

# ⌘ ONE ⌘

## INTRODUCTION

### *The Pervasiveness (and Limitations) of Measurement*



*[Numbers] can bamboozle and not enlighten, terrorize not guide,  
and all too easily end up abused and distrusted. Potent but shifty,  
the role of numbers is frighteningly ambiguous.*

—Blastland & Dilnot (2009, p. xi)

**O**n September 23, 1999, NASA fired rockets that were intended to put its Mars Climate spacecraft into a stable, low-altitude orbit over the planet. But after the rockets were fired, the spacecraft disappeared—scientists speculated that it had either crashed on the Martian surface or had escaped the planet completely. This disaster was a result of confusion over measurement units—the manufacturer of the spacecraft had specified the rocket thrust in pounds, whereas NASA assumed that the thrust had been specified in metric system newtons (Browne, 2001).

Measurement is obviously very important in the physical sciences; it is equally important in the social sciences, including the discipline of criminology. Criminologists, policy makers, and the general public are concerned about the levels of crime in society, and the media frequently report on the extent and nature of crime. These media reports typically rely on official data and victimization studies and often focus on whether crime is increasing or decreasing.

Both official crime and victimization data indicated that property and violent crime in the United States were in a state of relatively steady decline from the early 1990s to 2000. But in late May 2001, the release of Federal Bureau of Investigation (FBI) official crime data, widely publicized in the media, indicated that crime was no longer declining. This prompted newspaper headlines such as “Decade-Long Crime Drop Ends” (Lichtblau, 2001a) and led commentators such as James Alan Fox, dean of the College of Criminal Justice at Northeastern University, to assert, “It seems that the crime drop is officially over. . . . We have finally squeezed all the air out of the balloon” (as quoted in Butterfield, 2001a). However, some two weeks after the release of these official data, a report based on victimization data indicated that violent crime had decreased by 15% between 1999 and 2000, the largest one-year decrease since the federal government began collecting national victimization data in 1973 (Rennison, 2001). The release of these data prompted headlines such as “Crime Is: Up? Down? Who Knows?” (Lichtblau, 2001b) and led James Alan Fox to declare, “This is good news, but it’s not great news” (as quoted in Bendavid, 2001).

How do we reconcile the conflicting messages regarding crime trends from these two data sources? First, although most media sources commenting on the FBI data failed to mention this caveat, the official data report was in fact based on preliminary data: “[The report] does not contain official figures for crime rates in 2000” (Butterfield, 2001a). Second, and more important, the underlying reason for these differences is that the two data sources measured crime differently. Official crime data are based on reports submitted to the FBI by police departments, and they measure homicide, rape, robbery, aggravated assault, burglary, automobile theft, and larceny/theft. In contrast, victimization data are based on household surveys that question respondents about their experiences with crime, and they do not include homicide (for obvious reasons). However, the victimization survey does include questions about simple assaults, which are far more common than aggravated assaults or robberies and thus tend to statistically dominate the report. As Butterfield (2001b) pointed out, simple assaults accounted for 61.5% of all violent crimes identified in the victimization survey, and because they had declined by 14.4% in 2000 compared with 1999, they accounted for most of the decline in violent crime revealed in the victimization data. In short, and as the prominent criminologist Alfred Blumstein (as quoted in Butterfield, 2001b) noted, “[The data] are telling us that crime is very difficult to measure.”

A situation similar to the one noted above occurred in Britain in 2004, when the former Home Secretary commented, “the most reliable crime statistics—those recorded by the police—show that crime in England and Wales has risen by 850,000 in the past five years” (as quoted in Hough, 2004). However, the British Crime (Victimization) Survey indicated that, since 1995, house burglaries had declined by 47%, assaults by 43%, and wounding by

28%. It turned out that crime recorded by the police in Britain had increased due to changes in the way police counted crime—in particular, police data on violence included harassment and common assault charges that did not result in injury (Laurance, 2005). Commenting on the misrepresentation of crime data in Britain, especially in the popular media, Toynbee (2005) noted, “A vast industry of mendacity has a vested interest in scaring people witless with front-page shock, TV cops, and doom-laden moral panic editorials.”

Some five years after the alleged crime wave of 2001, another one was constructed in the United States. Based on Uniform Crime Report data from 2005, which indicated that, compared to 2004 figures, homicides had increased by 3.4%, robberies by 3.9%, and aggravated assaults by 1.8%, a Police Executive Research Forum (PERF) report asserted that “violent crime is accelerating at an alarming rate” (Rosen, 2006, p. 1). Los Angeles Police Chief (and President of PERF) William Bratton proclaimed, “we have a gathering storm of crime” (as quoted in Rosen, 2006).

As sometimes occurs in the construction of crime waves, the PERF report presented alarmist, and sometimes misleading, statistics in order to support the claim. For example, it noted that “last year, more than 30,600 persons were murdered, robbed, and assaulted than the year before” (Rosen, 2006, p. 2). Interestingly, however, the accompanying chart in the report indicated that more than 30,000 of that 30,600 total increase in crimes was for robbery and aggravated assault charges—the numerical increase in murders was 544. While acknowledging that several cities were not experiencing increases in crime, the report cautioned “but even in localities that continue to have flat or declining homicide rates, the escalating level of violence is manifesting itself in the rising number of reports of aggravated assaults and robberies in selected areas of cities” (Rosen, 2006, pp. 2–3).

The PERF report attributed this alleged increase in crime to a number of factors, including local, state, and federal cuts in funds allocated for crime fighting and prevention, an increasing number of prisoners being released back into society, and easy access to guns. In addition, Sheriff Bill Young of Las Vegas believed that “the influence of gangsta rap and some rap artists is having its effect on young people. He was not alone” (Rosen, 2006, p. 4).

In his foreword to the report, PERF executive director Chuck Wexler (2006) warned against complacency in addressing the alleged crime wave: “There are some in both academia and government who believe these increases in violent crime may represent just a blip and that overall crime is still relatively low. They argue that before we make rash conclusions we should wait and see if the violent crime rate continues to increase over time. This thinking is faulty. It would be like having a pandemic flu outbreak in a number of cities, but waiting to see if it spreads to other cities before acting. . . . The time to act is now” (p. ii).

Not surprisingly, the popular media devoted considerable attention to this alleged crime wave—several newspapers published articles on the issue,

frequently accompanied by alarmist headlines such as “Cities See Crime Surge as Threat to Their Revival” (El Nasser, 2007). An article in *USA Today* noted, “police are reporting spikes in juvenile crime as a surge in violence involving gangs and weapons has raised crime rates from historical lows early this decade” (Johnson, 2006).

However, the much ballyhooed crime wave did not materialize. While some cities, such as Philadelphia, did see increases in homicides in 2007 (Hurdle, 2007), nationally, violent crime decreased by 1.4% in that year compared to 2006, with most large cities showing the most significant declines (Sullivan, 2008). For example, New York City, which had 2,245 homicides in 1990, had 494 in 2007, the lowest total number since reliable statistics became available in 1963. As will be discussed in more detail in Chapter 3, crime in general, and violent crime in particular, has continued to decline in the ensuing years. But the impact of alarmist media reports regarding increases in crime cannot be understated. In 2009, despite continued decreases in crime in the United States, a Gallup poll found that 74% of Americans believed there was more crime that year than there was in the previous year. This represented the highest percentage of respondents believing that crime had increased since the early 1990s (Jones, 2009).

Statistics and numerical counts of social phenomena, including crime, have become a major fact of modern life. Countries are ranked in terms of statistical information on health, education, social welfare, and economic development. States, cities, counties, and individuals are compared on similar kinds of social indicators. Geographical areas, social groups, and individuals are judged as relatively high, low, or normal on the basis of various types of quantitative data.

Consider the increasingly pervasive rankings of cities on a number of dimensions, both in the United States and globally. For example, in 2010, *Men's Health* magazine included a list of “America's Fattest Cities,” with the rankings based on “the percentage of people who are overweight, the percentage with type 2 diabetes, the percentage who haven't left the couch in a month, the money spent on junk food, and the number of people who ate fast food nine or more times in a month” (Colletti & Masters, 2010). Using those criteria, Corpus Christi, Texas, was rated as America's fattest city, with Burlington, Vermont, and San Francisco, California, tied in 100th place (i.e., tied for the “least fattest” cities).

