

# INTRODUCTION TO Public Health

*Third Edition*



Mary-Jane Schneider

## The Role of Data in Public Health



Collecting Data

Just as a doctor monitors the health of a patient by taking vital signs—blood pressure, heart rate, and so forth—public health workers monitor the health of a community by collecting and analyzing health data. These data are called health statistics. Statistics are a vital part of public health's assessment function, used to identify special risk groups, to detect new health threats, to plan public health programs and evaluate their success, and to prepare government budgets. The statistics collected by federal, state, and local government are the raw material for research on epidemiology, environmental health, social and behavioral factors in health, and for the medical care system.

At the federal level, the primary agency that collects, analyzes, and reports data on the health of Americans is the National Center for Health Statistics (NCHS), part of the Centers for Disease Control and Prevention (CDC). NCHS collects its data in two main ways: First, states periodically transmit data they have compiled from local records; vital statistics, including virtually all births and deaths, are routinely collected this way. Second, NCHS conducts periodic surveys of representative samples of the population, seeking information on certain characteristics such as health status, lifestyle and health-related behavior, onset and diagnosis of illness and disability, and the use of medical care. Some of these surveys are conducted on a state-by-state basis, and the data are thus useful to states and local communities. In addition, other federal agencies that collect data for their own purposes share it with NCHS.

## Vital Statistics

---

Births and deaths are the most basic, reliable, and complete data collected. Virtually every birth and death in the United States is recorded on a birth certificate or death certificate. Certificates are filed with the local registrar by the attending physician, midwife, undertaker, or other attendant. The state health department is generally responsible for collecting these reports and transmitting them periodically to the NCHS.

Birth certificates contain information supplied by the mother about the child's family, including names, addresses, ages, race and ethnicity, and education levels. Medical and health information is supplied by the hospital, doctor, or other birth attendant concerning prenatal care, birth weight, medical risk factors, complications of labor and delivery, obstetrical procedures, and abnormalities in the newborn. In the past decade many states have added a question on the mother's use of tobacco to the birth certificate. Much of the information on the certificate is confidential, withheld even from the person represented by the certificate. Its main use is for public health research, providing the data that can be used to relate features of the mother and her pregnancy to the health of the child.

The information on death certificates is subject to a number of uncertainties, depending on how well the informant knew the deceased and the circumstances of the death. For example, information on parents, education, and occupation may not be known if the decedent is an elderly person with no surviving relatives. There is often difficulty in the accuracy and consistency with which causes of death are specified. Incorrect diagnoses are common; in the absence of an autopsy, the exact cause of death may not be known. If a number of conditions contribute to the fatal process, underlying causes and immediate causes may be confused. For some conditions such as AIDS or suicide, the cause of death may be misstated deliberately by the local official because of social stigma.

In addition to births and deaths, vital statistics include marriages and divorces, spontaneous fetal deaths, and abortions. Data on marriages and divorces are legal events that require universal reporting, but they are not very interesting from a public health point of view. Reporting of spontaneous fetal deaths is incomplete, especially for those that occur relatively early in a pregnancy; many may be unrecognized. Induced abortions are also probably somewhat under-reported. In some states, the name of the woman who had the abortion is not included in the report for reasons of confidentiality.

Because infant mortality is an important public health issue, the NCHS has set up a special computer system that links vital records of infants born during a given year who died before their first birthday. The linkage allows researchers to compare information on the death certificates with that on the birth certificates, providing insight into factors that contribute to infant deaths.

## The Census

---

As discussed in Chapter 7, the data collected through the vital statistics system and other methods must be converted into rates if they are to be useful for many public health purposes. The calculation requires information on the number of people in the population being referred to, the number that serves as the denominator when a vital statistic is used as the numerator. To calculate age-adjusted or age-specific rates, it is necessary to know how many people are in each age group. To determine sex-specific or race-specific rates, one needs to know how many males and females there are and how many blacks, whites, Hispanics, and people of other races in each sex and each age group. This information is collected by the U.S. Census Bureau, part of the Department of Commerce. Without an accurate count of the American population and all its characteristics, the government's health statistics would not be accurate.

