

CHANCE

A Magazine for People Interested in the Analysis of Data

- Articles
- Columns
- A Statistician Reads the Sports Pages

- Book Reviews
- Ethics and Statistics
- Goodness of Wit Test
- Here's to Your Health
- O Privacy, Where Art Thou
- Taking a Chance in the Classroom
- The Big Picture
- Visual Revelations
- Editor's Letter
- Letters to the Editor
- Supplemental Material

Open Data and Open Methods

• Columns, Ethics and Statistics



An ethics problem arises when you are considering an action that (a) benefits you or some cause you support, (b) hurts or reduces benefits to others, and (c) violates some rule. Other definitions are possible; there is a vast literature on professional ethics that I will not discuss, instead focusing here on my own perspective as a statistician.

Any ethical dilemma can be transformed into a close call by complicating the costs and benefits, making the rules violations less clear, and adding uncertainty. Ethical subtleties are often explained to children through questions such as, "Is it wrong to steal from a store?" Maybe not if the store is a drugstore that is closed for the night and you might need a certain drug to save a life right now.

Similar twists can be given to ethics problems in science. Consider Mark Hauser, the Harvard

psychologist who was found responsible for scientific misconduct after his research assistants became convinced, in the words of the *Chronicle of Higher Education*, that he was “reporting bogus data.” What if a researcher knows, simply knows, a certain theory is true—but, annoyingly, other researchers in the field disagree. Many scientists have had this experience: We make our point clearly, our reasoning is evidently correct, but others persist in believing the opposite. Is it ethical, then, to fake one’s data? I would say no, but one might argue that short-term fakery is the best way to advance scientific truth as he sees it—and isn’t truth more important than silly rules?

In other settings, behavior we would describe as unethical is considered by others to be simply part of the game. For example, a friend of mine, an academic statistician who occasionally does legal consulting, told us of a case in which he submitted an innocuous report on a random sample he had collected. Later, my friend learned the consultant on the other side of the case had written a rebuttal attacking his competence. The attack was baseless and my friend easily refuted it, but to just attack solid work, presumably for no other reason than you are being paid to do so, seems unethical to me. I suspect that consultant, however, merely saw this as standard practice, no less ethical than bluffing in a poker game or starting with a lowball offer in a negotiation.

Although any ethics violation can be framed to be ambiguous, this does not, and should not, negate the importance of ethics. Statisticians should be able to appreciate the necessity of decisionmaking under uncertainty and ambiguity.

In future columns, I would like to explore many dimensions of ethics, including those that arise in clinical research (e.g., concerns about randomly assigning patients to a control condition believed to be less effective than an available treatment) and statistical analysis (e.g., practices such as fishing with regressions to get statistical significance or, from the other direction, slicing data into small parts so as to lose significant comparisons in the noise) to problems involving probability and uncertainty (e.g., regulations that aim for an unrealistic zero tolerance for risk), as well as more general concerns such as plagiarism and misrepresentation of research findings.

Ethical challenges arise from many sources, including conflicts of interest, imprecise rules, uncertainty, and tradeoffs in values and consequences. As statisticians, our greatest contribution here may ultimately come from quantifying tradeoffs, as in evidence-based

medicine and evidence-based social policy.

Before attempting any sort of quantitative treatment, however, I will tell some stories. The story for the present column concerns the ethical imperative to share data.

An Unethical Refusal to Share Data

A bit more than 20 years ago, I attended—as a PhD student—a statistics conference on the health effects of low-frequency electromagnetic fields. At the time, there was controversy: elevated rates of leukemia had been found among children living near electric power lines, but it was difficult to see how the danger could arise from the underlying physics. There was no obvious resolution. On one hand, epidemiological studies are subject to confounding factors (the families in the study might have other problems, correlated but not caused by the power lines), but the human body is complicated and cancer is not well understood, so it was plausible that the electromagnetic fields could be causing a problem, even if their energy levels were too low to directly harm cells in the manner of microwaves or X-rays.

Several speakers at the conference mentioned a series of papers published in the journal *Bioelectromagnetics* by a biophysicist, Carl Blackman, and his colleagues at the Environmental Protection Agency. One possible pathway in which low-frequency radiation could cause cancer is by affecting signaling within the brain. Blackman et al. performed a series of experiments in which chicken brains were dissected and placed under electromagnetic fields and compared to a control of no radiation, and then the efflux of calcium was measured under the two conditions. The experiments were performed at a set of 38 frequencies ranging from 1 Hz to 500 Hz (the power grid in the United States is at 60 Hz, that is 60 cycles per second), and several chickens were used to provide a reliable estimate and standard error of the difference between electromagnetic field and control conditions at each.

The treatment appeared to have an effect, and it varied by frequency, not in any obvious way, but perhaps in some manner that made sense given the underlying biophysics. Figure 1a shows the basic findings of Blackman et al., in which they summarized their results based on the statistical significance level of their estimate at each frequency.