Another list included America's “craziest” cities, based on the number of psychiatrists per capita, the emotional and mental health of residents, eccentricity (“how crazy, wacky, and weird each city is, compiled with the help of a travel writer”), and the percentage of residents who were identified as “heavy drinkers” (“America's Craziest,” 2010). Cincinnati, Ohio, was ranked the craziest, and Salt Lake City, Utah, the least crazy, of the 57 cities rated on this list.

Cities have also been ranked with respect to “wastefulness” (based on a survey that asked residents about a range of behaviors from recycling to using

public transportation to turning off the lights when they leave a room)—of the 25 cities rated on this list, San Francisco was the least wasteful and Houston, Texas, the most wasteful (“Which U.S. Cities,” 2010). There are also media sources that rank cities for innovation (Fast Company), friendliness (NBC’s Today Show), for retaining Old West culture (American Cowboy; Briggs, 2008), and for being the best for singles (Sherman, 2009).

But as is the case with other forms of measurement, it is important to treat these numerous rankings with a degree of skepticism and to consider what factors are taken into account in establishing them. As Briggs (2008) commented, “with the proliferating pack of ‘best places’ lists, discrepancies are as common as corner coffee shops. One magazine or Web site may celebrate your city as a metro marvel, while another paints your burg as a gusher of civic flop sweat.”

He used Bethesda, Maryland, to illustrate his point. In 2008, *Fortune Small Business Magazine* ranked Bethesda as the fifth best place in the United States to “live and launch.” At roughly the same time, *Forbes* magazine ranked Bethesda as 104th of the “best places for businesses and careers.” Such differences are largely explained by differences in the dimensions and indicators used to establish the rankings.

## DATA FOR THOUGHT

Underlying many of the problems here is the simple fact that measurement is not passive, it often changes the very thing that we are measuring. And many of the measurements we hear every day, if strained too far, may have both caricatured the world and so changed it in ways we never intended. That limitation does not ruin counting by any means, but if you forget it, the world you think you know through numbers will be a neat and tidy illusion. (Blastland & Dilnot, 2009, p. 95)

As previously noted, we are constantly bombarded by statistics and data in the popular media. Consider the following examples, taken from a variety of sources.

- There were 41,518 injuries associated with a hammer in 1997. There were 44,335 injuries from toilets and 37,401 injuries from televisions in the same year (U.S. Bureau of the Census, 1999).
- Whooping cough deaths increased from 1,700 to 7,400 from 1980 to 1998. Deaths in the United States resulting from gonorrhea decreased from 100,400 to 35,600 in the same time frame (U.S. Bureau of the Census, 1999).
- Broccoli consumption increased from 1.4 pounds per capita in 1980 to 5.6 pounds per capita in 1998 (U.S. Bureau of the Census, 1999).

- Customers who buy premium birdseed are more likely to pay off their credit card bills than customers who buy chrome skull ornaments for the hood of their car (Flavelle, 2010).
- One out of 3 married women says their pets are better listeners than their husbands. Dog owners are more likely to declare their dog as the better listener than cat owners (25% vs. 14%; Petside Team, 2010).
- According to the American Association for Pet Obesity Prevention, over 45% of dogs and 58% of cats in the United States are currently estimated to be overweight or obese (Glynn, 2010). Dr. Ernie Ward's book *Chow Hounds: Why Our Dogs Are Getting Fatter—A Vet's Plan to Save Their Lives* (2010) provides instructions on how dog owners can "break the chow hound cycle."
- At the 2010 Winter Olympics in Vancouver and Whistler, British Columbia, 8,500 condoms were "airlifted" to the Olympic Villages after the initial supply of 100,000 distributed by the Canadian Foundation for AIDS research were nearly exhausted ("Go Figure," 2010).
- Under the No Child Left Behind Act, schools in the United States are required to report on their "adequate yearly progress" in achieving certain educational goals. In the 2008–2009 school year, the percentage of schools not making adequate yearly progress ranged from 6% in Wisconsin to 77% in Florida (Center on Education Policy, 2010).

At their face value, each of these types of statistical information may serve as a basis for social action. For example, this information may lead people to exhibit more care when using hammers or toilets, have greater concern with their nagging cough and less concern about particular sexually transmitted diseases, invest in broccoli, not issue credit cards or lend money to people who have skull ornaments on their car hoods, talk to their dogs when they have problems, put their pets on a diet regimen, attend the next Olympics if they wish to engage in sex, and enroll their children in schools in Wisconsin. It is also not uncommon for this type of numerical data to form the basis of public policy. In fact, public health programs, law enforcement, and other agencies rely on such descriptive statistics to implement various types of reform.

Before taking corrective actions based on such statistical information, however, it is important to consider several questions about its accuracy and how the data were collected. These questions about the measurement of social phenomena are often neglected in public discourse, but they ultimately will determine whether corrective action is necessary. For example, one's opinion about the presented statistics may change when one considers the following questions regarding the measurement of these social facts:

- How are injuries by hammers, toilets, and televisions counted? If a television repairperson hits a television with a hammer and it falls into the toilet and results in an electric shock to the repairperson, is this classified

as an injury by a hammer, toilet, or television? Do all agencies classify these injuries the same way? Note that these injury data are calculated from a sample of hospitals with emergency treatment departments. If people are injured by these products and do not go to an emergency ward, their injuries will not be counted. Under these conditions, the number of injuries by hammers, toilets, or televisions may be substantially higher or even lower, depending on how they are counted.

- Does the rise in whooping cough deaths reflect an actual increase in these fatalities or is it due to medical advances that have now made it easier to detect whooping cough as a medical problem? Is the dramatic decline in gonorrhea deaths due to improvements in medical care and early detection of this disease, or is it due to the reclassification of sexually transmitted disease (STD) deaths by medical personnel (i.e., a greater proportion of STD deaths are now attributed to AIDS)?
- Are we really measuring changes in the human consumption of broccoli or the amount of broccoli purchased per capita? For example, the increase in the last few decades in the number of exotic pets (such as iguanas) that eat broccoli may artificially inflate the estimates of human consumption. How are the figures for broccoli consumers who grow their own broccoli counted in data that are derived from grocery stores? Can an increase in a small number of super broccoli eaters underlie this increase instead of the apparent rise in the proportion of broccoli consumers over time?
- The data on credit card payments are collected by companies who analyze an increasing array of information on consumer habits in order to inform their own marketing strategies. But if a consumer has a skull ornament on the hood of his or her car and also purchases premium birdseed, is the consumer more or less likely to pay off their credit card bill?
- Why are women more likely to talk to their pets than men? Why are dogs deemed to be better listeners than cats?
- How is obesity in dogs and cats measured? What are the causes of obesity in dogs and cats, and why are cats more likely to be overweight and obese than dogs?
- Given that there were approximately 2,600 athletes participating in the 2010 Winter Olympics, and that the events were held over a 17-day period, can we assume that Olympic athletes had sex on average 2.5 times per day? Perhaps the condoms were adorned with Olympic logos and were taken home by athletes as souvenirs, or alternatively, perhaps they were used as balloons at celebration parties following competition.
- The vast discrepancy in data on schools making adequate yearly progress in Wisconsin and Florida (as well as schools in other states) may be only marginally related to the quality of schools in various states. Differences among states in the rigor of their standards, the content and difficulty of their tests, and the determination of the cut scores for proficient performance are more important factors.

As these examples illustrate, numerical measures of crime and other social phenomena have enormous potential to inform social scientists about their theories of human behavior, to provide politicians and legislators with an empirical basis for public policy decisions, and to help the general public structure their routine activities and how they live their lives. Unfortunately, however, many people who use these statistics are grossly uninformed about how they are collected, what they mean, and their strengths and limitations.

The goal of this book is to critically examine the various ways in which crime is measured and, thereby, to instill a healthy skepticism about the accuracy of current methods of counting crime. All social measurement involves human decisions, interpretations, and errors. By examining the sources of error in the measurement of crime, social scientists, legislators, and the general public will be in a better position to understand the utility of current theory and crime control practices that are derived from statistical data on crime. In later chapters, we address in considerable detail issues surrounding the three most commonly used measures of crime and delinquency: official data, self-report, and victimization studies. In this introductory chapter, we address the measurement of social phenomena in the context of the key concepts of reliability and validity.

