University of Michigan Journal of Law Reform

Volume 50

2017

Evidence of Child Abuse: Inferring the Causes of Effects

Stephen E. Fienberg
Carnegie Mellon University

https://doi.org/10.36646/mjlr.50.3.evidence

EVIDENCE OF CHILD ABUSE: INFERRING THE CAUSES OF EFFECTS

Stephen E. Fienberg*

When I was asked to come here and speak, I told the organizers I wasn't sure I had much to say beyond what Maria was going to say. I confess I was wrong. I've been thinking about this over the last week. I'm a statistician. But I also am deeply involved in the assessment of forensic evidence and for personal reasons have, over the last several years, come to know a lot about measurement error in CT scans. I thought I would just note that one of the things I do is that I serve as a member of the National Commission on Forensic Science with Judge McCormack. And at our forthcoming meeting next month, the commission will be considering a document that says that phrases like "reasonable degree of scientific certainty" and "reasonable degree of medical certainty" have no meaning and should not be allowed as part of expert forensic science testimony. I think that would be in accord with Peter Aspelin's thoughts.

The second reason why I realized that I actually had a lot more to say was that I've been involved in working with data that are directly relevant, in the sense that they've been used to describe work predicting child abuse. In 2013, a group of authors did a very elaborate analysis published in a statistical journal in which they attempted to predict the probability of abuse, given the evidence[1]—exactly the quantity that Maria was referring to before. In the subsequent paper which I've been involved with,[2] my coauthors and I referred to that as "probability forecasting" or "backcasting," because, in fact,

* Maurice Falk University Professor of Statistics and Social Science, Carnegie Mellon University. Professor Fienberg passed away on Wednesday, December 14, 2016 in Pittsburgh, PA. Professor Fienberg was elected as a member of the National Academy of Sciences in 1999 and served as a fellow of the American Academy of Arts and Sciences, the American Academy of Political and Social Science, and the Royal Society of Canada. Prior to his death, Professor Fienberg was the co-director of the Living Analytics Research Centre and the editor-in-chief of the ANNALS OF APPLIED STATISTICS as well as the JOURNAL OF PRIVACY AND CONFIDENTIALITY.

1. *See generally* Nicky Best et al., *A Bayesian Approach to Complex Clinical Diagnoses: A Case-Study in Child Abuse*, 176 J. ROYAL STATISTICAL SOC'Y 53 (2013) (using Bayes theorem to formulate a diagnosis of child abuse in terms of prior and inverse conditional probabilities and adapting systematic review methodology and Bayesian evidence synthesis to estimate these and to propagate the associated uncertainty).

2. *See* A. Philip Dawid, David L. Faigman & Stephen E. Fienberg, *Fitting Science into Legal Contexts: Assessing Effects of Causes or Causes of Effects*, 43 SOCIOLOGICAL METHODS & RESEARCH 359, 361 (2014) (noting that science tends to infer "the effects of causes" through experiences and observational studies but that legal fact finders need to infer "the causes of effects").

we're going backwards from the outcome to decide the issue of abuse. What we regularly hear in all of the arguments, including the ones you've heard today, is that violent shaking leads to the triad. And when we observe the triad, we can therefore go backwards and talk about whether or not the violent shaking was the cause of the outcome.

So as a statistician looking at this, I have many questions. My first question goes back to the kinds of things that Maria was describing and that you've heard this morning: What's the provenance of the data? How were they gathered? Who did the assessments? In the evaluation, what kinds of biases were possible? What kind of controls against those biases were made available? What about the interpretation of the evidence and biases in that? Was there any blinding in the assessment of any of the evidence at hand?

You heard a lot during Dr. Aspelin's keynote about the science. I would want to try to understand the science in order to create these predictive or backcasting models. But what I want to do in the rest of my time is tell you that that's the wrong question to ask. It's the one that I've been trained to ask over decades, and that we as statisticians have trained other people to ask. So is it predictions, and if so, what do you condition on? Because those are conditional probabilities. Or is it a question about causation? And if so, what do we mean?

I'm here to tell you that there really are two different questions. The question I posed before is one that I call the effects of causes.[3] Does the violent shaking of the infant cause the specific triad outcome? That's the traditional scientific question. You manipulate a treatment in an experiment and you look at the outcome, so you go from the causes to the outcome. But there's a different question: Was the violent shaking the cause of the triad symptoms that were actually observed? Can we eliminate the alternative possible causes? This was implicit in what Maria spoke about, but not distinguished with the care that I think we need to bring to this.

So the question is, are these the same? And if not, how do they differ? So I actually sat with her and we looked at the database she just described and we extracted some numbers, and we began to look at them and they made no sense. And once some of the numbers have problems, these problems propagate into everything else you can look at. So what I thought I was going to do, which was to make the traditional argument from such data is absolutely impossible from such data—the conclusion that Maria reached a few minutes ago with you.

---

3. *Id.*

Instead I'm going to ask a much simpler question. After Peter's talk, I had a headache. It wasn't the quality of the talk—it was wonderful. I learned a lot. I took an aspirin and my headache went away. When I got up after Maria's talk, I was feeling terrific. So I have a question: If I have a headache, will taking aspirin help? That's looking at the effects of causes. The causes of effects is the reverse question. I'm feeling so good standing up here and maybe I was inspired by Peter's talk and that made me feel better—or maybe it was the aspirin. I want to explore how they differ and what the implications are for the kind of research we've been talking about here.

