derived by the authors from a computer tape file furnished under a joint project sponsored by the U.S. Bureau of the Census and the Population Council and containing selected 1960 Census information for a 0.1 percent sample of the population of the United States. Neither the Census Bureau nor the Population Council assumes any responsibility for the validity of any of the figures or interpretations of the figures published herein based on this material." (From Supplement 1, p. 10.)

   The estimates shown for the Spanish Surname population in Tables 3 and 4 were drawn from this source. All categories in Table 3 are strictly comparable for the Spanish surname and other groups with the exception of the farm category. For Spanish surname families with heads whose occupation is farmer or farm worker we included employed, unemployed and those not currently in the labor force. For total and nonwhite families with heads who are farmers or farm workers, only those in the experienced labor force with income are included, because data are available only in this form.

   Family heads who have completed four or fewer years of education are included only for the Spanish-surname group in Table 3, since cross tabulations of education and income are published only for family heads with less than eight years of education for the total and nonwhite population.

8. The intricate strands of the web relating family structure and poverty have been analyzed for Negroes in several works. The Negro Family: The Case for National Action, Office of Planning and Research, U.S. Department of Labor, 1955, is one. Another is Thomas Pettigrew, A Profile of the Negro American, (Princeton, New Jersey; D. Van Nostrand Company, Inc., 1964).

9. Herman P. Miller, op. cit., pp. 64-70.

SAMPLING AND TELEVISION ELECTION COVERAGE

Irving Sivin, Louis Harris and Associates

I. Introduction

Good news reporting like good military tactics consists in "getting there fastest with the mostest." The traditional news "scoop" gets to the public first, and contains all the pertinent information - the who, what, when, where, why, and how of the event. The goal of television coverage of election returns is that of good reporting: to give the news with as great speed and accuracy as possible. The news, in this case, is who has won an election, how and why he achieved the victory, and what the victory means for the future.

Until 1948 newspapers and radio dominated national election news reporting. That year the opinion polls were somewhat in error, H.V. Kaltenborn's career as a radio political analyst came to an end, and an early edition of the Chicago Tribune proclaimed that Thomas E. Dewey had been elected president. (Harry Truman kept a copy of that one).

Not because of these events, but certainly soon after them, the television networks began to dominate election news coverage. Television grew up as a contemporary of the electronic digital computer. In 1952 and 1956 television networks consulted their computers during the course of election night broadcasting, but computer abilities were not truly integrated into the broadcast. The computers performed their limited news functions fairly well in those years. In both years Eisenhower won sweeping victories, carrying every state except some southern and border states. By 1960, as another generation of computers was evolving, the networks planned to make fuller use of their capacities, integrating them into the broadcast. The results were not altogether happy. Early in the evening, the computers failed to detect the narrow division that eventuated. When their later analysis proved reasonably accurate, it was too late. The public, like both John Kennedy and Richard Nixon, watched the total vote count and then went to bed -- still waiting for the undetermined outcome.

The networks' evaluation of their 1960 performance led to the introduction of scientific sampling as an important reporting tool for the presidential election in 1964. As a result of this change, all three networks produced remarkably accurate reports throughout election night that year. Some of the credit for this, however, must be given to the Johnson landslide. The fact that it was "LBJ all the way" enabled the presidential election to be reported quickly and accurately. But some of the best reporting in 1964 came about in calling the outcome of contests for Governor and U.S. Senator. Of the 107 contests (inclusive of President) in the old 48 states and the District of Columbia, 106 winners were called correctly by the National Broadcasting Company and the American Broadcasting Company, and all 107 were reported correctly by the Columbia Broadcasting System. ABC and NBC both failed on the same race, the tight contest for Senator in Ohio, which was won by the incumbent, Senator Young, against Robert Taft, Jr.[1]

The accuracy and speed of the 1964 election coverage in contrast with the fuzziness that characterized election night news coverage in previous years was due, in good measure, to the marriage of probability techniques and the statistical analysis of voting groups to computer capacities.

It is my purpose in this paper to formulate in statistical terms the problems of election news coverage and to describe one set of proposed solutions. In so doing, I shall point out some of the limitations of the proposals.

II. The Vote Collection Problem - Sampling, Complete Coverage or Both

Television coverage of the results of a presidential election generally begins between 6 and 7 pm Eastern Standard Time. By then the polls in Kentucky, Connecticut and Maryland have closed, the Eastern voters have returned home and are finished with their evening meal. The California polls close at 11 pm EST. Only Alaska and Hawaii, which have seven electoral votes between them, have polls open after 11 pm EST, a time when the Eastern audience begins to dwindle. A presidential election broadcast is continued on air until the outcome has been made clear, which frequently happens before the West Coast audience goes to bed, or around 3 am EST. The networks broadcast the effect of the presidential election on the senatorial, gubernatorial and congressional contests. In the past the coverage for senatorial and gubernatorial seats has been complete, while the congressional reporting has been selective. These contests are also reported by the networks in non-presidential years.

When the presidency is at stake, national patterns may become clear to the analysts early in the evening, as happened in 1956, when early returns from Connecticut showed that Adlai Stevenson was running behind his own 1952 mark. If the national story is late in breaking, as happened in 1960, the networks may lose large segments of their audience, particularly after 11pm EST, before they are able to announce a winner. In 1960 many people in the East went to bed believing that Kennedy had won a resounding victory, only to discover the next morning, when they turned on their sets, that the outcome was still in doubt.

In order to guarantee that they would be able to tell the news story to the public on the night of the election, the networks took two major steps between 1960 and 1964. First, they allocated a considerable part of their resources to select sample precincts in each state, from which they communicated the results directly to a central network computer input area. Second, the networks engaged in a vote collection effort on an unprecedented scale in order to speed the general reporting of returns.

In 1960 and earlier years the wire services did much of the raw vote collection and the television networks purchased this data from them. The raw vote figures displayed by the various networks could be quite different, since they could reflect different states and different reports within the same state. However, in states which do not have quick-counting automatic voting machines, the wire services' collection was too slow for the networks which might lose their audience before the outcome was known. In some of the slowly reporting states, like California or Massachusetts, a victory of landslide proportions was necessary to determine a winner before the audience went to bed. For these reasons the networks decided in 1964 to invest additional financial resources to speed the reporting of the vote.

The decision as to how much to invest in speeding up the coverage was pending when the California primary of 1964 was held. In that primary, on the basis of sample data, CBS and then ABC declared that Goldwater had won over Rockefeller. Hours later, press service returns, reflecting the vote in northern California, showed Rockefeller with a substantial lead. The California public was confused by the conflicting claims of television and the wire service reports. The sample data proved ultimately to be correct. This situation not only affirmed the networks in their decision to use samples, but also served to spur their efforts to insure that the vote count from all the precincts would be moved to the public as quickly as possible, in order to avoid any similar confusion.

The cost structures of the two types of data collection, sample and complete, are quite different. A very high unit price per precinct is required for the specially designated sample precincts. Election night costs for sample precincts involve the pay of the reporter covering each precinct, the cost of the phone message to the computer input area, and the payment of the operators at the input area. In addition, there may be extra costs for installing a telephone in the vicinity of the polling place. These costs are generally comparable to the complete coverage costs. The major expenses for sample precincts are the pre-election ones. These involve the selection of the sample and research into the voting history of the sample precincts. Such research also involves establishing whether or not the precinct has the same physical boundaries as it had in the recent past, and obtaining a detailed description of the socio-economic, ethnic and religious characteristics of the precinct. The descriptive material is integrated into the analysis on election night. Finally, there is a high cost per precinct associated with the programming of the input quality control and statistical analysis of the sample data. Unless the raw vote data are to be subjected to statistical or detailed political analysis, the cost of handling the raw vote is much smaller on an individual precinct basis.

