2

SOFT FACTS

Sources of Bad Statistics

A child advocate tells Congress that 3,000 children per year are lured with Internet messages and then kidnapped. Tobacco opponents attribute over 400,000 deaths per year to smoking. Antihunger activists say that 31 million Americans regularly "face hunger." Although the press tends to present such statistics as facts, someone, somehow, had to produce these numbers. But how? Is there some law enforcement agency that keeps track of which kidnappings begin with online seductions? Are there medical authorities who decide which lung cancer deaths are caused by smoking, and which have other causes, such as breathing polluted air? Who counts Americans facing hunger—and what does "facing hunger" mean, anyway?

Chapter 1 argued that people produce statistics. Of course they do. All human knowledge—including statistics—is created through people's actions; everything we know is shaped by our language, culture, and society. Sociologists call this the *social*

construction of knowledge. Saying that knowledge is socially constructed does not mean that all we know is somehow fanciful, arbitrary, flawed, or wrong. For example, scientific knowledge can be remarkably accurate, so accurate that we may forget the people and social processes that produced it. I'm writing this chapter on a computer that represents the accumulation of centuries of scientific knowledge. Designing and building this computer required that people come to understand principles of physics, chemistry, electrical engineering, computer science—who knows what else? The development of that knowledge was a social process, yet the fact that the computer works reliably reflects the great confidence we have in the knowledge that went into building it.

This is one way to think about facts. Knowledge is factual when evidence supports it and we have great confidence in its accuracy. What we call "hard fact" is information supported by strong, convincing evidence; this means evidence that, so far as we know, we cannot deny, however we examine or test it. Facts always can be questioned, but they hold up under questioning. How did people come by this information? How did they interpret it? Are other interpretations possible? The more satisfactory the answers to such questions, the "harder" the facts.

Our knowledge about society tends to be "softer" than our knowledge of the physical world. Physicists have far more confidence in their measurements of the atomic weight of mercury than sociologists have in their descriptions of public attitudes toward abortion. This is because there are well-established, generally agreed-upon procedures for measuring atomic weights

and because such measurements consistently produce the same results. In contrast, there is less agreement among social scientists about how best to measure—or even how to define—public opinion.

Although we sometimes treat social statistics as straightforward, hard facts, we ought to ask how those numbers are created. Remember that people promoting social problems want to persuade others, and they use statistics to make their claims more persuasive. Often, the ways people produce statistics are flawed: their numbers may be little more than guesses; or the figures may be a product of poor definitions, flawed measurements, or weak sampling. These are the four basic ways to create bad social statistics.

GUESSING

Activists hoping to draw attention to a new social problem often find that there are no good statistics available.* When a trouble-

*While activists are particularly likely to face this problem (because they often are the first to try to bring a problem to public attention), anyone trying to promote a new social problem—including experts, officials, and those representing the media or other institutions—may have the same difficulties. Just as I sometimes use the general terms "advocates" or "promoters" to refer to all the sorts of people who help create social problems, I use "activists" to suggest that they are the ones especially—but not uniquely—likely to handle statistics in particular ways.

some social condition has been ignored, there usually are no accurate records about the condition to serve as the basis for good statistics. Therefore, when reporters ask activists for facts and figures ("Exactly how big is this problem?"), the activists cannot produce official, authoritative numbers.

What activists do have is their own sense that the problem is widespread and getting worse. After all, they believe it is an important problem, and they spend much of their time learning more about it and talking to other people who share their concerns. A hothouse atmosphere develops in which everyone agrees this is a big, important problem. People tell one another stories about the problem and, if no one has been keeping careful records, activists soon realize that many cases of the problem—maybe the vast majority—go unreported and leave no records.

Criminologists use the expression "the dark figure" to refer to the proportion of crimes that don't appear in crime statistics.[1] In theory, citizens report crimes to the police, the police keep records of those reports, and those records become the basis for calculating crime rates. But some crimes are not reported (because people are too afraid or too busy to call the police, or because they doubt the police will be able to do anything useful), and the police may not keep records of all the reports they receive, so the crime rate inevitably underestimates the actual amount of crime. The difference between the number of officially recorded crimes and the true number of crimes is the dark figure.

Every social problem has a dark figure because some instances

(of crime, child abuse, poverty, or whatever) inevitably go unrecorded. How big is the dark figure? When we first learn about a problem that has never before received attention, when no one has any idea how common the problem actually is, we might think of the dark figure as being the entire problem. In other cases where recordkeeping is very thorough, the dark figure may be relatively small (for example, criminologists believe that the vast majority of homicides are recorded, simply because dead bodies usually come to police attention).

