Update 15/06/2018

Paper based on this work now published.

Update 09/11/2017

Someone mentioned the paper was a bit equation heavy to read. So I'll give a simple example. You don't have to download the data and code to get the idea but you can if you want to. The data comes from this paper. Stata code is here. There are three binary variables: the outcome, the exposure and the confounder. The mean of the confounder is 38.5% in the whole population, but it differs across the two levels of the exposure (58.3% for exposure = 0 and 32.5% for exposure = 1) so we need to balance the confounder when looking at the relationship between the exposure and the outcome. One method would be to use inverse probability weighting. This balances the confounder at the population mean in both levels of the exposure i.e. 38.5%. It is easy and standard to compare balance before and after adjustment when using inverse probability weighting, however you usually don't see this in papers using an outcome regression. An outcome regression is just your standard regression model where the regression includes the outcome as the “dependent” variable and the exposure and confounder as “independent” variables. We illustrate a method to check where the outcome regression is balancing the confounder over the two levels of the exposure. It is at 51.9%. This is not the population mean of the confounder. The “effect” of the exposure on the outcome is slightly different in the outcome regression compared to the inverse probability weighting as they are different populations (one with a mean of the confounder balanced at 38.5% and the other at 51.9%). In the code I show that it you run the outcome regression with an interaction between the exposure and the confounder (so it is saturated) and then calculate, using standardisation, an average effect of the exposure, if we standardise to a population where the confounder is balanced at 51.9% we get the effect of the exposure on the outcome we obtained from the outcome regression without the interaction. Put another way, the outcome regression isn’t balancing the interaction.

(Yes I am aware that in the code I use a linear regression for a binary outcome, but it doesn’t matter here, it was just a handy dataset!)

Original post

We've a new working paper (pre-print) and I would welcome comments either here on the blog or on the OSF site where the pre-print is hosted. There is also R and Stata code. We've not shown anything new statistically but think we have a method of checking confounder balance in outcome regressions.

The abstract is below.

"An outcome regression controlling for observed confounders remains a popular way to assess the causal effect of an exposure in epidemiology, despite more modern causal techniques for adjusting for observed confounders, such as inverse probability weighting. A feature of inverse probability weighting is that checking balance of confounders in the control and exposure groups after confounder adjustment is simple. However, researchers using outcome regressions commonly do not check confounder balance after controlling for confounders. Although outcome regressions will balance any confounder specified in the model, the confounder value the model balances at is not transparent. We show that a matrix representation of an outcome regression reveals that an outcome regression includes a weight similar to an inverse probability weight. We also show that outcome regressions may not be balancing at the sample mean of the confounders particularly if interactions are not included with the exposure, which is typically the case in outcome regressions. Finally, we show that the coefficient of the exposure in an outcome regression is simply the difference between two weighted counterfactuals. Thus, there is an important connection between traditional outcome regression and modern causal techniques."

Tweet