INFORMATION FOR THE NATION FROM A SAMPLE SURVEY

Conrad Taeuber *Associate Director, Bureau of the Census*

THE DOORBELL at the Robert Brown house rings as the Brown family is finishing lunch. Nick Brown, the oldest son of the family, goes to the door. "Mom," he calls back, "it's Mrs. Smith, the Census Bureau lady." This is not the Census lady's first visit to the Brown's, so no further introduction is needed. "Well, tell her to come on in," says Mrs. Brown, and Mrs. Smith is invited to join the Browns for coffee.

Courteously, but without wasting time, Mrs. Smith verifies that all members of the Brown family are still living at home and that no one is there to visit or stay. Assured that there are no changes, she begins her questions:

> Mr. Brown, what were you doing most of last week, working or something else? How many hours did you work last week at all jobs? For whom did you work? What kind of business or industry is this? What kind of work were you doing? Mrs. Brown, what were you doing most of last week, keeping house or something

else? Did you do any work at all, not counting work around the house? Did you have a job or business from which you were temporarily absent or on layoff? Have you been looking for work?

Mrs. Smith turns to Nick, who says he's been looking for work:

What have you been doing to find work? Why did you start looking for work? How many weeks have you been looking for work? Have you been looking for full-time or part-time work?

And similar questions are repeated for each member of the Brown household who is 14 years old or older.

Every month, about 50,000 families in the nation are interviewed in this way. Yours may have been one of them, although that is unlikely. Why is this program carried out? To understand this, we turn from the specific interview in the Brown household to the broad national scene.

A newspaper headline in the fall of 1969 announced "U.S. Jobless Rate Advances to 4 Percent, Highest Since '67." The same newspaper went on to observe, "Administration aides greet rise as sign of restraint on boom spurring inflation. Former Vice President [Humphrey] is reported as being critical, but a Treasury official said no change in tax policy is yet under consideration." The reporter had learned that the development was greeted by Administration officials as a welcome sign that their policies, aimed at ending inflation by slowing business expansion, were effective.

The rise in unemployment that month was spread throughout the work force. Unemployment rates increased for almost every category of workers—adult men, adult women, teenagers, Negroes, and Caucasians. The ones most seriously affected, however, were men in the 20 to 24 age group, blue collar workers, construction, trade and manufacturing employees, and agricultural workers.

These facts and many more gave the nation a major indication of the state of our economy in September 1969. Similar figures are available every month and they provide the Congress and the Administration with essential information to help in charting the course of the economy. The figures are widely published. Both the Administration and its critics find in these figures information about the levels of employment and unemployment and the persons who are most directly affected by changes in employment. At a time when there is serious concern with the possibility of inflation and with the consequences of efforts to slow down the rate of economic growth, the statistics on unemployment are carefully watched. Responsible officials want to be sure that the policies are working as desired and that the Government has an early warning if the effects are so great that action is needed to try to avoid dangerous levels of unemployment or recession.

Such statistics, and many others, are essential tools in the nation's efforts

to guide its economy. Information is needed on how many people are working, how many are unemployed, whether the workers are on overtime or on short time, whether the unemployed are primarily married men, single young men or women, and whether unemployment is hitting harder at young or older ages and at black or white. There is a need also for information on how many workers are on temporary layoffs, how many are young persons looking for their first jobs, and how many are women looking for jobs when their husbands have been thrown out of work.

Unemployment may mean many different things, depending on who is unemployed and for how long. The unemployed worker may be simply between jobs, with a short break before he takes on the new one. He may have been laid off because a plant has been shut down or relocated, or simply because a plant has temporarily reduced its work force. He may be a young person testing the labor market, looking for a job for a short time, or trying to decide whether to get a job or to stay in school. She may be a housewife who would like some additional work to help out with the family expenses, to buy some special item, or to finance the schooling of the children. She may be a housewife needing a job in order to supplement a husband's inadequate income or to make up for the fact that he has no income. The unemployed worker may be a man who finds it difficult or impossible to leave his home area where there is no job in order to go to some other location where one may be available. He may be an older worker whose skills have become obsolete and who faces continuing unemployment unless he can be retrained. Among three million unemployed, there are people in many different situations. Because of its concern with the welfare of all Americans and its commitment to a policy of full employment, the U.S. nation has developed a wide variety of measures to help meet the problems of unemployed workers. It needs a continuing source of reliable information on the actual situation at a given time and on the changes from one month and one year to the next.