So, when reporters or officials ask activists about the size of a newly created social problem, the activists usually have to guess about the problem's dark figure. They offer estimates, educated guesses, guesstimates, ballpark figures, or stabs in the dark. When *Nightline*'s Ted Koppel asked Mitch Snyder, a leading activist for the homeless in the early 1980s, for the source of the estimate that there were two to three million homeless persons, Snyder explained: "Everybody demanded it. Everybody said we want a number. . . . We got on the phone, we made a lot of calls, we talked to a lot of people, and we said, 'Okay, here are some numbers.' They have no meaning, no value."[2] Because activists sincerely believe that the new problem is big and important, and because they suspect that there is a very large dark figure of unreported or unrecorded cases, the *activists' estimates tend to be high,* to err on the side of exaggeration. Their guesses are far more likely to overestimate than underestimate a problem's size. (Activists also favor round numbers. It is remarkable how often their estimates peg the frequency of some social problem at one [or two or more] million cases per year.[3])

Being little more than guesses—and probably guesses that are too high—usually will not discredit activists' estimates. After all, the media ask activists for estimates precisely because they can't find more accurate statistics. Reporters want to report facts, activists' numbers look like facts, and it may be difficult, even impossible to find other numbers, so the media tend to report the activists' figures. (Scott Adams, the cartoonist who draws *Dilbert,* explains the process: "Reporters are faced with the daily choice of painstakingly researching stories or writing whatever people tell them. Both approaches pay the same."[4])

Once a number appears in one news report, that report is a potential source for everyone who becomes interested in the social problem; officials, experts, activists, and other reporters routinely repeat figures that appear in press reports. *The number takes on a life of its own, and it goes through "number laundering."*[5] Its origins as someone's best guess are now forgotten and, through repetition, it comes to be treated as a straightforward fact—accurate and authoritative. Soon the trail becomes muddy. People lose track of the estimate's original source, but they assume the number must be correct because it appears everywhere—in news reports, politicians' speeches, articles in scholarly journals and law reviews, and so on. Over time, as people repeat the number, they may begin to change its meaning, to embellish the statistic.

Consider early estimates for the crime of stalking.[6] Concern about stalking spread very rapidly in the early 1990s; the media publicized the problem, and most state legislatures passed anti-stalking laws. At that time, no official agencies were keeping

track of stalking cases, and no studies of the extent of stalking had been done, so there was no way anyone could know how often stalking occurred. After a newsmagazine story reported "researchers suggest that up to 200,000 people exhibit a stalker's traits,"[7] other news reports picked up the "suggested" figure and confidently repeated that there were 200,000 people being stalked. Soon, the media began to improve the statistic. The host of a television talk show declared, "There are an estimated 200,000 stalkers in the United States, and those are only the ones that we have track of."[8] An article in *Cosmopolitan* warned: "Some two hundred thousand people in the U.S. pursue the famous. No one knows how many people stalk the rest of us, but the figure is probably higher."[9] Thus, the original guess became a foundation for other, even bigger guesses (chapter 3 explores how repeating statistics often alters their meaning).[10]

People who create or repeat a statistic often feel they have a stake in defending the number. When someone disputes an estimate and offers a very different (often lower) figure, people may rush to defend the original estimate and attack the new number and anyone who dares to use it. For example, after activists estimated that there were three million homeless in the early 1980s and the Reagan administration countered that the actual number was closer to 300,000, the activists argued that the administration's figures could not be trusted: after all, the administration was committed to reducing expenditures on social programs and could be expected to minimize the need for additional social services.[11] Various social scientists set out to measure the size of the homeless population. When their findings confirmed that

the 300,000 figure was more reasonable, the social scientists came under attack from activists who charged that the research had to be flawed, that the researchers' sympathies must have been with the administration, not the homeless.[12] In general, the press continued reporting the large estimates. After all, activists and reporters knew that the actual number of homeless persons was much higher—didn't everyone agree that three million was the correct figure? This example suggests that *any estimate can be defended by challenging the motives of anyone who disputes the figure.*

In addition, *the dark figure often plays a prominent part in defending guesses.* There are always some hidden, unnoticed, uncounted cases and, because they are uncounted, we cannot know just how many there are. Arguing that the dark figure is large, perhaps very large ("The cases we know about are just the tip of the iceberg!"), makes any estimate seem possible, even reasonable. We know that some victims do not report rapes, but what proportion of rapes goes unreported? Is it two in three? Surveys that ask people whether they've been victimized by a crime and, if so, whether they reported the crime to the police, find that about two-thirds of all rapes go unreported.[13] But surely these surveys are imperfect; some rape victims undoubtedly refuse to tell the interviewer they've been victimized, so there still must be a dark figure. Some antirape activists argue that the dark figure of unreported rapes is very large, that only one rape in ten gets reported (this would mean that, for every two victims who fail to report their attacks to the police but tell an interviewer about the crimes, seven others refuse to confide in the

interviewer).[14] Such arguments make an impassioned defense of any guess possible.