## RELIABILITY, VALIDITY, AND SOURCES OF ERROR IN THE MEASUREMENT OF SOCIAL PHENOMENA

Stevens (1959) defined measurement as “the assignment of numerals to events or objects according to rule” (p. 25). The initial steps in measurement are to (1) clarify the concept one is interested in and (2) construct what is known as an operational definition of that concept. An individual’s social class is often operationally defined by income level, educational attainment is usually measured by years of formal education, sexual promiscuity is gauged by number of sexual partners, and political party preference (in the United States) is measured by one’s expressed attitudes toward Democrats and Republicans. As illustrated by these examples, the process of operationalization and measurement involves the attachment of a specific meaning to abstract concepts.

The accuracy of many measures of social phenomena, however, is both context and time specific. Sexual promiscuity, for example, was judged by different standards in the Victorian period of the 1800s, the “free love” era of the 1960s, and the current period. Similarly, our working definitions of crime are context and time specific. Prostitution, alcohol use, and drug use may be differently evaluated as “serious” crime, depending on the geographic location and historical period, the political circumstances, and the prevailing legal structures. Although illegal in the United States and punishable by death in some countries, prostitution is legal in the state of Nevada and in several other countries. The consumption and sale of alcohol are legal in most jurisdictions in the



present-day United States, but they were illegal during the Prohibition era (1919–1933). And although some of the most severe penalties in our criminal code are reserved for users of substances such as cocaine, marijuana, heroin, and methamphetamine, these substances were not illegal in the United States prior to the 20th century. Under these conditions, our choice of a particular working definition and unambiguous indicator of a concept such as crime becomes more difficult.

Selecting precise indicators of abstract concepts is a crucial step in attempting to operationalize any social phenomena. Within this process, two fundamentals of good measurement exist: *reliability* and *validity*. Reliability is concerned with questions related to the stability and consistency of measurement over repeated trials, and validity refers to the extent of congruence between the operational definition and the concept it purports to measure.

Reliability and validity are easily demonstrated when we consider the measure of intelligence. If a test of intelligence sometimes yielded a high intelligence quotient (IQ) and at other times a low IQ for the same individual, the test would be considered unreliable because it failed to achieve consistent results over repeated trials. An intelligence test would have questionable validity if there were differences in its ability to accurately measure the intellectual capacity of individuals from different cultures or races or both. In fact, one of the major criticisms of standardized intelligence tests is their low validity because they are not culturally sensitive (i.e., the test does not measure intelligence but instead indicates one's adaptation to middle class culture). Although a valid measure can be unreliable, a reliable measure is not necessarily valid (e.g., a thermometer is a reliable measure of temperature but an invalid measure of social class).

## RELIABILITY AND VALIDITY IN SURVEY RESEARCH

Many of the social measures and indicators we discuss in this chapter, and two of the most frequently used measures of crime and delinquency—self-report and victimization studies—rely on surveys of various segments of the general public to collect data and construct measures. A number of issues related to survey methodology should encourage caution in interpreting the results of studies employing this methodology. Among others, these include problems in sample and response rates to surveys, questionnaire format and wording, and interviewer effects.

Survey methodology is based on probability sampling theory. The basic principle is that a randomly selected, relatively small percentage of a population can be used to represent the attitudes, opinions, or behaviors of all people in the population if the sample is selected correctly. The key to being able to generalize to the larger population from a smaller sample is related to a fundamental principle in sampling theory known as *equal probability of selection*.

This simply means that each member of the population has an equal, or at least known, chance of being chosen to participate in the survey. It is instructive to discuss the principles of probability sampling in the context of the frequent public opinion polls conducted in the United States by organizations such as Gallup and Roper.

In telephone surveys conducted by such organizations, the usual goal is to generalize the results of the survey to all adults, 18 years of age and older, living within the continental United States (Newport, Saad, & Moore, 1997). However, such surveys generally do not cover individuals living in institutions, including college students who live on campus; armed forces personnel living on military bases; or prisoners, hospital patients, and others living in group settings or housing. The procedure that organizations such as Gallup use is to obtain a computerized list of all telephone exchanges in the United States, accompanied by estimates of the number of residential households attached to those exchanges. Then, through a procedure known as random digit dialing (RDD), a computer is used to generate a list of telephone numbers. This RDD procedure is important in the context of obtaining a representative sample because without it, the estimated 30% of the households in the United States that have unlisted phone numbers would not be included in the sampling frame. More recent challenges associated with conducting telephone surveys include caller identification, call blocking, "no call" lists, and the increasing number of individuals who use cell phones exclusively and do not have household phone lines (Dillman, Smyth, & Christian, 2009), among others. All of these changes have an impact on who will be reached via telephone surveys and, ultimately, the representativeness of the sample obtained.

The typical sample size for public opinion polls is between 1,000 and 1,500 respondents. However, the actual number of people interviewed for a survey is much less important than adherence to the equal probability of selection principle. As Newport et al. (1997) noted, if respondents are not selected according to equal probability of selection principles, it would be possible to conduct a survey with a million people that could turn out to be less representative of the population than a survey conducted with only 1,000 people.

The accuracy of estimates derived from these samples is also based on probability theory. With the typical sample size of 1,000, the results are highly likely to accurately represent the true population within a margin of error plus or minus three percentage points. For example, the results of a Gallup poll released in May of 1988 indicated that 64% of the U.S. public was familiar with the erectile dysfunction drug Viagra, which had been placed on the market only a few months earlier. This survey also indicated that 13% of the men interviewed indicated that they would like to try the drug within the next year. Interestingly, 15% of the women answered that they would like their husband to try Viagra within the next year (Saad, 1998). The margin of error indicates that the true rating of women who would like their husbands to try Viagra was somewhere between 12% and

18%. If the sample size for this survey was increased to 2,000, the results would be accurate within plus or minus two percentage points of the true population value, but the cost of conducting the survey would double.

Another important issue in assessing the reliability and validity of survey results is related to rates of response—what is also referred to as *contact and cooperation* (Singer & Presser, 1989)—the correspondence between the sample elements selected and those actually interviewed. In recent years, survey researchers have become concerned about the declining response rates to surveys, which can result in biased samples and, thereby, inaccurate measures or estimates. At least part of the reason for the general public's lack of willingness to participate in survey research is the proliferation of entities, both private and government, engaged in survey research. For example, the number of telemarketing firms in the United States increased from 30,000 in 1985 to more than 600,000 in 1995, and according to industry sources, as of the late 1990s, more than 25 million solicitation calls were made in a single day (Bearden, 1998). According to a 1994 study, one in three potential respondents refuses to participate in a survey, and even for respondents who do participate in surveys occasionally, 38% had refused to participate in at least one survey in the previous year. More generally, it is estimated that from 1990 to 2000, the response rate to telephone surveys declined from approximately 40% to 15% (Lewis, 2000), and it has generally stabilized at that rate since (Dillman et al., 2009). In most cases, data resulting from surveys with poor response rates can be assumed to be unrepresentative and biased because the respondents are likely to be self-selected and different in a number of unknown ways from those who do not respond. Unfortunately, many researchers take whatever data they collect, analyze it, and derive conclusions without any consideration of the issue of nonresponse bias.

A prime example of the problems that can result from inattention to issues of nonresponse bias occurred in 1985, when the Committee on Health and Long-Term Care issued a report that referred to the abuse of elderly persons in the United States as “a national disgrace.” This report cited research claiming that 4%, or 1 million elderly persons, were victims of abuse each year. However, the estimate was based on a survey of 433 elderly residents of Washington, D.C., of whom only 73, or 16% of the original sample, responded. Three of these respondents, representing 4.1%, reported experiencing some form of psychological, physical, or material abuse. The report then extrapolated from this small and undoubtedly unrepresentative sample to assert that 1 million elderly people were victims of abuse, “thereby constructing a national epidemic out of these three incidents” (Gilbert, 1997, p. 112).

Also in the context of nonresponse bias, consider the apparently increasingly popular (to students, although probably not professors) websites that allow students to rate their professors, such as [ratemyprofessor.com](http://ratemyprofessor.com) and [myprofessorsucks.com](http://myprofessorsucks.com). A search on the [ratemyprofessor.com](http://ratemyprofessor.com) website for the three authors of this book found that Clayton Mosher had an “overall quality” rating

of 2.9 (on a scale of 5); Terance Miethe an overall quality rating of 3.9, and Timothy Hart a rating of 4.0 (the latter was the only one of the three who received a “hotness” rating). However, it is important to note that Mosher’s rating was based on a total of 8 ratings from the several thousand students he has taught since this website became available to students; Miethe’s was based on 50 ratings from the potential pool of several thousand students, and Hart’s on 12 ratings from several hundred potential students. Are the students who take the time to post ratings to these websites representative of all students? Would it be wise for students to choose courses and professors based on such ratings?

