Appendix C: Selected Governmental Sources of Biostatistical Data

C

Three types of data collections are described here: (1) a population census, (2) a vital statistics system, and (3) sample surveys. In addition, (4) the sources of the data used in U.S. life tables are described. To understand the data resulting from these collection mechanisms, it is essential to be familiar with some definitions and the organization of the data collection systems.

I. Population Census Data

The *census* is a counting of the entire population at a specified time. In the United States, it occurs once every 10 years as required by the Constitution, and the latest census was taken on April 1, 2000. The U.S. Census attempts to count people in the place where they spend most of their time. Most people are counted at their legal residence, but college students, military personnel, prison inmates, and residents of long-term institutions are assigned to the location of the institutions.

The information available from the U.S. Census is derived from two types of questionnaires. The questions on the short form are intended for everybody in every housing unit, and the form includes such basic data items as age, sex, race, marital status, property value or rent, and number of rooms. The long form is intended for persons in sampled housing units and includes, in addition to the basic items, income, education, occupation, employment, and detailed housing characteristics. Data are tabulated for the nation and by two types of geographic areas: administrative areas (states, congressional districts, counties, cities, towns, etc.) and statistical areas (census regions, metropolitan areas, urbanized areas, census tracts, enumeration districts, block groups, etc.).

The tabulated census data are made available in several different forms: printed publications and electronic data files. To access the data, it is necessary to consult documentation for the data media of your choosing. The racial classification in the 2000 census data needs special attention since multiple choices of racial categories were allowed. The Census Bureau modified the race data and produced the Modified Race Summary File.

The census data are used for a variety of purposes: by the federal, state, and local governments for political apportionment and allocation of federal funds for planning and management of public programs; by demographers to analyze population changes and the makeup of the nation's population; by social scientists to study social and economic characteristics of the nation's population; and by statisticians to design sample surveys for the nation and local communities. The census data, most importantly, provide the denominator data for the assessment of social and health events occurring in the population — for example, in calculating the birth and death rates.

Postcensal population estimates are available from the Census Bureau. The postcensal estimates are made for the resident population as of July 1 of each year. These data are available from the U.S. Census Bureau website at www.census.gov.

II. Vital Statistics

Vital statistics are produced from registered vital events including births, deaths, fetal deaths, marriages, and divorces. The scope and organization of vital events registration system varies from one country to another. In the United States, the registration of vital events has been the responsibility of the states primarily and of a few cities. The federal government's involvement is to set reporting standards and to compile statistics for the nation. Each state is divided into local registration districts (counties, cities, other civil divisions) and a local registrar is appointed for each district. The vital records are permanently filed primarily in the state vital statistics office. The local and state vital registration activities are usually housed in public health agencies. The National Center for Health Statistics (NCHS) receives processed data from 50 states and other local vital registration offices.

Vital events are required to be registered with the registrar of the local district in which the event occurs. The reporting of births is the direct responsibility of the professional attendant at birth, generally a physician or midwife. Deaths are reported by the funeral directors or person acting as such. Marriage licenses issued by town or county clerks, and divorce and annulment records filed with the clerks or court official provide the data for marriage and divorce statistics. The data items on these legal certificates determine the contents of vital statistics reports. These certificates are revised periodically to reflect the changing needs of users of the vital statistics.

Vital statistics are compiled at the local, state, and federal levels. Data are available in printed reports and also on electronic files. Data are tabulated either by place of occurrence or by place of residence. Data by place of residence from the local vital statistics reports are often incomplete because the events for residents may have occurred outside the local registration districts and may not be included in the local data base.

What uses are made of vital statistics? In addition to calculating the birth and death rates, we obtain such well-known indicators of public health as the infant mortality rate and life expectancy from vital statistics. Much epidemiological research is based on an analysis of deaths classified by cause and contributing factors which comes from the vital statistics. Birth data are used by local health departments for planning and evaluation of immunization programs and by public health researchers to study trends in low-birth-weight infants, teenage birth, midwife delivery, and prenatal care.

There are several special data files that are useful for biostatistical use, including multiple cause-of-death data file, linked birth/infant mortality data file, and the Compressed Mortality File (CMF). Multiple cause data give information on diseases that are a factor in death whether or not they are the underlying cause of death and associated other diseases and injuries (for more information, see www.cdc.gov/nchs/products/elec_prods/subject/mortmcd.htm). National linked files of live births and infant deaths are especially useful for epidemiologic research on infant mortality (for more information, see www.cdc.gov/nchs/linked.htm). The CMF is a county-level national mortality and population data base. This data file is especially useful to epidemiologists and demographers, since mortality data and population data are available in the same file (for more information, see www.cdc.gov/nchs/products/elec_prods/subject/mcompres.htm).

III. Sample Surveys

To supplement the census and vital statistics, several important continuous *sample surveys* have been added to the statistics programs of the Census Bureau and the NCHS. Unlike the census and vital statistics, data are gathered from only a small sample of people. The sample is selected using a complex statistical design. To interpret the sample survey data appropriately, we must understand the sample design and the survey instrument.

The Current Population Survey (CPS) is a monthly survey conducted by the Census Bureau for the Department of Labor. It is the main source of current information on the labor force in the United States. The unemployment rate that is announced every month is estimated from this survey. In addition, it collects current information on many other population characteristics. The data from this survey are published in the Current Population Reports which include several series: Population Characteristics (P-20); Population Estimates and Projections (P-25); Consumer Income (P-60); and other subject matter areas. Public use tapes are also available (for more information, see www.census.gov).