Activists are by no means the only people who make statistical guesses. It is difficult to count users of illicit drugs (who of course try to conceal their drug use), but government agencies charged with enforcing drug laws face demands for such statistics. Many of the numbers they present—estimates for the number of addicts, the amounts addicts steal, the volume of illicit drugs produced in different countries, and so on—cannot bear close inspection. They are basically guesses and, because having a big drug problem makes the agencies' work seem more important, the officials' guesses tend to exaggerate the problem's size.[15] It makes little difference whether those promoting social problems are activists or officials: when it is difficult to measure a social problem accurately, guessing offers a solution; and there usually are advantages to guessing high.

There is nothing terribly wrong with guessing what the size of a social problem might be. Often we can't know the true extent of a problem. Making an educated guess—and making it clear that it's just someone's best guess—gives us a starting point. The real trouble begins when people begin treating the guess as a fact, repeating the figure, forgetting how it came into being, embellishing it, developing an emotional stake in its promotion and survival, and attacking those who dare to question what was, remember, originally just someone's best guess. Unfortunately, this process occurs all too often when social problems first come to public attention, because at that stage, a guess may be all anyone has got.

DEFINING

Any attempt to talk about a social problem has to involve some sort of definition, some answer to the question: "What is the nature of this problem?" The definition can be—and often is—vague; sometimes it is little more than an example. For instance, a television news story may tell us about a particular child who was beaten to death, and then say, "This is an example of child abuse." The example takes the place of a precise definition of the problem. One difficulty with this practice is that media coverage usually features dramatic, especially disturbing examples because they make the story more compelling. Using the worst case to characterize a social problem encourages us to view that case as typical and to think about the problem in extreme terms. This distorts our understanding of the problem. Relatively few cases of child abuse involve fatal beatings; comparatively mundane cases of neglect are far more common. But defining child abuse through examples of fatal beatings can shape how we think about the problem, and child-protection policies designed to prevent fatalities may not be the best way to protect children from neglect. *Whenever examples substitute for definitions, there is a risk that our understanding of the problem will be distorted.*

Of course, not all definitions of social problems depend on dramatic examples. People promoting social problems sometimes do offer definitions. When they do so, they tend to prefer general, broad, inclusive definitions. Broad definitions encompass more cases—and more kinds of cases. Suppose we want to define sexual violence. Certainly our definition should include

rapes. But what about attempted rapes—should they be included? Does being groped or fondled count? What about seeing a stranger briefly expose himself? A narrow definition—say, "sexual violence is forcible sexual contact involving penetration"—will include far fewer cases than a broad definition—for example, "sexual violence is any uninvited sexual action."[16] This has obvious implications for social statistics because *broad definitions support much larger estimates of a problem's size.**

No definition of a social problem is perfect, but there are two principal ways such definitions can be flawed. On the one hand, we may worry that a definition is too broad, that it encompasses more than it ought to include. That is, broad definitions identify some cases as part of the problem that we might think ought not to be included; statisticians call such cases *false positives* (that is, they mistakenly identify cases as part of the problem). On the other hand, a definition that is too narrow excludes cases that we might think ought to be included; these are *false negatives* (incorrectly identified as not being part of the problem).[17]

In general, *activists trying to create a new social problem view false negatives as more troubling than false positives.* Remember that activists often feel frustrated because they want to get people concerned about some social condition that has been

*Activists often couple big statistics based on broad definitions with compelling examples of the most serious cases. For example, claims about child abuse might feature the case of a murdered child as a typical example, yet offer a statistical estimate that includes millions of less serious instances of abuse and neglect.

ignored. The general failure to recognize and acknowledge that something is wrong is part of what the activists want to correct; therefore, they may be especially careful not to make things worse themselves by defining the problem too narrowly. A definition that is too narrow fails to recognize the problem's full extent; in doing so, it continues ignoring at least a part of the harm and suffering that ought to be recognized. Thus, activists might point to an example of a woman traumatized by a flasher exposing himself, and then argue that the definition of sexual violence needs to be broad enough to acknowledge the harm suffered by that woman. Activists sometimes favor definitions broad enough to encompass every case that ought to be included; that is, they promote broad definitions in hopes of eliminating all false negatives. Remember, too, that broad definitions make it easier to justify the big numbers advocates prefer.

