Data Collection: A Statement for Administrators (1981)

Pages
35

Size
5 x 9

ISBN
0309330564

Panel on Data Collection; Committee on Population and Demography; Assembly of Behavioral and Social Sciences; National Research Council

COMMITTEE ON
POPULATION AND DEMOGRAPHY Report No. 7

Data Collection: A Statement for Administrators

Panel on Data Collection
Committee on Population and Demography
Assembly of Behavioral and Social Sciences
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1981

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the Committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the federal government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

Printed in the United States of America

PANEL ON DATA COLLECTION

WILLIAM SELTZER (Chair), Statistical Office, United Nations, New York
LESLIE KISH, Institute for Social Research, University of Michigan
FORREST E. LINDER, International Institute for Vital Registration and Statistics, Bethesda, Maryland
SAMUEL H. PRESTON, Population Studies Center, University of Pennsylvania
CHRISTOPHER SCOTT, World Fertility Survey, London
CONRAD TAEUBER, Kennedy Institute, Center for Population Research, Georgetown University
JACQUES VALLIN, Institut National d'Etudes Demographiques, Paris
H. BRADLEY WELLS, Population Laboratories, University of North Carolina (currently, Oncology Research Center, Bowman Grey School of Medicine)

ROBERT J. LAPHAM, Study Director

COMMITTEE ON POPULATION AND DEMOGRAPHY

ANSLEY J. COALE (Chair), Office of Population Research, Princeton University
WILLIAM BRASS, Centre for Population Studies, London School of Hygiene and Tropical Medicine
LEE-JAY CHO, East-West Population Institute, East-West Center, Honolulu
RONALD FREEDMAN, Population Studies Center, University of Michigan
NATHAN KEYFITZ, Department of Sociology, Harvard University
LESLIE KISH, Institute for Social Research, University of Michigan
W. PARKER MAULDIN, Population Division, The Rockefeller Foundation, New York
JANE MENKEN, Office of Population Research, Princeton University
SAMUEL PRESTON, Population Studies Center, University of Pennsylvania
WILLIAM SELTZER, Statistical Office, United Nations
CONRAD TAEUBER, Kennedy Institute, Center for Population Research, Georgetown University
ETIENNE VAN DE WALLE, Population Studies Center, University of Pennsylvania

ROBERT J. LAPHAM, Study Director

NOTE: Members of the Committee and its panels and working groups participated in this project in their individual capacities; the listing of their organizational affiliation is for identification purposes only, and the views and designations used in this report are not necessarily those of the organizations mentioned.

CONTENTS

PREFACE

The Committee on Population and Demography was established in April 1977 by the National Research Council, in response to a request by the Agency for International Development (AID) of the U.S. Department of State. It was widely felt by those concerned that the time was ripe for a detailed review of levels and trends of fertility and mortality in the developing world, and for a review of the ways fertility and mortality data are collected and analyzed. Although most people in the demographic community agree that mortality has declined in almost all developing countries during the last 30 years, there is uncertainty about more recent changes in mortality in some countries, about current levels of fertility, and about the existence and extent of recent changes in fertility. Much of the uncertainty is associated with the amount and nature of the data collected to measure fertility and mortality.

The causes of reductions in fertility in low-income areas--whether they are the effect primarily of such general changes as lowered infant mortality, increasing education, urban rather than rural residence, and improving status of women or of such particular changes as spreading knowledge of and access to efficient methods of contraception or abortion--are strongly debated. There are also divergent views of the appropriate national and international policies on population in the face of these changing trends. The differences in opinion extend to different beliefs and assertions about what the population trends really are in many of the less-developed countries. Because births and deaths are recorded very incompletely in much of Africa, Asia, and Latin America, levels and trends of fertility and mortality must be estimated, and

disagreement has arisen in some instances about the most reliable estimates of those levels and trends.

The Committee on Population and Demography was established within the Assembly of Behavioral and Social Sciences of the National Research Council, and funded by AID under Contract No. AID/pha-C-1161. The Committee has undertaken three major tasks:

1. To evaluate available evidence and prepare estimates of levels and trends of fertility and mortality in selected developing nations;
2. To improve the technologies for estimating fertility and mortality when only incomplete or inadequate data exist (including techniques of data collection);
3. To evaluate the factors determining the changes in birth rates in less-developed nations.

The Committee approaches the first task through careful assessment, by internal and external comparison, and through analysis, by application of the most reliable methods known, of all the data sources available. Each of the country studies therefore consists of the application of a range of methods to a number of data sets. Estimates of levels and recent trends judged to be the best that are feasible with available resources are then developed on the grounds of their consistency and plausibility and the robustness of the individual methods from which they were derived. For several studies a panel or working group of experts was established, including nationals of the country and others knowledgeable about the demography and demographic statistics of the country.

The Committee's second task, refinement of methodology, is seen as a by-product of achieving the first. The application of particular methods to many different data sets from different countries and referring to different time periods will inevitably provide valuable information about the practical functioning of the methods themselves. Particular data sets might also require the development of new methodology or the refinement of existing techniques.