The United States Census Bureau has a high level of respect and is admired for the quality of its data collection policies and procedures. Census Bureau staff are well trained, many of the leading experts in research methodology have direct contact with the national agency, sampling designs are among the most sophisticated in the world, statisticians that work with Census staff possess state-of-the-art knowledge about population estimation, and rigorous pretesting is conducted before actual data collection begins. But even the census, conducted every 10 years in the United States, which is intended to represent a full enumeration of the population, is subject to nonresponse bias and other problems in counting the population.<sup>1</sup> As Barry (2010) noted, enumerating all residents in the United States is “something akin to counting the granules in an ever-filling, ever-leaking bucket of sand.”

In 1970, the first year that government officials administered the initial part of the census by mail, 83% of households returned the questionnaire. In 1980, the rate of return declined to 75%, and by 1990, it was only 65%. For the 2000 census, 67% of households that received the form returned it (Holmes, 2000). More important, these response rates vary across geographical regions of the United States and across different sociodemographic categories of the population. For instance, the Midwestern states of Iowa, Nebraska, and Wisconsin had mail participation rates of approximately 80% in the 2000 census (Davey, 2010), probably because of the fact that these states have a higher proportion of older white residents, who may see participation in the census as a civic duty (Yen, 2010). In particular areas of the Richmond Hill neighborhood of New York City, which has a large South Asian and Indo-Caribbean population, only about 40% of residents mailed back their census form in 2000 (Semple, 2010).

---

<sup>1</sup>Although the task of enumerating the entire population of a country such as the United States is enormous, consider the situation of China. In that country, with an estimated population of 1.3 billion (in the late 1990s), the census is conducted through face-to-face interviews by more than 10 million volunteers and government workers. In the past, social and political issues have affected the accuracy of population counts in China. Demographers estimate that between 20 million and 100 million people are left out of the count due to evasion of household taxes and child-bearing restrictions, which limit urban families to a single child and rural families to two children if their firstborn is a girl (Chang, 2000).

Due to nonresponse bias and other problems in enumerating the entire population, it is estimated that the 2000 census did not count between 1.6% and 2.7% of black residents and between 2.2% and 3.5% of Hispanics. A further 2.8% to 6.7% of Native Americans living on reservations were also not counted (Holmes, 2001). Interestingly, the population of one town in rural Pennsylvania was missed entirely in the 2000 census. The 14 people who live in the town of Slovenska Narodna Podpora Jednota apparently were not around when the census taker visited—they thought she would come back, but she did not. As a result, the town's population for the year 2000 is listed as zero ("A Pennsylvania Town," 2001). In total, it is estimated that the 2000 census did not count between 6.4 and 8.6 million people living in the United States.

The reverse problem with census data is that of overcounting. It was estimated that more than 4 million people were in fact counted twice in the 2000 census (Holmes, 2001). Those who are counted twice tend to be children of divorced parents, college students living away from home who independently fill out census forms but are also listed by their parents, and people with two homes who receive forms in the mail at both of their dwellings. This potentially large overcount is related to the fact that for the 2000 census, forms were available at convenience stores and government agencies, and respondents were able to provide information over the telephone.

The issues associated with an accurate enumeration of the population are by no means trivial because census data are used to determine how seats in the U.S. House of Representatives will be apportioned, to draw Congressional and state legislative district boundaries, to allocate more than \$400 billion in federal funds (Reamer, 2010) and significant amounts of state funds, to formulate a wide array of public policies, and to assist with planning and decision making in the private sector.

At the time of writing of this chapter, forms for the 2010 census were being mailed to approximately 120 million households in the United States. It is worthwhile to consider some of the changes in the 2010 census, as well as the continuing challenges associated with counting the U.S. population.

The estimated cost of the 2010 census was \$14 billion, with close to \$25 million of that amount devoted to advertising to encourage higher rates of participation. This advertising campaign included a \$2.5 million ad in the Super Bowl, ads that appeared during telecasts of the 2010 Winter Olympic games (Fahri, 2010), and sponsorship of a car in the NASCAR Sprint Cup race series (El Nasser, 2010). The Census Bureau claimed that the communications strategy "[would be] one of the most extensive and far-reaching marketing campaigns ever conducted in this country" (Saker, 2010), and they justified this advertising cost by noting that each percentage point increase in the number of households who mail back their census forms saves approximately \$85 million in follow-up costs (Fahri, 2010).

The Census Bureau also engaged in an aggressive and extensive outreach campaign to encourage participation among minority groups and others who

frequently are not counted in the census. Among the efforts associated with this campaign was the creation of more than 100,000 partnerships with church groups, a variety of ethnic associations, and service and fraternal organizations (Saker, 2010). In addition, census questionnaires were made available in English, Spanish, Chinese, Korean, Vietnamese, and Russian, and instructions on how to complete the forms were available in 59 languages (O'Keefe, 2010).

Despite the outreach efforts, among the concerns surrounding the 2010 census were that members of minority groups, especially Muslims and illegal immigrants, would be even more reluctant than in the past to answer and return the forms because of their fears that the Patriot Act, passed in response to the September 11, 2001 terrorist attacks, would be used to obtain individuals' census data (Bahrapour, 2010; O'Keefe, 2010; Semple, 2010). While census officials tried to allay these concerns, pointing out that individual information provided on census forms is not available to the public for 72 years and that any census employee who shares confidential information is subject to a \$250,000 fine and five years in prison (Mack, 2010b), it is worth noting that there is historical precedent for the improper sharing of census data. In World War II, the Census Bureau identified concentrations of people of Japanese ancestry in geographic units as small as city blocks and shared those data with War Department officials who used the information to select people of Japanese ancestry for internment in war camps (Holmes, 2000; Kopel, 2000; Seltzer & Anderson, 2001).

There were also concerns that young adults and college students would not fill out and return their census forms—many young adults who were living with their parents in 2000 would have no experience filling out census forms (Yen, 2010). College students are supposed to be counted as residents in the community where they attend college. However, they are particularly difficult to count because many of them are on spring break when census forms are mailed, and when census employees follow up with those who have not responded in May, many students may have returned to their community of residence (Marklein, 2010).

More generally, a significant number of conservatives, libertarians, and Tea Party supporters subscribe to the idea that the census should be nothing more than a count of the population and should not collect personal information (Mack, 2010a). In fact, a survey conducted by the Pew Research Center in March of 2010, approximately one week before census forms were mailed out, revealed that 12% of those surveyed did not intend to fill out and return their census forms (Pew Research Center, 2010). A *Washington Post* columnist (Dvorak, 2010) commented on the irony associated with this reluctance to fill out and return census forms: "We are a nation of people who will turn over our credit card numbers to someone on television guaranteeing rock-hard abs in 2 minutes a day. All too many of us are inclined to believe that a Nigerian lawyer will pay us handsomely if we just let him use our bank account to transfer a small fortune. And we have

no problem facebooking, twittering, or YouTubing our toe fungus issues, binge-drinking episodes, or childrens' transgressions to millions of others online. So what explains why some fear the U.S. census" (p. 1).

## RELIABILITY AND VALIDITY ISSUES RELATED TO THE QUESTIONNAIRE AND RESPONDENTS

A number of factors related to the survey instrument itself and to the individuals responding to survey questions affect the reliability and validity of results from this method of data collection. Three of these will be covered here: question wording effects, question order effects, and response effects.

### Question Wording Effects

A study of question wording effects using data from the General Social Survey (conducted annually in the United States by the National Opinion Research Center at the University of Chicago) compared two different versions of questions on government spending priorities and revealed systematic differences in responses. When respondents were asked if they supported increased spending on "welfare," only 32% answered in the affirmative. However, when respondents were asked whether there should be "more assistance for the poor," 62% favored increased spending (Smith, 1989). Similarly, four opinion polls conducted by different organizations in the summer of 2009 on support for a "public option" national health care plan in the United States revealed interesting question wording effects. A *New York Times/CBS News* survey found that 66% of Americans supported the plan; a *Time* magazine poll reported that 56% were in favor; a Pew poll found 52% supported it, while Fox News found 44% in favor. In the *New York Times/CBS* poll, the plan was explained as a "government administered health insurance plan—something like the Medicare coverage that people 65 and older get" (the other three surveys did not make reference to Medicare), the question in the *Time* survey asked about "a government sponsored public health insurance plan," Pew asked about a "government health insurance plan," while the question used in the Fox poll referred to a "government-run health insurance plan" (Sussman, 2010). Given these differences in question wording, the reported rates of approval across the four polls are perhaps not all that surprising. Another example of the effects of question wording and response options comes from studies examining support for capital punishment in the United States. An opinion poll conducted by Gallup in February of 2001 found that 67% of the U.S. population favored capital punishment. However, when interviewers asked whether the penalty for murder should be execution *or* life in prison with no possibility of parole, support for capital punishment declined to 54% (Jones, 2001).