However, broad definitions invite criticism. Not everyone finds it helpful to lump rape and flashing into a single category of sexual violence. Such broad definitions obscure important differences within the category: rape and flashing both may be unwanted, but classifying them together may imply they are equally serious. Worse, broad definitions encompass cases that not everyone considers instances of social problems; that is, while they minimize false negatives, they do so at the cost of maximizing cases that critics may see as false positives. Consider the long-running debate over the definition of pornography.[18] What ought to be considered pornographic? Presumably hardcore videos of people having sex are included in virtually all definitions. But is *Playboy* pornographic? What about nude sculp-

tures, or the annual *Sports Illustrated* swimsuit issue? Some antipornography activists may favor a very broad, inclusive definition, while their critics may argue that such definitions are too broad ("*That's* not pornography!").

Clearly, the definition of a social problem will affect statistics about that problem. The broader the definition, the easier it is to justify large estimates for a problem's extent. When someone announces that millions of Americans are illiterate, it is important to ask how that announcement defines illiteracy.[19] Some might assume that illiteracy means that a person cannot read or write at all, but the speaker may be referring to "functional illiteracy" (that is, the inability to read a newspaper or a map or to fill out a job application or an income tax form). Does illiterate mean not reading at all? Not reading at the third-grade level? Not reading at the sixth-grade level? Defining illiteracy narrowly (as being unable to read at all) will include far fewer people and therefore produce far lower statistical estimates than a broad definition (being unable to read at the sixth-grade level).

Often, definitions include multiple elements, each of which can serve to make the definition broader or narrower. Consider homelessness again. What should a definition of homelessness encompass? Should it include the *cause* of homelessness? If a tornado destroys a neighborhood and the residents have to be housed in temporary emergency shelters, are they homeless, or should we count only people whose poverty makes them homeless? What about the length of *time* spent homeless? Does someone who spends a single night on the streets count, or should

the label "homeless" be restricted to those who spend several (and if so, how many?) nights on the streets? Each element in the definition makes a difference. If we're counting homeless persons, and we count only those whose poverty made them homeless, we'll find fewer than if we include disaster victims. If we count those who were without a home for thirty days in the last year, we will find fewer homeless people than if our standard is only ten days, and using ten as a standard will produce a lower number than if we agree that even a single night on the streets qualifies someone to be considered homeless.

In fact, some advocates for the homeless argue that definitions based on these elements are far too narrow, and they offer even broader definitions.[20] They suggest that people who stay in the homes of friends or relatives—but who have no homes of their own—ought to be counted as homeless. Under this definition, an impoverished mother and child who never spend a night on the streets or in a shelter but who "double up" and live with relatives or another poor family ought to be counted as homeless. Obviously, using this broader standard to count cases will produce higher numbers than definitions that restrict homelessness to those living on the streets. Still other advocates argue that people whose housing is inadequate or insufficient also ought to be counted as homeless. This still broader definition will lead to even larger numbers. Calculating the number of homeless people (or illiterate people or acts of sexual violence) inevitably reflects our definitions.

In other words, statistics about social problems always de-

pend on how we define the problem. The broader the definition, the bigger the statistic. And, because people promoting social problems favor big numbers (because they make the problem seem bigger and more important), we can expect that they will favor broad definitions. Often, advocates justify broad definitions by emphasizing the importance of being inclusive. People who spend a single night on the streets (or who have to stay with friends, or who live in substandard housing) also suffer. Who are we to decide that their suffering shouldn't count? Clearly, advocates argue, these people deserve to be included when we speak of "homelessness."

There are, then, two questions about definitions that ought to be asked whenever we encounter statistics about social problems. First, how is the problem defined? It is all too easy to gloss over definitions, to assume that everybody knows what it means to be homeless or illiterate or whatever. But the specifics of definitions make a difference, and we need to know what they are. Second, is the definition reasonable? No definition is perfect. Definitions that are too narrow exclude false negatives (cases that ought to be included), while definitions that are too broad include false positives (cases that ought to be excluded). It is difficult to have a sensible discussion about a social problem if we can't define the problem in a way that we can agree is reasonable. But even if we cannot agree, we can at least recognize the differences in—and the limitations of—our definitions.

MEASURING

Any statistic based on more than a guess requires some sort of counting. Definitions specify what will be counted. Measuring involves deciding how to go about counting. We cannot begin counting until we decide how we will identify and count instances of a social problem.*

To understand the significance of measurement, let's begin by considering one of the most common ways social scientists measure social concern—survey research. Surveys (or polls) involve asking people questions, counting their answers, and drawing general conclusions based on the results. (Choosing which people to survey presents special problems that we'll consider in the next section, on sampling.) For example, we might ask 1,000 people whether they favor or oppose a new law; if we discover that 500 of the people asked favor the law and 500 oppose it, we might generalize from those findings and conclude that public opinion is about evenly split.