As of early 1981, 162 population specialists, including 89 from developing countries, have been involved in the work of the Committee as members of panels or working groups. The Committee, the Assembly, and the National Research Council are grateful for the unpaid time and effort these experts have been willing to give. A small staff assists the Committee panels and working groups.

As part of the second major task and given the importance of data collection in estimating fertility and mortality, the Committee established the Panel on Data Collection, chaired by William Seltzer, a member of the Committee. The specific objective of the Panel on Data Collection is to assist the Committee on Population and Demography in developing recommendations for improving and standardizing techniques for estimating fertility and mortality levels, trends, and patterns in developing nations. To this end, the panel has focused on methods of improving the accuracy, coverage, timeliness, and reliability of such estimates through improvements in the collection of underlying data. As indicated in this report, the methods may range from improved data collection procedures that can be applied immediately to long-term processes that take many years to bring into fruitful operation.

This report is No. 7 in series produced by the Committee on Population and Demography of the National Research Council, and it is the third report by the Panel on Data Collection. The first Panel report was a study of one of the major problems related to the collection of data for fertility and mortality estimation, that of age misreporting and age-selective underenumeration. The study was prepared by Douglas C. Ewbank and published as Report No. 4 of the Committee on Population and Demography. The second product of the Panel on Data Collection reviews experience on collecting data using population censuses, civil registration and vital statistics systems, and sample surveys.

Collecting Data for the Estimation of Fertility and Mortality (National Research Council, 1981) published as Report No. 6 of the Committee, is a state-of-the-art report on available procedures that focuses on questions of special relevance to the estimation of fertility and mortality. It is intended to provide detailed guidance and information for people who are concerned about the collection of data for demographic measurement in developing nations. These people include administrators and, especially, technicians and survey practitioners in government and international statistical agencies, teachers and researchers in universities and demographic training centers, policy makers who have to decide on the allocation of funds and personnel for data collection activities, and data users who wish to know more about the difficulties, opportunities, and consequences associated with various data collection processes used to

DATA COLLECTION:
A STATEMENT FOR ADMINISTRATORS

INTRODUCTION

Reliable information about fertility and mortality is essential to any planning for social and economic development. The relationship between births and deaths determines the rate of change in the size of the population and fertility is the major influence on the age composition of the population, both of which have wide-ranging implications for any society. Sharp changes in birth or death rates are reflected in the size and age distribution of the population for many years.

Much time and effort has been devoted to finding ways that governments and the administrators of development programs can collect the needed information about the population for whom such programs are planned and carried out. Three basic methods are available: (1) a census of the population, which provides a count of all the people in the specified area at a specified time; (2) a civil registration system, which seeks to record births and deaths as they occur; and (3) sample surveys, which collect the relevant information from a sample of the population. The three methods are complementary, each providing some information that is not available from the others. (A more detailed discussion of the procedures is found in National Research Council (1981).)

In evaluating the advantages, limitations, and costs of these data collection methods, it is useful to keep in mind a number of general considerations. First, different users of data on fertility and mortality have quite different needs in terms of the degree of detail, the timeliness, and the accuracy of the data. Second, it is important to know how accurate and how complete the data are in order to assess the degree of confidence one may

have in conclusions drawn from them. Third, while each of the procedures produces useful information, no one by itself can satisfy all needs for information about the levels and trends of fertility and mortality. Fourth, all the procedures require careful planning and pretesting. Fifth, each procedure calls for a certain level of technical competence if it is to produce accurate and usable results at reasonable cost.

METHODS OF DATA COLLECTION

Population Censuses

The United Nations defines a population census as the total process of collecting, compiling, evaluating, analyzing, and publishing or otherwise disseminating demographic and other data pertaining at a specified time to all persons in a country or a well delimited part of a country. Although a census is in effect a snapshot of a population at a given time, it also provides a basis for inferences concerning past and future changes. Comparisons of censuses taken at different times provide information about the growth or decline of the total population or groups within the population, as well as changes in subareas of the country. Censuses tend to serve many purposes and provide information needed by a wide range of government agencies.

In planning for social and economic development, the first thing the administrator of a development program must ascertain is how many people there are in the development area and what changes are occurring or are likely to occur: How many of the people are children? How many are of working age? How many are older persons? For some purposes it is essential to know how many are men and how many are women, and for the latter, how many are in prime childbearing years. It makes a difference if the people are evenly distributed throughout the area, or if they are clustered in one region or in villages and small cities. In some countries there are important differences between ethnic groups; they may occupy different regions of the country, engage in different occupations, and have different rates of population growth. Often the degree of literacy in the population is an important element in planning. The number of children of school age and preschool age must be known to be able to project accurately the need for teachers and school rooms. The

levels of infant and child mortality are important indicators of health and the need for health services.

In developing countries, census data can be particularly useful for estimating fertility and mortality because a number of indirect estimation techniques, which have been developed only recently, are particularly suitable to census data. These analytical methods take advantage of the interrelationships of demographic events to estimate fertility and mortality when direct measures are not available or are incomplete or unreliable. In many cases direct questions about how many births or deaths have occurred during the past 12 months have been found to yield inaccurate, misleading estimates.

