PROJECT MUSE®

Lessons we are still learning

Jennifer L. Hill

Observational Studies, Volume 1, Number 1, 2015, pp. 196-199 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2015.0019>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/793418/summary>

Lessons we are still learning

Jennifer L. Hill

jennifer.hill@nyu.edu

Department of Humanities and Social Sciences

New York University

New York, NY 10003, USA

I thoroughly enjoyed re-reading Cochran's commentary on observational studies. In particular, Cochran captured my feelings towards the topic of observational studies quite aptly in his final sentence, "observational studies are an interesting and challenging field which demands a good deal of humility, since we can claim only to be groping toward the truth." Scientific inquiry would benefit from greater humility among researchers pursuing causal answers today.

In this comment I will briefly highlight some of the knowledge that has been gained about causal inference since that time (apologies in advance for not referencing all the scholars who have contributed – there are too many people to do it equitably). I will then focus on what I feel has been lost in the past decades and point out what I see to be important directions for the future.

1. Causal inference without randomized experiments

Causal inference typically requires satisfying both structural and parametric assumptions. Randomized experiments have the advantage of addressing both of these types of assumptions. The most problematic structural assumption, ignorability, (also referred to as all confounders measured, selection on observables, conditional independence assumption, exchangeability, etc) is trivially satisfied in a pristine randomized experiment. Randomized experiments also ensure common support across treatment and control groups.

Randomized experiments have the added advantage that they do not require conditioning on confounders for unbiased estimation, thus eliminating dependence on parametric assumptions. Moreover, even if we use a model to estimate treatment effects in this setting (for instance with goal of increasing efficiency) it is likely that our estimates will be robust to violations of the parametric assumptions of the model.

Of course in practice, noncompliance, missing data, measurement issues and other complications can still wreak havoc with treatment effect estimation even in the context of a randomized experiment. Even more problematic, randomized experiments are often not possible due to ethical, financial, or logistical reasons.

In the absence of a randomized experiment (or natural experiment) the structural assumptions required to identify a causal effect become more heroic, requiring appropriate conditioning on confounding covariates. Unfortunately our dependence on the parametric assumptions grows as well since we now must appropriately estimate expectations conditional on the set of proposed confounding covariates.

Much of the work in causal inference methodology in the decades since Cochran’s paper was first published has focused on relaxation of parametric assumptions. Cochran discusses matching, subclassification, and covariance adjustment. Since that time, however, use of propensity scores for matching, inverse probability weighting using propensity scores have yielded improvements in our ability to estimate causal effects with less bias due to a reduced reliance on parametric assumptions. More recently, more sophisticated matching methods that capitalize on advances in computer science have increased our ability to find good balance targeted to particular balance criteria without undue investment of researcher time. Along another vein, it has been proposed that use of flexible modeling of the response surface using Bayesian nonparametrics along with appropriate checks for overlap may largely obviate the need for such preprocessing methods. All in all, our ability to condition on potential confounders without making extreme parametric assumptions has increased greatly in the past few decades.

Other groundbreaking work has been done since Cochran’s paper around causal inference in longitudinal settings, greater understanding of the role of double robustness in estimation, mediation and approaches to SUTVA violations. Moreover our awareness about and ability to exploit rigorous quasi-experimental designs has grown far stronger.

Most of these advances have been facilitated by the creation of a shared formal language for describing both causal estimands and assumptions. Actually, two “languages” currently hold sway: the potential outcome framework, and directed acyclic graphs, and their extensions. I will not advocate for one or another of these frameworks but rather suggest that to the extent that researchers interested in pursuing causal estimation understand these languages we can learn from each other more easily and advance science with greater efficiency and clarity.

2. A broader perspective

One of the most refreshing aspects of the Cochran article is his concern over many different parts of the research process, including: the framing of the research question (how often do we formally talk about that?), study design (too often statisticians are not involved!), measurement (we largely ignore... that’s for the psychometricians!), non-response (we sometimes address...), power (boring!), and generalizability (there is renewed interest in this topic). While all of these topics are still actively pursued as research, it is rare to see a study that does due diligence to all of these concerns. Perhaps this is due to the pressure in academia to be specialists, and the lack of reward to working in large research teams (in many, though not all, fields). Few academics rigorously address more than two or three of these concerns in their work. Moreover, these issues are typically addressed in separate papers rather than understanding the complexity of the relationships between them.