Although the media sometimes report survey results as though public issues involve clear-cut splits in opinions—implying that people either favor or oppose gun control, that they're either pro-choice or pro-life—this is an oversimplification. *Public attitudes toward most social issues are too complex to be*

*In fact, researchers recognize that what I'm calling measurement actually is a type of definition. They refer to *operational definitions,* that is, the operations one goes through to identify an instance of whatever is being defined.

classified in simple pros and cons, or to be measured by a single survey question. For example, surveys find that about 90 percent of Americans agree that legal abortions should be available to women whose health would be endangered by continuing their pregnancies (see Table 1). Pro-choice advocates sometimes interpret such results as evidence that most Americans support legalized abortion. However, surveys also find that only about 45 percent of Americans support abortion regardless of the woman's reason for wanting it, and pro-life advocates sometimes view this as evidence that most Americans oppose abortion on demand. Combining the responses to these questions (as well as others measuring attitudes toward abortion under various circumstances) reveals a more complex pattern of public opinion: there is a small, hard-core antiabortion faction (roughly 10 percent of the population) that opposes abortion under any circumstances; a larger minority (roughly 45 percent) that accepts women's right to choose abortion under almost any circumstances; and another large minority (roughly 45 percent) that occupies a territory between these extremes, that approves of abortion for "good" reasons but does not approve of all abortions, regardless of the circumstances. Attitudes toward abortion are too complicated to be measured by a single survey question or to be described in terms of simple pro/con categories. Obviously, then, measurement makes a difference. The choice of questions used to measure abortion attitudes affects what public opinion surveys discover.

Survey researchers know that *how questions are worded affects results.* Advocates who can afford to sponsor their own surveys

Table 1. *Percentages of Americans Favoring Legal Abortions under Different Circumstances, 1996*

If the woman's own health is seriously endangered by the pregnancy	92
If she became pregnant as a result of rape	84
If there is a strong chance of serious defect in the baby	82
If she is married and does not want any more children	47
If the family has a very low income and cannot afford any more children	47
If she is not married and does not want to marry the man	45
If the woman wants it for any reason	45

SOURCE: Data from the 1996 General Social Survey, from "The American Survey—Release 1997" (CD-ROM; Bellevue, Wash.: Micro-Case, 1997).

can shape the results; usually they try to demonstrate widespread public support for their position. (This is sometimes called *advocacy research.*) Advocates word questions so as to encourage people to respond in the desired way. For example, surveys by gun-control advocates may ask: "Do you favor cracking down against illegal gun sales?" Most people can be counted on to oppose illegal acts, and such questions routinely find that (according to the gun-control activists' interpretations of the results) more than three-quarters of Americans favor gun control. On the other hand, the National Rifle Association opposes gun control, and it sponsors surveys that word questions very differently, such as: "Would you favor or oppose a law giving

police the power to decide who may or may not own a firearm?" Not unexpectedly, most people answer that they oppose giving the police so much power, and the NRA can report that most Americans (roughly three-quarters) oppose gun control.[21] As in the example of abortion discussed above, public opinion seems to divide into minorities at the two extremes (some favoring a ban on all guns; others opposed to any gun control), and a large middle mass that, presumably, favors keeping guns out of the hands of "bad" people while letting "good" people have guns. However, the complexity of public opinion can be hard to recognize when our information comes from surveys sponsored by advocates who word questions to produce the results they desire.*

In addition to wording questions to encourage some responses, *advocates who conduct their own surveys can decide how to interpret the results.* A few years ago, the press reported that a national survey estimated that 2 percent of adult Americans (nearly four million people) had been abducted by UFOs. How did the researchers arrive at this figure? Did they ask: "Have you ever been abducted by a UFO?" No. The researchers argued that such a straightforward question would be a poor measure because many UFO abductees do not realize they've been abducted (or are unwilling to talk about the experience);

*Although I have chosen to focus on question wording, there are many other ways researchers can design surveys to encourage particular responses. For example, the order in which questions are asked can make a difference in how people respond.

therefore they could not (or would not) answer a direct question accurately. (Note that this is another instance of advocates trying to avoid false negatives; in this case, they did not want to measure abduction in a way that might exclude some cases they felt ought to be included.) Instead, the researchers devised a very different measure: they identified five indicators or symptoms that often figured in the accounts of people who say they've been abducted, and then asked whether respondents had experienced these more innocuous symptoms, for example: "Waking up paralyzed with a sense of a strange person or presence or something else in the room?" They then concluded that anyone who reported four or more symptoms probably had been abducted. Two percent of the survey respondents fell into this group, leading to the researchers' conclusion that 2 percent of the population had been abducted.[22]

