That one weird third variable problem nobody ever mentions: Conditioning on a collider

the100.ci/2017/03/14/that-one-weird-third-variable-problem-nobody-ever-mentions-conditioning-on-a-collider

Julia Rohrer March 14, 2017

Scroll to the very end of this post for an addendum.[1]

Reading skills of children correlate with their shoe size. Number of storks in an area correlates with birth rate. Ice cream sales correlate with deaths by drowning. Maybe they used different examples to teach you, but I'm pretty sure that we've all learned about confounding variables during our undergraduate studies. After that, we've probably all learned that third variables ruin inference, yadda yadda, and obviously the only way to ever learn anything about cause and effect are proper experiments, with randomization and stuff. End of the story, not much more to learn about causality.[2] Throw in some "control variables" and pray to Meehl that some blanket statement [3] will make your paper publishable anyway.

Here is the deal though, there is much more to learn about causal inference. If you want to invest more time in this topic, I suggest you take a look at Morgan and Winship's Counterfactuals and Causal Inference: Methods and Principles for Social Research.[4] If you don't have the time to digest a whole book[5], read Felix Elwert's chapter on Graphical Causal Models and maybe also his paper on colliders.

Causal inference from observational data boils down to assumptions you have to make[6] and third variables you have to take into account. I'm going to talk about a third variable problem today, conditioning on a collider. You might not have heard of this before, but every time you condition on a collider, a baby stork gets hit by an oversized shoe filled with ice cream[7] and the quality of the studies supporting your own political view deteriorates.[8]

Let's assume you were interested in the relationship between conscientiousness and intelligence. You collect a large-ish sample of N = 10,000[9] and find a negative correlation between intelligence and conscientiousness of r = – .372 (see Figure 1).