Despite the differences in unit costs, so few sample precincts are needed to determine the outcome of an election, that for most states the total cost of sampling is far below that of the complete vote collection effort. Since reporting from sample precincts provides accurate and quick results of the outcome of an election, and provides the possibility of detailed analysis, the real question is why have the networks invested in both types of data collection.

There is a threefold answer. I have stated one already: speeding up the general vote collection process means that there will be less possibility that the public will be confused by different reporting sources. Second, and most important, the public knows that an election is really determined only when all the votes are counted, and the public continues to demand that count. Finally, because of sampling error, samples experience difficulties in making determinations of the outcome in close contests. (In such a case a sample can report early in the evening that a contest will be a close one, and a sample will be able to indicate which socio-economic groups and which geographic areas are supporting each contestant, but it still will not be able to name the winner.) In a very close state election, data from all the precincts in states which use automatic (mechanical) voting machines in every precinct will be available before the Eastern audience has gone to bed, and therefore an investment in full collection is worth while. (In states where paper ballots or electronic voting machines are used in part or in whole, the full tabulation of races at the top of the ticket has not yet proved to be really fast enough to determine the outcome of a close election before the audience goes to bed. But the progress of the count has been speeded.)

It is perhaps somewhat ironic that the installation of electronic voting devices in many areas has slowed down the determination of the outcome of elections. With the electronic machines the vote must be counted at a county center. The very act of transporting the cards several miles delays the count. In many counties electronic machine tabulation has meant that several days elapsed before the exact outcome was known. The delay has been due to the fact that the public has not been educated to mark the new ballots properly. The sorting and tabulating equipment have therefore experienced difficulties in reading the ballots.

In sum, between 1960 and 1964 the networks discovered that state by state sampling was an indispensable aid in election coverage. Sampling speeded the determination of the outcome of an election, and this speed, in turn, spurred the effort to obtain even faster raw vote returns.

III Sampling Considerations

A. Initial Considerations - Precision, Risk, and the Underlying Variance

In each race that is being covered the sample must be able to discriminate between a victory and a defeat. It is not feasible for a sample to distinguish between 50.1 and 49.9%, but a sample may be required to distinguish between 51.0% (or at most 52.0%) and 49.9%. By the word "distinguish" I mean that the difference between the sample estimate and 49.9% is statistically significant at the 1 or 2 % level.

Whatever other value an election sample may have, its principal function is to determine the outcome of an election, or the investment in the particular race will not have fully paid off. The risk the networks are willing to take in being publicly in error can be gauged by their 1964 performance, a level of about one or at most two chances out of one hundred of being wrong.

The unit which must be sampled is, of course, the voting precinct. The number of voters in a precinct varies enormously from state to state and almost as much within some states. New Hampshire precincts may have as few as 8 or as many as 5000 voters. The largest precinct in Massachusetts contains fourteen thousand voters, while many towns have fewer than 300 voters. On the average, the 180,000 precincts in the United States in 1964 had about 390 voters each.

A precinct is not a randomly formed cluster of voters. Since precincts represent small geographic areas, they tend to be far more distinct from each other than might be expected from a random formation. In terms of economic characteristics, there are many precincts in which most of the voters have extremely high or low incomes. There are many precincts that are predominantly Negro, or Puerto Rican, Italian, Scandinavian , Polish, or Jewish. As an effect of social, economic, ethnic, and religious clustering, voting precincts show a greater diversity in their support of candidates than could be expected were they randomly formed from the population at large. In a close (50-50) election within a state, some of the precincts give 30% of the vote to the winner while others give 70%. This is such a common phenomenon that we hardly think anything of it. Yet, were precincts randomly formed, few of them (with 400 voters) would differ from the state's average by more than 5%, and almost none of them would differ by more than 8 or 10%. The fact that such vote clustering takes place increases the cost of sampling by requiring a larger number of sample precincts then would be needed were there no clustering. On the other hand, the clustering effect provides an enormously useful tool in the analysis of the voting behavior of the groups which together constitute the American body politic. I will return to the analytical advantages of clustering effects later. For the moment I will concentrate on their effect on sample size.

In a close election it would take a sample of about 10,000 persons selected at random to differentiate an estimate of 51.0% from a value of 49.9% at a 2% (one-sided) significance level. Were precincts randomly formed and of uniform size with 400 voters each, twenty-five precincts per state would be sufficient to determine the outcome of each race. The same precincts could be used for all the contests which one wished to report in a state.

Studies of the outcome of elections indicate that the intra-class correlation coefficient, the measure of voting homogeneity within a precinct, is rather high, attaining values of .05 - .06, for precincts of some 400 voters. This may be translated by saying that the standard deviation in the population of precincts in some states is of about 10 to 12 percentage points. If we suppose a simple random sample of precincts were possible within a state and that we wished to discriminate between 52.0% and 49.9% taking the chance of being in error about 2 percent of the time, the minimum size of sample needed in a state is 100 precincts. A sample size of 100 per state would give rise to a national stratified sample which would estimate the presidential electoral vote with good accuracy. The popular vote for the presidency could be estimated to well within half a percentage point.

An heuristic argument showing the precision to be expected in the national popular vote is shown below:

Let 1) K be a subscript standing for a state

2) $\hat{P}_k$ be the estimated percentage for a presidential candidate within state k

3) σ_k^2 be the variance in a state percentage

4) n_k be the size of sample in a state

5) W_k be the weight (proportion) which a state contributes to the national vote

The estimated national percentage, $\hat{P} = \sum^{51} W_k \hat{P}_k$

The variance of the estimate is $\sigma^2_{\hat{p}} \doteq \sum^{51} W_k^2 \sigma_k^2 / n_k$

If n_k is chosen so that $\sigma_k^2 / n_k = 1$ for all states, then

$$\sigma^2_{\hat{p}} \doteq \sum^{51} W_k^2$$

But in the America of the 1960's only two states, New York and California, have weights that approach one tenth of the total vote. Therefore $\sigma_p^2 < .3$.

B. Criteria for Stratification in State Samples

Simple random sampling of precincts is seldom possible in any state. Were it possible, it would seldom be employed, because stratified sample designs are more efficient, particulary when the stratification is accomplished on a correlate of the variable being measured, the percentage vote given to the Democratic (or Republican) candidate. The most efficient stratification is that in which the strata means are widely different. In some states, like New Hampshire, it is possible to accomplish this by allocating precincts individually to strata, because detailed data are published for each precinct, and because the number of precincts is small. For many states, however, the only available published data are those provided by Richard M. Scammon in America Votes or by the Bureau of the Census in its technical studies. The fundamental unit of publication in these sources is the county, although Scammon provides the vote for large units in the principal cities of the country, - the wards or state assembly districts. The large political units, for which past data are available, are the fundamental units used in stratification. Once grouped together in strata, the counties, wards, assembly districts, etc. serve as the primary sampling units for stratified two-stage cluster sampling.

For each primary sampling unit (PSU) the percentage obtained by a Democratic candidate in four recent statewide elections is computed. A time span of six years is used, in order to include the vote in the two most recent presidential election years and in two non-presidential years. The PSU's are then ranked on each election, the rank of 1 being arbitrarily assigned to the PSU with the highest Democratic percentage in a particular election. The PSU ranks are then summed over the four elections used. This sum is used as a basis for the final grouping of the PSU's into strata.