### Question Order Effects

The order in which questions are asked can also have an impact on responses. For example, in a poll conducted before the 2000 U.S. presidential election to determine the popularity of candidates Al Gore and George W. Bush, respondents were asked to state their preference for president after having responded to a question that asked them to evaluate then-President Clinton “as a person.” This ordering of questions resulted in a lower level of support for Gore, probably because the question about Clinton reminded respondents of the Monica Lewinsky scandal and led them to disapprove of his vice-president as well. However, when the company conducting the poll reordered the question and surveyed a new sample, support for Gore increased (Harwood & Crossen, 2000).

### Response Effects

Data from the U.S. censuses are also relevant to the issue of response effects. One of the most important characteristics of the U.S. population that the census attempts to measure accurately is its racial composition.<sup>2</sup> Although race is a social construct, the racial composition of various jurisdictions in the United States has important implications for economic and social policies. The 2000 census was the first in which people in the United States were allowed to identify themselves as belonging to more than one racial group: The six racial categories created a total of 63 possible racial combinations for respondents to self-identify. Results from the 2000 census indicated that fully 6.8 million people identified themselves as multiracial, and although 93% of these classified themselves into only two racial categories, 823 respondents actually checked all six racial categories (Kasindorf & El Nasser, 2001). Interestingly, in the 2010 census, President Obama, who, given his racial background, had more than a dozen options in filling out the race question checked “African American,” prompting the *New York Times* to proclaim: “It is official, Barack Obama is the nation’s first black president” (Roberts & Baker, 2010). With respect to the same question, in both the 2000 and 2010 censuses, people who indicated that they were “Some other race” were asked to write in a particular race. Answers to the “Some other race” question in the 2000 census included Bolivian, Bushwacker, Cosmopolitan, and Aryan (Scott, 2001). One respondent to a *USA Today* article (“Our View,” 2010) on the 2010 census noted that for the race category, they had checked “Other” and wrote in “race for a cure.”

The American Indian category offers an interesting glimpse into the complications created by the change in census racial classifications. The number of

---

<sup>2</sup>In the first U.S. census, conducted in 1790, blacks were counted as three-fifths of a person and Native Indians were not counted at all (Anderson & Fienberg, 1999).



American Indians and Alaska Natives who defined themselves only by that racial category increased by 26% between 1990 and 2000. However, when the number of people who claimed they were part Indian is added, the total increased to 4.1 million, representing a 110% increase in the number of American Indians since 1990 (Schmitt, 2001). However, it is not clear whether all of those who identified themselves as Native American legitimately fall into that category. An informal survey conducted by a newspaper in Spokane, Washington, for example, found that some individuals marked the Native American category “as a way to tell the U.S. Census Bureau to mind its own business.” Others apparently identified themselves as Native American “because they were born in the United States” (McDonald, 2001). More important, racial composition data from the 2000 and 2010 censuses will not be directly comparable with previous census figures, and the ability to track the progress of racial groups with respect to their educational, occupational, health, and income characteristics will become far more problematic.

Although it may seem straightforward, even the classification of gender in a census or survey can be ambiguous. In Canada, a transsexual person refused to answer the question, “Are you male or female,” on that country’s 2001 census. This individual, who was born a male but was taking hormones and had breasts and male genitals, noted that “my gender was not listed” (Raphael, 2001).

A related problem has characterized the U.S. census with respect to identifying the number of households occupied by gay couples. In 1990, a person who shared a household with an individual of the same sex and also reported being married created a problem for census data-coders because the Census Bureau did not recognize same-sex marriages. To make the responses consistent, the Census Bureau changed either the person’s sex or his or her relationship to the other person because “if they said they were married and had a spouse of the same sex, the simple thing was to change the spouse’s sex. We made them a married couple” (Spencer, as quoted in Peterson, 2001). At least partially as a result of changes in this procedure in the 2000 census, such that gay and lesbian households could claim an unmarried partner and then identify his or her sex, there was a huge increase (Peterson, 2001) in the number of gay households identified in 2000, to approximately 600,000 (Crary, 2010).

It was predicted that a change in the 2010 census form that allowed same sex couples to check the “Husband or wife” boxes on the census form (rather than unmarried partner) would result in a further increase in the count of gay and lesbian couples (Turnbull, 2010). The Census Bureau also deployed a team of professional field workers to reach out to gays and lesbians and produced public service videos encouraging members of these groups to respond (Crary, 2010).

Errors in questionnaire data are also associated with response styles—the tendency to choose a certain category when responding to a question—regardless of the content of the item. For example, in the frequently used agree-disagree

format on questionnaires, some respondents may be characterized by an acquiescence response set: the tendency to agree with a question, regardless of its content (Singleton & Straits, 1999). A second response style is referred to as social desirability: the tendency to choose those response options most favorable to an individual's self-esteem or in accord with prevailing social norms, regardless of one's real position on the given question. Some have argued that social desirability effects may explain why comparisons of survey data over time reveal a general decline in overt expressions of racially prejudiced attitudes (Quillian, 1996).

Additional response problems are related to issues of memory, and in this context, two types of errors can be distinguished: forgetting and telescoping in time. With respect to telescoping, events and behaviors are reported as having happened more recently than they actually did. This form of response error is particularly relevant in the context of self-report and victimization surveys, which are addressed in Chapters 4 and 5 of this book.

The very real possibility also exists that respondents, for a number of different reasons, may be somewhat less than truthful in responding to questionnaires: The evidence regarding lying on questionnaires is well documented. In a 1950 study, Parry and Crossley asked individuals a number of questions in situations where the accuracy of their answers could be assessed. The proportion of honest answers ranged from 98% on a question asking whether the respondent had a telephone to approximately 50% on one that asked about their voting behavior. McCord (1951) similarly demonstrated that people sometimes lie when they are asked questions about things that do not exist: One-third of his sample claimed they had voted in a special election that was never held. An additional example suggesting that some respondents may be less than truthful in responding to questions comes from surveys of sexual behavior that ask respondents to estimate how many sexual partners they have had over the course of their lifetime—these surveys typically find that men report 2 to 4 times as many sexual partners as women (Brown & Sinclair, 1999). But if such surveys are eliciting accurate reports from respondents, heterosexual men and women should, on average, report having the same number of partners (because each new sexual partner for a male is also a new sexual partner for a female). Studies also suggest that between 33% and 45% of respondents will lie when they are asked about their level of education, about half will lie when they are asked whether they have received welfare assistance (Nettler, 1978), and fairly large percentages will lie when asked to report their age. In addition, some studies have suggested that the tendency to be less than truthful in answering questions may vary according to the racial or ethnic and gender characteristics of respondents (Mensch & Kandel, 1988; see also Chapter 4).

Some surveys of criminal behavior and drug use, which will be addressed in more detail in Chapter 4, have discovered that minority groups have a greater tendency to underreport these behaviors. One explanation of this tendency is that minorities feel more threatened or are made uneasy when asked to report on involvement in delinquent activities. Whatever the possible reasons for this

underreporting, researchers conducting studies and those reporting on the results of such studies need to be aware of the possibility of biases resulting from these tendencies.

A more general concern with respect to survey research is related to respondents' general knowledge. Public opinion polls have shown that many people in the United States are unaware that there are three branches of government; significant numbers of the U.S. population believe that Brazil is the capital of Ohio, and approximately 18% believe that the sun circles the earth ("Public Opinion," 1997). In a survey conducted by the National Campaign to Prevent Teen and Unplanned Pregnancy, 18% of American men aged 18–29 indicated they believed that standing up during sex is an effective form of contraception (Harper's, 2010).

In the 1989 General Social Survey, 61% of respondents did not feel they were able to rank the social standing of the "Wisian" ethnic group. However, 39% were able to rank this group, and they provided Wisians with a rather low average rating of 4.12 on a 9-point social ranking scale ("Wisians," 1992). Wisians were a fictitious ethnic group, added by designers of the General Social Survey to test the honesty of respondents in answering questions.

