

Statistics as squid ink: How prominent researchers can get away with misrepresenting data¹

Andrew Gelman² and Alexey Guzey³

10 Jan 2020

Under the heading of “research misconduct,” the National Institutes of Health defines “falsification” as “manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.”

We recently came across an example of falsification in a book, *Why We Sleep*, by neuroscientist and psychologist Matthew Walker. The falsification was uncovered by Finnish fitness blogger Olli Haataja, and then Guzey added it to his long list of errors in that book (Guzey, 2019).

Here is a graph from Milewski et al. (2014), a source cited in *Why We Sleep*:

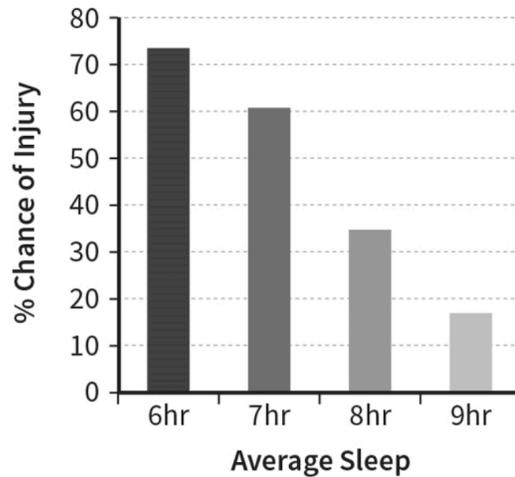


And here is how Walker presents these data in his book:

¹ To appear in *Chance*.

² Department of Statistics and Department of Political Science, Columbia University, New York.

³ Independent researcher, <http://guzey.com>



The 5 hours of sleep column—which is associated with a lower chance of injury than 6 hours of sleep—has simply disappeared. As an aside, that 9 hours of sleep column is based on exactly 1 child being injured out of 6 children who reported sleeping for 9 hours.

Matthew Walker is a prominent researcher, author of dozens of published research articles, professor at the University of California, and also employed by Google. He has delivered a successful TED talk, and his book received rave reviews from leading newspapers and an endorsement from Bill Gates. At first, it might seem surprising that such a leading figure in his field could think it was OK to falsify data. But, as demonstrated by the careers of primatologist Marc Hauser, behavior researcher Brian Wansink, and the entrepreneurs behind the blood-testing company Theranos, it's possible to go far in the worlds of academia, publicity, and technology with bold claims backed by fake data. And these now-disgraced figures didn't do it alone: they had the backing of top figures in the academic and business establishments.

Clear benefits can be gained from falsifying data: you accrue the financial and reputational benefits from research breakthroughs without having to actually do the work. In the case of *Why We Sleep*, it was not a claimed breakthrough but rather a claimed preponderance of evidence that was never there, as revealed in the point-by-point review by Guzey (2019).

This is a column on ethics and statistics, and we assume that all readers will agree that it is unethical to reproduce a graph, removing the one bar on the graph that contradicts your story. Indeed, when it comes to unethical statistical moves, hiding data is about as bad as it gets. At best, this error was introduced by some copy editor by mistake, but that seems like a very unlikely scenario, given the many other errors also found in the book.

To us, the interesting statistical ethics questions here are not, Is unethical to falsify data? (yes, that was unethical) and not, Why did Walker do it? (no great surprise given that he and others have achieved academic and worldly success by exaggerating and misrepresenting evidence), but rather, Why do leading figures in academia and business

play along? and What is the role of statistics in such episodes? We consider each of these questions in turn.

Why do academic and business leaders endorse and then stand by scientific and technological claims that are based on falsified data? The short answer is that they don't know about the falsification—but that doesn't fully answer the question, for two reasons. First, whether or not these people know about the falsification, often it doesn't seem that they look very hard for it, even when real money is on the line as for the investors in Theranos. Second, leaders often continue their support even after the problems are revealed. Noam Chomsky continued to defend Hauser even after the revelations of his misconduct came out, Cornell University stood by Wansink for over a year until finally letting him go, and, at the time of this writing, we have not heard of any of Walker's colleagues or employers expressing concern at his research practices.