Questions that can provide indirect measures of current or past fertility have been included in many censuses in developing countries. One set of questions includes asking women about the number of live-born children they have ever had, the number of children still living at the time of the interview, and whether those children are living at home or elsewhere. Where consensual unions are common, they are usually treated in the same way as marriages that have been recognized legally.

For a number of reasons, censuses are generally less effective in providing data on mortality than on fertility. Death is often considered a topic to be avoided, which makes it difficult to gather information. In addition, death often leads to the breakup of a household, with the survivors leaving the original household for some other location or being dispersed to other households. For some developing countries, comparison of the reported age-sex structure of the population with model stable populations has provided better estimates of mortality than have direct questions about deaths in the past year or other recent period. Censuses can also obtain other types of data that permit indirect estimates of mortality: of child mortality by questioning mothers about the survival of their children; of adult mortality by asking people about the survival of their parents or of their first spouse.

Although this document focuses on the collection of data to estimate fertility and mortality, it must be noted that censuses collect useful information on a number of other topics, notably migration and current residence, and characteristics of individuals, such as educational attainment. A census can provide information on who was where at a given time. In places where problems of internal mobility--such as movement to cities or

industrial areas--are of great concern, census or sample surveys are necessary to generate the required information.

Civil Registration Systems

A major source of information about births and deaths is a civil registration system that records vital events (also known sometimes as a vital statistics system, CR/VS). A major purpose of a civil registration system is to issue certificates on vital events. In addition, the United Nations defines a registration system as the total process of (a) collecting, by registration, enumeration, or indirect estimation, of information on the frequency of occurrence of births, deaths, marriages, and divorces, as well as relevant characteristics of the events themselves and of the persons concerned, and (b) compiling, analyzing, evaluating, presenting, and disseminating these data in statistical form.

Registration systems differ from censuses and sample surveys in a number of important ways. They record the occurrence of certain events during a specified period of time and do so on a continuing basis. The primary purpose of registration is to establish legally certain facts about an individual, i.e., that he or she was born at a specified place on a specified date or that death occurred at a specified time and place. In many areas, the registration of a death is a prerequisite to the issuance of a burial permit and legal disposition of the corpse. If the registration certificates carry information about the cause of death, they may also be used to provide information about the state of public health and the increase or decline of certain diseases, including those that reach epidemic proportions.

The recording of the events normally provides certain benefits to the individual or to society, such as establishing that the individual meets minimum requirements for such diverse matters as entering school, accepting employment, marrying, voting, being liable for military duty, being eligible for assistance available to senior citizens, and other rights and duties.

Although the primary purpose of a registration system generally is to meet juridical needs, it is also a source of current data on births and deaths, which can be made available as often as monthly if necessary. If the system operates throughout the country, it can provide data

on both a national and regional basis. Some countries have found it desirable to establish a registration system gradually, that is, making it fully effective in one province or district before moving to establish it in another area. India has found it expedient to establish a fully operational system in a set of sample areas and use the reports from the sample areas as the basis for national estimates of births and deaths.

It is not necessary that the registration of vital events be complete in order to provide usable estimates of fertility and mortality. Indirect estimation procedures have been developed that enable analysts to adjust some sets of faulty and incomplete registration data, and to make reasonably good estimates. For example, techniques developed in recent years make it possible in many cases to estimate adult mortality (mortality of those aged 10 and older) even when only 60 or 70 percent of the deaths to persons over age 10 are registered.

Sample Surveys

There are two main types of surveys: those conducted on an occasional or ad hoc basis, typically used to gather special information about a particular topic or region; and those conducted as part of a program of surveys or carried out in conjunction with another activity such as a census. (The distinction is not absolute: for example, an ad hoc survey may take advantage of a recent census for recruitment of personnel and sample design, but have no other relationship to the census.) A sample survey conducted in connection with a census may be done at the time of the census enumeration, by asking a sample of the total population to respond to additional, more detailed questions; or a sample of households or of the population may be drawn from the census universe for additional separate questioning, usually at some time after the census. Often such a survey is intended primarily to provide information on the quality of the census (as in the case of a post-enumeration survey that can be used to estimate census coverage); but it may be a separate activity, simply scheduled to take advantage of the proximity in time to the full-scale census operation. Surveys as part of a continuing program also include the "current population survey" type in which periodic surveys (e.g., quarterly or annually) are carried out with rotating

samples or panels, sometimes with different questions included in each survey round.

The sample survey normally allows more time for each interview than is possible during a census so that any single topic can be explored in greater depth. In a survey done independently of the census there is even more latitude for detailed questioning. However, in any survey there are limits to the amount of time that can be devoted to a single interview. The patience of respondents imposes some limits, and cost considerations and the major aims of the survey may impose others.