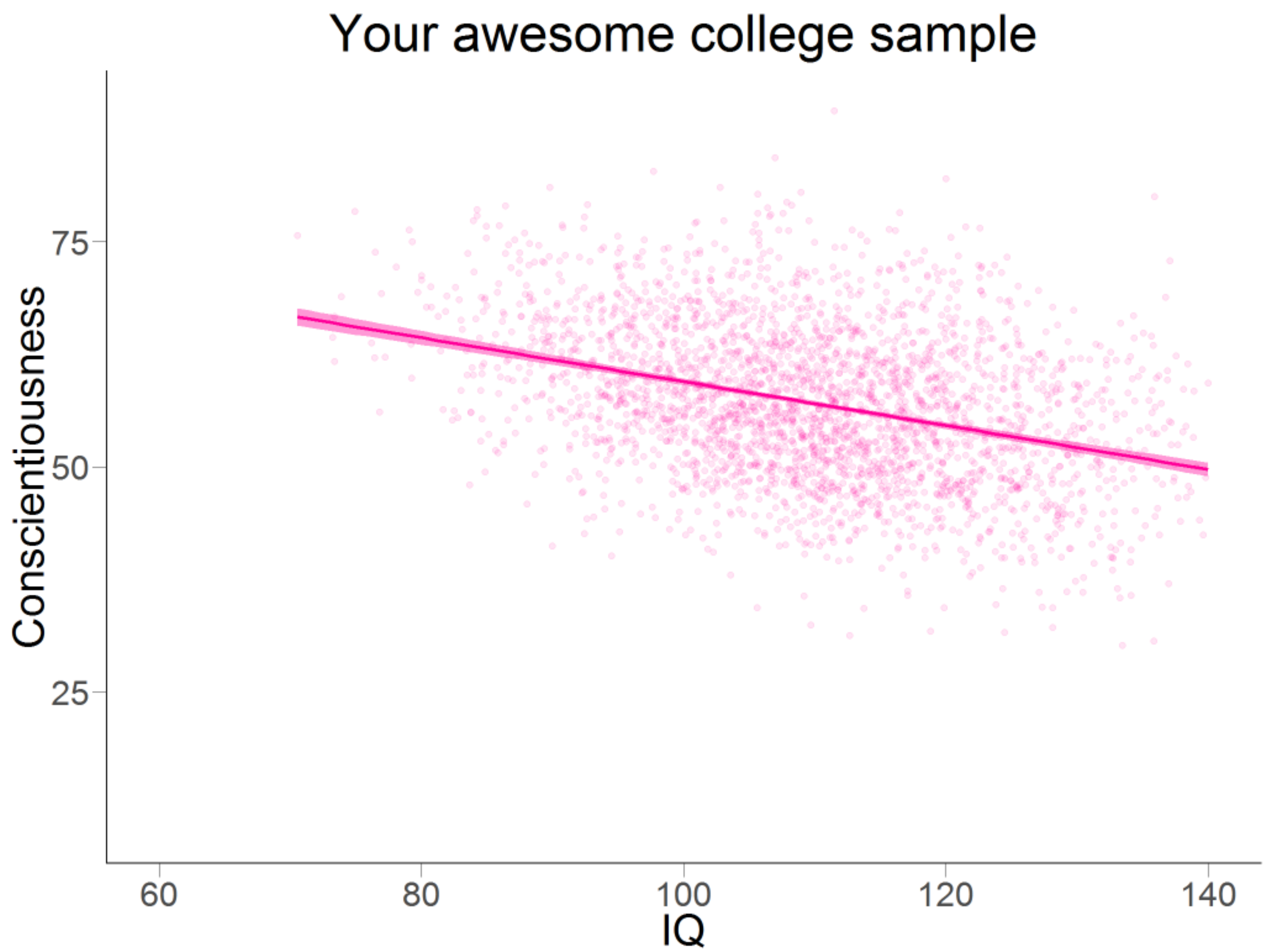

Figure 1: The relationship between IQ and conscientiousness in your hypothetical college sample. Anne expressed doubts regarding college students with an IQ of 75-85 and she might be right about that, but that's what you get for sloppy data simulations.

However, your sample consisted only of college students. Now you might be aware that there is a certain range restriction in intelligence of college students (compared to the overall population), so you might even go big and claim that the association you found is probably an underestimation! Brilliant.

The collider – being a college student – rears its ugly head. Being a college student is positively correlated with intelligence (r = .426). It is also positively correlated with conscientiousness (r = .433).[10] Let's assume that conscientiousness and intelligence have a causal (non-interactive) effect on college attendance, and that they are actually not correlated at all in the general population, see Figure 2.

