A Hard Unsolved Problem?
Post-Treatment Bias in Big Social Science Questions

Gary King

Institute for Quantitative Social Science
Harvard University

Talk at the "Hard Problems in Social Science" Symposium,
Harvard University 4/10/2010

A Few Big Social Science Questions We Can't Answer

- Does democratization reduce international conflict?
- Does ethnic diversity in developing countries cause civil war?
- Does trade openness reduce the risk of state failure?
- Do strong international institutions lead to or result from international cooperation?
- How to distinguish between (1) the effect of your health on others and (2) the effect of interventions to improve your health on others?
- Is individual student achievement caused by (1) students' socioeconomic characteristics or (2) their peers' average achievement? (I.e., does tutoring a subgroup help others and then feedback?)
- Thousands of other examples in most areas of the social sciences.

Experiments are terrific, but that's not the issue

- Experimental research: requires random treatment assignment
- (Perhaps the most important methodological idea in the last century)
- Random assignment: occasionally possible, usually not
    - The vast majority of research is observational
    - The vast majority of knowledge learned is from observation
    - Most knowledge learned *from experiments* is observational
- ⇝ So yes
    - experiments are terrific
    - we can't run them that often
    - but we can often learn plenty without experiments

Progress in Causal Inference

- **Inference:** using facts you have to learn about facts you don't have
- **Causal effect:** Your outcome with the treatment minus your outcome if you had not received the treatment
- **Causal inference:** estimating the causal effect
  - Get equivalent treatment and control groups
  - Apply or observe the treatment
  - average(outcome for treateds) − average(outcome for controls)
- Biggest problem we know about: **Omitted variable bias**
  - What if those who get the medicine are healthier than those who don't?
  - What if those who participate in a job training program are less educated?
- **Well-known solutions to omitted variable bias:**
  - Randomize treatment (all confounders are unrelated to treatment)
  - Control (physically or statistically) for the potential confounders

So What's Post-Treatment Bias?

- It occurs:
  - when controlling away for the consequences of treatment
  - when causal ordering among predictors is ambiguous or wrong
- Resulting biases can go in either direction
- The problem is obvious once you think about it
- For many big social science questions, we have no solution

Examples of Avoidable Post-Treatment Bias

- Causal effect of Party ID on the Vote
  - Do control for race
  - Do not control for voting intentions five minutes before voting
- Causal effect of Race on Salary in a firm
  - Do control for qualifications
  - Don't control for position in the firm
- Causal effect of medicine on health
  - Do control for health prior to the treatment decision
  - Do not control for side effects or other medicines taken later

⤳ Post-treatment bias in these examples: easy to see & avoid

Examples of Unavoidable Post-Treatment Bias

- Causal effect of democratization on civil war; do we control for GDP?
  - Yes, since GDP→democratization we must control to avoid omitted variable bias
  - No, since democratization→GDP, we would have post-treatment bias
- Causal effect of legislative effectiveness on state failure; do we control for trade openness?
  - Yes, since trade openness→legislative effectiveness, we must control to avoid omitted variable bias
  - No, since legislative effectiveness→trade openness, we would have post-treatment bias
- Causal effect of tutoring on individual test scores; do we control for average test scores?
  - Yes, since peers→individual scores, and so must control to avoid omitted variable bias
  - No, since individual scores→averages, we would have post-treatment bias

Solutions for Post-Treatment Bias?

- Is there a statistical fix? Nope.
- What about trying it both ways (include and exclude)? Nope. The estimates do not bound the truth!
- Is there identifying information in the ideal case? In many situations, no; in others we don't know.
- Can we redesign data collection to avoid the problem? Usually not.
- Is there progress on related issues? See Kosuke Imai, Adam Glynn, Chuck Manski, Jamie Robins, Paul Rosenbaum, Don Rubin, Tyler VanderWeele, and others.
- Are there areas of social science scholarship not affected? Yes, but the bigger the question the more likely you'll find the problem
- Is there hope? There's always hope; just no answers!

http://GKing.Harvard.edu