We now have much more up-to-date information than we did in the early thirties when the nation was deeply concerned with the problems of unemployment. Then everyone knew that there was an unbearably large number of unemployed, but hard information was lacking on how many there were and what types of persons they were. Different authorities made widely differing estimates. No one had a real count; people had to rely on guesses. Though the Administrator of the Federal Emergency Relief Administration could state emphatically that hunger is not debatable, even he did not have a good measure of the size or nature of the problems that he had set out to correct. It was the early efforts of the Works Progress Administration to get some reliable measure of changes in the levels of unemployment that led to the establishment of the survey which now provides the official monthly figures on employment and unemployment.

These figures, of course, do not settle the arguments about how much unemployment is normal, tolerable, or dangerous. Nor do they entirely satisfy the people who feel that a person on part-time work, but who would like full-time employment, should be counted as unemployed. Some think that 14- and 15-year-olds who want work should be counted as unemployed. There can be differences of opinion about the proper classification of persons on temporary layoffs, those who are not working because of some labor dispute, seasonal workers in the off-season, the "discouraged workers" who have given up looking for a job because they are sure none is available, and the persons who aren't sure but say they would like a job if an attractive one comes along. It is sometimes argued that the primary concern should be about married men with dependents and that social economic policy need not be seriously concerned about others who may be unemployed.

Statistics cannot settle such policy questions, but they can supply information that helps in identifying essential issues and in narrowing the debate to the policy issues. Persons advocating different policies can carry on useful discussions or debates only if they are agreed on certain basic propositions. If they can agree on the statistics, they can then match their different interpretations against each other. If, however, they do not agree on what the basic facts are, much of the argument may be fruitless. Even though they may disagree on precisely who should be included and who should not, they can narrow the policy argument if they can agree on how many and what kinds of people are included in the groups about which they disagree.

How do we know about the number of unemployed and whether they are young or old, men or women, black or white, blue collar or white collar workers? How do we know how many of them have just become unemployed and how many have been looking for work for three months or more? How do we know that unemployment in one month is more or less than it was in the preceding month or in the same month a year ago? There are important seasonal changes in some kinds of work—the harvesting of crops, canning, retail selling with its large need for temporary help before Christmas and Easter, and many others. We need to take all of these seasonal changes into account in assessing whether a change from one month to the next is really significant or whether it is simply a reflection of the normal seasonal development.

THE CURRENT POPULATION SURVEY

The basic source of such information, both in this country and in some others, is the Current Population Survey (CPS). In the U.S., every month some 1100 persons leave their housework or other duties and call at some 60,000 addresses to ask the occupants a number of carefully worded questions. The

answers (usually there are some 50,000 of them) are promptly assembled, and from them, the nation has its monthly report on employment and unemployment.

Every month, in the week that includes the 19th, these interviewers begin their rounds. Each one has about 60 addresses on his or her list. The questionnaires are so designed that many of the answers can be recorded simply by filling in little circles. Some of the answers, however, must be written out, for example, the kind of work that each worker was doing. The interviews begin on Monday and always relate to activity during the previous week, that is, the week that includes the 12th of the month. Completed questionnaires go to a regional office of the Census Bureau, where they are given a quick review, and then promptly shipped to the Census Bureau's processing office in Jeffersonville, Indiana. There trained persons translate the answers from ordinary language into a language that the computer can read. If the interviewing begins on the 15th of the month, all of it is completed by the 22nd, and the computer receives all of the forms by the 29th. Overnight it does the work of combining the data into usable tables. Approximately two weeks after that month's enumeration began, the first statistics are available. They need to be reviewed to be sure all of the processing, including that done by the computer, has been done in accordance with the instructions. Within two days more, the results are ready for distribution to newspapers, radio, and television, which pass on the information to the public. Analysts in the Government and outside then have a new set of figures to analyze in order to arrive at proper policy decisions in the light of the changes revealed that month. This process is repeated month after month, as it has been in essentially similar form for more than 25 years.