Outside of the South, which is now undergoing what amounts to a political revolution, the rank order correlations of the PSU's have been over .80 for most states and for most pairs of elections. In the states of the Old Confederacy the effect of recent political change has been to produce much lower coefficients, even negatives ones. In Virginia, for example, where the Byrd machine was entrenched until the primary of July 1966, some of the counties which were the strongest supporters of the Democratic party, when Byrd candidates were contending for office, gave relatively low percentages to the Presidential nominees of the Democratic party in 1960 and 1964 when the Byrd machine did not support the Democratic party. Nevertheless, even in the South, the rank order provides a useful stratification device. In some states the rank order goes hand in hand with geography and the degree of urbanization. In New York State and Massachusetts a simple partitioning of the state into a few geographic areas accomplishes a very effective reduction of the electoral variance. The final stratification used in sampling recognizes the three basic stratification variables: 1) average rank order, 2) degree of urbanization, 3) and geographic location.

When the PSU's and strata are formed, an effort is made to keep them equal in size. Counties or wards with very small voting populations are joined together with adjacent counties or wards for this purpose.

Most strata consist of only a few PSU's. There may be 10 to 20 strata in a state. The results of the stratifying process have been to produce PSU's of relatively small size and with small variance between the PSU's. The chief contributor to sampling error has been the variation in the percentage Democratic within the PSU's (the variation among the precincts within the counties, wards and assembly districts). The typical PSU contains somewhere between 25 and 100 voting precincts depending on the state and the number of strata. Stratification of this type has brought the substantial reductions in sampling error. In some states the sampling variance has been cut in half by stratification. This reduction is, of course, equivalent to a reduction of sampling error of about 30% below that of a simple random sample of the same number of precincts.

These estimates of the efficiency of stratification are based on samples used in 1964, in which there was an equal allocation of the number of sample precincts to each stratum. Since all the strata have approximately the same number of voters, and since the within-stratum variances are also about equal, an equal number of sample precincts in each stratum is close both to proportional and to optimum allocation. Although the samples have not departed greatly from the optimum in any state, the gains from stratification vary markedly from state to state.

The effect of the stratification is to reduce the number of precincts required to make a 52.0-49.9% distinction down from 100 to the neighborhood of 50-70 precincts. These gains are translated directly into cost savings.

IV Estimation Procedures

A. Procedures Without Historical Data

The fact of federalism almost insures that there will be no single sampling or estimation procedure that will uniformly produce the smallest possible mean square error for all contests within all states in any year. Even within a single state, a variety of factors make certain estimates more useful for one contest than for another.

Whenever possible both sample design and estimation should take full advantage of historical data. The creation of strata from the PSU's by ranking them is one such use. But below the county level, it may not be possible to find relevant historical voting data. Let me assume for a moment that as a result of 1) the Supreme Court's "one man, one vote" decision, 2) civil rights legislation abolishing the poll tax, 3) the creation of a viable Republican party in the South, 4) a Decennial or local Census, 5) land annexation, or 6) the redrawing of precincts to accommodate new types of voting machines, not a single precinct in the United States has the same land boundaries today as it had in the most recent past election. Revised precinct boundaries do not permit historical data to be used on a precinct level in an unbiased way, and I shall examine the type of estimates which can then be used. (A similar lack of relevant precinct historical data arises when one wishes to determine the outcome of a primary election through sampling, particularly in a state where there has been no primary for the office for that party in the past few years.)

In a simple random sample of the current precincts in a state, or in a stratified self-weighting sample with equal probabilities of selection of precincts at all stages, the mean per precinct estimate would produce an unbiased estimate, if all precincts had the same total number of voters. The term "mean per precinct" refers to the average of the percentages of the vote cast for a candidate in each precinct. Until now my discussion of the size of sample necessary to solve the election night coverage problem has been predicated upon the assumption of equality in the number of voters in each precinct of a state. However, precincts do differ in size, and PSU's within a state differ on three key size variables: the total number of voters in the PSU, the total number of precincts in the PSU, and the average number of voters per PSU. Therefore, the mean per precinct estimate produced by equal probability selections (whether in one or in two stages) can be seriously biased. The bias is a function of the correlation between the percentage Democratic (or Republican) and the number of voters in the precinct. In many states these correlations are sufficiently large, so that, for samples of 40 to 100 precincts, the bias can be the largest contributor to the root mean square error. In New Hampshire, where great variation in the size of the precincts exists, there is a strong association between the number of voters in a precinct and percentage Democratic. This is shown in the table below.

The largest precincts are in the Democratic cities of Manchester, Nashua, and Portsmouth, while the small precincts are identical with the small towns, sturdily Yankee and ever Republican.

If it were possible to select precincts in a single stage process with probabilities proportionate to size, the simple mean per unit estimate (the average of the precinct percentages) from the resulting sample would be unbiased. The considerations which I have adduced with regard to the size of sample necessary to accomplish the election night task would remain substantially valid, and unchanged. But one can never sample precincts with probability exactly proportional to the unknown, future voter turnout. If good measures of the size of the voter turnout in each precinct are available, the mean per unit estimate will be biased, but the bias will be small in comparison to the sampling error for samples of 50 or more precincts. Even under the harsh conditions which I have assumed, the annihilation of virtually all useable past voting data, the current total voter registration, be it for one party in a primary or for all parties in a general election, is available a few months before the actual date of an election and could be used for precinct selection . (The cost of obtaining registration data for all precincts in the sample PSU's is extremely high and the likelihood of 100% coverage is small.)

In the event that no adequate precinct size measures are on hand or that one wishes to produce a consistent estimate, a ratio estimate of the percentage of the vote cast for the Democratic party is always available, regardless of the particulars of the selection procedure. This estimate takes the form

$P' = (X'/Y')100$ where X' is the estimated total Democratic vote
where Y' is the estimated total vote
and P' is the estimated percentage.

X' is an unbiased estimate of the true Democratic total and is derived by inflating the precinct vote to the population or stratum vote. The same is true for the Y' variable.

In stratified samples this estimate is the combined ratio.[2] Since a ratio estimate can be greatly affected by variations in the size of the denominator variable, in states where the precincts differ appreciably in total votes

SIZE AND PERCENTAGE CORRELATION
NEW HAMPSHIRE GUBERNATORIAL ELECTION, 1964

Number of Voters in Precinct	Percentage Democratic	Percent of State's Precincts	Percent of State's Vote
< 50	48.5	2.6	0.05
50- 99	52.7	5.3	0.4
100- 199	54.0	9.9	1.7
200- 399	51.8	16.9	5.0
400- 799	56.5	21.9	13.2
800-1599	67.2	26.2	31.7
1600-3199	70.6	12.9	28.7
3200+	73.0	4.3	19.2

cast, the ratio estimate may best be used in conjunction with sampling with probability proportional to estimated size (PPES). PPES will insure that the coefficient of variation of the sample mean of the denominator is small, less than .1, and therefore that the bias of this estimate will be small in comparison with the sampling error. The detailed form which the ratio estimate assumes in stratified two-stage cluster sampling is shown below:

$$\frac{\displaystyle\sum_{h=1}^{L} \frac{1}{m_h} \sum_{i=1}^{m_h} \frac{\frac{1}{n_{hi}} \sum^{n_{hi}} \frac{X_{hij}}{Z_{hij}}}{Z_{hi}}}{\displaystyle\sum_{h=1}^{L} \frac{1}{m_h} \sum_{i=1}^{m_h} \frac{\frac{1}{n_{hi}} \sum \frac{Y_{hij}}{Z_{hij}}}{Z_{hi}}} = \frac{X'}{Y'} = \hat{P}$$

In the above formulation

X_{hij} is the Democratic vote in precinct j of PSU i in stratum h

Y_{hij} is the total vote in precinct j of PSU i in stratum h

Z_{hij} is the (conditional) probability of selecting precinct j given the fact that PSU i has been selected in stratum h

n_{hi} is the number of precincts selected in PSU hi

Z_{hi} is the probability of selection of PSU hi

m_h is the number of sample PSU's in stratum h

Regardless of the sampling or estimation procedures used, when historical data is not available on a precinct level, stratification is unable to reduce the inherent variance by much more than 50%. The reduction in some states will be appreciably less than that. This fact is due to the high variance within the PSU's.

