Assessing the Total Effect of Time-Varying Predictors in Prevention Research

Bethany Cara Bray[1]
The Methodology Center

Rick S. Zimmerman
University of Kentucky

Donald Lynam
University of Kentucky

Susan Murphy
University of Michigan

Society for Prevention Research Annual Conference
Washington, D.C.
June 12, 2003

[1]Corresponding Author: Email: bcbray@psu.edu, Telephone: 814.865.1225, Address: The Methodology Center, Pennsylvania State University, S-159 Henderson Building, University Park, PA 16803

Frequently Asked Questions

1. What are time-varying and non-time-varying variables?

We consider a variable (i.e. peer pressure resistance) that has different values through out time to be time-varying. A variable (i.e. sex) that does not have different values through out time is non-time-varying.

2. What are confounders?

For the purposes of this work, a confounder is a variable that is correlated with both the predictor and the response. Confounders are alternate explanations of the observed relationship between the predictor and response. For instance, if the response is marijuana initiation and the predictor is conduct disorder initiation, peer pressure resistance is one confounder.

3. What are compositional differences?

We call the unequal distribution of levels of the confounder between the types of individuals that initiate the predictor and those who do not compositional differences.

4. Why are confounders so problematic?

Since confounders are correlated with both the predictor and response they offer alternate explanations of the observed relationship between the predictor and response. When confounding is not controlled, the unequal distribution of levels of the confounders among levels of the predictor (called compositional differences) causes bias in the estimated total effect of the predictor on the response. When confounding is not controlled, the estimated coefficient of the effect of the predictor on the response reflects the difference between the predictor groups, in addition to the causal effect of the predictor on the response. In other words, it is unclear whether the estimated effect of the predictor on the response represents the consequence that delayed predictor initiation has on the initiation of the response, or whether the estimated effect merely reflects compositional differences in the confounder, or if the estimated effect reflects a combination of the two.

5. Why are we worried about confounders only in observational and not experimental settings?

In experimental settings compositional differences are minimized by randomization. In observational settings, statistical methods and scientific assumptions to adjust for compositional differences are required.

6. What is a response regression model?

We call a regression model of the response (i.e. marijuana initiation) on the predictor (i.e. conduct disorder) and possibly other covariates (i.e. sex and race) a response regression model. The goal of this model is to estimate the total effect of the predictor on the response.

7. What is a total effect?

The entire effect of the predictor on the response through all direct and indirect influences.

8. What is the standard response regression model and what does it do?

We call a response regression model that includes confounders as covariates in the model the standard model. The standard model attempts to do two things simultaneously. The first is to control for confounding. The second is to estimate direct effects. We should worry when we are using one model to do two different things. Here we are going to focus on the problems with using the standard model to control for confounding while estimating the total effect of a predictor on a response when confounders are affected by the predictor, as often happens when the predictor and confounders are time-varying.

9. What is a spurious correlation?

A false, accidental correlation. For our purposes, a spurious correlation is an accidental correlation between the predictor and response, created by including confounders in the response regression model. Spurious correlations make the relationship between the predictor and response appear different than what is actually true.

10. Why are spurious correlations so problematic?

When confounders are included as covariates in the response regression model, a pathway opens between the predictor at time 1 and the response at time 3. This pathway is a spurious correlation that makes the relationship between the predictor and response (the estimated effect) appear different than what is actually true. When these spurious correlations cause the estimated effect to be different than what it actually is (a biased estimate), false conclusions regarding the consequences that the timing of the predictor has on the timing of the response may be made, leading to inaccurate conclusions, treatment, and intervention decisions.

11. How dangerous can spurious correlations be?

Simulations done by Barber et al. showed that, when using the standard model, in 80% of the data sets, false conclusions were reached. The strength of the correlations between the unmeasured confounder and the predictor, and the predictor and the confounder, is associated with the degree of bias when using the standard model. As these correlations decrease so does the bias from the standard model. No bias was found using the response regression model that has been weighted (if no assumptions are violated), no matter the strength or weakness of the correlations, or if a small number of covariates that are not confounders are included in the weights.

