The use of negative controls to detect confounding and other sources of error in experimental and observational science

Marc Lipsitch
Eric Tchetgen Tchetgen
Ted Cohen

eAppendix

1. Use of negative control outcomes to detect uncontrolled confounding

We briefly provide an analytical basis for the use of an ideal *U*-comparable negative control outcome to detect the presence of uncontrolled confounding. For the sake of simplicity, we limit our treatment to mean regression with identity link function, where confounding adjustment is done by conditioning on potential confounders in the model, and we provide details only for the case of dichotomous (A,U). Nonetheless, with additional technical arguments, our results can be generalized to other types of regression models, and to settings with continuous (A,U).

The fact that *A* has no effect on the mean of *N* within levels of *L* and *U* implies that $E(N|A,U,L)=E(N|U,L)$. Thus, marginalizing over *U* given *A* and *L*, we obtain the following expression

$$E(N \mid A,L) = E[E(N \mid U,A,L) \mid A,L] = E[E(N \mid U,L) \mid A,L]$$

In the case of dichotomous *A* and *U*, this result immediately leads to the following simple and intuitive expression for the confounded conditional effect of *A* on *N*for each level of *L*,

$$\Delta(L) = E(N \mid A=1,L) - E(N \mid A=0,L)$$
$$= [E(N \mid U=1,L) - E(N \mid U=0,L)] \times [E(U \mid A=1,L) - E(U,A=0,L)]$$

This formula confirms that in the absence of uncontrolled confounding, $\Delta(L)=0$ for all values of *L*. This is because under the assumption of no uncontrolled confounding, any random variable *U* not in (*L,A,Y*) either is independent of *N* given *L*, making $E(N \mid U=1,L) - E(N \mid U=0,L) = 0$, or is independent of *A* given *L*, making $E(U \mid A=1,L) - E(U,A=0,L) = 0$. However, in the presence of uncontrolled confounding, we can generally expect that $\Delta(L) \neq 0$ for some value of *L*.

The above observation leads to the following general strategy to detect the presence of uncontrolled confounding by testing the null hypothesis: $\Delta(L)=0$ for all values of *L*. In practice, as *L* may contain 2 or more continuous components, or multiple categorical variables, this is achieved by fitting a regression model $E(N \mid A,L,\boldsymbol{\theta})$)with unknown parameter $\boldsymbol{\theta}$, for the mean of *N* given covariates *A* and *L*; and subsequently testing for a significant effect of *A* in this model. For instance, in the simple case where one correctly specifies the model $E(N \mid A, L, \boldsymbol{\theta}) = \theta_0 + \theta_1 A + \theta_2 ' LA + \theta_3 ' L$ where $L = (L_1 ', L_2 ')'$ and $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2 ', \theta_3 ']'$, a test for the presence of uncontrolled confounding thus involves a standard statistical test of the null hypothesis of zero values for the coefficients for the regression of N on A, i.e. a test that θ_1 and θ_2 are both zero. We note that none of the above arguments require any additional restriction on the relationship between *U* and *L*, which can generally be dependent. Furthermore, the usual modeling caveats equally apply to this situation, in particular, the validity of the proposed test is reliant on the analyst's ability to appropriately account for measured confounders *L*; a misspecified model $E(N \mid A, L, \boldsymbol{\theta})$ will often lead to incorrect conclusions about the presence of uncontrolled confounding.

eAppendix2.Use of Negative control exposures to detect uncontrolled confounding.

In the case of an ideal *U*-comparable negative control exposure, knowledge that *B* does not affect the mean of *Y* within levels of *A,L* and *U* implies that $E(Y \mid A, B, U, L) = E(Y \mid A, U, L)$. Thus, upon marginalizing over *U* given *B,A* and *L*, we obtain $E(Y \mid B, A, L) = E[E(Y \mid A, U, L) \mid B, A, L]$. When *B* and *U* are binary, this expression provides a simple and intuitive formula for the confounded conditional effect of *B* on *Y*:

$$\delta(L, A) = E(Y \mid B = 1, A, L) - E(Y \mid B = 0, A, L)$$
$$= [E(Y \mid A, U = 1, L) - E(Y \mid A, U = 0, L)] \times [E(U \mid B = 1, A, L) - E(U \mid B = 0, A, L)]$$

Thus, similar to our result for negative control outcomes, in the presence of uncontrolled confounding, we expect that $\delta(L, A) \neq 0$ for some joint level of A and L. A test for uncontrolled confounding is thus a test of the null hypothesis $\delta(L, A) = 0$ for all values of *A* and *L*. This is easily operationalized by regressing *Y* on *A,L* and *B* in a parametric mean regression model, and subsequently performing a standard statistical test for the effect of *B* on *Y* given *A* and *L* in the model.

eAppendix3. A remark on the use of negative control variables to infer the direction and magnitude of confounding bias.