From my statistical training, I was suspicious of using significance levels in this way—indeed, several years later, Hal Stern and I wrote a paper, “The Difference Between ‘Significant’ and

‘Not Significant’ Is Not Itself Statistically Significant”—and so I made a new graph showing estimates and confidence intervals, shown here as Figure 1b.

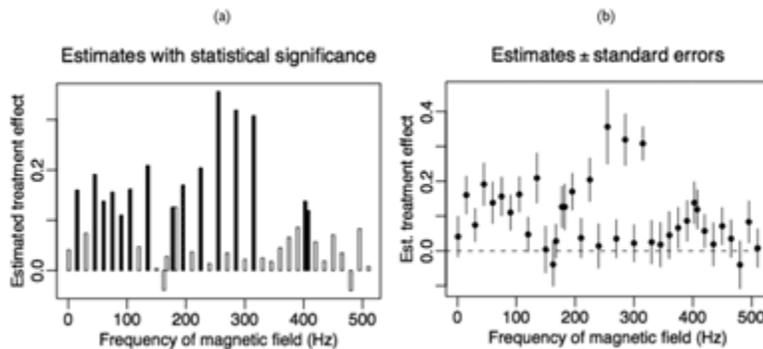


Figure 1. (a) Estimates of the effects of low-frequency electromagnetic fields on calcium efflux from chicken brains. This is a redrawn version of a graph from Blackman et al., who summarized their results based on statistical significance and then went on to make sharp distinctions between frequencies in which effects were observed at different significance levels. (b) The same information used in the first graph, but redrawn as estimates with ± 1 standard error. The new graph makes it clear that the data show some variation, but that it would be a mistake to sharply distinguish between significant and nonsignificant results.

It would be a mistake to draw scientific conclusions, or even conjectures, based on distinctions of whether a result hits the 5% or 1% level of significance, but that is what Blackman et al. did in their published papers. For example, they wrote, “certain frequencies are effective ($P < .05$) in causing enhanced calcium-ion efflux while others are not.” And they went on to present what they called a “three-model analysis” based on partitioning the data into cases with p -values less than 0.01, cases with p -values near 0.05, and nonsignificant cases. They used this completely inappropriate statistical distinction as the basis for several pages of speculation. This was an unfortunate case in which a poor statistical analysis added noise to scientists’ understanding and led to a waste of effort.

A more appropriate analysis would respect the uncertainty in the estimates and characterize

the responses in Figure 1b as being in a continuous range, rather than being sharply divided into categories. (When assigning classroom grades, a teacher might need a bright line separating As from Bs, and Bs from Cs, but a scientific paper should be able to acknowledge uncertainty in inferences.)

Looking more closely at the articles, I noticed another serious flaw. The experiments were set up with a treatment and control condition, and each of these was performed under active and “sham” conditions (the identical setup but with the radiation turned off). The result was, essentially, one treatment group and three controls. Call these A1, A2, B1, B2, in which A1 is the active treatment, A2 is the active control (but receiving no electromagnetic field), and B1 and B2 are sham treatment and control (thus receiving no field either). This is not the optimal design, but given that the data had already been collected, the best estimate of treatment effect is $A1 - (A2+B1+B2)/3$. This is not, however, what the researchers did. Instead, they took a difference of differences: $(A1-A2) - (B1-B2)$. Under some circumstances (e.g., an experiment on humans who might respond to expectations of treatments), such a conservative analysis is appropriate, but it hardly seems necessary for a study of the brains of dead chickens!

In addition, I did a quick analysis of the summary statistics presented in the published research articles and found that (i) the differences B1-B2 (the effects under the sham conditions) were not statistically significantly different from zero and (ii) the correlation of the differences A1-A2 and B1-B2 across the 38 experiments was not statistically significantly different from zero. That is, the data were consistent with the sham conditions being pure noise. By taking a difference in differences, Blackman et al. were, amazingly, reducing their statistical efficiency by more than a factor of three. Even the simple comparison A1-A2 would be twice as efficient as what they did.

At this point, there has been no ethics violation. The lead researcher earned his PhD in biophysics and would not be expected to be aware of alternative statistical analyses, and the statistician on the study had a master’s in statistics, but no further degrees. I feel awkward for relating these qualifications—it makes me sound like a credentials snob—but that is not my point. There is a role for statisticians at all levels in research, and I would expect that someone with basic statistical training would use the most standard methods, which, in this case, might well be statistical significance levels and a very conservative data analysis.

I repeat that I do not consider it an ethics violation in any way for a group of scientists to design

an experiment and analyze their data using standard approaches. And the error of drawing conclusions from differences between significance levels is a subtle trap: In a recent survey of neuroscience articles, psychologist E. J. Wagenmakers and his colleagues found this mistake to have been made about half the time by this statistically sophisticated research community.

The ethics violation, as I see it, by Blackman and his statistician colleague came not in their design, data collection, or even their flawed analysis, but when they had the opportunity to subject their data to an outside analysis.