As every schoolchild knows, the Constitution requires that the population of the United States be counted every ten years to determine each state's representation in the House of Representatives. Based on that simple mandate, the Census Bureau has developed a national survey that provides data not only on the geographical distribution of the population and its sex, age, and ethnic characteristics, but also on a wide variety of social and economic characteristics, including education, housing, and health insurance status. Furthermore, because the population is always in flux and its circumstances tend to change fairly quickly, the Census Bureau tracks trends in the population between the decennial censuses, using polls and surveys and other sources of data such as birth and death records, immigration and emigration records, and school statistics. Census Bureau data are vital for the operation of the nation's social, political, economic, and industrial systems, and they are essential for the practice of public health.

Because census data can determine the political composition of the U.S. Congress and the distribution of federal funds to states and communities, various interest groups carefully monitor how the data are collected. An issue that was particularly controversial in preparing for the year 2000 census concerned how a person's race is determined. The broad categories previously used in the census were white, black, Hispanic, Asian and Pacific Islander, and American Indian and Alaska native. Individuals identify their own race and ethnic category. The issue has been further complicated by the fact that interracial marriage and parenthood has become increasingly common in the United States, and many of mixed racial parentage wanted an "interracial" category to be included on the year 2000 census. After considerable debate, the Census Bureau decided against such a category, but it allowed individuals to check more than one racial category for themselves.<sup>1</sup> This policy affects race-specific health statistics, but the effect is still small. Only 2.4 percent of the population chose to check more than one race in 2000.<sup>2</sup> That number is likely to increase in the 2010 census, especially now that President Obama has perhaps made self-identification as mixed race more acceptable.

An even more politically controversial issue is the chronic problem of how to count every individual person in the United States. The census is mandated by the U.S. Constitution, and the Supreme Court has interpreted the mandate to mean that every person in the country must be enumerated; no statistical corrections are allowed.<sup>3</sup> The process for the 2010 census begins in March 2010, when a short questionnaire will be mailed or delivered to every household. The head of household is asked to complete the form, providing information on all the residents in the household, and return it to the Census Bureau. If the form is not returned, a second questionnaire is mailed. If there is still no response, the household will be called or visited by a census worker to collect the information.

Inevitably, people will be missed or counted twice. The missing ones are likely to be the poorest and most marginal members of the population—the homeless, illegal immigrants, fugitives from the law. Wealthy people who own more than one home might be counted twice. The Census Bureau estimates that the 2000 census missed at least 6.4 million people and counted at least 3.1 million people twice.<sup>4</sup> Such errors can lead to systematic inaccuracies in health statistics. For example, blacks tend to be undercounted in the census, while black births and deaths are more accurately recorded, meaning that birth and death rates calculated for blacks tend to be higher than their true value would be if correct population numbers were used for the denominator.

Preparations for the 2010 census, according to *The New York Times*, have been a shambles.<sup>5</sup> The Census Bureau had planned that workers would use hand-held computers equipped with GPS to collect information from the nonresponders, but the plan fell through because of technical problems and budget shortfalls. The agency's director and deputy director resigned in 2006 over the Bush administration's lack of support for the census, and it took over a year for a



new director to be nominated and confirmed.<sup>6</sup> There have been partisan battles in Congress about how much effort should be made to count racial and ethnic minorities: Republicans tend not to care that inaccurate counts affect congressional representation, because hard-to-count groups, like minorities, immigrants and the poor, tend to vote Democratic.<sup>7</sup> And because census numbers determine allocation of hundreds of billions of dollars in federal funds, cities and states whose populations are undercounted tend to suffer.

A major change in the way the 2010 census will be conducted is that only the most basic data will be collected from everyone, using what used to be called the short form, which asks for name, age, sex, race and ethnicity, and relationship of everyone living in the household. Previous censuses have sought to gain a fuller understanding of population characteristics by using a long form for about one in six addresses, asking questions about education, housing, employment, transportation, language, ancestry, and other issues useful for governments and businesses. In an attempt to make the collection of this detailed information more efficient and more timely, the Census Bureau in 2005 launched a new ongoing survey called the American Community Survey (ACS), which collects the same kind of information previously collected on the long form. The long form will no longer be used in the decennial census. The ACS is sent each year to about three million households selected to be representative of the populations of local jurisdictions. The ACS is designed to help communities plan transportation systems, zoning, schools, healthcare facilities, and housing, as well as the need for social services.<sup>8</sup>