This example illustrates the importance of measurement decisions. Measurement involves choices. Had the UFO researchers decided that only one or two symptoms indicated abduction, they would have found more abductees. Had they decided to insist that respondents report all five symptoms, they would have found fewer. (And, of course, had they decided to only count people who reported having been abducted, they presumably would have found fewer yet.) Such choices shape the results of many surveys. Based on affirmative answers to such questions as "Have you had sexual intercourse when you didn't want to because a man gave you alcohol or drugs?" one survey concluded that roughly a quarter of female college students had been raped.[23] Critics challenged this finding; they

argued that the questions were ambiguous, and noted that nearly three-quarters of the respondents identified as rape victims indicated they did not consider the incident a rape.[24] But, as these examples demonstrate, it is the advocates conducting the surveys—not the respondents—who create the measurements and interpret the results, who identify the victims of UFO abduction or rape.

Activists justify such measurement decisions as efforts to reveal the true scope of social problems. Remember that activists usually believe that the problem they seek to bring to public attention is both large and largely unrecognized, that there is a substantial dark figure of hidden cases. They design their research to shed light into this darkness; they try to collect data that will reveal the true (substantial) extent of the problem. Therefore, *they devise measurements that will minimize false negatives.* Survey researchers know many techniques for encouraging respondents to give the desired answers. In addition to wording questions carefully, for example, one researcher advises that surveys can elicit more reports of violence against women by: defining violence broadly; asking women about violent incidents throughout their lives (rather than focusing on only, say, the previous year); asking multiple questions about victimization (rather than having a single question); asking open-ended questions that invite wide-ranging responses; and employing sympathetic female interviewers.[25] And, of course, the researcher has the ultimate power to decide which responses indicate the presence of the social problem.

Often, *measurement decisions are hidden.* The media report

statistics ("Research shows that . . .") without explaining how the study measured the social problem. These reports usually ignore controversies about measurement, and even well-established measures can be controversial. For example, the U.S. Bureau of the Census calculates the poverty line—the income levels below which people are considered poor. First established in 1964, the poverty line involves a set of calculations: first, the government determined a food budget for a nutritionally sound diet for a family of four; then, assuming that families spend one-third of their income on food, the bureau multiplied that figure by three. This produced the total income for the base poverty line (which was adjusted for families of different sizes). Each year, the poverty line is multiplied by the change in the Consumer Price Index to take inflation into account. Obviously, the poverty line is an arbitrary standard; different assumptions could be used to justify setting the line higher or lower. In fact, every element in the current formula has come under attack. Critics have argued that the food budget is unrealistic, that the poor do not spend one-third of their income on food, that the Consumer Price Index does not accurately measure inflation, that maintaining the same formula for calculating the poverty line fails to take changes in the standard of living into account, and so on.[26] There is, for instance, a debate over what ought to count as family income: if a family gets food stamps, should the value of its food stamps be considered income? Imagine a family that earns an income just below the poverty line; if the value of the food stamps it receives plus its income produces a total above the line, should it still be considered poor?

How to measure poverty has been a contentious issue for decades. In general, political liberals favor measurement choices that raise the poverty line or make it harder to meet the threshold; they oppose, for example, treating the value of food stamps as income. Measuring poverty in this way means that more people will be considered poor (and therefore deserving additional social services). In contrast, most conservatives argue that the poverty line ought to be set low and that it should be easy to exceed; thus, they favor counting food stamps as income. Such measurement choices will identify fewer poor people (therefore supporting arguments that there is less need for social services). In this debate, liberals argue that the established poverty line produces too many false negatives (that is, too many people who are "really" poor fall above the line), while conservatives charge that the poverty line designates false positives (too many people who "really" aren't poor fall below the line).

Like definitions, measurements always involve choices. Advocates of different measures can defend their own choices and criticize those made by their opponents—so long as the various choices being made are known and understood. However, when measurement choices are kept hidden, it becomes difficult to assess the statistics based on those choices.

SAMPLING

Virtually all social statistics involve generalizing from samples. It is usually too difficult and too expensive to count every instance

of some social condition; it is cheaper, faster, and more efficient to select some cases, examine them, and then generalize to the larger social problem. In the specialized vocabulary of statisticians, the cases examined are a *sample* chosen to represent the larger *population* of all cases.

There are two problems with sampling—one obvious, and the other more subtle. The obvious problem is sample size. Samples tend to be much smaller than their populations. Even national surveys (such as the Gallup Poll) usually interview only 1,000 to 2,000 people. Many studies by social scientists use much smaller samples; it is not uncommon for social research to be based on interviews with only a few dozen people. Obviously, it is possible to question results based on small samples. The smaller the sample, the less confidence we have that the sample accurately reflects the population.