And in applied papers we typically pick our battles. How often do we apply a new method in an applied problem that might eliminate a small percentage of the bias only to ignore measurement error or missing data issues that might be causing bias that would easily swamp this gain? If we address only some of the statistical issues what can we expect to achieve overall?

I advocate that we could increase our impact by focusing on being more broad than deep. Rather than trying to act as surgeons working in a specialized field, we would reach

more people by thinking of ourselves as army surgeons at the front trying to extract some meaning from the messy and imperfect world of empirical science.

3. Beyond “correlation is not causation”

Understanding what situations allow for causal inference is non-trivial. This challenge is complicated by the fact that most people intuitively assign causal attribution without thinking too hard about it; it feels natural. Psychological experiments have demonstrated time and again, however, that humans are easily misled into drawing causal conclusions even when none are warranted. Yet understanding the answers to causal questions is critical for making progress in science, assessing policy implications, and even trying to better understand the implications of our actions in our individual lives.

While we have made advances in creating a broader understanding of causal issues it is mostly captured by the phrase “correlation is not causation”. This is useful and has led to more broadly visible manifestations of this advice. For instance, science writer Gary Taubes wrote a helpful (if slightly imperfect) New York Times magazine article about the limitations of observational studies (Taubes, 2007). As a sillier example Tyler Vigen has a website <http://www.tylervigen.com/spurious-correlations> and a book (Vigen, 2015) that highlight amusing real world examples of correlations that quite clearly do not reflect causal relationships (for instance the 94% correlation between per capita cheese consumption and the number of people who died by becoming tangled in their bedsheets).

This is helpful but we need to do much more. Too many scientific journals promote the practice of using the word association (rather than effect) as a panacea. Consequently, authors dutifully describe relationships between variables as associations (typically not even as conditional associations with a very particular conditioning set!) and then proceed in the discussion section of the paper to make recommendations for policy or practice (clearly interpreting their own results causally). That is hardly less damaging than just using the word cause throughout.

Instead, we need to encourage a culture of transparency about causal claims. After all, almost all interesting scientific questions are causal in nature. We’re all trying to do it. So let’s be honest about it and be clear about our assumptions and explore how far we might stray from those assumptions and what the implications of those violations might be.

In terms of education, we need to move beyond the catch-all “correlation is not causation” admonition and help create a deeper understanding about the broader populace about what it means to make a causal claim and what kind of research strongly supports such claims. That would mean starting to teach about counterfactuals and a wider range of designs even in introductory statistics courses (I would like some of these concepts to be taught in grade school). It would also mean totally rethinking the way we typically teach regression.

4. Embrace your inner scold

Returning to Cochran’s final sentence. Humility may be one of the least exhibited traits among academics today. We see a fair amount of hubris with regard to causality, both within the academy and in industries that rely heavily on “data science”. For instance, in a

now infamous article in Wired magazine, Chris Anderson wrote “There is now a better way. Petabytes allow us to say: ‘Correlation is enough.’ We can stop looking for models. We can analyze the data without hypotheses about what it might show.” (Anderson, 2007). These types of overstatements may seem amusing and nonthreatening, but this type of hype about “big data” or new technologies could set us back decades.

Unfortunately, statisticians are often seen as scolds or Chicken Little. They are less apt to say “Yes you can!” and more apt to say “Oh my goodness you can’t do that!!!” This does not always make us popular either as cocktail party guests or collaborators. I don’t suggest that we entirely give up this role – it’s too important and nobody else seems prepared to take it on. However, if we embrace the role of scold we also need to point out what can be done to fix the problems. Focus on better designs (bring statisticians in to a study from day one!) and sensitivity analyses that explore how far from the mark our estimates may be if our assumptions are violated are a start to this process.

References

- Anderson, C. (2007). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*
- Taubes, G. (September 17, 2007). Do We Really Know What Makes Us Healthy? *New York Times Magazine*.
- Vigen, T. (2015). *Spurious Correlations*. Hatchett Books, New York.