## **NCHS Surveys and Other Sources of Health Data**

---

As noted previously, the NCHS, in addition to collecting data from the states, actively conducts a number of surveys to gather additional information on the health of the American population. Follow-back surveys are a way to expand on the vital statistics data the NCHS has received. For example, in a survey conducted in 1988, NCHS chose a sample of birth certificates to investigate further, sending questionnaires to the mothers, doctors, and hospitals to learn more about family characteristics as well as the circumstances of the pregnancy and birth. Called the National Maternal and Infant Health Survey, the survey followed back a sample of fetal deaths and infant deaths, allowing researchers to study factors related to poor pregnancy outcomes.<sup>9</sup> Similar surveys are periodically conducted on a sampling of deaths; the person who filled out the death certificate is asked to provide more information on the lifestyle of the deceased as well as what medical care he or she received. The most recent mortality follow-back survey was conducted in 1993.

Two ongoing NCHS surveys aim to assess the health of the population as a whole, estimate the prevalence of selected diseases and risk factors, and look for trends. Each year, interviewers for the National Health Interview Survey (NHIS) contact 35,000 to 40,000 households and ask questions about illnesses, injuries, impairments, chronic conditions, utilization of medical resources, and other health topics.<sup>10</sup> The National Health and Nutrition Examination Survey (NHANES) is designed to obtain even more detailed and accurate information; doctors and nurses are sent in vans to conduct physical and dental examinations and laboratory tests on a carefully selected sample of the population. Each year, fifteen counties are visited, and about 5000 individuals of all ages are selected to undergo the tests. Data are collected on the prevalence of chronic conditions, including cardiovascular disease, diabetes, kidney disease, respiratory disease, osteoporosis, and hearing loss, as well as risk factors for those conditions, such as smoking, alcohol consumption, sexual practices, physical fitness and activity, weight, and dietary intake.<sup>11</sup>

NCHS is also collaborating with the National Institute on Aging on several follow-up studies of the population surveyed in previous NHIS surveys. In one follow-up study, over 7000 individuals who were 70 years of age or older in the 1984 NHIS were re-interviewed in 1986, 1988, and 1990. In a similar follow-up study, over 9000 individuals who were 70 years or older in 1994 were re-interviewed in 1997–98 and 1999–2000. Since then, subjects have been re-interviewed at two-year intervals. The interview data are linked to Medicare records and death certificates. The purpose of these studies is to describe the process by which older people progress from functioning in the community, becoming dependent, being institutionalized, and dying. Information is also collected on use of medical care and services. The second study looks for trends in healthy aging. The information provided by these follow-up studies is expected to be very valuable in relating the clinical, nutritional, and behavioral factors identified more than two decades ago to subsequent health status as people age, including their need for hospitalization or institutionalization in a nursing home.<sup>12</sup>

The Behavioral Risk Factor Survey (BRFSS) conducted by the states, which report their findings to the CDC, is another way of obtaining information on health-related behavior. It asks questions about health, including high blood pressure, high blood cholesterol, diabetes, and weight, as well as about high-risk behaviors such as cigarette smoking, excessive alcohol consumption, drinking and driving, and physical inactivity. It also asks whether people get preventive medical care such as mammograms, Pap smears, colon-cancer screening, and immunizations.<sup>13</sup> The BRFSS gathers some of the same information as NHANES, but it has the advantage of surveying many more people, and it allows analysis of how the factors vary from one state to another. However, the information is self-reported and may be less reliable than that obtained in NHANES. For example, the BRFSS finds that, according to people's own re-

ports, about 25 percent of adults are obese,<sup>14</sup> while the NHANES survey, using direct measurements, found a rate of about 34 percent.<sup>15</sup> This finding accords with previous observations that overweight people generally report that they weigh less than they do.

The NCHS conducts a variety of other surveys, some in collaboration with other agencies. Some of these include the National Asthma Survey, in collaboration with the CDC's National Center for Environmental Health, the National Infant Feeding Practices Study, in collaboration with the Food and Drug Administration, the National Health Interview Survey on Disability, in collaboration with several other agencies including the Social Security Administration, and the National Nursing Home Survey.

