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 raison d'être 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.