Statistical Science
2005, Vol. 20, No. 3, 210–214
DOI 10.1214/088342305000000232

Lies, Calculations and Constructions: Beyond *How to Lie with Statistics*

Joel Best

Abstract. Darrell Huff's *How to Lie with Statistics* remains the best-known, nontechnical call for critical thinking about statistics. However, drawing a distinction between statistics and lying ignores the process by which statistics are socially constructed. For instance, bad statistics often are disseminated by sincere, albeit innumerate advocates (e.g., inflated estimates for the number of anorexia deaths) or through research findings selectively highlighted to attract media coverage (e.g., a recent study on the extent of bullying). Further, the spread of computers has made the production and dissemination of dubious statistics easier. While critics may agree on the desirability of increasing statistical literacy, it is unclear who might accept this responsibility.

Key words and phrases: Darrell Huff, social construction, statistical literacy.

In the spring of 1965, I was a freshman taking Sociology 25, the introductory course in social statistics at the University of Minnesota. One day the TA in charge of our lab mentioned that this stuff could actually be interesting. There was, he said, a pretty good book called *How to Lie with Statistics*. I perked up; any book with that title promised to be fun. As a high-school debater, I'd had a favorite opening for rebuttals: "Disraeli[1] said, 'There are lies, damned lies, and statistics.' While I certainly don't want to accuse our opponents of lying, they have presented a lot of statistics. . . ." I checked Darrell Huff's little book out of the library and I'd have to say it made as big an impression on me as anything else I read during my freshman year.

I recommended the book to friends and, once I began teaching sociology myself, to countless students. I don't think I read it again until the early 1990s. By that time, I'd encountered other, more sophisticated books on related topics, such as John Allen Paulos' *Innumeracy* (1988), Edward Tufte's *The Visual Display of Quantitative Information* (1983) and Mark Monmonier's *How to Lie with Maps* (1996). *How to Lie with Statistics* remained a wonderful primer but, as a sociologist, I now realized that there was much more to say.

In particular, I'd become interested in the role statistics play in drawing attention to social problems. During the early 1980s, the campaign to call attention to the problem of missing children used a simple, familiar recipe to mobilize public concern: (1) present terrifying examples (e.g., the most notorious case involved a six-year-old boy who wandered away from his mother in the local mall and disappeared until, weeks later, the authorities recovered part of his body); (2) explain that this example is but one instance of a larger problem and name that problem (e.g., that boy was a *missing child*); and (3) give shocking statistics about the problem's extent (e.g., each year, activists claimed, there are nearly two million cases of missing children, including 50,000 abducted by strangers). It was years

Joel Best is Professor and Chair, Department of Sociology & Criminal Justice, University of Delaware, Newark, Delaware 19716-2580, USA (e-mail: joelbest @udel.edu).

[1]This aphorism also gets attributed to Mark Twain. So far as I know, no one has been able to locate it in Disraeli's writings, but it does appear in Twain's autobiography, where Twain ascribes it to Disraeli. Given that Twain was not unwilling to take credit for a funny line, I had come to assume that he at least believed that it originated with Disraeli. However, Peter M. Lee of the University of York's Department of Mathematics has traced the aphorism to Courtney's (1895) reference to "...the words of the Wise Statesman, 'Lies–damned lies–and statistics' ..." (for a full discussion, see Lee's Web page: www.york.ac.uk/depts/maths/histstat/lies.htm).

before reporters began to challenge these widely circulated numbers, in spite of their obvious implausibility. (At that time, there were roughly 60 million Americans under age 18. Was it really possible that one in thirty—think of a child from every schoolroom in the nation—went missing each year?)

Once I'd noticed the three-part (atrocity tale/problem name/inflated statistic) recipe for problem building, I began to appreciate just how often it was used. To be sure, the bad guys—that is, those with whom I disagreed—regularly adopted this combination of claims to arouse public opinion. But then, so did advocates for positions I personally supported. And, while undoubtedly some claims featuring bad statistics were disingenuous—Huffian *lies*, as it were—others seemed to be sincere—albeit innumerate—claims. People trying to draw attention to some social problem tend to be convinced that they've identified a big, serious problem. When they come upon a big numeric estimate for the problem's size, they figure it must be about right, so they decide to repeat it. Since everybody in this process—the advocates making the claims, the reporters covering the story, and the audience for this media coverage—is likely to be more-or-less innumerate, it is easy for bad numbers—especially bad big numbers—to spread. And, of course, in today's world the Internet guarantees a figure's continued circulation. Ridiculous statistics live on, long after they've been thoroughly debunked; they are harder to kill than vampires.