Other governmental agencies collect health-related data according to the focus of their responsibilities. For example, the Environmental Protection Agency carries out surveillance for health hazards in the environment, including air pollutants and releases of toxic chemicals, as discussed in Chapter 20. The National Cancer Institute coordinates a program called Surveillance, Epidemiology, and End-Results (SEER), used to monitor long-term trends of cancer incidence and mortality. The Centers for Medicare and Medicaid Services has billing records for the Medicare program, which are useful for research on utilization and outcomes of medical care, as discussed in Chapter 27. The Food and Drug Administration collects reports of adverse reactions to drugs after they have been approved and are on the market, sometimes recommending recalls if a serious problem appears that was not noted during preapproval testing. Surveillance for product-related injuries is conducted by the Consumer Product Safety Commission.

## Is So Much Data Really Necessary?

---

While it seems that the government collects enormous amounts of information on its citizens, there is never too much. These data are critically important in making up the surveillance systems that form the basis of effective public health practice as well as the planning and evaluation efforts that are increasingly being used in public health programming.

The statistics collected by federal, state, and local agencies are used in all areas of public health. Early notification of communicable disease cases is a classic use of public health information to protect the public's health. The need for public health intervention to control other problems may not be obvious without an analysis of data. This explains the Institute of Medicine committee's insistence on the importance of assessment as a core function of public health<sup>16</sup> (see Chapter 1). As discussed in Chapter 30, public health leaders are increasingly stressing the importance of planning, setting goals, and managing public health programs to meet these goals, a process that requires data at the local, state, and federal levels. For example,



a community may not recognize that it has a problem with unintended pregnancy and low birth weight unless it analyzes the data from birth certificates, comparing local data with statewide or national averages. Recognition of the problem might persuade local public health leaders to consider school-based birth control education and services.

Throughout this book, during discussions of public health issues (including biomedical, social and behavioral, environmental, and medical care issues), problems are defined according to the data that are available. In any area of public health, problems are identified in terms of statistics. The success of intervention programs to confront a problem is evaluated based on whether they improve the statistics.

In an era when people tend to frown on “big government” and yearn for lower taxes, there is always pressure to cut back fiscal support for data collection and analysis, activities that seem less urgent than fighting a known epidemic, for instance. Yet without data, experts cannot recognize that an epidemic is beginning. Inspired by the recommendations of *The Future of Public Health*,<sup>16</sup> the CDC has taken a lead in coordinating and encouraging the use of data in public health assessment. Recent events, including the emergence and resurgence of infectious diseases and the fear of bioterrorism, have stimulated the development of new surveillance systems within the United States and around the world.

With or without adequate data, decisions affecting public health policy and the allocation of scarce resources from government budgets must be made. It is increasingly important that these policy and fiscal decisions be made on the basis of timely and accurate information.

## Accuracy and Availability of Data

---

British economist Sir Josiah Stamp (1880–1941) wrote in 1929, “The Government [is] very keen on amassing statistics. They collect them, add them, raise them to the *n*th power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of those figures comes in the first instance from the village watchman, who just puts down what he damn well pleases.”<sup>17</sup>

The process of data collection is always imperfect. Even data for births and deaths, the most accurately reported health events, may be flawed. The census produces errors, and there are political difficulties in trying to rectify them. Most other sources of health information, relying as they do on surveys or voluntary reports, are even more incomplete or subject to bias. For example, the Youth Behavioral Factor Risk Survey of high school students, conducted by states and reported to the CDC, misses adolescents who have the highest risks—those who have dropped out of school.

Errors in reporting cause of death on death certificates, a prime example of the errors about which Stamp warns, are especially worrisome for public health in that mortality data have such a strong influence on planning and priority setting for public health programs. Autopsies are being done with declining frequency, in part because of cost concerns, but also because doctors may believe that sophisticated diagnostic technology has rendered autopsies obsolete. In 1961, autopsies were performed in 41 percent of hospital deaths; in the mid-1990s, that number had fallen to 5 percent to 10 percent<sup>18,19</sup> Cause of death information is still subject to uncertainty in many cases, however, and several studies have found that evidence obtained from an autopsy contradicted the clinical judgment of doctors in 15 percent to 32 percent of cases. Information gained from an autopsy answers the question, did the patient receive the correct treatment for the correct disease? This information can improve the quality of medical care for future patients as well as improve the accuracy of vital statistics.