Having been supplied free travel and housing to that conference and having spent several days more reading a key source article and analyzing its summary statistics, I felt both an obligation and an inclination to help. So I looked up Blackman's address in North Carolina and sent him a polite letter saying I was a statistician who had attended a conference in which his work was mentioned, that I had two ideas of how he could analyze his data better (I gave some details here and maybe a graph or two), and that I would like to see his raw data so I could do more. I used Harvard letterhead, but was careful not to identify myself as a PhD student—I think I called myself a “researcher”—and I ran the letter by some of my fellow students to make sure I was being sufficiently polite.

A few days later, I followed up the letter with a phone call—older readers of *CHANCE* might recall these primitive technologies—at which point Blackman told me he had discussed the matter with his statistician and they decided their analysis was just fine and it would be too much trouble for them to copy the data from their logbooks and send it to me.

That was the unethical step. Refusing to share your data is improper, and the lead researcher and his statistician should have realized that, given their lack of expertise in statistics, it was at least plausible that an outsider could improve on their analysis.

You might consider my ethical judgment too harsh—maybe these guys were busy that week, or just in a bad mood—but sharing data is central to scientific ethics. If you really believe your results, you should want your data out in the open. If, on the other hand, you have a sneaking suspicion that maybe there's something there you don't want to see, and then you keep your raw data hidden, it's a problem. I don't think the dead chickens had any confidentiality issues. And what sort of researcher is so sure of the analysis of his MS-level statistician that he won't even consider the possibility that an outside analysis might reveal something new? (Again, I'm

not trying to be a PhD snob here. There's a lot statisticians at all education levels can do, but one should also recognize one's limitations. I have a PhD, myself, but try always to be open to the possibility that I might be making a mistake—which is a good thing, because I make mistakes a lot!)

I regret not following up with a more formal request—I assume the Environmental Protection Agency is ultimately subject to the Freedom of Information Act—but at this point, the lab notebooks are probably lost forever.

The ethics of a decision depend on one's state of knowledge. In this case, I would not blame a team of insufficiently trained researchers for a weak but conventional statistical analysis—but I do blame the principal investigator for violating the principle of openness in scientific research. This is hardly an ethical lapse on the scale of tobacco executives who commissioned research studies and then buried their findings, but it is something statisticians must always be aware of: Do not be so tied to your analyses that you are afraid that others might, with the same data, find something different.

Complications

Data sharing is an obligation, but the practicalities are a bit tricky. Twenty years ago, it was a lot harder to extract your data in a nice format for sharing. You could photocopy your lab notebooks, but the raw numbers might have needed some explanation. And it also took a lot more effort to ask someone for their data. Today, people don't need to write a letter and make a follow-up phone call, but can simply fire off an email to any researcher in the world asking for data.

Especially for high-stakes policy questions (such as the risks of electric power lines), transparency is important, and we support initiatives for automatically making data public upon publication of results so researchers can share data without it being a burden. When requested to supply information from our own past research, I have sometimes been unable to find some data files or replicate analyses written in archaic computer languages. This is embarrassing—I can still share much of my data, but cannot always share the steps of the analysis—and motivates me to be more systematic with our computations in the future.

As a statistician, I think the key point is to recognize that different analyses can give different

perspectives on a data set. I am not suggesting that researchers be regularly subjected to forensic analyses of all their decisions in data collection and analysis, explaining every email exchange or every new version of a data set that had a transformation or data exclusion. But openness should be the norm.

I look forward to continuing the discussion of ethics and statistics. Feel free to send me any ethical dilemmas you have observed or experienced—and also any disagreements with anything you see here—and we will try to discuss these in future columns.

Ethics and Statistics will discuss ethical issues that come up in the cross-disciplinary work of statisticians, as well as in the application of statistical methods by nonstatisticians. If you are interested in submitting an article, please contact Andrew Gelman, column editor, at gelman@stat.columbia.edu.

Further Reading

Blackman, C. F., S. G. Benane, D. J. Elliott, D. E. House, and M. M. Pollock. 1988. **Influence of electromagnetic fields on the efflux of calcium ions from brain tissue in vitro: A three-model analysis consistent with the frequency response up to 510 Hz.** *Bioelectromagnetics* 9:215–227.

Gelman, A., and J. Hill. 2007. **Data analysis using regression and hierarchical/multilevel models.** Cambridge University Press: Cambridge, UK.

Gelman, A., and H. S. Stern. 2006. **The difference between ‘significant’ and ‘not significant’ is not itself statistically significant.** (PDF download) *The American Statistician* 60:328–331.

Nieuwenhuis, S., B. U. Forstmann, and E. J. Wagenmakers. 2011. **Erroneous analyses of interactions in neuroscience: A problem of significance.** *Nature Neuroscience* 14:1105–1107.

Back to Top

Tagged as: **ethics**, **sharing data**

Leave a Response