SOME COMPARISONS OF THE METHODS

The three major methods of collecting data should be viewed as complementary. The industrially developed countries generally use all three, since they serve somewhat different but related purposes. In a country that has little or no data on the level or trends of births and deaths, it may be necessary initially to choose one method in order to begin collecting at least some of the needed information. An understanding of the strengths and limitations of each method is the key to determining which should be given priority if all cannot be implemented at once and which is best suited to collecting a given type of information. However, recognizing that information on fertility and mortality is a long-term interest, it would be best to plan for a national program of data collection involving all three methods. Table 1 classifies the three major data collection methods in terms of their suitability for providing data to estimate fertility and mortality. Table 2 lists some advantages and disadvantages of each of the three methods. The sections that follow describe briefly how the methods compare in terms of administrative demands, the area detail and timeliness of the data they provide, the variety and depth of subjects they can cover, the response rates and errors to which they are subject, and their costs.

Administration

A census normally is administered by an arm of the central government. It requires a large work force for a relatively short period of time. Enumerators who collect the information must reach every village and every home throughout the country. Since the census relates to a specific point in time, often a single day, it is essential that there be enough enumerators to reach each potential respondent within a short time period. Whether the recruitment and training of the enumerators is done entirely by the central government or by provincial or local authorities depends on the relationships that exist among the various governmental units.

Once the census questionnaires have been collected, they normally are transferred to a central unit for processing. The work force to do that processing must be recruited and trained by the central government, which is also responsible for publishing the results. Processing may also be done on a provincial basis. In such cases, the central government is responsible for assuring coordination.

Most countries now have electronic computers that can tabulate the data more quickly than was possible previously. However, if the computer system is not under the direct control of the central statistical office efficient operation requires that there be clear agreements on the priority to be given the census work. The use of computers also requires adequate technical staff to do the necessary programming and to monitor all aspects of tabulation and publication. A number of "canned" tabulation and analysis programs are now available from the United Nations and other agencies, such as the U.S. Bureau of the Census. While they simplify the programming tasks, they do not eliminate the need for technically trained personnel who can monitor the operation and make the modifications needed to tailor the programs to the local situation.

The administration of a civil registration system requires that registrars be stationed in every area of the country and be readily accessible to the persons who are to report the events. Normally registration is a part-time function for people who have other duties that occupy most of their time. A registrar may be a local official with other governmental duties or a person who performs only this duty for the local government but depends on private activities for a livelihood. Finding

TABLE 1 Intrinsic Characteristics of Data Collection Methods in Providing Data Needed to Estimate Fertility and Mortality: A Comparison Using Seven Criteria

Criteria	Data Collection Method		
	Census	Civil Registration	Sample Survey
Topical detail (richness and diversity of subject matter)	Moderate	Weak	Strong
Accuracy	Moderate	Strong	Moderate
Precision (absence of sampling errors)	Strong[a]	Strong[a]	Weak
Timeliness of data	Weak	Strong	Strong
Geographic detail (subgroups, etc.)	Strong	Strong	Weak
Obtaining information on population at risk[b]	Strong	[c]	Strong
Ease of organization in a developing nation	Moderate	Weak	Strong

[a]An important qualification must be noted. With respect to arriving at inferences, censuses and CR/VS systems are subject to sampling errors. For example, if one has a death rate for a city or county (based on complete registrations and a complete census) or for a specific cause, based on a small sample, and it differs from the death rate in another area (or for another cause), the difference may well be due to the number of observations involved and may not represent any real underlying difference in the cause systems. Thus, when the inference is to a cause system, as is common, census and complete registration results are subject ot sampling errors.

[b]The "population at risk" refers to the group of persons who are subject to the events that are counted, measured, or analyzed. For example, the population at risk for the crude death rate is the entire population in the area under measurement or study. The population at risk for each age-specific fertility rate consists of all women in that age group. The population at risk of infant mortality includes all live-born children during their first year of life. (The infant mortality rate is frequently approximated by dividing the deaths to children under one year of age during a year by the number of births occurring in that year.)

[c]In general, CR/VS systems do not provide information on the population at risk. However, for some measures, such as infant mortality, CR/VS systems do provide data on the population at risk. Also, historical analysis is possible when CR/VS data from earlier periods are available.

Table 2 Some Advantages and Limitations of Population Censuses, Civil Registration Systems, and Sample Surveys

Data Collection Method	Advantages	Limitations
Population census	1. Data can be tabulated for many local geographic areas. 2. Detailed cross-tabulations are not subject to sampling errors for complete-count items (except when arriving at inferences--see note a in Table 1) and are subject to relatively low sampling errors for sample items. 3. Simultaneously obtains information related to enumerated events and population at risk. 4. Useful for time series covering long periods of time.	1. Infrequent. 2. Limited range and depth possible in the collection of data on fertility and mortality as well as on classifying variables. 3. Information on "flow" variables (for example, income, births, deaths) and data from proxy respondents are subject to increased levels of response error. 4. Persons not at their usual place of residence are subject to high non-response rates (a lesser problem in de facto censuses). 5. Comparatively difficult to control conditions of observation (because it is extensive in space). 6. Costly and massive in scale, so relatively inflexible.
Civil registration system	1. Data can be tabulated for many local geographic areas. 2. Detailed cross-classifications often not subject to sampling error (see note a in Table 1).	1. Need for separate estimates of population at risk. 2. Limited range and depth possible in the collection of data on classifying variables.