Because some of the inaccuracies on birth and death certificates may result from carelessness on the part of the busy health professionals who file them, new electronic methods of filing that are being introduced in some states are expected to improve the quality of the data. For example, maternal deaths are suspected of being underreported because doctors often fail to check off on a women's death certificate whether she was pregnant or gave birth in the time period prior to her death. If an electronic death certificate is used, the computer will refuse to accept the form—will not “send” it—until that question is answered.<sup>20</sup>

Computers are extensively used in the analysis of public health data, of course, and new applications are continually improving the timeliness and accessibility of the data. Weekly reports of notifiable diseases from state and local health agencies are transmitted electronically to the CDC, allowing prompt response to new outbreaks. Laboratory results are also reported electronically, facilitating the rapid identification of bacterial and viral strains that may be causing illness in scattered locations around the country. Databases that are kept up-to-date by electronic filings can provide rapid feedback on the effectiveness of new public health interventions as well as help detect emerging problems.

The new information technology—or public health informatics as it is sometimes called—has vastly improved the accessibility of public health information to public health workers and the general public. The CDC and most other federal and state public health agencies make information available over the Internet. For example, *Morbidity and Mortality Weekly Reports* is searchable online, and articles can be downloaded from the CDC's Web page. The National Cancer Institute provides the latest information on cancer therapies and prognoses tailored for doctors and for patients. Most of the information is freely available to all, although some data sets require users to have special passwords before they are allowed access; others are available to authorized users only on diskettes or CDs.

## Confidentiality of Data

---

When anyone collects information on other people, questions always arise about how the information is going to be used and who is going to be allowed access to it. In general, all information collected from individuals by governments for whatever purpose is considered confidential and cannot be divulged without the consent of the individual. In most cases, the information is entered into a massive database from which individual names and addresses are removed. For research purposes, an identifying number may remain attached to the data to enable researchers to match information in one database with that in another. This technique is used, for example, in matching birth and death records as described above in order to learn more about the factors that contribute to infant mortality.

There is always concern that a determined snoop who works in an agency or knows an employee could obtain confidential information on an individual and use it to that individual's detriment. Agencies that handle confidential data impose stringent rules on access. Researchers must explain and justify their need for the data and promise to safeguard its confidentiality. Most agencies have an institutional review board or data protection committee, often including members from the community, which weighs the researchers' claims and decides whether to grant permission for access. Other than its use for research, the only exception made to the promise of confidentiality is when people must be notified that they have been exposed to a communicable disease.

The conflict between the need for confidentiality and the need for open access to information has been played out over various aspects of the AIDS epidemic, as discussed in the Prologue. Because HIV-positive individuals feared, with good reason, that they might be discriminated against if employers, landlords, and others learned of their infection, public health practitioners were concerned that patients would refuse to be tested unless confidentiality was ensured. Hence, the rules for reporting HIV were handled differently from other communicable diseases: anonymous testing was allowed, and the system for reporting cases to many state health departments and the CDC was modified to maintain anonymity. Recently, however, with the advent of new drugs that can clearly help AIDS patients and slow the onset of AIDS in HIV-infected individuals, HIV's exempt status has, for the most part, been discontinued, and it is treated like other communicable diseases.

## Conclusion

---

Statistics are the vital signs of public health. Local, state, and federal governments collect data on their citizens, starting with birth certificates and ending with death certificates. The U.S. census, conducted every ten years, provides information on the age, sex, and ethnic composition of communities, information that allows the calculation of birth rates, death rates, infant mortality rates, life expectancies, and other data that form the basis for public health's assessment function.

The NCHS is the repository for the vital statistics data received from the states. The NCHS also conducts a number of periodic and ongoing surveys to collect additional information on Americans, including information on family structure, specific health conditions, behavioral risk factors, and other data useful in planning public health intervention programs.

Health statistics are used for all aspects of public health policy development and evaluation. Uses of the data include health needs identification, analysis of problems and trends, epidemiologic research, program evaluation, program planning, budget preparation and justification, administrative decision making, and health education.<sup>21</sup>

Increasingly, electronic means are being used to collect, transmit, store, and analyze data and to make the data available to public health workers and the general public. Strict precautions are taken to ensure confidentiality of information about individuals.