B. Estimation Using Historical Data

1) The difference or change estimate in single stage sampling.

In any state the relationship between the percentage of the vote cast for a Democratic candidate for office in 1966 and that cast for a Democratic candidate for the same (or similar) office in a contest in 1964 can be expressed as $P_{66} = P_{64} + C$. In this statement P_{66} and P_{64} are the true Democratic percentages in 1966 and 1964 respectively. C is the change in the Democratic percentage in the two years.

Exactly the same relationship holds true on a precinct basis. That is $P_{66j} = P_{64j} + C_j$. Here j is a subscript used to indicate a precinct.

If all precincts had the same number of voters, the arithmetic mean of C_j over all precincts would be the true population value C. A simple random sample of precincts would produce unbiased estimates of the change, C, by using the mean per unit estimate. Since precincts do vary in size, sampling in one stage, with replacement and with probabilites proportional to the turnout (PPES) in the precincts would lead to an unbiased difference estimate of the form:

$$P''_{66} = P'_{66} - P'_{64} + P_{64} = C' + P_{64}$$

P'_{66} is the arithmetic mean of the sample precinct percentages and is an (unbiased) estimate of P_{66}, the true percentage in 1966.

P'_{64} has a similar meaning.

P_{64} is the true Democratic percentage in 1966.

P''_{66}, the difference estimate, will be a better estimate than P'_{66}, the ordinary mean per unit estimate, whenever the variance of the Democratic percentages are about the same in the two years and the correlation between the percentages on a precinct basis is greater than 0.50.

Empirical research indicates that correlations around .85 exist in most elections in the United States. The variance of the difference (change) estimate would be only three-tenths that of the ordinary mean per unit estimate, assuming equality of the variances in the two election years. In general, the variance of the estimate P''_{66} is

$$\sigma^2_{P''_{66}} = \sigma^2_{P'_{66}} + \sigma^2_{P'_{64}} - 2\rho_{66,64}\, \sigma_{P'_{66}}\, \sigma_{P'_{64}}$$

If the variance of the past race is larger than that of the present race, the correlation coefficient must be high for the difference estimate to be useful. For example, if the past race had a standard deviation of 15 percentage points and the current race one of 10, a correlation of .75 would have to be present in order for this difference estimate to be as useful as the mean per unit estimate. Conversely if the past race had a smaller variance than the present one, lower values of the correlation are required.

The following table indicates what values of the variance of the sample estimate can be achieved with different levels of correlation and different relationships between the underlying standard deviations in different years. The sample variances are compared to 100, taken as the variance of the mean per unit estimate based solely on current data. The comparisons are made for the same size of sample.

VARIANCE OF THE DIFFERENCE ESTIMATE
In Comparison to the Mean Per Unit Estimate (Considered as 100) For Different Relations of the Underlying Standard Deviations

RHO	$\sigma_{66} = \sigma_{64}$	$\sigma_{66} = 1.5\,\sigma_{64}$	$\sigma_{66} = 2/3\,\sigma_{64}$
1.00	0.0	11.1	25.0
.95	10.0	17.8	40.0
.90	20.0	24.4	55.0
.85	30.0	31.1	70.0
.80	40.0	37.8	85.0
.75	50.0	44.4	100.0
.70	60.0	51.1	115.0
.65	70.0	57.8	130.0
.60	80.0	64.4	145.0
.55	90.0	71.1	160.0
.50	100.0	77.8	175.0
.45	110.0	84.4	190.0
.40	120.0	91.1	205.0
.35	130.0	97.8	220.0
.30	140.0	104.4	235.0
.25	150.0	111.1	250.0

In Georgia in 1964, the difference estimate had a much higher sampling error than the mean per unit estimate based on PPES sampling. In this state many of the traditionally Democratic precincts voted strongly for the Republican party, while the urban Negro of Atlanta, who had historically given his vote to the party of Abraham Lincoln (Nixon did well among Atlanta's Negro population), turned overwhelmingly against Barry Goldwater. The estimated Johnson-Kennedy correlation was -0.2.

The question of whether or not the correlations between two elections will be sufficiently high to justify the use of a difference estimate need not be answered as late as election night. An opinion poll taken just before the election can determine whether the correlations will be high enough to justify the use of the difference estimate. On election night itself the correlation coefficient may be estimated from the precinct data.

2.) Bias in the Difference Estimate

If simple random sampling of precincts were used, the difference estimate could be biased because of the persistence of the correlations between the number of voters in a precinct and the change in the precinct's percentage. In 1964, outside of the South, Republican voters did not turn out to vote to the same degree that they had in the previous two elections. Some New Hampshire precincts, which were small and heavily Republican in 1960, had an even smaller voter turnout in 1964. The Democratic percentage in these precincts moved upward by a far greater degree than it did in the large, normally Democratic precincts of Manchester and Portsmouth. A precinct which had given Kennedy only 35% of the vote could easily give Johnson 30% higher and remain relatively close to the state and national percentage. But a precinct which had given Kennedy 75% could not possibly give Johnson an increase of more than 25%, and generally did not give him an increase of more than 10 or 15%. Single stage equal probability selections would produce too many large increases and too few small ones, wherever large precincts had been heavily Democratic in the past. In simple random sampling with equal selection probabilities the difference estimate, based on the average change in the precinct percentages, is biased because of the existence of these correlations between the past percentage, the change in the percentage, and the size of the precinct. The mean per unit estimate with PPES sampling reduces this bias to the point where it is small in comparison to the sampling error, for samples of 50 or more precincts.

3.) The Difference Estimate in Stratified Sampling

If no historical data were available for use in estimation, the stratified estimate of the percentage of the vote received by the Democratic candidate might take the form of $P = \sum W_h P'_{66_h}$. With historical data available, the estimate would be:

$$P''_{66} = \sum_{h=1}^{L} W_h (P'_{66_h} - P'_{64_h}) + P_{64} = \sum_{h=1}^{L} W_h C'_h + P_{64}$$

That is, the estimated Democratic percentage is the weighted sum of the estimated changes within the strata. W_h represents a weight and the P'_h is an estimate of the stratum percentage, while C'_h is an estimate of the change in the stratum percentage.

The magnitude of the correlations within most strata tends to be at about the same level as that for the state as a whole. Thus, where the initial introduction of stratification reduces the underlying variance in a state from 30 to 50% of the total variance, the use of a stratified difference estimate within each stratum can accomplish a substantial further reduction. Using a single random sample with no historical data available, 100 sample precincts may be necessary to differentiate an estimate of 52.0% from 49.9%. Using stratified difference estimates, where precincts are selected directly in a single stage by PPES, may mean that only 35-40 precincts are necessary to make that distinction.

