Harvard University
Harvard University Biostatistics Working Paper Series

Year 2015 *Paper* 192

Negative Outcome Control for Unobserved Confounding Under a Cox Proportional Hazards Model

Eric J. Tchetgen Tchetgen* Tamar Sofer†
David Richardson‡

*Harvard T.H. Chan School of Public Health, etchetge@hsph.harvard.edu
†University of Washington, tsofer@uw.edu
‡University of North Carolina, david.richardson@unc.edu

Commentary

Negative outcome control for unobserved confounding under a Cox proportional hazards model

Eric J. Tchetgen Tchetgen[1,2], Tamar Sofer[3] and David Richardson[4]

Departments of Biostatistics[1] and Epidemiology[2], Harvard University

Department of Biostatistics[3], University of Washington,

Department of Epidemiology[4], University of North Carolina

Corresponding author: Eric J. Tchetgen Tchetgen, Department of Epidemiology, Harvard School of Public Health 677 Huntington Avenue, Boston, MA 02115.

Unobserved confounding can seldom be ruled out with certainty in observational studies. An approach sometimes used to determine the extent to which a treatment effect on a primary outcome may be subject to confounding bias, is to evaluate whether the treatment of interest is associated with a so-called negative control outcome after adjusting for all observed confounders.[1–6] A secondary outcome is then said to be a valid negative control variable, to the extent that it is influenced by unobserved confounders of the treatment effects on the primary outcome, while not causally influenced by the treatment.[3] Thus, a negative control outcome found to be empirically associated with the treatment may provide compelling evidence of unobserved confounding for the primary outcome, provided that no unobserved confounder of the treatment-negative control outcome fails also to confound the treatment-primary outcome relation.[3] Suppose that in a certain application, beyond assessing the presence of an association between a negative control outcome and the treatment to detect confounding bias, one may wish to use the negative control outcome to correct for confounding bias. A natural next step might be to consider the magnitude of the estimated association as an unbiased estimate of bias due to unmeasured confounding, and one might be tempted to simply correct the confounded estimate of the treatment-outcome association by subtracting the estimated bias. While this ad-hoc bias correction is intuitively appealing, it has previously been noted that it is sensitive to the relative scales of the primary and negative control outcomes.[5] A notable difficulty with the approach is that it requires interpreting the bias observed for the negative control outcome as somehow equivalent to the bias one would have observed between the treatment and the primary outcome under the null hypothesis of no causal effect of the treatment (more precisely between the treatment and the treatment-free potential outcome). A prerequisite for such "bias equivalence" is that the primary and negative control outcomes are measured on scales of comparable magnitude. An important case where bias equivalence can be expected to hold is when the control outcome is a baseline measure of the outcome process prior

to treatment. In such settings, the well known difference-in-differences approach to control for unobserved confounding may formally be justified as a negative outcome control approach.[5] However, outside of this special situation, the assumption of bias equivalence may not be appropriate if the outcomes are clearly measured on different scales, such as, say if the negative control outcome is dichotomous while the primary outcome is continuous. The assumption would then likely be violated since an additive association of the exposure with the control outcome would a priori be restricted by the binary nature of the outcome, while the additive association of the outcome in view with exposure would not.[5]

In this note, we describe another prominent setting where indirect adjustment for unobserved confounding can be achieved by differencing regression estimates of the effect of treatment on the primary and negative outcome. Specifically, we show that when the primary and negative outcomes are time to event outcomes that follow a certain Cox proportional hazards model respectively, then under fairly reasonable monotonicity assumptions, a consistent estimate of the treatment causal effect can be obtained by simply subtracting the estimated log-hazards ratio for the treatment-negative outcome association from that of the treatment-primary outcome association. Thus, we show that for the Cox proportional hazards regression model, the intuitive indirect adjustment for unobserved confounding by differencing effect estimates is in fact sound for a wide range of settings of common interest. To formally state the result requires introducing some notation. Let T denote the primary time to event outcome, N denote the negative control time to event outcome, A denote the exposure of interest, and X denote a set of pre-exposure covariates. In addition, suppose that the hazard of N conditional on A and X, $h_{N|A,X}(n)$ is given by:

