A

SPOTTING QUESTIONABLE NUMBERS

The billion is the new million. A million used to be a lot. Nineteenth-century Americans borrowed the French term *millionnaire* to denote those whose wealth had reached the astonishing total of a million dollars. In 1850, there were 23 million Americans; in the 1880 census, New York (in those days that meant Manhattan; Brooklyn was a separate entity) became the first U.S. city with more than one million residents.

At the beginning of the twenty-first century, a million no longer seems all that big. There are now millions of millionaires (according to one recent estimate, about 8.6 million U.S. households have a net worth of more than $1 million, not counting the value of their principal residences).[1] Many houses are priced at more than a million dollars. The richest of the rich are billionaires, and even they are no longer all that rare. In fact, being worth a billion dollars is no longer enough to place someone on *Forbes* magazine's list of the four hundred richest Americans; some individuals have annual *incomes* exceeding a billion dollars.[2] Discussions of the U.S. economy, the federal budget, or the national debt speak of trillions of dollars (a trillion, remember, is a million millions).

The mind boggles. We may be able to wrap our heads around a million, but billions and trillions are almost unimaginably big numbers. Faced with such daunting figures, we tend to give up, to start thinking that all big numbers (say, everything above 100,000) are more or less equal. That is, they're all *a lot*.

Envisioning all big numbers as equal makes it both easier and harder to follow the news. Easier, because we have an easy way to make sense of the numbers. Thus, we mentally translate statements like "Authorities estimate that HIV/AIDS kills nearly three million people worldwide each year" and "Estimates are that one billion birds die each year from flying into windows" to mean that there are *a lot* of HIV deaths and *a lot* of birds killed in window collisions.

But translating all big numbers into *a lot* makes it much harder to think seriously about them. And that's just one of the ways people can be confused by statistics—a confusion we can't afford. We live in a big, complicated world, and we need numbers to help us make sense of it. Are our schools failing? What should we do about climate change? Thinking about such issues demands that we move beyond our personal experiences or impressions. We need quantitative data—statistics—to guide us. But not all statistics are equally sound. Some of the numbers we encounter are pretty accurate, but others aren't much more than wild guesses. It would be nice to be able to tell the difference.

This book may help. My earlier books—*Damned Lies and Statistics* and *More Damned Lies and Statistics*—offered an approach to thinking critically about the statistics we encounter.[3] Those books argued that we need to ask how numbers are socially constructed. That is, who are the people whose calculations produced the figures? What did they count? How did they go about count-

ing? Why did they go to the trouble? In a sense, those books were more theoretical; they sought to understand the social processes by which statistics are created and brought to our attention. In contrast, this volume is designed to be more practical—it is a field guide for spotting dubious data. Just as traditional field guides offer advice on identifying birds or plants, this book presents guidelines for recognizing questionable statistics, what I'll call "stat-spotting." It lists common problems found in the sorts of numbers that appear in news stories and illustrates each problem with an example. Many of these errors are mentioned in the earlier books, but this guide tries to organize them around a set of practical questions that you might ask when encountering a new statistic and considering whether it might be flawed. In addition, all of the examples used to illustrate the various problems are new; none appear in my other books.

This book is guided by the assumption that we are exposed to many statistics that have serious flaws. This is important, because most of us have a tendency to equate numbers with facts, to presume that statistical information is probably pretty accurate information. If that's wrong—if lots of the figures that we encounter are in fact flawed—then we need ways of assessing the data we're given. We need to understand the reasons why unreliable statistics find their way into the media, what specific sorts of problems are likely to bedevil those numbers, and how to decide whether a particular figure is accurate. This book is not a general discussion of thinking critically about numbers; rather, it focuses on common flaws in the sorts of figures we find in news stories.

I am a sociologist, so most of the examples I have chosen concern claims about social problems, just as a field guide written by

an economist might highlight dubious economic figures. But the problems and principles discussed in this book are applicable to all types of statistics.