However, large samples aren't necessarily good samples. This leads to the second issue: *the representativeness of a sample is actually far more important than sample size*. A good sample accurately reflects (or "represents") the population. If tens of thousands of readers fill out a questionnaire they find in a magazine and mail in their responses, it is not at all clear what population they represent. Some kinds of people will be much more likely to read the magazine than others, and readers who complete the questionnaire will differ from those who don't; the resulting sample—however large—is not representative.

Selecting a representative sample is a key challenge in social science. Ideally, researchers know the full extent of the population they want to study, and they can select a sample from this

population at random. Statisticians can calculate the probability that such *random samples* represent the population; this is usually expressed in terms of *sampling error* (for example, there might be a 95 percent probability that the distribution of responses in a sample will be within 3 percent of the distribution in the population).

The real problem is that few samples are random. Even when researchers know the nature of the population, it can be time-consuming and expensive to draw a random sample; all too often, it is impossible to draw a true random sample because the population cannot be defined. This is particularly true for studies of social problems. Because social problems always have hidden cases (the dark figure), the actual dimensions of the population always are uncertain. Suppose we want to study teenage runaways. How can we identify the runaway population (which, after all, changes from minute to minute as some teenagers run away and others return home)? Some runaways are gone for minutes or a few hours; other leave for years, even permanently. Some go to stay with friends or relatives; others live on the streets. With such a fluid, diverse population, drawing a random sample presents terrible challenges. Researchers have to make compromises, to draw the best samples they can; they might, for example, combine different samples—perhaps randomly sampling shelters to identify long-term runaways *and* sampling teenagers drawn from the general population to try and identify some former short-term runaways who have returned home. The best samples are those that come as close as possible to being random.

But *statistics about social problems usually are based on samples that fall far short of randomness.* People who want to promote new social problems may not think critically about the representativeness of the samples that serve as the basis for their statistics. Remember that activists tend to spend their time among people intensely concerned about their social problem; they share stories about the severity of the problem and generally reaffirm one another's concern. They probably feel that this experience gives them a pretty good sense of the nature of the problem, and that the cases they encounter are fairly typical. Thus, someone who works in a particular runaway shelter and sees the teenagers who come there may decide to collect statistics on the shelter's clients. The great advantage of this method is its convenience—the runaways come to you. This is called *convenience sampling;* it is inexpensive, it is easy, and it is by far the most common way to study social problems.

The disadvantage with convenience samples, of course, is that it is hard to know whether they reflect the population. Suppose we keep track of all the runaways who pass through one shelter in a month. What population do they represent? Certainly we cannot argue that they represent all runaways, because many runaways never go near a shelter. Well, do they at least represent the population of all runaways who visit shelters? Maybe—but maybe not: there may well be differences in the runaways who come to different shelters, depending on the region where the city is located (warm versus cold climate), shelter policies (shelters may have different rules for admitting runaways), whether a city's runaways have a choice of places they can go to get help,

and so on. An apparently conservative answer might be that our sample at least represents the population of runaways who pass through that shelter. But does it? It might depend on which month we decided to collect data: there are probably more runaways in summer months (because it is warm and school is out); winter runaways might be different (more committed, or more desperate). The potential limitations of convenience samples always need to be considered.

A related problem emerges when activists select particular cases to illustrate a social problem. Recall that activists often choose vivid examples to raise concern; they pick these cases precisely because they are not typical, but are especially dramatic and arouse our shock, horror, anger, or outrage. There is, of course, nothing random about the choice of these examples. However, advocates may talk about these cases as though they are somehow typical, representative of the larger problem. A murdered runaway is a tragic story that may capture media attention and mobilize public concern, but the case is a poor basis for generalizing to the larger problem of teenage runaways. Terrible examples should not be treated as though they are representative samples of a social problem.

The difficulties of drawing accurate samples also invite another sort of generalizing claim. Activists sometimes take advantage of a social problem's dark figure to speculate about the problem's scope. For example, they may argue that the problem threatens "everyone," that it affects "people of all sorts," even that it strikes "at random." Thus, an advocate might argue that teenagers run away from all sorts of families, that it is impossible

to predict which teenagers might become runaways. Such claims are powerful because they raise everyone's concern: if any teenager might run away, then runaways could happen in families you know; even someone in your family might run away. But there is a difference between saying that some teenagers run away from upper-middle-class families that have both parents present ("It happens!") and saying that teenagers from such families are just as likely to run away as other teenagers. All teenagers may be at *some* risk of running away, but that does not mean that all teenagers have the *same* risk of becoming runaways. In general, social problems are *patterned;* people do not run away—or commit crimes, become homeless, or become infected with HIV—at random. But *people promoting social problems often find it advantageous to gloss over these patterns,* to imply that everyone shares the same risks and therefore we all have the same, substantial stake in solving the social problem.[27]