THE TROUBLE WITH LYING

In at least one way, Huff's book may have made things worse. His title, while clever and—at least to former debaters—appealing, suggests that the problem is lying. Presumably lying with statistics involves knowingly spreading false numbers, or at least deceptive figures. Others have followed Huff's lead. A surprisingly large number of book titles draw a distinction between statistics and lies. In addition to *How to Lie with Statistics* [also, *How to Lie with Charts* (Jones, 1995), *How to Lie with Maps* (Monmonier, 1996), etc.], we have *How to Tell the Liars from the Statisticians* (Hooke, 1983), *The Honest Truth about Lying with Statistics* (Homes, 1990), *How Numbers Lie* (Runyon, 1981), *Thicker than Blood*: *How Racial Statistical Lie* (Zuberi, 2001), and (ahem) my own *Damned Lies and Statistics* (Best, 2001) and *More Damned Lies and Statistics* (Best, 2004). Other books have chapters on the theme: "Statistics and Damned Lies" (Dewdney, 1993), "Lying with Statistics" (Gelman and Nolan, 2002), and so on. Folk wisdom draws on the same theme: "Figures may not lie, but liars figure"; "You can prove anything with statistics." You get the idea: there are good statistics, and then there are bad lies. Let's call this the statistic-or-lie distinction.

Of course, this is an appealing interpretation, particularly when the numbers bear on some controversy. I have statistical evidence. My opponent (the weasel) has lies. It has been my experience that almost everyone enjoys criticizing the other guy's bad statistics. I have appeared on conservative radio talk shows where the hosts focused on dubious figures promoted by liberals, and I have been on shows with liberal hosts (they do exist!) who pointed to the bad numbers favored by conservatives. Our critical faculties come into play when we confront a statistic that challenges what we believe; we become analytic tigers pouncing on problems of improper sampling, dubious measurements, and so on. On the other hand, we tend to be more forgiving when we encounter numbers that seem to support what we'd like to think. Oh, maybe our figures aren't perfect, but they're certainly suggestive, so let's avoid quibbling over minor matters. . . .

It is my impression that the statistic-or-lie distinction is often implicitly endorsed in statistics instruction. Statistics courses naturally gravitate toward matters of *calculation*; after mastering each statistic, the class moves on to the next, more complicated one. If "lies" are mentioned, it tends to be in terms of "bias." That is, students are warned that there are biased people who may deliberately choose to calculate statistics that will lend support to the position they favor. This reduces lying to a variant of the larger problem of bias—simply another flaw to be avoided in producing sound calculations.

As a sociologist, I am not sure that the statistic-or-lie distinction is all that helpful. It makes an implicit claim that, if statistics are not lies, they must be true—that is, really true in some objective sense. The image is that statistics are real, much as rocks are real, and that people can gather statistics in the way that rock collectors pick up stones. After all, we think, a statistic is a number, and numbers seem solid, factual, proof that somebody must have actually counted something. But that's the point: people count. For every number we encounter, *somebody* had to do the counting. Instead of imagining that statistics are like rocks, we'd do better to think of them as being like jewels. Gemstones

may be found in nature, but people have to create jewels. Jewels must be selected, cut, polished and placed in settings so that they can be viewed from particular angles. In much the same way, people create statistics: They choose what to count, how to go about counting, which of the resulting numbers they will share with others, and which words they will use to describe and interpret those figures. Numbers do not exist independently of people; understanding numbers requires knowing who counted what, why they bothered counting and how they went about it.

SOCIAL CONSTRUCTION AND STATISTICS

This is what sociologists mean when they speak of *social construction*. I know this term has gotten a bad rap. After being introduced by the sociologists Peter Berger and Thomas Luckmann in their 1966 book, *The Social Construction of Reality*, the notion of social construction was hijacked and put to all sorts of uses—some of them rather silly—by an assortment of literary critics and postmodernist thinkers. Ignore all that. Berger and Luckmann's key observation is very simple: Without doubting that the real world (rocks and such) exists, it remains true that we understand that world through language, and we learn words and their meanings from other people, so our every thought is shaped by our culture's system for categorizing the world. This means that everything we know is necessarily a social construction. Calling something a social construction doesn't mean that it is false or arbitrary or wrong. When I think, "This rock is hard," my notions of rockness and hardness derive from my culture, they are social constructions. But this does not imply that the thought is false or illusionary, that other members of my culture won't agree that it's a hard rock, or that if I whack my head with the rock, it won't hurt. Much of what we know—of our social constructions—provides essential help in getting through the world.

In my view, it helps to think about statistics in terms of construction, as well as calculation. Understand: I am not suggesting we replace the statistic-or-lie distinction with a calculation-or-construction distinction. Rather, my point is that every number is inevitably *both* calculated and constructed, because counting is one of the ways we construct the world. Anyone who has done any sort of research is painfully aware that this is true. All research involves choosing what to study and how to study it. This is why scientists include methods sections in their papers. When we say that science is a social construction, this does not mean that science is fanciful or arbitrary; instead, it means that scientific knowledge is the result of people's work.