Civil registration system	3. If properly functioning, provides contemporaneous reporting for substantially all events regardless of household status. 4. Institutional continuity. 5. Well-suited for providing both long-term and short-term time series.	3. Relatively inflexible to changes in content and procedures. 4. Very difficult to administer and supervise (because extensive in both time and space). 5. Difficult to establish occurrence of events when births and deaths (or knowledge of them) are not associated with individuals who can serve as informants (for example, health workers or religious personnel).
Sample survey	1. Simultaneously obtains information related to enumerated events and population at risk (with the exception of surveys conducted as parts of dual-record systems). 2. Topical flexibility (that is, the depth and range of topics investigated can be altered relatively easily). 3. Conditions of observation are subject to control in a well-designed and administered survey because of the limited geographic scope of collection operations (that is, because a sample is used). 4. Relatively easy to initiate given availability of a survey-taking infrastructure. 5. Can be useful for time-series analysis, given comparability in data collected.	1. Inability to produce estimates for local areas. 2. Detailed cross-classifications are subject to large sampling errors. 3. Information on "flow" variables (for example, income, births, deaths) and data from proxy respondents are subject to increased levels of response error. 4. Coverage for the nonhousehold population is very poor and it varies markedly for those who are not members of a primary family (for example, members of secondary families, secondary individuals, and distant relatives of the household head). 5. Comparisons over time of estimates based on different ad hoc surveys are subject to many uncertainties. 6. Requires close supervision of field work.

a qualified person for this task in every locale may be difficult. If the registration system places too heavy a burden on respondents (for example, if they must travel long distances on foot to find the registrar), then some events are likely to go unreported. Likewise, requiring respondents to pay a fee to register an event can contribute to incomplete registration.

If a registration system is to work effectively, the public must be convinced of the need for full and complete registration, and the local registrars must be employed on a continuous basis. Adequate recognition and payment of the registrars is one way of motivating them to do their assigned task and to transmit reports promptly. Laws or regulations that confer certain benefits through registration are also very important to the functioning of the system. For example, requiring that a child produce a birth certificate before being allowed to enter school can be a strong incentive to the parents to make sure the birth is registered.

A system of registration of vital events requires careful coordination of the several agencies that are usually involved. Normally the local registrars report to a district or provincial office, which in turn reports to a higher level or to the central office. In addition to the several levels of government that become involved, there may be a transfer of data between different ministries, as when the local registration is done through the ministry of health but the preparation of the statistical reports is the responsibility of the statistical office. Details of the needed coordination will vary depending on the nature of the relations between the central government and the local governmental units.

A sample survey designed to provide national data makes demands on the central administration similar to those of a census, but the number of individuals who need to be recruited and trained is far smaller than for a census. However, since sample surveys normally seek to obtain more detail on the topics that are covered, more intensive training and supervision of the interviewers is required. The selection of the sample and the translation of sample results into totals for the country require well-trained statistical staff. Supervisors must be available to watch for deviations from the sampling plan and for errors by the field workers, which can seriously bias the results. Although in a survey there are far fewer questionnaires to be collected and processed than in a census, the technical requirements of a sample

survey are more rigid than those that can be applied to a full census. An ongoing program of repeated or multiround sample surveys can profit from the carry-over effect of staff training from one round to the next; however it may also face special technical problems, such as using the same sample area boundaries in each round and finding new households that move into the sample areas between rounds.

Area Detail

In principle, a census covers the entire country and thus can provide data for every identified village or other subareas. In fact, some censuses omit certain tribal groups, nomads, or inhabitants of remote areas, or deal with them by procedures that differ from those used with the majority of the population. Nevertheless, when data are needed for the smallest recognized areas, the census is usually the major source of statistical information.

As with census data, the data from a civil registration system can be made available for the smallest identified governmental units in the country and summarized for any desired area. A special precaution applies to these data, however: it must be clearly specified whether the events are recorded in the place where they occurred or in the registrant's usual place of residence. The residence of the mother should be the controlling element in the case of a birth, the residence of the decedent in case of death. If events are reported in both place of occurrence and place of residence, care must be taken to avoid double counting.

The size of the area for which sample survey data can be made available depends on the size of the sample, but surveys normally cannot provide data for small areas. In designing a sample survey, the needs for area detail in the final results must be a major consideration. This holds for surveys conducted as part of a census as well as those conducted independently of the census.

Timeliness

A census is a massive undertaking that is done relatively infrequently. Ten-year intervals between censuses are the rule, although a number of countries take censuses at five-year intervals. Even in a country with a relatively

small population, the census places exceptionally heavy demands on the statistical office responsible for the preparation and dissemination of the results. Careful planning is required to make the results available within a reasonable time span. In most cases the most efficient approach is to identify the most urgently needed data and release them first. Even using electronic tabulating equipment, the complete processing of a census may require several years. This is true even if the country has a long history of census taking.