In short, the process of generalization is at the center of sampling. People who talk about social problems almost inevitably want to generalize from some cases—a sample—to the larger problem. The key question is what sorts of generalizations the sample permits. Researchers able to draw random samples from populations that are well understood can make convincing generalizations. But this is almost never possible in analyses of social problems, especially when the problem is first coming to public attention. At that early stage, the dark figure is usually unknown, but probably large, and advocates often don't even have a clear sense of what they don't know. They generalize on the basis of the minimal evidence they do have: perhaps they

know about some dramatic examples that might serve to arouse public concern; perhaps they have conducted studies based on convenience samples of cases that have come to their attention; perhaps they believe that the largely hidden problem is so widespread that it reaches every corner of society. Their limited knowledge, coupled with their enthusiasm for the cause, encourages them to make generalizations that cannot be supported by the evidence.

The media often fail to question activists' generalizations. Frightening examples and claims that a problem threatens everyone make good, compelling news stories. And, if reporters do try to check the activists' generalizations, they may be unable to find anyone with better evidence. Only later, after the problem is defined as a matter of serious public concern, are experts likely to design careful, authoritative research projects. Such research can be expensive, particularly when it tries to examine a random—or at least, an approximately random—sample of a reasonable size. Usually, careful research of this sort requires funding from major institutions—government agencies, foundations, or industries—and they are unlikely to sponsor such work until activists have convinced them to make the problem a priority.

Again, virtually all claims about social problems involve generalizing from a sample of cases. We need to ask how much confidence we should have in these generalizations, and the answer to this question will depend on the nature and quality of the sample. Advocates ought to clearly explain the nature of their sample, so that others can evaluate the strength of the generalizations the sample will support.

CHARACTERISTICS OF GOOD STATISTICS

This chapter's focus has been the production of bad statistics through guessing, dubious definitions, questionable measurement, and poor sampling. At this point, you may be wondering whether all statistics are bad, nothing more than "damned lies." Are there any good statistics? How can we tell the good numbers from the bad?

The problems identified in this chapter suggest some standards that good statistics meet. First, *good statistics are based on more than guessing.* The most basic question about any statistic is: How did someone arrive at this number? All statistics are imperfect, but some flaws are worse than others. Obviously, we should not place too much confidence in guesses (even educated guesses). Watch for the danger signs of guessing: Do the people offering the statistic have a bias—do they want to show that the problem is common (or rare)? Is the statistic a big, round number? Does the statistic describe an unfamiliar, hidden social problem that probably has a large dark figure (if so, how did the advocates manage to come up with their numbers)?

Second, *good statistics are based on clear, reasonable definitions.* Remember, every statistic has to define its subject. Those definitions ought to be clear and made public. An example—particularly a dramatic, disturbing example, a horror story, a worst case—is not a definition. Anyone presenting a statistic describing a social problem should be able and willing to explain the definition used to create the statistic. Definitions usually are broad: they encompass kinds of cases very different

from (and usually less serious than) the examples. We need to ask: How broad? What does the definition include? Again, ask yourself whether the people offering the statistic favor broad (or narrow) definitions, and why. Consider whether their definition might exclude too many false negatives or include too many false positives.

Third, *good statistics are based on clear, reasonable measures.* Again, every statistic involves some sort of measurement; while all measures are imperfect, not all flaws are equally serious. People offering a statistic should be able and willing to explain how they measured the social problem, and their choices should seem reasonable. If the people offering the statistic have some sort of bias (in favor of big—or small—numbers), that bias may be reflected in the way they've measured the problem. For example, they may have worded survey questions to encourage certain responses, or they may interpret responses in peculiar ways. Be suspicious of statistics based on hidden measurements, and consider how measurement choices might shape statistics.

Finally, *good statistics are based on good samples.* Clear, reasonable definitions and clear, reasonable measurements are not enough. Almost all statistics generalize from a sample of cases to a larger population, and the methods of selecting that sample should be explained. Good samples are representative of that larger population; ideally, this means the sample has been selected at random. Watch out for statistics based on small, nonrandom, convenience samples; such samples are easier and cheaper to study, but they are a poor basis for sweeping general-

izations. Ask yourself how the sample chosen might skew the resulting statistics.

One sign of good statistics is that we're given more than a number; we're told something about the definitions, measurement, and sampling behind the figure—about how the number emerged. When that information remains concealed, we have every reason to be skeptical.