This book is divided into major sections, each focusing on a broad question, such as: Who did the counting? or What did they count? Within each section, I identify several problems—statistical flaws related to that specific issue. The discussion of each problem lists some things you can "look for" (that is, warning signs that particular numbers may have the flaw being discussed), as well as an example of a questionable statistic that illustrates the flaw. (Some of the examples could be used to illustrate more than one flaw, and in some cases I note an example's relevance to points discussed elsewhere in the book.) I hope that reading the various sections will give you some tools for thinking more critically about the statistics you hear from the media, activists, politicians, and other advocates. However, before we start to examine the various reasons to suspect that data may be dubious, it will help to identify some statistical benchmarks that can be used to place other figures in context.

B

BACKGROUND

Having a small store of factual knowledge prepares us to think critically about statistics. Just a little bit of knowledge—a few basic numbers and one important rule of thumb—offers a framework, enough basic information to let us begin to spot questionable figures.

B1 Statistical Benchmarks

When interpreting social statistics, it helps to have a rough sense of scale. Just a few benchmark numbers can give us a mental context for assessing other figures we encounter. For example, when thinking about American society, it helps to know that:

- The U.S. population is something over 300 million (about 312 million in 2011).
- Each year, about 4 million babies are born in the United States (the 2011 total was 3,953,593).[1] This is a surprisingly useful bit of information, particularly for thinking about young people. How

many first graders are there? About 4 million. How many Americans under age 18? Roughly 4 million × 18, or 72 million. Young people are about evenly divided by sex, so we can calculate that there are around 2 million 10-year-old girls, and so on.

- About 2.5 million Americans die each year (there were 2,513,171 deaths recorded in 2011). Roughly one in four people dies of heart disease (23.7 percent in 2011), and cancer kills nearly as many, so that about half (1,171,652 deaths in 2011, or 46.6 percent) die of either heart disease or cancer. In comparison, some heavily publicized causes of death are much less common: for instance, traffic accidents killed roughly 35,000 people in 2011, breast cancer 41,000, suicide 38,000, homicide 16,000, and HIV/AIDS 8,000. That is, each of these causes accounted for less than 2 percent of all deaths.[2]
- Statistics about race and ethnicity are complicated because these categories have no precise meaning. In general, however, people who identify themselves as blacks or African Americans account for just about 13 percent of the population—about one person in eight. (Remembering that the overall population is more than 300 million, we can figure that there are about 40 million black Americans: 300 million ÷ 8 = 37.5 million.) Slightly more—over 16 percent, or about one in six—identify themselves as Hispanic or Latino. But people cannot be divided neatly into racial or ethnic categories. Most government statistics treat Hispanic as an ethnic rather than a racial category, because Hispanics may consider themselves members of various races. Thus, in a 2007 press release announcing that "minorities" now accounted for one-third of the U.S. population, the census bureau announced that "the non-Hispanic, single-race white population [is] 66 per-

cent of the total population."[3] Note the awkward wording: "non-Hispanic" is used because some people who classify their ethnicity as Hispanic also list their race as white; "single-race" because some people report mixed ancestry (such as having an American Indian ancestor). In short, the bureau is classifying as minority-group members some people who may consider themselves white. No single, authoritative method exists for classifying race and ethnicity. Still, a rough sense of the ethnic and racial makeup of the U.S. population can be useful.

Having this small set of basic statistical benchmarks for the overall population can help us place the numbers we hear in context. Sometimes, when we compare a statistic to these benchmarks, alarm bells may ring because a number seems improbably large or small. For instance, all other things being equal, we might expect blacks to account for about one-eighth of people in various circumstances: one-eighth of college graduates, one-eighth of prison inmates, and so on. If we learn that the actual proportion of blacks in some group is higher or lower, that information might tell us something about the importance of race in that category.

It isn't necessary to memorize all of these figures. They are readily available. One of the most useful sources for basic statistics—just crammed full of official figures—is the annual *Statistical Abstract of the United States*. It is accessible online, and most libraries have a printed copy.[4] Whether you can remember these basic numbers or whether you need to look them up, they can help you critically evaluate new statistics. We will have occasion to use these benchmarks (and we will identify a couple of others) later in this book.

LOOK FOR

Numbers inconsistent with benchmark figures

EXAMPLE: BATTERING DEATHS

A Web site claims that "more than four million women are battered to death by their husbands or boyfriends each year."[5] Right away, our benchmarks help us recognize that this number can't be correct. With only about 16,000 homicides annually, there is no chance that there could be 4 million women killed in battering incidents. In fact, 4 million exceeds the nation's annual 2.4 million death toll from all causes. We have no way of knowing what led the creator of the Web site to make this error, but there can be no doubt that this number is simply wrong.