12. If the standard method produces biased results because of these spurious correlations what should I do instead?

Hernán, Brumback, and Robins (2000) use sample weights to statistically control for time-varying confounders. This method can produce unbiased estimates of the effect of a predictor on a response. Weights are created in a specific way and then used in a response regression model we call the weighted model.

13. Why does weighting work?

Weighting attempts to do what randomization does - equalize the compositional differences in the confounders among levels of the predictor. This makes the groups of people in the different predictor levels comparable. By equalizing the compositional differences between the predictor levels, the confounders are controlled and the correlation between the confounder and predictor is eliminated. Thus, the confounder does not need to be controlled by including it as a covariate in the response regression model, which eliminates the spurious correlation. In other words, weighting eliminates the path of the spurious correlation by not conditioning on the confounder in the final response regression model while controlling for confounders by equalizing the compositional differences between initiators and non-initiators of the predictor. Hence, the estimates from the final response regression model are unbiased.

14. What does the weighting method assume?

The weighting method assumes the following:

Assumption 1: there are no (direct) unmeasured confounders. This means that there are no direct causal effects (arrows) from any unmeasured variable (i.e. parent-child relationship quality) to the predictor (conduct disorder). This assumption is called "sequential ignorability."

Assumption 2: none of the past confounder patterns exclude particular current levels of the predictor. For example, even if an individual has very low peer pressure resistance it is still possible that the individual has initiated conduct disorder and vice-versa. The standard model can still produce unbiased estimates if the patterns that don't exist currently can exist in the future.

15. How robust is the weighting method to assumption violations?

So far, we know that if Assumption 1 is violated:

1. The response regression model that does not control for confounders at all has the most biased estimates, the weighted model has the least biased estimates, and the standard model's estimates are somewhere in between.

2. The more confounding we adjust for in the weights, the less and less biased the estimators become.

3. Adjusting for even a little bit of confounding is better than not controlling at all (e.g. including only some of the confounders is better than not including any of them).

4. The bias in the total effect estimate increases with the standard model as more confounders are included as covariates, if those confounders are affected by the predictor and are time-varying.

So far, we know that if Assumption 2 is violated:

1. Biased estimates result from the weighted model, due to some very high weights. This problem is unlikely as the probability must be very low for the unlikely combination of past confounder and current predictor levels.

16. In what types of prevention settings might I expect confounding to be particularly troublesome?

Confounding is an issue that should be considered in all prevention settings. When constructing models of prevention settings, researchers should always consider whether other measured and unmeasured variables are possibly related to both predictors and responses. If these other variables are time-varying, researchers need to think carefully about how to control for them in their analyses.

This project show that in substance use prevention common time-varying variables like peer pressure resistance, parent-child relationship quality, or family dysfunction can affect relationships between predictors and responses like conduct disorder, alcohol, cigarette, and marijuana use.

17. What are some future directions?

Some future directions include generalizing the weighting methodology to multilevel data structures, such as data organized into neighborhood groupings; examining procedures to detect assumption violations; and investigating the weighting method's robustness to assumption violations.

18. What are some additional references I can use?

Barber, J. S., Murphy, S. A., & Verbitsky, N. (2002). Adjusting for time-varying confounding in survival analysis. Manuscript submitted for publication.

Hernán, M., Brumback, B., & Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology, 11*, 5, 561-570.

Robins, J. M., Hernán, M. & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*, 5, 550-560.

Notes:

For more information or copies of the manuscript please contact Bethany Cara Bray.

Preparation of this article and poster was supported by Grant # P50 DA10075 from the National Institute on Drug Abuse to the Methodology Center at Pennsylvania State University and by the National Institute on Drug Abuse award #1 K02 DA15674-01.