So, what do we gain when we think about statistics as socially constructed? For one thing, we can get past the statistic-or-lie distinction. Talking about lies leads us to concentrate on whether people knowingly, deliberately say things they know aren't true. Thus: "Those tobacco executives knew full well that smoking was harmful; we can prove this because we have uncovered internal memoranda that make it clear they knew this; therefore they were lying when they said smoking was harmless." Well, yes. But few bad statistics involve this sort of egregious bad faith. Misplaced enthusiasm is probably responsible for more dubious figures than conscious lying.

Consider the case of anorexia deaths. Someone active in the struggle against eating disorders estimated that perhaps 150,000 Americans suffer from anorexia nervosa, and noted that this disease can be fatal (Sommers, 1994). Someone else—probably inadvertently—garbled this claim and announced that anorexia kills 150,000 each year. This dramatic number was repeated in best-selling books, in news stories and—here I speak from experience—countless term papers. It was a patently ridiculous number: most anorexics are young women; the annual death toll *from all causes* for women aged 15–44 was about 55,000; so what were the odds that 150,000 of those 55,000 were anorexic? Yet, were the various advocates, authors and journalists who repeated this very dubious number *lying*? I presume most of them thought it was true. After all, they believed that anorexia is a big problem, and 150,000 is a big number; moreover, other people said that was the right number, so why not repeat it? Does it help to call the sincere, albeit credulous, dissemination of a bad number a lie?

Or what about a recent, widely publicized report that 30% of students in sixth through tenth grades have moderate or frequent involvement in bullying? This was the highlighted finding from an article in the *Journal of the American Medical Association* (Nansel et al., 2001), mentioned in the article's abstract, in *JAMA*'s news release about the article, and in the extensive media coverage that resulted (Best, 2004). This article survived the peer review process in one of medicine's premier journals; the study, conducted by researchers in the federal government, surveyed more than 10,000 students. But of course the researchers had to make choices when analyzing their data. Respondents were asked whether they had been bullied or had themselves bullied others and, if so, how often. Bullying that was

reported occurring "sometimes" was designated "moderate," while bullying at least once a week was labeled "frequent." This produced a pie of data that could be sliced in various ways. The researchers carved the data to show that 30% of the students reported moderate or frequent involvement in bullying. But consider other possible slices: "involvement" meant either as a bully or a bullying victim; only 17% reported being *victims* of moderate or frequent bullying; and only 8% reported being victims of *frequent* bullying. All of this information is included in the text of the article.

In other words, the claim that the study found 30% of students with moderate or frequent involvement in bullying was no lie. But it would have been equally true to state that 8% were frequent victims of bullying. The former statement was featured in the abstract and the press release; the latter was buried in the article. We can imagine that everyone involved in disseminating the results of this study—the newspaper editors trying to decide whether to run a story about this research, the wire-service reporter trying to write a story that would seem newsworthy, *JAMA*'s editors preparing news releases about that week's articles, the authors hoping that their paper would be accepted by a top-tier journal and that their research would attract attention, even the funders who wanted to feel that their money had been well spent—found a statistic that implicated 30% of students in bullying more compelling than one that suggested 8% were frequent targets of bullies. If there is publication bias against studies with negative findings, so, too, is there a publicity bias favoring studies with dramatic results. But drawing a distinction between statistics and lies ignores this pattern in disseminating research results.

TOWARD STATISTICAL LITERACY?

While many of Huff's critiques remain perfectly applicable to contemporary statistics, there have been important developments during the intervening 50 years. In particular, personal computers have transformed the production and dissemination of statistics. The PC's effects—including inexpensive software for generating sophisticated statistical analyses, bundled spreadsheet programs that allow users to create an extraordinary array of graphs and charts and professional designers able to create eye-catching graphics—have democratized the means of statistical production. Philosophers speak of the Law of the Instrument (originally stated, in an era less concerned with sexism, as: "If you give a small boy a hammer, he'll find things to pound."). Tens of millions of people have been given statistical and spreadsheet software. We can hardly be surprised that we find ourselves surrounded by statistical presentations.

Interpreting these numbers, however, requires two distinct sets of statistical skills. The first set concerns matters of *calculation*—the sort of lessons taught in statistics classes. But in order to assess, to criticize those numbers, we also need to appreciate issues of *construction*. That is, we need to worry about how statistics were brought into being. Who did the counting? What did they decide to count, and why? How did they go about it?

There is a great deal of discussion these days about the desirability of increasing numeracy, quantitative literacy and particularly statistical literacy. Who can disagree? Certainly, part of the problem is that many people aren't particularly adept at calculation. But, I would argue, genuine statistical literacy requires that people also become more alert to what I've called matters of construction.