In many contests the gains from stratification unfortunately overlap those made by the difference estimate. Samples smaller than 40 precincts may not achieve the needed precision.

I should point out that if stratification is performed on the rank order of the component PSU's, the bias of the ordinary mean per unit or of the difference estimate tends to be small, even when a simple random sample is taken within each stratum. That is, stratification on the basis of the past vote tends to reduce the correlation between the size of the precinct and the change in the percentage. Stratification of precincts by size would also reduce the bias.

C) Estimation Procedures Using More Than One Past Race

In discussing this subject I shall limit my remarks to the simple model of single stage using sampling PPES and the mean per unit estimate.

Estimates of the following linear forms are under discussion:

1.) $P'''_{66} = P'_{66} - 1/2\,(P'_{64} + P'_{62})$

$\qquad + 1/2\,(P_{64} + P_{62})$

2.) $P''''_{66} = P'_{66} - 1/3\,(P'_{64} + P'_{62} + P'_{60})$

$\qquad + 1/3\,(P_{64} + P_{62} + P_{60})$

In the above difference formulae, the primes indicate estimates and the P's without primes are true values. The subscripts are chosen to represent different years, but two or more contests from a previous election year may be used instead. The coefficients are pre-assigned values, giving equal weight to each of the past elections used. They may be regarded as regression coefficients. The variance of the above two estimates can be given as follows:

1.) $$\sigma^2_{P'''_{66}} = \sigma^2_{P'_{66}} + 1/4\,(\sigma^2_{P'_{64}} + \sigma^2_{P'_{62}}) + 1/2\,\rho_{64,62}\,\sigma_{P'_{64}}\,\sigma_{P'_{62}} - \rho_{66,64}\,\sigma_{P'_{66}}\,\sigma_{P'_{64}} - \rho_{66,62}\,\sigma_{P'_{66}}\,\sigma_{P'_{62}}$$

and

2.) $$\sigma^2_{P''''_{66}} = P'^{2}_{66} + 1/9\,(\sigma^2_{P'_{64}} + \sigma^2_{P'_{62}} + \sigma^2_{P'_{60}}) + 2/9\,(\rho_{64,62}\,\sigma_{P'_{64}}\,\sigma_{P'_{62}} + \rho_{64,60}\,\sigma_{P'_{64}}\,\sigma_{P'_{60}} + \rho\,\sigma_{P'_{62}}\,\sigma_{P'_{60}}) - 2/3\,(\rho\,\sigma_{P'_{66}}\,\sigma_{P'_{64}} + \rho\,\sigma_{P'_{66}}\,\sigma_{P'_{62}} + \rho\,\sigma_{P'_{66}}\,\sigma_{P'_{60}})$$

The following tables indicate the efficiencies of these estimates in comparison with the mean per unit estimate, the variance of which is considered as one. The tables assume that the variances of all the races are the same.

VARIANCE ACHIEVED BY MULTIPLE DIFFERENCE ESTIMATES IN COMPARSION TO THE MEAN PER UNIT ESTIMATE (Considered 100) (All Correlations Assumed to be Equal)

RHO	1 Past Race	2 Past Races	3 Past Races
1.00	0.0	0.0	0.0
.95	10.0	7.5	6.7
.90	20.0	15.0	13.3
.85	30.0	22.5	20.0
.80	40.0	30.0	26.7
.70	60.0	45.0	40.0
.60	80.0	60.0	53.3
.50	100.0	75.0	66.7
.40	120.0	90.0	80.0

The next table indicates the effect on the variances if one past race has a correlation with the current race which is smaller by 0.1 than the value(s) of the correlation(s) of the other past race(s).

VARIANCE ACHIEVED WHEN ONE CORRELATION WITH THE CURRENT RACE IS 0.1 BELOW OTHERS

Highest RHO	2 Past RACES	3 Past RACES
1.0	10.0	6.7
.9	25.0	20.0
.8	40.0	33.3
.7	55.0	46.7
.6	70.0	60.0
.5	85.0	73.3
.4	100.0	86.7
.3	115.0	100.0

It is evident from the above tables that where high correlations exist among all races, their use improves the estimates. The most significant reduction in variance comes about from the introduction of the first historical correlate. If a very high correlation is present in one past race, additional historical correlates have only marginal utility. The use of two or three past races, however, lowers the correlation threshold level that is needed to keep the variance of the estimate below that of simple mean per unit estimate.

When one correlate is below that of the others, the advantages of using the multiple correlates are attenuated.

D. Analytic Estimates

One of the most important functions of the sampled precincts is to report how the component groups within the electorate have voted. By classifying the precincts along several analytic dimensions, e.g. degree of urbanization, ethnic and religious make-up, socio-economic status, historical voting patterns and issue orientation, reliable estimates are available for those groups with sufficiently large sample representation.

In addition to providing "hard" news on how these groups have voted, the classification process permits alternative estimates of the state-wide percentages to be made. This is possible through the use of weights associated with each group within each analytic dimension. In effect, this is stratification after sampling.

For those interested in the behavioral aspects of political science, the analytic estimates are among the most interesting and important produced on election night. Far more information is generated on the computer then can be relayed to the public through the television medium on election night. It is to be hoped that the information already culled and reduced to tractable form will be fully explored by political scientists in the future.

V Final Determination of Sample Size and Sampling Procedures

A. A sample must be able to determine the outcome of a contest. Depending upon the type of historical data and on the nature of the analytic estimates that are desired, the required sample size may be as small as 30 precincts or as large as 100 precincts. Generally there is more than a single statewide contest in a state and the same set of sample precincts may be used to determine the outcome of more than one contest. Usually the networks do not cover more than three contests in any state. The sample size necessary is then determined by that contest which looks as if it will be the closest of the three. If none of the contests seem to be close, very few sample precincts are used to determine the victors. In order to maintain the required speed of reporting, however, a minimum of thirty precincts are used in all states for state-wide contests. In most states the information necessary to decide whether contests will be close or not, and whether historical data will prove beneficial for the difference estimates, can be obtained from pre-election polls. Polls determine the optimum solution to the sampling problem in an individual state. Polls also provide the background of issues and personalities as seen by the electorate, and permit interpretation of the "why" of the election returns.

B. Final Sampling Procedures

Within each stratum two PSU's are selected systematically without replacement. Within each sampled PSU two precincts are generally sampled. In states where smaller samples are used , only one precinct may be sampled within each sample PSU. The sampling of PSU's and the sampling of precincts within PSU's are both accomplished by sampling with probability porportional to the size measures assigned. (PPES)

The use of two precincts per PSU is a device to reduce the cost of sampling, which simultaneously permits an unbiased estimate of the within PSU variance to be made. The costs of sampling arise because within PSU precinct data are generally not published. Such data can be obtained only at the county clerk's or town clerk's office, and often only by personal and persuasive visits. Precinct vote history is a matter of public record, but old precinct maps lie, at times, in dusty drawers. County clerks have often asserted that particular precincts have not changed boundaries in a decade, only to have their maps belie them.