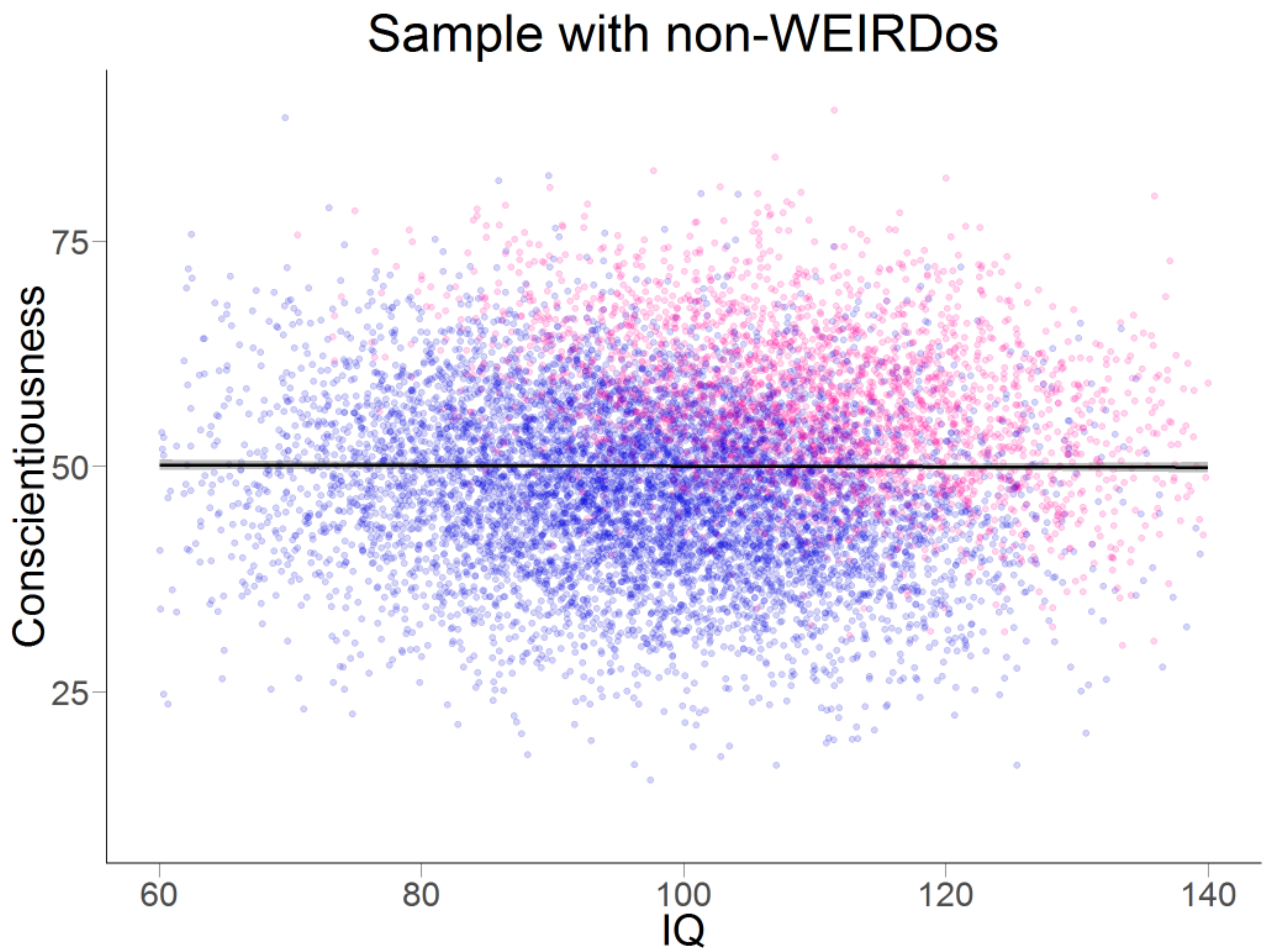

Figure 2. Oh no, where did your correlation go?

If you select a college sample (i.e. the pink dots), you will find a negative correlation between conscientiousness and intelligence of, guess what, exactly r = -.372, because this is how I generated my data. There is a very intuitive explanation for the case of dichotomous variables:[11] In the population, there are smart lazy people, stupid diligent people, smart diligent people and stupid lazy people.[12] In your hypothetical college sample, you would have smart lazy people, stupid diligent people, smart diligent people but no stupid lazy people because they don't make it to college.[13] Thus, in your college sample, you will find a spurious correlation between conscientiousness and intelligence.[14]

By the way, additionally sampling a non-college sample and finding a similar negative correlation among non-college peeps wouldn't strengthen your argument: You are still conditioning on a collider. From Figure 2, you can already guess a slight negative relationship in the blue cloud,[15] and pooling all data points and and estimating the relationship between IQ and conscientiousness while controlling for the collider results in r = -.240. Maybe a more relevant example: If you find a certain correlation in a clinical sample, and you find the same correlation in a non-clinical sample, that doesn't prove it's real in the not-so-unlikely case that ending up in the clinical sample is a collider caused by the variables you are interested in.

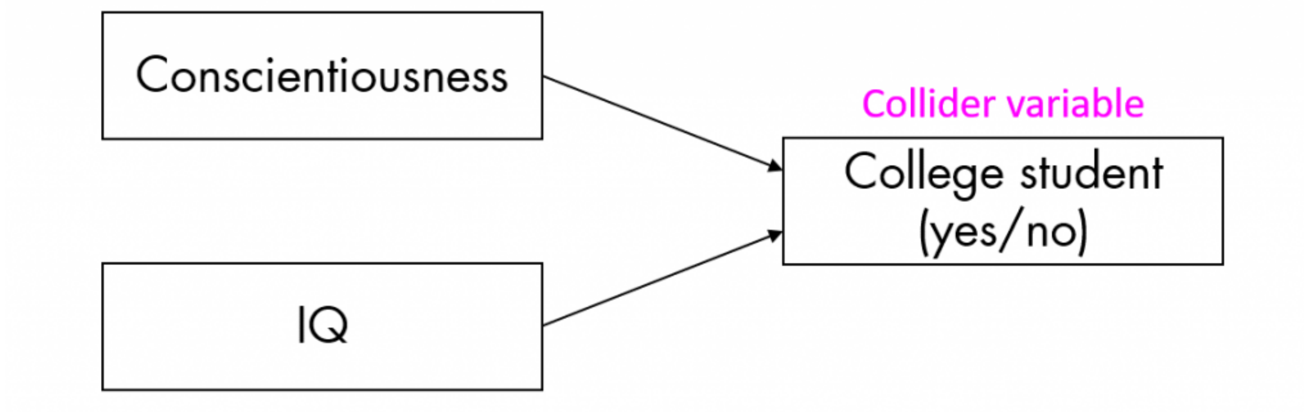

Figure 3. A variable constellation with a collider: The effects of conscientiousness and IQ "collide" in the joint outcome.