Anyone using such statistics is likely to ask how reasonable it is to draw conclusions about all of the 60 million households in the country from a sample of a little more than 50 thousand (about 0.1%). To answer this natural question, we must look at the method by which the sample of households is drawn. If that method is highly restrictive (for example, if only families in major industrial cities were selected or only farm families), then it would obviously be inappropriate to reach conclusions about all the country from the sample. Similarly, if families were chosen by asking each interviewer to interview 50 families who are his friends or neighbors, we might properly doubt inferences from the sample. We want to be sure that the sample has appropriate numbers of rural and urban families, of high- and low-income families, and so on. All parts of the country should be represented in proportion to their share of the population.

The major assurance that the sample will reflect conditions for the entire country comes from the fact that the households are probabilistically selected from within homogeneous strata. Not only does this help to keep the process free from personal bias on the part of the persons choosing the sample, but

it also permits accurate knowledge of the precision of the estimates that come from the sampling. This also means that it is possible to determine the sample size that will assure the degree of precision in the data needed to meet the requirements of public policy. The details of how the sample is selected are described below.

There are more than 60 million households in the U.S., and the problem is how to select an appropriate sample each month of about one out of every 1240 for interview. If there were an up-to-date list of all households in the U.S., one out of every 1240 of them might be selected in an appropriate manner, but there is no such list.

Because the information is to be collected by interviewers visiting the households, it is desirable to select the sample in clusters of neighboring households. This helps the interviewer get to all of them within the few days each month in which the interviewing must be done. In this survey, clusters of about six households are used.

Another requirement is that the sample should be selected in such a way that it is possible to estimate from the sample itself how much the results differ from those which would have been secured if every household in the country had been included. It is well known that the results from a sample are rarely precisely the same as those from a complete enumeration, but from a properly designed sample, we can measure the chance that these deviations are small enough so that the major findings from the sample can be used with confidence. The same type of consideration applies to the comparison of changes from month to month. For them, too, it is necessary to know whether a difference is real or is simply the kind of chance difference that could be the result of "sampling error."

The problem for the statisticians was to develop a way of drawing the sample from the entire country in such a way that an economical and reliable survey could be conducted each month. A measure of the degree of confidence to put in the figures also should be obtained from the sample.

It was decided to give each interviewer a fixed set of addresses at which to call. If no one is living at one of the addresses or if the persons found there do not actually live there, that address will not contribute data to the survey for that month. If a household moves away from an address between two of the monthly visits, it is dropped from the survey and the household which has moved to the sample address is included. The interviewer is responsible for completing an interview at each of the assigned addresses, or explaining why no interview was required or possible. If no one is found at home after several calls or if the residents at the assigned address refused to provide information, it is so noted. No substitutions for sample households are allowed.

The Current Population Survey, which is the source of this information, is carried out in 449 sample areas, which include 863 counties and independent cities. They are located in every state and the District of Columbia. In all, about 60,000 residential addresses are designated for the sample each month; about 52,000 of them, containing about 105,000 persons 14 years of age and over, are actually interviewed. The survey is limited to the civilian population and excludes members of the Armed Forces, as well as persons living in institutions, such as prisons, long-stay hospitals, and so on. Most of the 8000 addresses for which interviews are not secured are for housing units that are vacant, units that have been converted to nonresidential use, units whose usual occupants are temporarily away, or those that are temporarily occupied by persons who actually live somewhere else (for example, persons occupying a home on their vacation). Answering the questions is voluntary, but fewer than 2% of the persons interviewed refuse to answer.