In short, all data derived from survey research are subject to reliability and validity problems. An intelligent consumer of such data will pay attention to these issues before uncritically accepting the findings from survey research.

## MEASURING CRIME AND DEVIANCE

We now move on to a consideration of issues that are more directly relevant to the main topic of this book: the measurement of crime, delinquency, and deviant behavior. We begin with a discussion of the problems associated with measuring crime on college campuses, followed by a consideration of how questionable measures of the extent of drug consumption have been used to create alleged drug epidemics with resulting policy changes.

### Measuring College Campus Crime

Since the 1990s, numerous states and the federal government have enacted laws requiring colleges and universities in the United States to publish crime statistics. (These statistics are available online at <http://ope.ed.gov/security>.) The first federal law related to this requirement, known as the Crime Awareness and Campus Security Act, was passed in 1990 (Port & Lesser, 1999). As is often the case with legislative proposals in the United States, this law was enacted primarily in response to the occurrence of a single event: the murder of 19-year-old Jeanne Clery at Lehigh University in Pennsylvania in 1986. Clery was a freshman who was assaulted and murdered while asleep in her residence room. When Clery's parents investigated the situation, they discovered that

Lehigh University had not informed students about 38 violent crimes that had been committed on the campus in the three years prior to their daughter's murder. The Clerys joined with other campus crime victims and persuaded Congress to enact legislation requiring all colleges and universities to publish statistics on the amount and type of crime occurring on their campuses.

As a result of subsequent amendments to this legislation in 1998, institutions must report the incidence of homicide, manslaughter, arson, rape, robbery, aggravated assault, burglary, motor vehicle theft, drug offenses, liquor law violations, and illegal weapons possession. In addition, institutions are required to provide greater detail regarding alleged hate crimes, defined by federal law as incidents that "manifest evidence of prejudice based on race, religion, sexual orientation, or ethnicity (Port & Lesser, 1999)." Campuses that do not comply with the legislation face the possibility of significant fines and the loss of federal student aid.

When data on college crime were first released in the early 1990s, several media outlets invoked rather alarmist language to describe the situation. For example, *U.S. News and World Report* ("Campus Crime," 1994), commenting on the 1993 statistics, alleged that there was an "epidemic" of college campus crime. Similarly, *USA Today* (Henry, 1996) referred to "steep increases in crime" in describing the 1994 campus crime statistics. But serious crime on college campuses is exceedingly rare when compared to overall crime rates in the United States—there is less than one homicide for every million students on campus in any given year in the United States.

Problems in the reliability and validity of campus crime data became apparent soon after the federal legislation was enacted. These problems ranged from confusion surrounding how to code particular crimes to outright manipulation of the statistics. A study conducted by the National Center for Education Statistics found that 40% of the colleges and universities were using federal definitions of crime to classify their data, 45% were using state definitions, and 15% were using definitions of their own design (Port & Lesser, 1999). A 1997 audit conducted by the U.S. General Accounting Office discovered that only 2 of the 25 colleges examined were correctly reporting their crime statistics. Among other omissions, some colleges were routinely excluding rapes and other sexual assaults that were reported to school officials but not to the police. For example, in September of 1999, the University of Florida admitted to withholding 35 rapes from its annual crime reports for the years 1996, 1997, and 1998. Instead of the 12 rapes that were recorded in the official report for this period, the university was aware of 47; however, university officials claimed that they believed that rapes reported to a victims' advocacy group should not be counted (Port & Lesser, 1999).

Perhaps the most notorious example of the manipulation of campus crime statistics occurred at the University of Pennsylvania. In 1996, this university reported 18 robberies in its federally mandated campus security report, whereas the police blotter indicated that 181 robberies had occurred. The apparent reason for this gross discrepancy was that the university had chosen to exclude crimes

that had occurred on sidewalks and streets that crossed the campus and in buildings located on campus that it did not own (Port & Lesser, 1999).

Anomalies in the officially recorded data and incidents such as the one that occurred at the University of Pennsylvania resulted in further amendments to the legislation. Beginning in 1998, institutions were required to report crimes occurring on public property that was “reasonably contiguous” to their campuses. Not surprisingly, there was initially considerable confusion on the part of university officials regarding what constituted reasonably contiguous property; it has since been defined as *public sidewalks, streets, and parking lots adjacent to a campus, or any public property running through the campus*.

Comparisons of crime data across college campuses in the United States suggest that universities are not adopting the same definitions of contiguous areas, however. For example, campus police at the University of Washington in Seattle expressed skepticism when the 1998 figures on campus crime were released. In that year, the University of Southern California, located in the middle of a high-crime area of South Central Los Angeles, recorded only 4 assaults, whereas the University of Washington recorded 93 (Rivera, 2000). In 1999, the University of Washington’s 127 drug arrests placed it fourth in the nation. However, campus police noted that the arrests sometimes involved street people and individuals who wandered onto the campus (Rivera, 2001). The perils associated with uninformed comparisons of these data are also revealed when we consider the situation of colleges and universities with branch campuses. The 1997 report for the University of Idaho, located in a rural area of the state, indicated that seven rapes had occurred on campus that year. However, the rapes had actually occurred at a smaller branch campus of the university, located in Coeur d’Alene. Similarly, Eastern Washington University, located in a largely rural area of Washington State, recorded 74 aggravated assaults in 1997, but the overwhelming majority of these had occurred in a contiguous area of the university’s branch campus in the heart of downtown Spokane (deLeon & Sudermann, 2000).

Two additional categories of campus crime to examine are those of alcohol and drug arrests. Between 1997 and 1998, alcohol arrests on college campuses increased by 24.3% nationally, whereas arrests for violations of drug legislation increased by 11.1%. However, campus law enforcement officials attributed these increases to tougher enforcement of existing drug and alcohol guidelines and changes in the previously mentioned reporting categories stipulating that colleges had to include crimes taking place in reasonably contiguous areas.

At the University of Wisconsin, where arrests for alcohol violations increased from 342 in 1997 to 792 in 1998, the campus police chief claimed that the 132% change was due to the hiring of more campus police officers who were more vigorous in enforcing the laws. At the University of North Carolina at Greensboro, which experienced more than a 700% increase in drug arrests between 1997 and 1998, the increases were attributed to the expanded geographical area for which crimes were recorded; of the 132 drug arrests in

1998, 88 occurred on public property near the campus and in 17 residence halls, areas the campus had not included in its 1997 report (Nicklin, 2000).

There has also been considerable confusion regarding the procedures for counting these drug and alcohol arrests. The University of New Hampshire at Durham was unable to meet the Department of Education's reporting deadline of October 24th for their 1997 and 1998 drug-arrest data. When officials at the university asked the Department of Education how to deal with this problem, they were told to record no offenses for these categories. As a result, an uninformal perusal of the official data for the University of New Hampshire would lead one to believe that the campus had no drug arrests in 1997 and 1998 and 124 in 1999, instead of what actually occurred—56 arrests in 1997 and 85 in 1998 (Nicklin, 2000).

In addition to the problems with respect to counting drug and alcohol crimes or offenses, stipulations in the legislation requiring institutions to report the number of campus disciplinary referrals for violations of alcohol, drug, and weapons violations have created further confusion. In the 1998 report, several institutions placed arrests and referrals in the same category, creating the illusion of a significant increase in these arrests. For example, Wake Forest University reported an increase from 8 to 298 for alcohol-related arrests between 1997 and 1998; however, officials at the university claimed they had made only one alcohol-related arrest—the remaining 297 were referrals (Nicklin, 2000).

More recently, a series of reports by the Center for Public Integrity (Lombardi, 2009, 2010; Lombardi & Jones, 2009) documented a wide discrepancy between universities' official data on sexual assaults and records kept by sexual assault counseling centers or other places on campuses where victims sought assistance. The Center conducted a survey of 152 crisis service programs and clinics on or near college campuses and received responses from 58 facilities. Forty-nine of these programs reported higher numbers of sexual offenses than were recorded in the universities' data. Institutions with some of the most glaring discrepancies included the University of West Virginia, whose sexual assault prevention program documented 46 sexual assaults, none of which were recorded in the university's annual security report, and the University of Iowa, whose victim advocacy program served 62 students, faculty, and staff who reported being raped or almost raped in the previous year, also none of which showed up in the official university report (Lombardi & Jones, 2009). More generally, in 2006, 3,068 colleges and universities (77% of the total) reported zero sexual offenses—it is likely that many of these institutions misclassified or simply chose not to report those crimes.