On an abstract level (Figure 3): Whenever X1 (conscientiousness) and X2 (intelligence) both cause Y (college attendance) in some manner, conditioning on Y will bias the relationship between X1 and X2 and potentially introduce a spurious association (or hide an existing link between X1 and X2, or exaggerate an existing link, or reverse the direction of the association…). Conditioning can mean a range of things, including all sort of "control": Selecting respondents based on their values on Y?[16] That's conditioning on a collider. Statistically controlling for Y? That's conditioning on a collider. Generating propensity scores based on Y to match your sample for this variable? That's conditioning on a collider. Running analyses separately for Y = 0 and Y = 1? That's conditioning on a collider. Washing your hair in a long, relaxing shower at CERN? You better believe that's conditioning on a collider. If survival depends on Y, there might be no way for you to not condition on Y unless you raise the dead.

When you start becoming aware of colliders, you might encounter them in the wild, aka everyday life. For example, I have noticed that among my friends, those who study psychology (X1) tend to be less aligned with my own political views (X2). The collider is being friends with me (Y): Psychology students are more likely to become friends with me because, duh, that's how you find your friends as a student (X1->Y). People who share my political views are more likely to become friends with me (X2->Y). Looking at my friends, they are either psych peeps or socialist anti-fascist freegan feminists.[17] Even though those two things are possibly positively correlated in the overall population,[18] the correlation in my friends sample is negative (X1 and X2 are negatively correlated conditional on Y).

Other examples: I got the impression that bold claims are negatively correlated with methodological rigor in the published psychological literature, but maybe that's just because both flashy claims and methodological rigor increase chances of publication and we just never get to see the stuff that is both boring and crappy?[19]
At some point, I got the impression that female (X1) professors were somewhat smarter (X2)

than male professors, and based on that, one might conclude that women are smarter than men. But female professors might just be smarter because tenure (Y) is less attainable for women (X1->Y)[20] and more likely for smart people (X2->Y), so that only very smart women become professors but some mediocre males can also make it. The collider strikes again!

Tenure and scientific eminence are nice examples in general because they are colliders for a fuckload of variables. For example, somebody had suggested that women were singled out as instances of bad science because of their gender. Leaving aside the issue whether women are actually overrepresented among the people who have been shamed for sloppy research,[21] such an overrepresentation would neither tells us that women are unfairly targeted nor that women are more prone to bad research practices.[22] Assuming that women (X1) have worse chances to get into the limelight than men, but overstating the implications of your evidence (X2) helps with getting into the limelight; we could find that women in the limelight (conditioning on Y) are more likely to have overstated their evidence because the more tempered women simply didn’t make it. That’s obviously just wild speculation, but in everyday life, people are very willing to speculate about confounding variables, so why not speculate a collider for a change?

Which leads to the last potential collider that I would like you to consider. Let’s assume that the methodological rigor of a paper (X1) makes you more likely to approve of it as a reviewer. Furthermore, let’s assume that you – to some extent – prefer papers that match your own bias (X2).[23] Even if research that favors your point of view is on average just as good as research that tells a different story (X1 and X2 are uncorrelated), your decision to let a paper pass or not (Y) will introduce a negative correlation: The published papers that match your viewpoint will on average be worse.[24]

So peeps, if you really care about a cause, don’t give mediocre studies an easy time just because they please you: At some point, the whole field that supports your cause might lose its credibility because so much bad stuff got published.

The End. (credit: Martin Brümmer)

Addendum: Fifty Shades of Colliders

Since publishing this post, I have learned that a more appropriate title would have been "That one weird third variable problem that gets mentioned quite a bit across various contexts but somehow people seem to lack a common vocabulary so here is my blog post anyway also time travel will have had ruined blog titles by the year 2100."

One of my favorite personality psychologists,[25] Sanjay Srivastava, blogged about the "selection-distortion effect" *before it was cool*, back in 2014.

Neuro-developmental psychologist Dorothy Bishop talks about the perils of correlational data in the research of developmental disorders in this awesome blog post and describes the problems of within-groups correlations.

Selection related to phenotypes can bias correlations in genetic studies which has be pointed out by (1) James Lee in Why It Is Hard to Find Genes Associated With Social Science Traits by Chris Chabris et al. and (2) in Collider Scope by Marcus Munafò et al.

Last but not least, Patrick Forscher just started a series of blog post about causality (first and second post are already up), starting from the very scratch. I highly recommend his blog for a more systematic yet entertaining introduction to the topic![26]

Update: In the meantime, I have written an article loosely based on this blog post which gives a somewhat more formal introduction to issues of causal inference. Check out the preprint, Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data