In the event of redistricting between the previous and the current elections, precincts are selected within the PSU on the basis of the total current registration. When no current data are available except the number of precincts in the PSU, the sample precincts are chosen with equal probability within the PSU. Except for states like California and Rhode Island, where deliberate care is taken to equalize the number of registered persons per precinct, the variations in the number of voters per precinct is so great that equal selection probabilities for precincts within the PSU's is disadvantageous.

Although the strata are formed in such a way that the total number of voters in previous elections are about the same in each stratum, when only one or two precincts are in sample from each precinct, sampling fractions vary from PSU to PSU, and from stratum to stratum. A five digit weighting factor is therefore associated with each precinct. While some computer storage capacities are therefore occupied by the weights used, this is a small price to pay to avoid the biases that develop by ignoring the different probabilities of selection.

VI Biases Their Origins and Importance

In samples of the size that are necessary to produce reliable estimates of the election outcome, bias may be a factor as large or larger than the sampling error. In samples where the precincts are specified for selection because they have been barometric in the past or because they have proved to move up and down with about the same percentages that the states have moved, the danger of bias is quite large. Only a random sample, employing the known probabilities of selection of the individual precincts can avoid such biases.

The desire to use a good estimate, such as a difference, a ratio or a regression estimate, can induce the sampler to give zero probabilities to precincts whose boundaries have changed since the most recent past election. Since one of the reasons boundaries change is the fact that a precinct may have gained or lost voting population, ignoring such precincts can cause serious biases in the estimate, particularly where the growth (or loss) areas vote differently from the more stable population areas. Given some flexibility in computer programming and in field reporting, it is unnecessary to incur this bias. The field procedures require reporting all the precincts that are now within the boundaries of the old precinct (or to record all the historical vote of the old precincts which now are included in the new precinct). In case the boundaries have severely changed and reconstruction is not possible, the estimate of the past stratum totals may be derived only from those sample precincts with past data. This will increase the variance of the past data and may decrease the correlation between the estimates of past and present derived from the set of sample precincts. Nevertheless, most of the advantages of the use of the past data are retained, without bias.

A different type of bias arises because of the absentee vote. In most states the absentee vote is comparatively small, yet it has been different in terms of the Republican-Democratic vote split from the regular vote. It was the absentee vote which shifted California from the Kennedy to the Nixon column in 1960. In some states the absentee ballot is counted at the same time as the regular vote. Then it can be sampled without bias. In other states the absentee ballot is not counted early on election night, and this causes the problem. In Texas the absentee vote in the larger counties is considerable, amounting to 6% in Harris County and 7% in Ector county in 1964. In a close race in Texas, when this vote is not available one must say that the outcome is undecided. The absentee vote also can prove troublesome when the method of counting it changes. The absentee vote may have been included in the official count of the individual precincts in the previous election, and it may be counted as an absentee box at the town or county level in the current election or vice versa. When this occurs it is difficult to reconstruct the past vote of the individual precinct, although the bias of the resulting procedure should usually be rather small.

VII. The Determination of the Outcome of an Election

When the results of the election have been reported from the sample precincts, a determination must be made of whether or not the sample percentage is significantly above (or below) 50%. This is a question of making an estimate of the sampling error, of judging whether or not the distribution of the normalized variate follows Student's "t" distribution and of judging how much bias, if any, may be in the sample statistics.

For the full model, the estimate of the sampling error is made for all the components of the difference estimate, where the estimated percentage for the two elections in question are themselves ratio estimates. For election night purposes the simple biased method proposed by Des Raj for estimating the variance of stratum totals based on PPES Sampling and systematic selection of PSU's has proved quite useful. [3]

It is often necessary to make an estimate of sampling error before the complete sample has reported on election night. Such an estimate is possible, using the collapsed stratum technique. [4] In this scheme, all the strata in the state have been paired *a priori*. Errors are calculated whenever one precinct in each of the paired strata has reported. Pairing estimates in this fashion leads to an overestimate of the sampling error, but this has proved to be a useful safeguard against inaccurate declarations of the winner.

Another estimate of sampling error which has proved to be useful, completely disregards the effects of stratification and measures only the unstratified differences in percentages. This estimate is an overstatement of the error, it is not a consistent estimate, but it has proved accurate in calling the outcome of all but the closest races (under 52.0%)

VII. Summary

Election news broadcasting requires quick and accurate determination of the results. Broadcasting is enhanced by the analysis of the voting behavior of different groups within the electorate, and even more so by an analysis of the impact of issues and personalities on the voting groups. Probability sampling is the most dependable way to meet these requirements, and is the least expensive way. No single sampling method or estimation procedure suffices to embrace the variety of situations produced by the volatile American political scene. But sampling techniques are sufficiently flexible to enable one to come close to the optimum solution for each contest in each state.

Notes and References

1) By a correct call I mean a) one in which the winner was correctly named during the broadcast or b) a contest which was stated to be too close to call and which in fact showed a 52-48% split or closer. The counts are taken from reviews of network performance and from published advertisements. Neither ABC nor NBC used sample data alone to make their calls in 1964.

2) Hansen, M.H., Hurwitz, W.N. and Madow, W.G. -- Sample Survey Methods and Theory, pages 189 - 200.

3) Raj, Des, "Variance Estimation in Randomized Systematic Sampling with Probabilities Proportionate to Size", Journal of the American Statistical Association, 60 page 278.

"The Use of Systematic Sampling with Probability Proportionate to Size in a Large Scale Survey", Journal of the American Statistical Association 59. page 251.

4) Sukhatme, P.V., Sampling Theory of Surveys with Applications pages 399-404

Hansen, M. H., Hurwitz, W.N. Annals of Mathematical Statistics. 14, pages 333-362

Appendix

Estimated Johnson-Kennedy Correlations

State	Correlation	State	Correlation
Maine	.91	Missouri	.82
New Hampshire	.90	Arkansas	.22
Vermont	.80	Oklahoma	.83
Massachusetts	.91	Texas	.94
Rhode Island	.85		
Connecticut	.96	Ohio	.91
		Indiana	.86
New York	.85	Michigan	.91
New Jersey	.87	Illinois	.88
Pennsylvania	.90		
		Wisconsin	.88
Maryland	.86	Minnesota	.85
Delaware	.91	Iowa	.84
West Virginia	.90		
		Nebraska	.88
Virginia	.52	Kansas	.86
North Carolina	.87	North Dakota	.71
South Carolina	.74	South Dakota	.88
Georgia	.20		
Florida	.76	Montana	.84
		Wyoming	.90
Kentucky	.87	Colorado	.89
Tennessee	.80		
		Idaho	.67
Alabama	.40	Utah	.94
Mississippi	.00	Nevada	.87
Louisiana	.67		
		Arizona	.95
		New Mexico	.92
		Washington	.85
		Oregon	.84
		California	.90

A SIMPLE APPROXIMATION FOR THE HYPERGEOMETRIC PROBABILITY: CASE $(0|N,n,k)$.

Irving Gedanken, Board of Governors of the Federal Reserve System.

The hypergeometric probability distribution is given for exactly x occurrences in a sample of n items without replacement.

$$p(x) = p(x|N,n,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} =$$

$$\frac{k!n!(N-k)!(N-n)!}{(k-x)!(n-x)!x!N!(N-n-k+x)!} \quad . \qquad (1)$$

Computations involving the hypergeometric probability distribution have been notoriously tedious. Most authors present the formula, give some simple examples and then advise the reader to approximate, using the binomial, Poisson or normal distributions. Unfortunately, for simple solutions of the hypergeometric, the user generally would prefer equation (1) solved for n. The problem becomes more complex when it is not the probability $p(x)$ that needs estimating but rather the sample size n, given the other parameters of the distribution.