The first step in selecting the households is to select a sample of the counties or equivalent governmental units. These are then subdivided into subareas, and a selection is made among these subareas. Within each of the selected subareas, a sample of addresses is selected for interviewing.[1] In making the final selection of the specific addresses to be used, two different procedures are utilized. In urban areas it is generally possible to work with addresses that give specific house numbers and streets and even apartment numbers. When such lists are available, a cluster of about 18 consecutive addresses is selected from the census enumeration districts (ED). Every third address within the cluster is taken for the current sample; the remaining 12 addresses are saved for use in future samples. Arrangements are made to take into account new construction since the last census, chiefly by checking building permits.

In rural areas and other areas where such addresses are not available, *area sampling* is used. The sample EDs are subdivided into small land areas with well-defined boundaries. Insofar as can be determined from available information, each such area segment has about six housing units. If it is not possible to define area segments of that size, larger segments may be defined. These six addresses are drawn by a systematic sampling of all housing units.

It is desirable to have a household in the sample for consecutive months, and for the same months in successive years in order to secure measures of month-to-month and year-to-year change. To avoid overburdening the households who cooperate in the interview, it has been arranged that interviews are conducted at a sample address for four successive months, then that address is omitted for eight months, and after that, it is interviewed again for four consecutive months. This rotation occurs in such a way that each month one-eighth of the sample addresses are entirely new, one-eighth consists of addresses that are starting on the second round of four interviews, and three-

[1] See the appendix to this essay for a more detailed description of the sampling procedure.

fourths were interviewed in the preceding month. Thus one-half also were interviewed in the same month a year before.

PRECISION AND ERROR CONTROL

Modern sampling theory makes it possible to measure the size of the fluctuations arising from the sampling process, for the probability of including any unit in the sample is known. Of course, as in any survey, there are also other, nonsampling sources of error that must be investigated and controlled. Interviewers must be carefully trained both in the techniques of interviewing and in the content of the questionnaire they are using. Every effort must be made to assure that respondents understand the questions and provide correct answers. Controlling such a series of interviews requires attention to all possible sources of error, and taking appropriate steps to control them. In the case of the Current Population Survey, there is continuing training of the interviewers, careful review of their work each month, periodic observation, and a program of reinterviews by supervisory personnel.

On the average of twice a year, a subsample of the addresses assigned to each interviewer is visited a second time by a supervisor and the same questions are asked again to make sure that the correct information has been obtained. The interviewers do not know when their work will be checked or which addresses will be selected for the reinterview. The supervisors do not know at which addresses they are to reinterview until after the initial interviews have been completed. If the information secured at the two interviews differs, an effort is made to find out which of the two answers is correct and why the differences occurred. The reinterview program serves as a basis for further training of the interviewers and gives a measure of the quality of the survey in general.

All steps in the office processing of the interviews are similarly kept under constant control. The preparation of the estimates in which the results of the sample are projected to the entire population also requires application of modern statistical principles. The results are published with a measure of the sampling variability of each of the major figures.

QUESTIONS ASKED

Selecting the sample properly and training and supervising the interviewers carefully to make sure that they interview at the designated addresses and report the replies accurately would not be adequate if the questionnaire itself were not well designed or if the questions to be asked were left to the interpretation of each interviewer or each respondent. A question such as "Were you unemployed?" would give results of little value, for there would be wide differences in the interpretations placed on such a question. Instead, over the years a battery of questions has been developed to secure information

on what a person actually did during the survey week and to classify him or her on that basis. If he reported that he was working, he is asked how many hours he worked during the week, and because people are likely to report some standard number such as 40, he is also asked whether he worked any overtime or lost any time or took off any time. He is asked also to state what his job was and for whom he worked, as well as the kind of business or industry in which he worked.