Although this particular figure is clearly outlandish, I have seen it repeated on a second Web site. Statistics—both good and bad—tend to be repeated. People assume that numbers must be facts; they tell themselves that somebody must have calculated the figures, and they don't feel obliged to check them, even against the most obvious benchmarks. For example, neither whoever created the 4-million-battering-deaths statistic nor the people who repeated that figure thought to ask: "Does this number for battering deaths exceed the total number of deaths from all causes?" Instead, folks feel free to repeat what they understand to be factual information. As a result, bad numbers often take on a life of their own: they continue being repeated, even after they have been thoroughly debunked. This is particularly true in the Internet age, when it is so easy to circulate information. A bad statistic is harder to kill than a vampire.

B2 Severity and Frequency

In addition to having our small set of statistical benchmarks, it is useful to keep in mind one rule of thumb: in general, the worse things are, the less common they are.

Consider child abuse and neglect. Cases of neglect far outnumber cases of physical abuse, and only a small fraction of cases of physical abuse involve fatal injuries. Now, one can argue that every case of child abuse and neglect is bad, but most people would probably agree that being beaten to death is worse than, say, not having clean clothes to wear to school.

Or take crime. In 2011, there were about 700,000 motor vehicles stolen, but fewer than 15,000 murders.[6] Stealing a car and killing someone are both bad, but almost everyone thinks that murder is worse than car theft.

Most social problems display this pattern: there are lots of less serious cases, and relatively few very serious ones. This point is important because media coverage and other claims about social problems often feature disturbing typifying examples: that is, they use dramatic cases to illustrate the problem. Usually these examples are atrocity stories, chosen precisely because they are frightening and upsetting. But this means they usually aren't typical: most instances of the problem are less troubling than the example. Still, it is easy to couple a terrible example to a statistic about the problem's scope: for instance, a report of an underage college student who died from acute alcohol poisoning (a terrible but rare event) might be linked to an estimate of the number of underage college students who drink (doubtless a big number).[7] The im-

plication is that drinking on campus is a lethal problem, although, of course, the vast majority of student drinkers will survive their college years.

LOOK FOR

Dramatic examples coupled to big numbers

EXAMPLE: THE INCIDENCE OF BEING INTERSEX

A person's sex—male or female—strikes most people as the most fundamental basis for categorizing people. Classification usually occurs at the moment of birth (if not earlier, thanks to ultrasound imagery): "It's a girl!" or "It's a boy!" This seems so obvious and natural that most of us rarely give it a thought.

Still, there are babies who don't fit neatly into the standard male/female framework. Some babies have ambiguous genitalia; they can be recognized as hermaphrodites at birth. Others have less visible conditions that may take years to be recognized. People with androgen insensitivity syndrome, for instance, have the XY chromosomes found in males, but because their cells do not respond to testosterone, they develop female genitalia; the condition is usually not discovered until puberty. There are several such conditions, and people with any of them may be categorized as *intersex.*

Some advocates argue that intersex people are common enough to challenge the naturalness of the male/female distinction and that we ought to reconceptualize sex as a continuum rather than a dichotomy. Just how common is intersexuality? One widely cited estimate is that 1.7 percent of people are intersex: "For example, a city of 300,000 would have 5,100 people with varying degrees of intersexual development."[8] (The Internet circulates claims that the actual proportion may be closer to 4 percent.)[9]

However, many of the people included in these estimates live their entire lives without discovering that they are intersex. The most common form of intersexual development is late-onset congenital adrenal hyperplasia

(LOCAH—estimated to occur in 1.5 percent of all people, and therefore accounting for nearly 90 percent of all intersex individuals: 1.5 ÷ 1.7 = .88). Babies with LOCAH have normal genitalia that match their chromosomes; their condition may never be identified.[10] In other words, the most common variety of intersex—accounting for the great majority of cases—is subtle enough to go undiscovered. In contrast, "true hermaphrodites"—babies born with obviously ambiguous genitalia—are in fact rare; there are only about 1.2 per 100,000 births.

Intersexuality, then, displays the pattern common to so many phenomena: the most dramatic cases are relatively rare, whereas the most common cases aren't especially dramatic.