One of the chief advantages of a properly functioning civil registration system is that it can provide timely data on births and deaths at all levels of area detail: national, regional, and by districts, cities, and subdistricts. In analyzing data from very small areas, one must be alert to the effects of random fluctuations; however, civil registration data can provide the basis for timely analysis of levels and trends of fertility and mortality by regions, states, urban and rural areas, and other reasonably large geographic areas. A properly functioning system should be able to provide initial summaries of tabulated data relatively soon after collection of the basic information. The availability of communication and transportation facilities and the effectiveness of relations among the several governmental units involved all influence the timeliness of the results, but for summary statistics, monthly or quarterly reports that reflect a relatively short lag between recording and publication of results usually can be achieved after a break-in period.

The results of sample surveys normally can be made available more quickly than the results of a census. There are far fewer cases to process and this alone makes some difference. However, the number of records is not the main factor governing how quickly the findings can be released. Sample surveys often are intended to provide analytical results and these often call for more intricate tabulation routines than those used for a census, for which the major concern is with relatively simple totals.

It cannot be emphasized too strongly that for censuses and surveys the data collection and processing must be viewed as an integrated operation, and the desired products should be clearly in mind before the questionnaire is designed or the field work is begun. The costs of failing to specify the desired tabulations at the outset of a data collection procedure are high in terms of both time and money. A common error is to allow too little

lead time for development of plans and procedures before actual operations are begun. The price is invariably counted not only in terms of time and money but also in foregone opportunities for insightful tabulations. The history of censuses and sample surveys provides far too many cases that illustrate the truth of the aphorism "Haste makes waste."

One of the most cost-effective practices is to run pretests of the questionnaires and trial runs of the full set of procedures that are to be used during the entire census or survey activity. Errors and gaps can be identified during such trial runs, and corrections can be made before the procedures become so well established that any alterations would be disruptive and costly.

Subject Coverage

A census of population is a general-purpose survey that must meet many different needs. Although the amount of probing it can do on any one topic is necessarily limited, many censuses provide useful data about the fertility of the population. As was pointed out above, a census is a less satisfactory source of information about mortality than other methods (except about child mortality, if certain questions are included). The fact that it is intended to cover every area within the country means also that the data it collects can serve as a basis for studying regional, rural-urban, ethnic, or social-class variations in fertility and the reasons that these variations occur. Comparison with previous censuses, when they exist, can supply information concerning long-range trends in fertility and mortality. The broad scope of a census may also provide the basis for studies of interrelations between fertility and the other characteristics of the population, such as education, labor force participation, household relationship, rural-urban residence, and others. Studies of the interrelationships between child mortality and these other characteristics are also possible.

For censuses to provide the data for these studies, it is essential that a small set of questions on fertility and child mortality be asked of all women aged 15 and over regardless of marital status (except where cultural constraints make it impossible to ask women who have never been married whether they have had any children). The set includes three basic questions: Of the children

you have ever born alive, how many are still living at home with you? Of the children you have ever born alive, how many are still alive but living elsewhere? Of the children you have ever born alive, how many have died? Ideally these questions should be asked for children of each sex, because that not only provides data for estimating child mortality by sex but may also reveal patterns of error. This may not always be possible in censuses, which must cover many topics in limited scope, but most surveys can include questions about both male and female children and should do so.

Civil registration systems are oriented to providing information for legal and administrative uses. As a result, the variety of information requested by those systems tends to be somewhat restricted, although some items desired by health authorities are often included, for example: where a birth occurred (home, maternity center, hospital); who delivered the woman of her baby (traditional midwife, trained midwife, physician, etc.); and the extent of medical attention received prior to a birth or death. In some instances, civil registration records have been used to secure indirectly social and economic information related to births and deaths by using the registered birth or death as the initial information leading to an interview in the household that has experienced the event. For example, a special survey of health and infant care and feeding may use a sample based on births in selected areas of a country during a specified time period.

Civil registration systems also collect some information very useful in the estimation of fertility and mortality and in the analysis of the determinants of fertility and mortality. For example, as health systems develop and as information on causes of death are included in death registration records, the civil registration system provides information for the study of differential causes of death by age and sex, among regions, and, eventually, over time. Fertility analysis is possible when birth registration includes data on age of mother, marital status, duration of marriage, parity, time elapsed since previous pregnancy ended, and in some cases, on ethnic characteristics and educational attainment of the mother and father, all among regions and over time. Thus a well-functioning civil registration system can provide a large body of information for planners and others who want to study fertility and mortality levels and trends.

In a sample survey focused on fertility, a significant portion of the total interview time is naturally devoted to questions relating to fertility, including marriage and pregnancy histories, prenatal care, knowledge and practice of contraception, sterilization, and, in some settings, abortion. The survey questionnaire can include items on attitudes toward family size. It can also include questions that can be used for the investigation of differences in family size of mothers in relation to where they grew up, the number of siblings they had, and other background characteristics.