If he did not work during the survey week (i.e., the week preceding the interview) he is asked whether he had a job or business from which he was temporarily absent or on layoff. If he was absent, he is asked why and whether he was paid for the time off. If he was not working, but looking for work, he is asked to indicate what he has done during the preceding four weeks to find work, such as checking with an employment agency, with employers, or with friends and relatives, placing or answering advertisements, and so on. He is also asked why he started looking for work, whether because he lost his job, quit his job, left school, or wanted temporary work. There are questions to ascertain how long he has been looking for work, whether he was looking for full-time or part-time work, and whether there was any reason, such as illness or school attendance, why he could not have taken a job last week. There are questions on when he last worked for pay, and the kind of work he did on his last job.

If he is not looking for work, he is asked when he last worked for pay, why he left his last job, and whether he intends to look for work of any kind in the next 12 months. He is also asked why he is not looking for work, whether because he believes that no work is available in his line of work or in the area in which he lives, that he lacks the necessary training or skills, that employers think he is too old or too young, or that there is some personal handicap that stands in his way. Other possible reasons include family responsibilities, the inability to arrange for child care, ill health, a physical disability, or the fact that he is going to school. The answers to these questions provide a basis for meaningful classification of persons as employed or unemployed. Information on the respondent's sex, age, color or race, marital status, relationship to the head of the household, and years of schooling is also available to assist in providing some meaningful classification of persons as employed, unemployed, or not in the labor force.

The published results provide a monthly measure of national employment and unemployment. From time to time a few questions are added to the questionnaire. These special questions, in addition to those that are asked every month, provide the basis for much of the information about important social trends in the years between the major censuses.

Because the sample continually reflects the growth or decline of population in the sample areas, it provides a basis for estimating the major geographic shifts of the population. Between 1960 and 1968 the total population of the central cities of our metropolitan areas remained almost the same, but

there was a major change in the makeup of the population of these cities, including a net loss of about 2.1 million Caucasians and a gain of about 2.6 million Negroes. The continuing movement of people to the suburbs is clearly reflected in the statistics from this survey, with about four-fifths of the national growth occurring in the suburban areas. The survey has also reflected a slowing down of interregional migration during the sixties, compared to the higher rates of migration in the fifties.

The annual report on the number of families and persons in poverty comes from this same survey, for once a year each person in the survey is asked to report the amount of income he received during the preceding year. The number of persons living in poverty declined from nearly 40 million in 1959 to about 25 million in 1968. The percentage of the population living in poverty declined from 22 in 1959 to 13 in 1968. There were differences in regard to income and poverty between Caucasians and Negroes, as shown by these figures. In 1959, more than half the Negroes were classified as living in poverty; in 1968, the comparable figure was 35%. Average family income has been increasing. In 1968, it amounted to $8600, which represented a real gain of about 3.5% over the previous year, even after allowing for the increased level of prices.

An important measure of educational attainment is the proportion of young adults who have completed high school. Between 1960 and 1969 the proportion of white men 25 to 29 years old who had completed at least four years of high school rose from 63 to 78%. For black men in the same age bracket, the percentage went up from 36 to 60% in the same nine years.

Americans are a mobile people, and the survey provides a measure of that mobility. One person in five changes his residence in the course of a year; most of these moves are within the same county. There has been very little change in this rate of mobility over the last 20 years for which information is available.

The survey also indicates that in homes with children that include only one parent (usually the mother) the family income is likely to be relatively low. In 1969, 89% of white children and 59% of black children were living with both parents. In recent years, there has been little change in this percentage among whites, but some decrease among blacks.

Such facts about life in the U.S. are available annually because the families included in the survey are willing to answer the questions put to them by the interviewers. A census covering the entire population is taken only once in 10 years. Between censuses, we now have statistics that reveal changes for the nation as a whole and for the major regions. A survey of this size, however, cannot provide figures of the same reliability for states, individual cities, or for metropolitan or smaller areas.

From time to time, the Bureau of the Census is asked to add some other questions to the questionnaire. As a result, it has been able to supply statistics

on the number of persons who smoke, and the proportions of young men and women and older men and women who do so, as well as the numbers who quit smoking. The most recent survey showed that about 2.7 million persons had quit smoking between 1966 and 1968. Some persons began smoking during that time, but the total number reporting that they smoke dropped by about 1 million. Information has also been supplied on the number and proportion of persons who have had immunization for polio, smallpox, diptheria, and other communicable diseases.