$$h_{N|A,X}(n) = \exp\{\beta(A,X)\}\, h_{N|A=0,X}(n) \tag{1}$$

where the log hazards ratio function $\beta(A, X)$ is independent of N, satisfies $\beta(0, X) = 0$, and therefore the proportional hazards assumption holds conditional on X. Likewise, suppose that the hazard function of T conditional on A, X also satisfies the proportional hazards assumption within levels of X:

$$h_{T|A,X}(t) = \exp\{\eta(A, X)\} h_{T|A=0,X}(t) \tag{2}$$

In addition, let N_a and T_a denote the potential outcomes under treatment a for the negative and primary outcomes respectively. Under consistency, we have that $N_a = N$ and $T_a = T$ if $A = a$. Under the assumption that N is a valid negative control outcome, we also have that $N_a = N_{1-a} = N$, that is A has no individual causal effect on N. We further let U denote an unmeasure confounder which is a continous unobserved common cause of A, T and N. Unobserved confounding can naturally be incorporated algebraically by supposing that there exist functions h_0 and h_1 such that given X, the potential outcomes in the absence of treatment T_0 and $N_0 = N$ satisfy

$$T_0 = h_0(U, X) \tag{3}$$

$$N = h_1(U, X) \tag{4}$$

with $h_j(u, x)$ monotone increasing in u for all x, but otherwise unrestricted functions. We then have the following result.

Result : Under assumptions $(1) - (4)$, *we have that*

$$\psi(X) = \frac{h_{T_1|A=1,X}(t)}{h_{T_0|A=1,X}(t)} = \exp(\eta(1, X) - \beta(1, X))$$

and therefore the causal log hazards ratio $\psi(X)$ *encoding the effect of* A *for the exposed conditional*

on X can be obtained by simply subtracting the log- hazards ratio of A for the negative control outcome from that for the primary outcome.

The proof of Result is provided in the Supplemental Materials, where we show that the result continues to hold even if the unobserved confounder of the A-T association is distinct from that of the A-N association, provided that the association between the former and treatment is equivalent to that of the second in a sense made precise in the Supplemental Materials.

APPENDIX

Proof of Result. Let $S_{N|A,X}(n) = P\{N \geq n|A,X\}$ and $F_{N|A,X}(n) = P\{N < n|A,X\}$. Let $F_{T_0|A,X}$ and $F_{U|A,X}$ be likewise defined. Then, by Theorem 1 of Sofer et al,[11] we have that under monotonicity assumptions (3) and (4), for all v in the unit interval,

$$F_{T_0|A,X} \circ F^{-1}_{T_0|A=0,X}(v) \tag{5}$$

$$= F_{N|A,X} \circ F^{-1}_{N|A,X}(v) \tag{6}$$

$$= F_{U|A,X} \circ F^{-1}_{U|A=0,X}(v) \tag{7}$$

Note that , under the proportional hazards assumption, we have that

$$S_{N|A,X}(n) = \exp\left\{-\exp(\beta(A,X))\, H_{N|A=0,X}(n)\right\}$$

, where $H_{N|A=0,X}(n) = \int_0^n h_{N|A=0,X}(m)\, dm$, and

$$\begin{aligned} F^{-1}_{N|A,X}(v) &= S^{-1}_{N|A,X}(1-v) \\ &= H^{-1}_{N|A=0,X}\{-\log(1-v)\exp(-\beta(A,X))\} \end{aligned}$$

therefore

$$\begin{aligned} & F_{N|A=0,X} \circ F^{-1}_{N|A,X}(v) \\ &= 1 - \exp\left\{-H_{N|A=0,X}\left(F^{-1}_{N|A,X}(v)\right)\right\} \\ &= 1 - \exp\left\{-H_{N|A=0,X}\left(H^{-1}_{N|A=0,X}\{-\log(1-v)\exp(-\beta(A,X))\}\right)\right\} \\ &= 1 - \exp\{\log(1-v)\exp(-\beta(A,X))\} \end{aligned}$$