It is certainly true that, both historically and in the current context, sexual assault is one of the most underreported crimes for a number of reasons, including self-blame and the frequent insensitive handling of such cases by law enforcement. In the specific context of college campuses, Lombardi and Jones (2009) noted that several clinics reported higher sexual assault statistics than appeared in official data because they served clients beyond the student

population and also received reports from students who might have experienced sexual assault during spring break, which did not fall under the reporting requirements of the Clery Act. As James Alan Fox of Northeastern University commented, “Crime is difficult to measure anyway, but rape is the most difficult. On campus, a large share of the crimes are not stranger rape, they are date rape. I don’t think we’ll ever get a precise statistic. I don’t think colleges know, and I don’t think they’ll ever know. We’ll have an estimate which is an undercount” (as quoted in Mulvihill & Bergantino, 2010). However, the result is that the true incidence of such crimes on college campuses is minimized, and campuses on the surface may appear to be safer than they actually are.

In an apparent attempt to discourage the underreporting of crime by universities, the Department of Education has issued fines against some institutions in recent years. For instance, in April of 2005, Salem International University in West Virginia was fined \$200,000 after not reporting a single sexual offense in its Clery reports even though the school was aware of such offenses; and in June of 2008, Eastern Michigan University agreed to pay a fine of \$350,000, the largest fine ever under the Clery Act, for several violations, including the miscoding of rapes (Lombardi & Jones, 2009).

### **Crime Reporting in Public Schools**

While reports of crime on college campuses are clearly subject to accuracy problems, crime reports from public schools in the United States are arguably even less reliable. Under the 2001 No Child Left Behind Act, schools are required to report offenses so that the government can identify “persistently dangerous schools,” and parents are allowed to transfer their children out of schools designated as such. A report on schools’ reporting of crime noted the following: “Federal statistics grossly underestimate the extent of school crime and violence. Public perception tends to overstate school crime and violence. Reality exists somewhere in between—but statistically, nobody knows exactly where this ‘somewhere’ is in numbers” (National School Safety and Security Services, n.d.).

Underreporting of crime has been uncovered in numerous school districts across the United States—we provide just a few examples here. In Colorado in 2003–2004, the largest school district (Jefferson County, with 85,000 students) reported 644 assaults and fights, but in the following year, the district reported zero assaults and fights (Olinger, 2005). In addition, one middle school in Colorado reported more assaults than all the schools in the state’s eight largest school districts combined, and one grade school in Denver reported three times as many assaults as any high school in the same city. In New Jersey, one in five school districts reported no violent offenses in 2004–2005, and in Philadelphia, a 180,000 student district reported only one incident of theft to the state but listed more than 1,000 in its own annual report (Hardy, 2006). School crime data from Seattle also appear to be inaccurate. An analysis of two years of

school district databases on crime listed more than 1,000 violent incidents, including assaults, threats, robberies, and weapons possession, that were not reported to the police (Heffter, 2007).

The reasons for this underreporting are myriad, and they include differences in the definition of crime across various school districts and schools. In addition, given the implications of being labeled a persistently dangerous school—that is, having students transfer out of the school, resulting in a loss of funding—school administrators and principals may be pressured to underreport or simply not report school crime and violence. It is also important to note that the persistently dangerous component of the No Child Left Behind Act does not provide funding to assist schools identified as such to improve their safety programs (National School Safety and Security Services, n.d.).

### **Drugs and Drug Epidemics**

Illegal drugs have been a major concern of policy makers in the United States since the beginning of the 20th century. And as is the case in other areas of social, economic, and crime policies, competing interests rely on both official and unofficial data to support their respective agendas.

Prior to the 1996 presidential election, incumbent President Bill Clinton presented data from victimization surveys to suggest there had been a 9% decrease in violent crime in the United States, and he claimed that the decline was due to the effectiveness of his administration's crime policies. Republican candidate Bob Dole saw things differently, and he used self-report data from the Federal Department of Health and Human Services to blame Clinton for a doubling of drug use among teenagers. However, the questions used in the 1994 survey that led Dole to attack Clinton were very different from those used in previous surveys of drug use, and the agency could not ensure that it had successfully adjusted for those differences. Even more important, many of the increases in drug use to which Dole referred were not statistically significant. Heroin use by teenagers, for example, superficially doubled from 0.3% in 1994 to 0.7% in 1995, but the actual number of youth reporting heroin use in the sample of 4,600 surveyed had only increased from 14 to 32 (Schoor, 1996).

An additional example of the confusion that can be caused by uninformed comparisons of drug use statistics comes for the 1999 report of the Office of National Drug Control Policy. That report claimed that there were 1.5 million people in the United States who had used cocaine in the previous month. However, the same document claimed that 3.6 million people in the United States had used cocaine in the past week (Caulkins, 2000). Clearly, these estimates are highly inconsistent and difficult to reconcile. The explanation for the large discrepancy in these estimates is that the first was based exclusively on data from the National Household Survey on Drug Abuse, whereas the latter included data from the Drug Use Forecasting program, which collects self-reports of drug use among arrestees in local jails in a number of jurisdictions in



the United States—such individuals are much more likely to use drugs. (For further discussion, see Chapter 4.)

Questionable official and unofficial data on drug use are frequently used to justify changes in drug policies. An interesting example of this phenomenon occurred in 2000 and 2001, when the popular media published hundreds of articles on an alleged epidemic in the use of the drug ecstasy (MDMA). A March 5, 2001 editorial, written by former federal drug czar William Bennett (2001), claimed that “while the crack cocaine epidemic of the 1990s has passed, methamphetamine and ecstasy are growing in popularity, especially among the young.” Bennett did not provide statistics, official or otherwise, to support his claim of this increase in the use of ecstasy. However, a survey that was cited widely in the media, conducted under the auspices of the Partnership for a Drug Free America, reported that the percentage of teenagers using ecstasy had doubled between 1995 and 2000—from 5% to 10%.

Given the paucity of additional self-report data on the use of ecstasy, especially by adults, media sources relied on alternative measures, such as reported seizures of ecstasy tablets, reports of law enforcement officials, and emergency room admission data, to support their claim of an “alarming explosion” (Rashbaum, 2000) in the use of MDMA. The commissioner of the U.S. Customs Service claimed that seizures of ecstasy by his agency had increased from 350,000 pills in 1997 to 3.5 million in 1999, then to 2.9 million in just the first two months of 2000. He projected that seizures would amount to 7 or 8 million by the end of 2000. An *Associated Press* article (Hays, 2000) suggested that “seizures of the tablets . . . have multiplied like rabbits.” An article in *USA Today* (“Crackdown,” 2001) noted that “ecstasy, a drug once used primarily at nightclubs, has expanded beyond the club scene and is being sold at high schools, on the street, and even in coffee shops in some cities.” The source of these claims of ecstasy use spreading to previously unknown contexts was an informal convenience survey of officials in 20 cities in the United States, 80% of whom said that ecstasy was “more available than ever.” An additional measure of the alleged increase in ecstasy use came from the federal Drug Abuse Warning Network (DAWN), which tracks hospital emergency room admissions. Rashbaum (2000) reported that mentions of ecstasy in this source increased from 60 in 1983 to 637 in 1997 (the latest year for which statistics were available at the time).

Despite the questionable validity of the statistics used to document this ecstasy epidemic, in March of 2001, the U.S. Sentencing Commission enacted harsh new penalties for MDMA. These penalties treat ecstasy offenders more severely than cocaine offenders, resulting in a five-year sentence of incarceration for individuals selling 200 grams (approximately 800 pills) of the substance and a 10-year sentence for those selling 2,000 grams or more (Lindesmith Center, 2001). These legislative changes were enacted despite the opposition of many medical experts and researchers, who argued that the use of the substance was far less likely to cause violence than drugs such as alcohol and was less

addictive than cocaine or tobacco. Advocates of the increased penalties argued that these were necessary to curb ecstasy use by teenagers and young adults ("Sentencing Guidelines," 2001).