This paper presents an easy solution for sample size n for that special case when none with the characteristic in question are found in the sample, that is $x = 0$ in equation (1). What size sample (n) need one take from the population (N), so that if a characteristic is absent from the sample, one can feel $1-p(0)$ per cent confident that there are no more than k such events in the population?

It is in the very nature of sampling that one can only surmise from an absence in a sample, $x = 0$, a similar absence in the population.

Having recorded the sunrise each day, from a size n sampling of recorded history, if no other data were available concerning the movements of the heavenly spheres, one could only hypothesize as follows,--in all of recorded history it is highly unlikely that the failure of the sun to rise occurred more than k times. I cite this example to emphasize that in most instances zero in the sample will be because of zero in the population. However, it is comforting on a cold winter night to be able to set some outside limits to the possibility that there may be events in the population missed by the sample--especially when one has other bases for questioning the zero results obtained from the sample.

The hypergeometric formula stated previously provides us with the exact probabilities of 0, 1, 2, etc., of the k events in the population showing up in a sample of size n. If the chance is $p(0)$ of zero events in the sample, when there are k events in the population, then we can be $1-p(0)$ per cent confident that true k is at least no more than hypothesized k.

The probability of exactly zero events in the sample when there are k events in the population is

$$p(0) = p(0|N,n,k) = \frac{(N-k)!(N-n)!}{N!(N-n-k)!} \quad . \qquad (2)$$

Herbert Arkin in his "Handbook of Sampling for Auditing and Accounting"* using only selected values of N, n, and k, has over 25 pages of tables covering the condition $x = 0$. This paper will now derive an approximation which will simplify sample size determination for the zero condition. This condition is appropriate for zero acceptance sampling and for discovery or exploratory sampling.

Formula (2) can be expanded as shown in formula (3) as the product of k terms in both the numerator and the denominator. Each term is one less than the preceding term.

$$p(0) = \frac{(N-k)!(N-n)!}{N!(N-n-k)!} =$$

$$\frac{(N-k)!(N-n)(N-n-1)\ldots(N-n[k-1])(N-n-k)!}{N(n-1)\ldots(N-[k-1])(N-k)!(N-n-k)!} =$$

$$\frac{(N-n)(N-n-1)\ldots(N-n-k+1)}{N(n-1)\ldots(N-k+1)} \quad . \qquad (3)$$

One approach, as our chairman Mr. Raff pointed out, is that a product of k equally spaced positive numbers may be approximated by the kth power of their arithmetic mean if the spread of the numbers is not too great.

If you recall, the sum of an arithmetic progression is

$$S = \frac{n}{2}(A + L)$$

and

$$\text{Arith. Mean} = \frac{S}{n} = \frac{A + L}{2} \quad .$$

---

* Volume I - Methods, McGraw-Hill, 1963, pp. 613.

Therefore, taking the first and last terms,

$$p(0) \leq \left(\frac{N-n-\frac{k-1}{2}}{N-\frac{k-1}{2}}\right)^k \leq \left(1-\frac{n}{N-\frac{k-1}{2}}\right)^k, \quad (4)$$

where p(0) is the <u>k</u>th power of the geometric mean of the <u>k</u> terms in both numerator and denominator.

This result was also developed by D. B. Owen, E. J. Gilbert, G. P. Steck and D. A. Young in "A Formula for Determining Sample Size in Hypergeometric Sampling When Zero Defectives are Observed in the Sample."* Solving for <u>n</u> in equation (4)

$$n \leq [1-p(0)^{1/k}][N-\frac{k-1}{2}]. \quad (5)$$

The key simplification is to go one step further and drop the reducing term $-(k-1)/2$. Hence

$$\frac{n}{N} \leq (1-p(0)^{1/k}) \quad (6)$$

or

$$p(0) \leq (1-\frac{n}{N})^k$$

$$p(0) = (1-\frac{n'}{N})^k \quad (7)$$

where $n' \geq n$

and $n' \leq N-k$.

Another approach which may, perhaps, be more evident to some, is to pair the terms of the numerator and denominator of equation (3).

$$p(0) = \left(\frac{N-n}{n}\right)\left(\frac{N-n-1}{N-1}\right)\left(\frac{N-n-2}{N-2}\right)\cdots\left(\frac{N-n-(k-1)}{N-(k-1)}\right)$$

$$= \left(1-\frac{n}{N}\right)\left(1-\frac{n}{N-1}\right)\left(1-\frac{n}{N-2}\right)\cdots\left(1-\frac{n}{N-(k-1)}\right). \quad (8)$$

Since each succeeding term on the right in equation (8) is less than its predecessor--

$$p(0) \leq (1-\frac{n}{N})^k$$

$$p(0) = (1-\frac{n'}{N})^k \quad \text{then} \quad (7)$$

$$\frac{\log p(0)}{k} = \log (1-\frac{n'}{N}). \quad (9)$$

Equation (9) provides a simple formula for solving for <u>n'</u> given <u>N</u>, <u>k</u> and the confidence level desired $(1-p(0))$.

On log log paper k, $\frac{n'}{N}$ and $1-p(0)$ show a simple linear relationship that is easy to plot and easy to read.

How much different is equation (7) from equation (4) or

$$\left(1-\frac{n'}{N}\right)^k \text{ from } \left(1-\frac{n}{N-\frac{k-1}{2}}\right)^k \text{ ?}$$

Since both are functions of p(0):

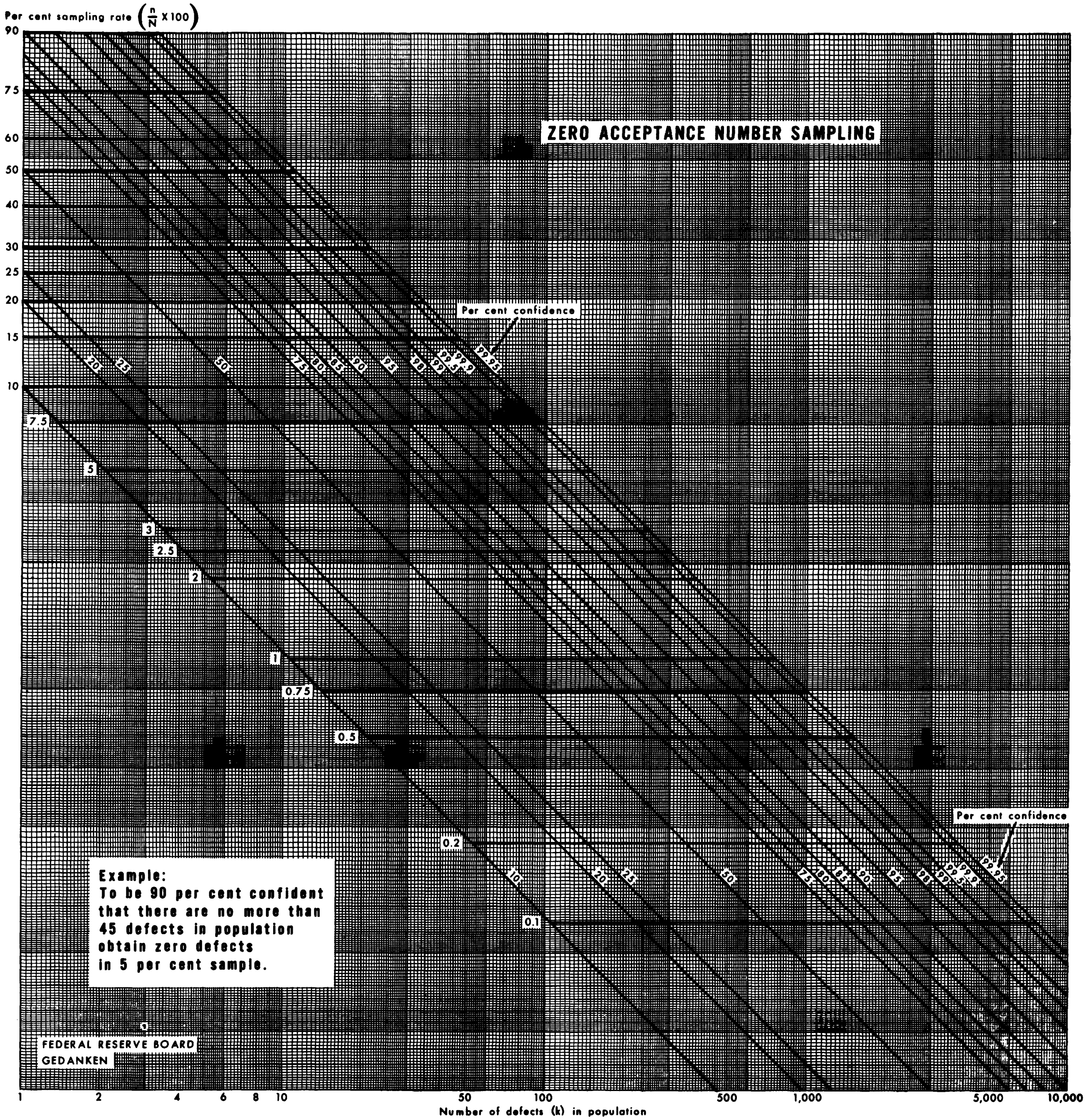
$$\frac{n'}{N} \leq \frac{n}{N-\frac{k-1}{2}} \quad (10)$$