does not depend on the conditional cumulative hazard function $H_{N|A=0,X}$. Consider the causal quantile-quantile transformation

$$\psi(v, A, X) = F_{T_0|A,X} \circ F^{-1}_{T_A|A,X}(v)$$

Under the Cox model

$$h_{T|A,X}(t) = \exp(\eta(A, X))\, h_{T|A=0,X}(t)$$

holds, we then have that,

$$\begin{aligned}
\psi(v, A, X) &= F_{T_0|A,X} \circ F^{-1}_{T_0|A=0,X} \circ F_{T_0|A=0,X} \circ F^{-1}_{T_A|A,X}(v) \\
&= F_{U|A,X} \circ F^{-1}_{U|A=0,X} \circ F_{T_0|A=0,X} \circ F^{-1}_{T_A|A,X}(v) \\
&= F_{N|A,X} \circ F^{-1}_{N|A=0,X} \circ F_{T_0|A=0,X} \circ F^{-1}_{T_A|A,X}(v) \\
&= 1 - \exp\{\log(1-v) \exp(\eta(A, X) - \beta(A, X))\}
\end{aligned}$$

or equivalently

$$\frac{h_{T_A|A,X}(t)}{h_{T_0|A,X}(t)} = \exp(\eta(A, X) - \beta(A, X)),$$

proving the result.

The result continues to hold if we relax the assumption of a common unmeasured confounder for T as for N, thus allowing the unobserved confounder U of the $A-T$ association to be distinct from the unobserved confounder W of the $A-N$ association, provided that the association between U and A is similar to the association between W and A in the following sense. For all u in the unit interval,

$$F_{U|A,X} \circ F^{-1}_{U|A=0,X}(v) = F_{W|A,X} \circ F^{-1}_{W|A=0,X}(v) \tag{8}$$

This is the assumption of quantile-quantile equi-confounding introduced in Sofer et al[11]. The assumption is weaker and thus implied by the more easier interprete assumption of distributional equi-confounding which states that the conditional distribution of U given A and X matches the conditional distribution of W given A, X, i.e.

$$F_{U|A,X}(\cdot) = F_{W|A,X}(\cdot).$$

The result then follows from the fact that equation (8) implies equations (6) and (7).

References

[1] Rosenbaum P. The Role of known effects in observational studies. Biometrics.1989, 45(2), 557-569.

[2] Rosenbaum P. Observational Studies. 2002. Second Edition. Springer-Verlag, New York.

[3] Lipsitch M, Tchetgen Tchetgen EJ, Cohen T. Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies Epidemiology. Epidemiology. 2010; 21(3): 383–388.

[4] Flanders WD, Klein M, Strickland M, Darrow L, Sarnat S, Sarnat J, Waller L, Winquist A, Tolbert PE. A Method for Detection of Residual Confounding in Time-Series and Other Observational Studies. Epidemiology. 2011;22(1):59–67.

[5] Tchetgen Tchetgen EJ. The control outcome calibration approach (COCA) for unobserved confounding (2013) American Journal of Epidemiology, 179(5):633-40.

[6] Schuemie, Martijn J., et al. "Interpreting observational studies: why empirical calibration is needed to correct p-values." Statistics in medicine 33.2 (2014): 209-218.

[7] Groenwold, Rolf HH. "Falsification End Points for Observational Studies." JAMA 309.17 (2013): 1769-1771.

[8] Richardson D, Cole S, Tchetgen Tchetgen EJ, Laurier D. Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models. 2014. American Journal of Epidemiology.180(9): 933-40.

[9] Richardson D, Tchetgen Tchetgen EJ, Keil A,Cooper G. Negative Control Outcomes and the Analysis of Standardized Mortality Ratios. (2015). Epidemiology. In Press.

[10] Dusetzina, S, Brookhart A, and Maciejewski, M. Control Outcomes and Exposures for Improving Internal Validity of Nonrandomized Studies. Health services research (2015). In Press.

[11] Sofer T, Colicino E, Schwartz, Richardson, D,Tchetgen Tchetgen EJ.(2015) On simple relations between difference-in-differences and negative outcome control: identifying assumptions and a generalization. Submitted.