In a sample survey on fertility, the respondent is ordinarily asked to report on his or her own experience or attitudes. In contrast, a census interview is often conducted by proxy, with one person reporting for all members of the household. Self reporting is generally more accurate than proxy reporting, especially when the questions deal with information that may not be well known by others, as is the case with fertility. While proxy reporting is cheaper and generally provides a higher response rate than self reporting, both the variety and detail of questions that can be asked and the accuracy of the responses are less than for self reporting. Furthermore, because the time required to plan and complete a sample survey is ordinarily much shorter than that required for a census and because a survey requires fewer resources than a census, surveys are conducted more frequently, thus providing a more up-to-date basis for measuring current levels and recent trends.

The conditions and the auspices under which information is solicited affect not only the response rate but also the appropriateness of certain questions and people's willingness to answer them accurately or at all. Both censuses and civil registration systems usually are conducted in accordance with laws that require the respondents to answer fully and to the best of their ability. Sample surveys, on the other hand, are more often conducted on the basis of voluntary participation, which gives the respondent the right to refuse to answer any or all of the questions. Items that are considered sensitive may not be included in a census questionnaire but be included among the questions in a survey: for example, questions on contraceptive practice or on abortion.

Frequently the information collected in surveys of fertility and mortality is not completely analyzed. In some cases very little information is provided to planners

even though quite detailed subject matter has been collected. When planning surveys, more attention should be given to the personnel and budgetary requirements for tabulation, analysis, and publication of the data collected. One of the most effective ways of having the results of all data collection systems fully analyzed is to make the data available (after a suitable time lapse) to the scientific and academic communities. Users' samples of censuses and data tapes containing survey results are particularly desirable forms of data for diffusion. In addition, the archival conservation of data sets beyond their immediate use by governments is necessary for future data analysis, especially for research on and analysis of population trends.

Errors

An error-free statistical undertaking has not yet been invented. Every effort to collect data is subject to a number of errors: mistakes made by the designer, interviewer, or respondent; mistakes made at various stages of processing; mistakes in printing; and errors in the geographic work underlying a census or a sample survey. Careful training and supervision and the following of quality-control procedures at each step can substantially reduce the frequency of errors, but since it is impossible to eliminate such errors completely, it is not worth the cost and effort to attempt to do so. However, it is essential to identify and measure as well as possible what such errors are and what effects they have on the final results. Without information on limitations of the data, planners and other users of the data may be led to erroneous conclusions: for example, concluding that there has recently been a large decline in fertility when in fact no decline or a much smaller decline has occurred.

Age is a particularly critical item that is subject to error. In any analysis of fertility, the age of the mother is of primary significance; in analyses of mortality, the age of the decedent is important. (For some estimation procedures, the age of the respondent who is reporting whether or not a parent or spouse is still living can be used for analyzing mortality.) However, errors in age reporting are widespread, especially in the many cultures for which the measurement of age does not have the significance that is taken for granted in western and some East Asian societies. Ingenious approaches have

been developed to overcome the difficulty of dating events according to a conventional calendar, but results have been mixed. Age data often require some adjustment before they can be safely used in demographic analysis. Statisticians and demographers trained in such adjustment procedures are an essential part of any data collecting effort, in both the planning and analytical stages. (For an in-depth discussion of the problem of age misreporting, see Ewbank (1981); for a detailed description of adjustment and estimation procedures that can be applied to faulty demographic data, see Hill et al. (1981).)

In censuses and universal civil registration systems, the results are not subject to sampling errors because the entire population is included. Results of sample surveys, however, are subject to sampling errors, and these may be especially large when detailed cross-classifications are undertaken. Sampling errors can be calculated and should always be made available with the survey results. They form an integral part of the total error, which must be recognized in any analytical or interpretive use of tabulated data--whether from a census, a sample survey, or a registration system.

A second major source of error is bias, which affects censuses and civil registration systems as well as surveys. Biases refer to systematic errors that affect any data collection process operating under specified design conditions with the same constant error. Examples include noncoverage or consistent underreporting of infant deaths or other vital events; some average understatement of abortion among a particular set of interviewers in a survey; consistent miswording of certain questions through questionnaire design or by some interviewers so that replies do not really respond to the questions on the questionnaire; telescoping of events in time; coding or programming errors that assign persons of "unknown" ages to the age group 85 and older; and so on. The values of biases remain largely unknown, unless measured by methods external to the census, registration system, or survey. Because errors resulting from bias may be substantial, attempts should be made to estimate their magnitude or at least to recognize their existence and possible effects on measures of fertility and mortality.

Data from civil registration systems in developing nations are affected by one major type of error: in many of these countries the registration is currently incomplete. Indirect estimation techniques frequently make it possible to make adjustments and base estimates on data

from the vital events that are registered, but the degree to which civil registration data can be used in estimation will increase as this major source of error decreases.

Two other types of errors are unique to civil registration systems. The first type of error concerns the need for careful distinction between place of residence of the person experiencing an event and the place where that event is registered. For example, a seemingly high death rate in a city may be due to the practice of registering deaths by place of occurrence coupled with the existence in that city of a major hospital that attracts patients from surrounding rural areas, especially persons who are near death. Similarly, errors arise when a system tabulates births by place of occurrence and there is the practice of city women returning to their villages of origin to give birth or of rural women going to city maternity facilities to give birth. These situations lead to an overestimate of rural fertility, an overestimate of urban fertility, or some combination of these and other errors.