$$n'(1-\frac{k-1}{2N}) \leq n. \quad (11)$$

It is evident that $n' = n$ when $k = 1$, and that if <u>k</u> is substantial, the reducing factor is approximately $1-\frac{1}{2}k/N$. It would be highly unlikely that this type of sampling would be used if <u>k</u> exceeded .2N or even .1N. However, even at .2N, the sample size <u>n'</u>, as read from the chart or equation (10), would be only 10 per cent too large. In most situations of practical interest, the two estimates would be approximately equal. In any event, it is worth noting that this is an upper-bound or a "fail-safe" approximation. If it errs at all in estimating sample size it errs conservatively by requiring a size larger than is necessary. Of course, if the elegance of refinement is necessary one can always multiply the <u>n'</u> estimate by $(1-\frac{k-1}{2N})$.

---

* Sandia Corporation Technical Memorandum, SCTM 178-59 (51). Available from the clearinghouse for Federal Scientific and Technical Information, U. S. Department of Commerce, Washington, D. C.



Copies of this chart may be obtained from the author at the Board of Governors of the Federal Reserve System, 20th and Constitution Avenue, N. W., Washington, D. C. 20551

For the user not interested in derivations, an understanding of the chart would be worthwhile. What would a one per cent sample which produces zero defects give you 95 per cent confidence in--(looking at the chart)--that there are no more than 300 defects in the population.

Example: $N = 100{,}000$
$n = 1{,}000$
$x = 0$
$k \leq 300$ or $\frac{3}{10}$ of one per cent.

Another example is indicated on the right hand side of the chart. This is worded the way the problem is usually raised. What size sample do I need to be 90 per cent confident that there are less than k occurrences in the population when no such events occur in the sample? Answer: a 5 per cent sample.

One might also refer to the p(0) values associated with the diagonals as the producer risk that a defect will show up in the sample when there are only k defects in the population. One minus the p(0) value indicates the risk the consumer takes that there are k defects in the population when none show up in the sample.

Consider again the 5 per cent sample--the consumer risk is just about 1/2 of one per cent that there are more than 100 defects when none are found in the sample. The producer risk on the other hand is that, even if there are as few as 4 defects, the chances are 20 per cent that one will be found in a 5 per cent sample.

You will note that the sampling rate is identically the chance of picking up the one defect when there is only one defect in the population or lot. In a 5 per cent sample, the producer risk is 5 per cent of failing lots with only one defect.

As I mentioned earlier, this chart should be a useful adjunct to zero acceptance sampling and to discovery or exploratory sampling.

MINUTES OF ANNUAL MEETING OF THE SOCIAL STATISTICS SECTION

Los Angeles, California, August 17, 1966

Meeting was opened at 7:50 a.m., 17 August 1966 by Margaret E. Martin, Chairman.

Con Taeuber, Section Representative, reported on the Board of Directors meeting. The American Statistical Association plans eventually to have a full-time Executive Director, at which time Don Riley will retire.

Ed Goldfield reported on the Proceedings. The Proceedings is now paying its own way (i.e., printing costs). Distribution is about 1000 copies. The coverage has improved in 1965--all but 2 of 40 papers and most of the discussions appeared in the Proceedings. The question of the policy of publication in the Proceedings not precluding publication elsewhere was discussed. Fred Stephan noted that most papers that appear in the Proceedings do not appear elsewhere or appear only after considerable time and extensive revision.

Margaret Martin announced the officers for 1967:

Chairman	Jack Feldman
Chairman-Elect	John Durand
1st Vice Chairman	Henry Shryock
2nd Vice Chairman	Elijah White
Secretary	Philip Sagi
Representative on Board of Directors	Margaret E. Martin
Council Representative	John Folger
Editor of Proceedings	Ed Goldfield

The 1967 program was discussed and suggestions offered as follows:

1) Improve the organizing of similar topics into sessions. Some papers on time series were not in the same session.

2) As a follow-up of this year's series on 1970 Census planning, a program on Canadian 1971 plans.

3) A session on censuses of state governments.

4) Another session on evaluation of public programs.

5) A session on data banks and the invasion of privacy.

6) Rediscover family statistics. People working on the "1 in 1000" Census tapes at Princeton, Duke, etc., could contribute to such a session.

7) Population projections for small areas.

8) A joint session on training of social statisticians with Training Section.

9) More attention to organized activity for accompanying spouses.

10) Complex sample design.

11) A session on Medicare and Medicare Statistics.

12) A session on Educational Statistics and Model building in education.

13) A session on survey designs which involve allocation of investment to reduce both sampling and non-sampling errors.

The problem of stimulating papers in Social Statistics was discussed. Margaret Martin noted that the dearth of social statistics articles in the JASA is due primarily to authors' not submitting, and not to rejections. Authors tend to think of the JASA in terms only of articles with a methodological orientation.

Fred Stephan stressed the need for expository articles to disseminate knowledge on non-mathematical aspects of statistics. People should be approached on why JASA doesn't get expository articles.

Carl Erhardt suggested that the JASA Editorial Board review the Social Statistics Section Proceedings and encourage authors of meritorious articles to prepare articles for JASA publication.

1966 Officers of the Social Statistics Section

Chairman:	Margaret E. Martin
Chairman-Elect:	Jacob J. Feldman
Vice-Chairmen:	John D. Durand (1965-66)
	Henry S. Shryock (1966-67)
Secretary:	Philip C. Sagi (1966-67)
Proceedings Editor:	Edwin D. Goldfield