## References

---

1. S. A. Holmes, "People Can Claim One or More Races on Federal Forms," *The New York Times*, October 30, 1997.
2. U.S. Census Bureau, "Census 2000 Shows America's Diversity," Press Release (March 12, 2001), [www.census.gov/Press-Release/www/releases/archives/2001\\_03.html](http://www.census.gov/Press-Release/www/releases/archives/2001_03.html), accessed January 31, 2010.
3. L. Greenhouse, "In Blow to Democrats, Court Says Census Must Be by Actual Count," *The New York Times*, January 26, 1999.
4. E. Schmitt, "Count of the 2000 Census Is Said to Err by Millions," *The New York Times*, March 16, 2001.
5. Editorial, "Census Damage Control," *The New York Times*, June 23, 2008.
6. Editorial, "The Census at 'High Risk,'" *The New York Times*, March 25, 2008.
7. Editorial, "Rescue the Census," *The New York Times*, December 4, 2008.
8. C. Holden, "New Annual Survey Brings Census into the 21st Century," *Science* 295 (2002): 2202-2203.
9. National Center for Health Statistics, "National Maternal and Infant Health Survey," [www.cdc.gov/nchs/about/major/nmihs/abnmih.htm](http://www.cdc.gov/nchs/about/major/nmihs/abnmih.htm), reviewed November 24, 2008, accessed December 5, 2008.

10. National Center for Health Statistics, "National Health Interview Survey: Celebrating the First 50 Years, 1957–2007," [www.cdc.gov/nchs/data/nhis/brochure2007june.pdf](http://www.cdc.gov/nchs/data/nhis/brochure2007june.pdf), June 2007, accessed January 31, 2010.
11. National Center for Health Statistics, "National Health and Nutrition Survey, 2007–2008: Overview," [www.cdc.gov/nchs/data/nhanes\\_07\\_08/overviewbrochure\\_0708.pdf](http://www.cdc.gov/nchs/data/nhanes_07_08/overviewbrochure_0708.pdf), January 2007, accessed January 31, 2010.
12. National Center for Health Statistics, "Longitudinal Studies of Aging," [www.cdc.gov/nchs/lsoa.htm](http://www.cdc.gov/nchs/lsoa.htm), reviewed August 25, 2009, accessed January 31, 2010.
13. U.S. Centers for Disease Control and Prevention, "BRFSS: Turning Information into Health," [www.cdc.gov/brfss/index.htm](http://www.cdc.gov/brfss/index.htm), updated December 3, 2008, accessed December 5, 2008.
14. P. P. Chowdhury et al., "Surveillance of Certain Health Behaviors Among States and Selected Local Areas—United States, 2005," *Morbidity and Mortality Weekly Report Surveillance Summary* (2007) 56 (SS04): Table 25.
15. C. L. Ogden et al., "Obesity among Adults in the United States—No Statistically Significant Change since 2003–2004. NCHS Data Brief no. 1, 2007," [www.cdc.gov/nchs/data/databriefs/db01.pdf](http://www.cdc.gov/nchs/data/databriefs/db01.pdf), November 2007, accessed December 5, 2008.
16. U.S. Institute of Medicine, Committee for the Study of the Future of Public Health, *The Future of Public Health* (Washington, DC: National Academy Press, 1988).
17. J. Stamp, *Some Economic Factors in Modern Life* (London: P. S. King, 1929), 258–259, quoted in R. Goodman and R. Berkelman, "Physicians, Vital Statistics, and Disease Reporting," *New England Journal of Medicine* 258 (1987): 379.
18. National Center for Health Statistics, "The Autopsy, Medicine, and Mortality Statistics," *Vital and Health Statistics Series 3*, No. 32 (2001), [www.cdc.gov/nchs/data/series/sr\\_03/sr03\\_032.pdf](http://www.cdc.gov/nchs/data/series/sr_03/sr03_032.pdf), October 2001, accessed January 31, 2010.
19. K. G. Shojana and E. C. Burton, "The Vanishing Nonforensic Autopsy," *New England Journal of Medicine* 358 (2008): 873–875.
20. Michael Zdeb and Mary Applegate, personal communication.
21. G. Pickett and J. J. Hanlon, *Public Health: Administration and Practice* (St. Louis, Missouri: Times Mirror/Mosby, 1990), 151–153.