We conjecture three reasons for leaders to stand by and not react to scientific misconduct. First is a don't-rock-the-boat attitude: whistleblowing can result in retaliation, and there is also loyalty to friends and colleagues. Indeed, it is possible that Gelman was receptive to Guzey's criticisms in part because they represented the views of a fellow-blogger. Second, when people like a story, they aren't always inclined to look carefully about the evidence. So what if Walker misrepresented data, the argument goes, if his larger point of a lack-of-sleep epidemic is still valid? Similarly, the authors of the book *Nudge* felt no need to look carefully at the veracity of Wansink's experiments, which they described as "masterpieces" (Thaler and Sunstein, 2008). If you already know the answer, the processes of science can seem like so much red tape.

This brings us to a familiar ethical dilemma of improper means used for the purpose of ostensibly worthy ends. We say "ostensibly" worthy because we don't actually know if sleeping six hours per night is bad for you, or if people really eat 50% more popcorn if it comes from a larger container, and so forth. The most common defense of Walker is some form of, "but he did a good thing by alerting people to the dangers of sleep loss," but we don't know that these purported dangers are real.

Daniel Davies (2004) has written, "Good ideas do not need lots of lies told about them in order to gain public acceptance." We don't know if this applies here—perhaps it really is unhealthy to sleep fewer than seven hours of sleep a night. But, given the evidence from *Why We Sleep*—and given that Walker in the writing of that book had the opportunity to use whatever data he could find to make his case—we are far from convinced. If getting a lot of sleep is so important, why are so many lies needed for people to pay attention to it?

In our view, the choice of whether to accept data falsification does not seem like much of an ethical dilemma, but perhaps the perspective of means and ends is valuable in understanding the attitudes of academic and business leaders who seem so unbothered by misrepresentation of data.

Data falsification can be seen as a failure of communication, in that it would not occur if there were a clean data-to-publication pipeline. We typically don't fact-check books—

even influential, best-selling books—because checking is not so easy: indeed, Guzey (2019) reports spending over 150 hours on his effort. We can also assume that one's errors will eventually get caught and cause reputational damage; the famed self-correcting nature of science provides a motivation for researchers to stay honest, which in turn might suggest that we should not be so bothered by episodes of research misconduct which ultimately come to light (as happened with *Why We Sleep* two years after the book's appearance). However, it was the general community of readers, not the academic community, that came to the rescue here: despite the book's massive sales, publicity, and positive reviews, we are not aware of any sleep scientists who went to the trouble of bringing the public's attention to the book's flaws.

One problem is that even when the particular evidence is ultimately dismissed—at least by those who read the right blogs and stay aware of internet discussions—the larger claims remain.

To put it another way, if the above Davies quote is roughly true, why do so many people place trust in statements supported by weak or discredited evidence? To draw another analogy, falsified evidence is a sort of scaffolding used to support an interlocking structure of beliefs—about the role of sleep, or the role of nudging in decision making, or some other topic that is important and difficult to study—and, once the belief structure has been built, the details of the evidence don't seem to matter.

Finally, how does this become a topic of statistics rather than of psychology or rhetoric? How is a misrepresented graph different from a cropped photo, fake quote, or misleading testimony? One difference, we believe, is that numbers and statistical arguments can give an air of authority to what might otherwise appear to be weak qualitative evidence. As the great baseball analyst Bill James wrote, you can lie with statistics just as you can lie with English, French, or any other language. But quantitative analysis when used unscrupulously can serve as a sort of squid ink that hides the holes in scientific reasoning, and it is the role of statisticians to be bothered by this when it happens.

References

Davies, D. (2004). The D-Squared Digest One Minute MBA - Avoiding projects pursued by morons 101. Economics and similar, for the sleep-deprived blog, 27 May. https://blog.danieldavies.com/2004_05_23_d-squaredigest_archive.html

Guzey, A. (2019). Matthew Walker's "Why We Sleep" is riddled with scientific and factual errors. Alexey Guzey blog, 15 Nov. <https://guzey.com/books/why-we-sleep>

Milewski, M. D., Skaggs, D. L., Bishop, G. A., Pace, J. L., Ibrahim, D. A., Wren, T. A., and Barzdukas, A. (2014). Chronic lack of sleep is associated with increased sports injuries in adolescent athletes. *Journal of Pediatric Orthopaedics* 34, 129-133.

Thaler, R. H., and Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

Walker, M. (2017). *Why We Sleep: The New Science of Sleep and Dreams*. Penguin Random House.