So let's analyze that aspirin problem. By the way, I made these numbers up, but don't worry about it. It's true: Lots of people who don't take aspirin, their headaches get better and we'll say that's going to be twelve percent. And if they take aspirin, it gets better at a higher rate, but not so high. What statisticians did, and if you read my book on categorical data, you might have done, was compute the odds ratio. That's what we've been telling you to do for decades: compute the odds ratio, adjust for all the other things you want to condition on, and then reason backwards because there are some magic statistical results that actually let you do that, if all you're doing is talking about probabilities.

But we're not just talking about probabilities; we're talking about causes. And so in this example, there are two potential causes and two kinds of responses. There's the response to aspirin, and there's the response to no aspirin. But in my aspirin hypothetical, I took the aspirin and so I don't have the data for "no aspirin," and therefore, I can't look at those two jointly. But what I'd really like to know is that given that I did take aspirin and I saw the outcome, what was the probability of what would have happened had I not taken aspirin? My colleagues and I refer to that as the probability of causation. It requires the joint distribution and we can't estimate it. We only know the marginal probabilities. That's the thing that we learn from the clinical trials, the definitive data on aspirin. So what can we say?

Well, we could actually create a 2-by-2 table[4]:

4. *Id.* at 373–74.

	R_0		
R_1	**0**	**1**	**Total**
0	88 – x	x – 18	70
1	x	30 – x	30
Total	88	12	100

- PC = Pr(R0=0 | R1=1) = x/30
- Know $18 \leq x \leq 30$
- So PC $\geq 60\%$

On the right is the information that I just told you about; I just displayed it as the margins, and it's what's inside the table that we don't know. The key thing of what we don't know is this quantity X, which is the probability of causation divided by the total thirty. And what we know because of the marginal totals—this is a little bit of algebra that the statisticians at your institutions can explain to you—is that X has to be greater than equal to eighteen and less than or equal to thirty. Those are the constraints that the marginal totals imply. We don't know anything else. We only can produce in a case where we have perfect data and no variation, upper and lower bounds for the probability of causation (PC). So the probability of causation in this case has to be equal or greater than sixty percent. Not bad.

In general, this argument actually can be used to show that if you looked at the ratio of these potential responses that PC, this quantity I talked about is equal or greater than one minus one over a risk ratio.[5] That means if that risk ratio is greater than two, something that epidemiologists in courts actually talk about in a slightly different way, then the probability of causation is greater than a half. That's often referred to as proof on the balance of probabilities, but that's a civil standard and not a criminal standard. And the converse is false.

So let's return to the problem of Shaken Baby Syndrome. We can just take the same thing and replace the words. The problem is we have no experimental data. There's lots of variation. We could estimate the marginal totals, but the data for that is deeply flawed. That's what Maria just told you about a little while ago. So I can work with bounds on PC, the probability of causation, but I have to take into account biases and uncertainty.

We went back to the Best et al. dataset[6] that I described at the outset, and I confess that I did this before I had done any reading,

5. *Id.* at 375.
6. *See generally* Best et al., *supra* note 1.

including Patrick Barnes's work and other work of people in the room, before Heather Kirkwood had sent me boxes of files. I took the data that were used in the Best et al. study at face value with my colleagues. For their assessments, the prediction probability of abuse, that quantity they computed, lay somewhere between twenty and forty percent with lots of noise.[7] So, variation. We did an extended analysis of this, trying to take into account the kinds of uncertainty, only that the data do not bring any of the biases people have talked about in our work. But now we're dealing with attribution, this probability of causation. When we conducted our full Bayesian analysis, our interval for PC runs from zero, our estimated lower bound, to 0.043 (PC* = (0, 0.043)). What looked like moderately high predictive value of abuse is very low evidence for the probability of causation. It's certainly so low that it wouldn't meet the criteria that courts bring to bear in other settings.[8]

I just have a little more for child abuse evidence here. Most of the statisticians and epidemiologists, when they have quality data, try to address what I call the traditional scientific question: looking at the effects of causes. For that, the gold standard is randomized experiments. As Peter Aspelin said this morning, "We're never going to see that in any of these domains. What we have are deeply flawed observational data." The real question isn't that traditional one at all; it's actually one of causes of effects.[9] For that we need a very different kind of statistical approach. It brings to bear much, much greater of an uncertainty. And it shifts the argument in many instances in substantial ways. This is an overlay to the kinds of issues and descriptions you heard this morning and from Maria. I think it's essential for the field to come to grips with it.

---

7. *Id.* at 65–66.

8. *See* Dawid et al., *supra*, note 3 at 366–73 (discussing ways that courts limit expert testimony to the evidence of causation, permit expert testimony on the effects of reputed causes and the cause of the observed effect, and permitting experts to reason from the existence of the effect to possible causes).

9. *See* Dawid et al., *supra*, note 3 at 361 (noting that science tends to infer "the effects of causes" through experiences and observational studies but that legal fact finders need to infer "the causes of effects").