The NCHS is responsible for two major national surveys: the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES). The sampling design and the estimation procedures used in these surveys are similar to the CPS. Because of the complex sample design, analysis of data from these surveys is complicated (see Chapter 15). These two surveys are described following. There are several other smaller surveys conducted by the NCHS, including the National Survey of Family Growth, the National Hospital Discharge Survey, the National Ambulatory Medical Care Survey, the National Nursing Home Survey and the National Natality and Mortality Surveys.

The NHIS, conducted annually since 1960, is a principal source of information on the health of the noninstitutionalized civilian population of the United States. The data are obtained through personal interviews covering a wide range of topics: demographic characteristics, physician visits, acute and chronic health conditions, long-term limitation of physical activity, and short-stay hospitalization. Some specific health topics such as aging, health insurance, alcohol use, and dental care are included as supplements in

different years of the NHIS. The data from this survey are published in the Vital and Health Statistics Reports (Series 10) and data tapes are also available (for more information, see www.cdc.gov/nchs/nhis.htm).

The NHANES, conducted periodically, is a comprehensive examination of the health and nutrition status of the U.S. noninstitutionalized civilian population. The data are collected by interview as well as direct physical and dental examinations, tests, and measurements performed on the sample person. Among the many items included are anthropometric measurements, medical history, hearing test, vision test, blood test, and a dietary inventory. Several health examination surveys have been conducted since 1960; the two most recent surveys are the Hispanic HANES (conducted in 1982–1984 for three major Hispanic subgroups: Mexican Americans in five southwestern states; Cubans in Dade County, Florida; and Puerto Ricans in the New York City area) and NHANES III (conducted in 1988–1994).

Beginning in 1999, the survey has been conducted continuously. With the continuous survey, new topics have been included. These include cardiorespiratory fitness, physical functioning, lower extremity disease, full body scan (DXA) for body fat as well as bone density, and tuberculosis infection. The data from health examination surveys are published in the Vital and Health Statistics Reports (Series 11). Electronic data files are also available (for more information, see www.cdc.gov/nchs/nhanes.htm).

IV. Life Tables

Life tables have been published periodically by the federal government since the mid-19th century. The first federally prepared life tables appeared in the report of the 1850 Census. Life tables prior to 1900 were based on mortality and population statistics compiled from census enumerations. The accuracy of these life tables was questioned, since mortality statistics derived chiefly from census enumeration were subject to considerable underenumeration. The year 1900 is the first year in which the federal government began an annual collection of mortality statistics based on registered deaths. Since then life tables have been constructed based on registered deaths and the enumerated population. Prior to 1930, life tables were limited to those states that were included in the death registration area. Until 1946, the official life tables were prepared by the United States Bureau of the Census. All subsequent tables have been prepared by the United States Public Health Service (initially by the Nation's Office of Vital Statistics and later by the NCHS).

Life tables provide an essential tool in a variety of fields. Life insurance companies largely base their calculations of insurance premiums on life tables. Demographers rely on life tables in making population projections, in estimating the volume of net migration and in computing certain fertility measures. In law cases involving compensation for injuries or deaths, life tables are used as a basis for adjudicating the monetary value of a life. Personnel managers and planners employ life tables to schedule retirement and pension programs and to predict probable needs for employee replacement. Applications are numerous in public health planning and management, clinical research, and studies dealing with survivorship.

There are three series of life tables prepared and published by the NCHS (for further information, visit www.cdc.gov/nchs):

1. *Decennial life tables*. These are complete life tables, meaning that life table values are computed for single years of age. These are based on decennial census data and the deaths occurring over three calendar years around the census year. The advantage of using a three-year average of deaths is to reduce the possible abnormalities in mortality patterns that may exist in a single calendar year. The decennial life tables are prepared for the United States and for the 50 individual states and the District of Columbia. This series also includes life tables by major causes of death; these tables are known as multiple decrement life tables.

2. *Annual life tables*. These are based on complete counts of deaths occurring during the calendar year and on midyear postcensal population estimates provided by U.S. Bureau of Census. From 1945 to 1996, the annual tables were abridged life tables, meaning that life table values are computed for age intervals instead of single years of age, except for the first year of life. The set of age intervals used are 0–1, 1–5, 5–10, 10–15, . . . , 80–85, and 85 or over. Beginning with 1997 mortality data, complete life tables are constructed using a new methodology (Anderson 1999) and extended the ages to 100 years of age. Vital statistics for old ages are supplemented by Medicare data. These are prepared for the United States total population by gender and race.

3. *Preliminary annual life tables*. Preliminary life tables, based on a sample of death records, are published annually before the final annual life tables become available. This series has been published annually since 1958. Only a 10 percent sample of registered deaths was used to construct an abridged table for early years. Preliminary tables are now based on a substantial sample (approximately 90 percent) to construct complete life tables for the total United States population only. These are published in the Monthly Vital Statistics Report.

REFERENCES

"History and Organization of the Vital Statistics System." In *Vital Statistics of the United States, 1950*, Vol. I, Chapter 1, pp. 2–19.

Anderson, R. N. "A Method for Constructing Complete Annual U.S. Life Tables." National Center for Health Statistics. *Vital and Health Statistics Report* 2 (129), 1999.