Anyone who reads the newspaper can find examples of stat wars, debates over social issues in which opponents lob competing numbers at each other. Statistical literacy ought to help people assess such competing claims, but that requires more than teaching them how to calculate and warning them to watch out for liars. It would help to also understand something about the place of statistics in contemporary policy rhetoric, about the processes by which numbers get produced and circulated and so on. But who's going to teach these lessons?

Here I think we might pause to consider the lessons from the "critical thinking" movement that became fashionable in academia in the late 1980s and early 1990s. It is no wonder that the cause of critical thinking gained widespread support. After all, virtually all academics consider themselves critical thinkers, and they would agree that their students need to become better critical thinkers. Yet, if you track the numbers of articles about critical thinking in the education literature, you discover a steep rise in the late 1980s, but then a peak, followed by a decline. This is a pattern familiar to sociologists: these dynamics characterize—in fact define—the fad. The celebration of critical thinking turned out to be just one more academic fad, a short-lived enthusiasm.

Why, if everyone agreed that critical thinking was very important, did interest in the topic fade? I think the answer is that no one assumed ownership of the critical thinking problem. Sociologists interested in how

particular social issues gain and then lose public attention argue that an issue's survival depends on someone assuming *ownership* of the problem, so that there are continuing, active efforts to keep it in the public eye. In the case of critical thinking, no discipline stepped up and took responsibility for teaching critical thinking. Rather, teaching critical thinking was seen as everybody's responsibility, and that meant, in effect, that nobody was especially responsible for it. Without an intellectual owner to promote it, critical thinking slipped quietly from view, to be replaced by the next next thing.

So—even if we agree that statistical literacy is important, and that we need to teach these skills, we still need to figure out who is going to do that teaching. I speak as an outsider, but I doubt that it will be statisticians. The statistics curriculum is based on mastering ever more complex matters of calculation. It may be desirable for students to learn, say, the principles for making good pie charts, but few Ph.D.s in statistics will be eager to teach those lessons. Statisticians are likely to consider teaching courses in statistical literacy beneath their talents, just as professors of English literature tend to avoid teaching freshman composition.

Even though I am a sociologist who believes that the idea of social construction has much to contribute to the cause of statistical literacy, I also doubt that sociologists will claim ownership of statistical literacy. After all, statistical literacy is only tangentially related to sociologists' core concerns. Similar reactions can be expected from psychologists, political scientists, and people in other disciplines.

In other words, its advocates are likely to wind up agreeing that statistical literacy is important, so important that it needs to be taught throughout the curriculum. Once we reach that agreement, we will be well along the faddish trajectory taken by critical thinking. We all know statistical literacy is an important problem, but we're not going to be able to agree on its place in the curriculum.

Which means that *How to Lie with Statistics* is going to continue to be needed in the years ahead.

REFERENCES

BERGER, P. L. and LUCKMANN, T. (1966). *The Social Construction of Reality*. Doubleday, Garden City, NY.

BEST, J. (2001). *Damned Lies and Statistics. Untangling Numbers from the Media, Politicians, and Activists*. Univ. California Press, Berkeley.

BEST, J. (2004). *More Damned Lies and Statistics. How Numbers Confuse Public Issues*. Univ. California Press, Berkeley.

COURTNEY, L. H. (1895). To my fellow-disciples at Saratoga Springs. *The National Review* (London) **26** 21–26.

DEWDNEY, A. K. (1993). 200*% of Nothing*. Wiley, New York.

GELMAN, A. and NOLAN, D. (2002). *Teaching Statistics. A Bag of Tricks*. Oxford Univ. Press.

HOMES, C. B. (1990). *The Honest Truth about Lying with Statistics*. Charles C. Thomas, Springfield, IL.

HOOKE, R. (1983). *How to Tell the Liars from the Statisticians*. Dekker, New York.

HUFF, D. (1954). *How to Lie with Statistics*. Norton, New York.

JONES, G. E. (1995). *How to Lie with Charts*. Sybex, San Francisco.

MONMONIER, M. (1996). *How to Lie with Maps*, 2nd ed. Univ. Chicago Press, Chicago.

NANSEL, T. R. et al. (2001). Bullying behaviors among U.S. youth: Prevalence and association with psychosocial adjustment. *J. American Medical Assoc.* **285** 2094–2100.

PAULOS, J. A. (1988). *Innumeracy. Mathematical Illiteracy and Its Consequences*. Hill and Wang, New York.

RUNYON, R. P. (1981). *How Numbers Lie*. Lewis, Lexington, MA.

SOMMERS, C. H. (1994). *Who Stole Feminism?* Simon and Schuster, New York.

TUFTE, E. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT.

ZUBERI, T. (2001). *Thicker than Blood. How Racial Statistics Lie*. Univ. Minnesota Press, Minneapolis.