In view of the public interest in the percentage of persons who vote, questions have also been asked about whether the individual had voted and, if not, whether he had been registered to vote. The survey found shortly after the 1968 election, that 68% of all persons of voting age reported that they had voted. Men had higher voting participation rates than women. Northern blacks had a higher voting rate than Southern whites. Persons over 65 and those under 35 had lower voting participation rates than persons between 35 and 64; persons with higher educational levels and those with higher incomes tended to have higher voting participation rates. Unskilled workers had lower voting participation rates than persons in the occupations that required higher levels of training.

THE ROLE OF THE SURVEY

This survey, which involves the willing cooperation of more than 50,000 households each month, has proven to be an important source of information about the U.S. This information is needed by the Government in planning its economic and other policies. It is also needed by many other agencies and by private organizations concerned with employment and unemployment, educational levels, poverty, incomes, health, and living conditions generally. Through the application of modern sampling theories and through a system of carefully developed training and supervision, it has become possible to provide this and much other information on a timely basis and at a fraction of the cost of conducting a complete census. Congress and other policy makers and administrators look to this source for up-to-date information, knowing that the results are reliable.

Such a survey is not a substitute for a census, which provides information for each state, county, and city, and for smaller areas, but it does provide essential information between censuses and is an indispensable source of data needed for a continuing appraisal of important developments in the nation.

APPENDIX: PROCEDURE FOR SELECTING SAMPLES FOR THE CURRENT POPULATION SURVEY

The first step in selecting the addresses to be visited was to determine the counties in which the interviewing was to be done. This was accomplished

by combining all of the counties in the U.S. into Primary Sampling Units (PSUs). Each of the Standard Metropolitan Statistical Areas (SMSA) was taken as a PSU. An SMSA is a city of 50,000 or over, plus the county in which it is located and adjoining counties that are closely tied to it economically and socially as determined by certain specified criteria. Outside the SMSAs, counties were grouped into PSUs (small groups of counties that are sufficiently compact so that a sample of households within a unit could be visited without undue travel cost). Whenever possible, a PSU was constructed to include both urban and rural residents of both high and low economic levels and, to the extent feasible, a variety of occupations and industries. When the current sample was selected in 1962, there were 1913 PSUs, including the 212 SMSAs defined in the 1960 census.

The next step was to combine these PSUs into 357 strata. Each of the SMSAs with 250,000 or more residents was treated as a single stratum. The other strata, in general, consisted of sets of PSUs as much alike as possible in various characteristics. These included geographic region, density of population, rate of growth between 1950 and 1960, the proportion of the population in 1960 that was not white, the principal industry, type of agriculture, and so on. Except for the strata in which an area represented itself (such as the larger SMSAs), the strata were so arranged that their populations in 1960 were approximately equal.

If a PSU was a stratum by itself, it automatically fell into the sample, and thus there were 112 strata that consisted of only one PSU. Within the other 245 strata, PSUs were selected for this sample at random in such a way that the probability of each being drawn was directly proportional to its 1960 population. Thus, within a stratum, the chance that a PSU with 100,000 persons would be drawn was twice as great as that of a PSU with only 50,000 persons.

The next step was to select the sample households within the designated PSU. The sampling rate within each PSU was determined in such a way that the overall chance of an address being included is equal to one in 1240. In this way, it is possible to reflect changes, such as new construction or demolition of housing units or changes in the characteristics of the population of the area that have occurred since the last census was taken. Within each designated PSU, the first step was to select a sample of the census enumeration districts (ED) that were used in 1960. These small administrative units contained approximately 250 households in the 1960 census. The EDs were arranged in geographical order to make sure that the sample EDs will be spread over the entire PSU. The probability of selection of any one ED is proportionate to its 1960 population.