The second type of error concerns the assigning of an event to a specific period of time. In some countries, many registrations, especially of births, are delayed; other events may never be registered at all. This can lead to errors in several ways. When a birth is followed by the early death of the infant, the parents may not bother to register either event. Or a birth may be registered only at the time a birth certificate is needed, such as for entrance to school, usually five or more years after the birth. If all births are tabulated by the year of registration, the data for any given year will include some births that occurred in previous years; there will also be missing events that may be registered in the future and events that may never be registered. Delayed registrations may be so large a proportion of the events that are registered in any one period that they affect all efforts to analyze trends over time.

Costs

It is difficult to estimate precisely the costs of collecting data on fertility or mortality. In the case of a census, those data constitute only a small fraction of the total cost of the census. The costs of a sample survey devoted entirely to fertility or mortality (or both)

can be computed fairly easily, but such estimates become trickier when a survey is conducted at the same time as other data collection efforts and the costs are being shared with some other operation. Joint costs are likely to be especially troublesome in assessing the costs of a registration system. The fact that all levels of government are involved in this activity makes it difficult to allocate costs precisely, and the fact that the registration serves chiefly legal and administrative purposes makes it difficult to assign costs to the data's use for statistical purposes. (For a more detailed discussion of costs, see Chapter 1 in National Research Council (1981).)

In addition to the administrative and labor costs of gathering statistics, there are social costs that any data collection system places on the public, especially on the respondents who are asked to supply the information. Successful collection of data depends on public cooperation. If the public willingness to supply information is overtaxed, it can affect all data gathering. While public willingness has not yet become an issue in most less-developed countries, it is a constraint in some developed countries in the implementation of censuses and surveys. On the positive side, there are social benefits: censuses, civil registration systems, and surveys provide information useful for businesses, various levels of government, organizations, and for research. In addition, civil registration systems create documents that are valuable to the person to whom the document relates or to other people closely associated with that person.

CONCLUSIONS

Accurate statistics are essential to any government and to the society that government serves. In many developing countries, reliable population data are still not available despite the fact that the need for accurate data is great in all nations. For example, in Bangladesh severe consequences may result from even a small miscalculation in food production statistics or ignorance of the true size of the population in a region.

Methods for collecting adequate data are available, and while they may appear to increase the costs of policy making, one must weigh those costs against the costs incurred when programs are undertaken on the basis of inadequate or incorrect information. The institution of a regular census, the operation of a civil registration

system, and the efficient use of sample surveys are essential and complementary steps in constructing and maintaining a picture of a country's status and progress in economic and social development. Rational, effective, and efficient policy making and program administration are enhanced by timely, reliable, and appropriate data. To this end, three points merit emphasis.

First, in assessing data collection plans and results, an understanding of technical issues is an essential complement to common sense and familiarity with the population to be studied.

Second, all data collection activities are affected by errors; although some types of errors tend to cancel out, others do not. The existence of errors in a body of data does not render such data useless, but it does impose responsibilities on those who produce and use the data. The data producers should provide as much information as possible about the type and size of errors affecting a given set of data. The data users must take into account the limitations in the data and examine the quality of the data in terms of the specific use intended. Identification of the causes, extent, and types of errors can lead to reductions in the errors, to adjustments that permit the construction of more reliable estimates of fertility and mortality, and to modifications in the future collection of fertility and mortality data.

Third, there are trade-offs concerning data collection systems that take into account the uses of the data and the resources required to collect them. As shown in Table 1, in terms of their suitability for providing data to estimate fertility and mortality, the performance of each major data collection method is strong on some criteria but weak or moderate on others, which reveals the essential complementarity of the systems. The extent to which one data collection method can substitute for another depends closely on such performance trade-offs. In some cases and for some purposes such substitution is possible: for example, incorporating questions on children ever born and on children surviving into the census questionnaire instead of conducting a special survey. Substitution is not always possible, however: for example, surveys usually cannot provide the local area data that are available from a census, and censuses do not provide individuals with legal documentation of vital events as does a civil registration system.

Censuses, civil registration systems, and sample surveys will remain the essential tools for generating in-

formation on the levels and trends of fertility and mortality. Knowledge of the strengths and weaknesses of each system, what each requires to operate efficiently, and what each can contribute to the estimation of fertility and mortality will enable administrators and policy makers to make more efficient and effective use of the data generated from all three methods.

REFERENCES

Ewbank, D.C. (1981) Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis. Report No. 4, Panel on Data Collection, Committee on Population and Demography, Assembly of Behavioral and Social Sciences, National Research Council. Washington, D.C.: National Academy Press.

Hill, K., H. Zlotnik, and T.J. Trussell (1981) Indirect Techniques for Demographic Estimation. Committee on Population and Demography, Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences. (To be published as Manual X in the series of manuals of demographic analysis of the Population Division, United Nations.)

National Research Council (1981) Collecting Data for the Estimation of Fertility and Mortality. Report No. 6, Panel on Data Collection, Committee on Population and Demography, Assembly of Behavioral and Social Sciences. Washington, D.C.: National Academy Press.