Apparently, ecstasy also became a serious problem in Canada in the late 1990s and early 2000s. In May of 2000, a drug enforcement officer from Toronto claimed, "I believe ecstasy has reached epidemic proportions in this country" (as quoted in Godfrey, 2000). Given similar problems with respect to the availability of current statistics on the actual extent of ecstasy use, the Canadian media also relied extensively on seizure figures to support the claim that ecstasy use had increased. In an article in the *National Post*, Grey (2000) reported that seizures of ecstasy in Canada had doubled between 1998 and 1999. Police across the country seized 712,000 ecstasy tablets in 1999, with an estimated street value of between \$17.8 million and \$28.5 million. The article also claimed that it was becoming "common knowledge" among law enforcement officials and researchers that ecstasy was "the drug of choice across demographic lines."

In May of 2000, several Canadian newspapers announced that the largest seizure of ecstasy in Canadian history had taken place at Pearson International Airport in Toronto. Police reported that they had seized 170,000 ecstasy tablets, valued at \$5 million. However, it turned out that police had made a mathematical error in their calculations, weighing the quantity of pills per pound instead of per kilogram. Thus, the actual seizure was 61,000 tablets, valued at \$1.8 million. Ben Soave, a superintendent for the Royal Canadian Mounted Police, noted, "It's one of those unfortunate situations. It was an error that we made and we're only human. So I apologize for that" (as quoted in Alphonso, 2000). The ecstasy problem was given further publicity when testimony given at an inquest into the death of a Toronto youth alleged that 13 deaths had been caused by the substance during a three-year period beginning in 1998. Although these ecstasy-related deaths were widely published in the media, it was eventually determined that seven of the deaths were the result of individuals using drug *cocktails*, mixtures of heroin, cocaine, and methadone (Freed, 2000). Although no specific federal or provincial legislation was enacted in Canada to deal with the ecstasy "problem," a Raves Act for the city of Toronto was proposed in May of 2000. This legislation would have defined a *rave* as a dance event occurring between 2:00 a.m. and 6:00 a.m. for which admission was charged. The law would have increased police powers of arrest in situations where drugs were sold at such events and allowed them to terminate the event if illegal acts were occurring (Freed, 2000). We need to question whether it is good public policy to change laws based on such questionable data.

Around the same time as claims of an ecstasy epidemic were being made in the United States, numerous media, government, and Internet sources were also reporting that a methamphetamine epidemic was occurring. Then, President Clinton referred to methamphetamine as the "crack of the 90s," and federal Drug Czar Barry McCaffrey commented, "Methamphetamine has exploded from a west coast biker drug into America's heartland and could

replace cocaine as the nation's primary drug threat" (as quoted in Pennell, Ellet, Rienick, & Grimes, 1999).

In addition to government assertions of an emerging methamphetamine epidemic, a number of popular media sources made similar claims. It was alleged that methamphetamine had "ravaged the state [of Missouri] for more than a decade, ensnaring young and old, businessmen, housewives, and entire families" (Pierre, 2003). Perhaps most prominently, a 2005 *Newsweek* article, "America's Most Dangerous Drug," used data from the U.S. National Household Survey on Drug Use and Health (also see Chapter 4) and claimed that in 2004, there were 1.5 million regular users of methamphetamine in the United States (Jefferson, 2005). However, this figure was based on survey respondents who reported that they had used methamphetamine at least once in the previous year. As noted by Gillespie (2005), it is questionable whether use of a substance in the past year is equivalent to "regular use": "Are you a regular user of liquor if you've had one drink in the past year?"

The *Newsweek* (Jefferson, 2005) article also reported on data from a telephone survey of 500 law enforcement agencies conducted by the National Association of Counties (NAOC): 58% of those responding said that methamphetamine was "their biggest drug problem." However, as Gillespie (2005) pointed out, responses were likely influenced by the preface to the survey, which stated, "As you may know, methamphetamine use has risen dramatically in counties across the nation." In addition, there are questions surrounding the methodology of the NAOC survey because it provided no information regarding response rates or how representative the sample of 500 counties was of the more than 3,000 counties in the United States.

A report on a second survey of hospital emergency rooms by the NAOC provided additional "evidence" of the emergence of the methamphetamine epidemic, with the claim that there was a 73% increase in meth-related emergency room visits between 2000 and 2005. However, this finding was based on 200 responses, representing less than 5% of the 4,079 emergency departments in the United States. And of these 200 responses, 161 were from emergency departments serving rural areas with populations of less than 50,000, despite the fact that 58% of all emergency departments are in metropolitan areas (Shafer, 2006).

In addition to the questionable use of data to construct a methamphetamine epidemic, the drug was also portrayed as a particularly dangerous substance<sup>3</sup> in both the popular media and government sources. For example, the federal government's Drug Enforcement Administration's website included a link to "Meth is Death," a site sponsored by the Tennessee District Attorneys General Conference. This site claimed that "one in seven high school students will try

---

<sup>3</sup>In our discussion of methamphetamine, we are by no means trying to minimize its devastating effects. Instead, our purpose is to focus on how the extent of the methamphetamine problem has been grossly overstated, primarily through the misrepresentation of data.

meth”; “99 percent of users are hooked after the first try”; “only five percent of meth addicts are able to kick it and stay away”; and “the life expectancy of a habitual meth user is only five years.” Sullum (2005) encouraged a critical examination of these statistics:

Do the math . . . and you will see that 13.4% of Americans die as a result of methamphetamine abuse within five years of graduating from high school. According to the Census Bureau, there are more than 20 million 15- to 19-year-olds in the U.S., so we are talking about hundreds of thousands of deaths a year, and that’s not even counting people who start using meth *after* high school.

In an opinion/editorial published in the *Los Angeles Times*, Kleiman and Satel (1996) noted that “in the case of methamphetamine, there is no need for the exaggeration that has created a credibility problem for other drug campaigns.” It is certainly important that we not underestimate the very real problems caused by substance use, but at the same time, our ability to effectively deal with these problems is not helped through the misrepresentation and manipulation of data.

## CONCLUSION

To conclude, it is clear that the data we have addressed in this chapter are subject to reliability and validity problems and are also subject to varying interpretations. And although such data are frequently used to draw attention to social problems and issues, to explain theoretically the causes of these problems, and to influence policies to deal with them, it is important to remain critical of the construction of these measures. As Campbell (as cited in Johnston & Carley, 1981) argued, “The more any social indicator is used for decision-making, the more subject it will be to corrupting pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.” More specific to crime data, Gurr (1977) noted that in the interpretation of crime data, it is necessary to “disentangle the social reality of behavioral change from the political and administrative reality of change in the institutions which respond to and record behavioral change” (p. 117).

## THE DESIGN OF THIS BOOK

How we measure, but too often mismeasure, crime and delinquency or deviance is the general topic of this book. This chapter introduced the primary constraints on constructing useful measurements of any social phenomena while illustrating those constraints with specific examples from research efforts on crime and deviance. Chapter 2 offers a historical overview of the measurement of crime,

paying particular attention to the evolution from official data to self-report studies to victimization surveys. We will see that many of the shortcomings characteristic of contemporary data on crime were identified by social scientists writing in the late 1800s and early to mid-1900s. Chapter 3 focuses on official measures of crime based on data compiled by local, state, and federal law enforcement agencies and examines problems associated with the collection and interpretation of these data. Chapter 4 reviews self-reported measures of criminal as well as deviant activity and, in the process, provides an overview of the methodology—survey research—used to collect both self-report and victimization data on crime. Chapter 5 describes victimization surveys in detail, highlighting the advantages as well as the disadvantages of measuring crime on the basis of victim accounts. The final chapter of the book moves us from the realm of describing various measures (and mismeasures) of crime and delinquency or deviance to the province of applying these measures to particular situations. Here we see that accurate and appropriate measurement is absolutely essential in testing explanations or theories of criminal behavior and in developing crime prevention or control policies.

### REVIEW QUESTIONS

1. Based on the principles of survey research identified in this chapter, describe the necessary factors to consider before you trust results that are reported in public opinion surveys.
2. As noted in this chapter, despite the fact that crime rates have generally been declining for the past few decades, a significant percentage of United States residents believe that crime is increasing. What explains this phenomenon?

### INTERNET EXERCISES

1. Go to the Internet and search for a ranking of cities. Look through the fine print and find a description of the methods that were used to create the rankings. Does a closer examination of the methods used affect your opinion about the reliability and validity of these rankings?
2. Go to the U.S. Census Bureau website (<http://www.census.gov>) and examine response rates for the 2010 census across states. What factors explain the higher and lower response rates across states?
3. Access the Department of Education website, which lists crime statistics (<http://ope.ed.gov/security>), and compare reported crime statistics for your institution with at least one other institution. Why is crime on your campus lower or higher than at the comparison institution?