In the event that the null hypothesis of no uncontrolled confounding is correctly rejected (i.e. $\Delta(L) \neq 0$ for some value of *L*, or $\delta(L, A) \neq 0$ for some value of *A* and *L*) , an interesting question arises: can the magnitude and/or the direction of confounding bias of an estimate of the conditional effect of *A* on *Y*, be inferred from an estimate of $\Delta(L)$, the regression of the negative control outcome on the exposure of interest, or $\delta(L, A)$, the regression of the outcome of interest on the negative control exposure? As we now argue, this is generally impossible, unless additional assumptions are made beyond those stated so far.

To illustrate this point, consider the case where a researcher identifies an ideal *U*-comparable negative control outcome *N*; furthermore, suppose that the data were generated by a process described by the model $E(Y \mid A,L,U;\boldsymbol{\beta}) = \beta_0 + \beta_1 A + \beta_2 ' L + \beta_3 U$ for the mean of *Y* given *A,L* and *U*, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2 ', \beta_3)$, so that β_1 encodes the unknown conditional causal effect of *A* on *Y*. Now, because *U* is not observed, suppose the analyst fits the reduced model that specifies $\beta_3 = 0$ to data *Y*,*A* and *L*. Then, by standard linear regression theory, the least-squares estimate of the effect of *A* in this reduced model can be shown to converge in probability (with increasing sample size) to the quantity $\beta_3 \pi + \beta_1$, where π denotes the asymptotic value of the least-squares estimate of the effect of *A* in a (possibly incorrect) linear regression of *U* on *A* and *L*. Therefore, the product $\beta_3 \pi$ quantifies the asymptotic bias of the ordinary least-squares estimate of the effect of *A* on *Y* due to uncontrolled confounding.

Next, suppose that in order to detect the presence of confounding bias, the analyst uses the regression model given in section 1 of this Appendix, $E(N \mid A,L,\boldsymbol{\theta}) = \theta_0 + \theta_1 A + \theta_3 ' L$ of the negative control outcome on *A* and *L*, where for simplicity, we set $\theta_2=0$. When this latter model is correct,as previously established, the confounded effect of *A*on *N* within levels of *L* equals $\Delta(L) = [E(N \mid U=1,L) - E(N \mid U=0,L)] \times [E(U \mid A=1,L) - E(U \mid A=0,L)] = \theta_1$. Therefore, we see that, although the formulae for the confounding bias θ_1 and $\beta_3 \pi$ in estimating the effect of *A*on *N* and the effect of *A* on *Y* respectively,are both a product of two terms, there is a priori no reason why these two expressions should be equal unless one is willing to make additional assumptions.

For instance, the equality $\theta_1 = \beta_3 \pi$ would indeed hold if the following two assumpions were met:

1) the mean function $E(U \mid A,L)$ is in the linear span of *A* and *L*, and
2) within levels of *A* and *L*, the effect of *U* on *N* is equal to the effect of *U* on *Y*, i.e.

$$\begin{aligned} & E(N \mid U=1,L) - E(N \mid U=0,L) \\ & = E(N \mid A,U=1,L) - E(N \mid A,U=0,L) \\ & = E(Y \mid A,U=1,L) - E(Y \mid A,U=0,L) \end{aligned}$$

This is because the first assumption implies that $\pi = E(U \mid A=1,L) - E(U \mid A=0,L)$ when the mean of *U* depends linearly on *A* and *L*, whereas the second assumption states that the magnitude of the association between *U* and *N* is equal to that between *U* and *Y,i.e* $\beta_3 = E(N \mid U=1,L) - E(N \mid U=0,L)$. By combining 1) and 2), we obtain $\theta_1 = \beta_3 \pi$. Clearly, both of these assumptions are empirically untestable as they directly involve the uncontrolled confounder and would generally be unrealistic unless they are based on very firm scientific understanding.

In summary, the regression results for a negative control outcome or a negative control exposure cannot be used in a simple way to “correct” the equivalent regressions for the outcome of interest or the exposure of interest, respectively, even in the simple linear setting considered here.

Nonetheless, upon rejecting the hypothesis of no uncontrolled confounding, the results of negative controls can be additionally informative in some simple settings without the need for potentially unrealistic assumptions. To illustrate, suppose that there is no *L,* so that a sample of independent and identically distributed data on *Y*,*A*,*U* is generated,where *A* and *U* can either be discrete or continuous. However data on *U* are not observed, and the following models hold:

$$E(Y \mid A,U;\boldsymbol{\beta}) = \beta_0 + \beta_1 A + \beta_3 U; \qquad E(A \mid U;\boldsymbol{\rho}) = \rho_0 + \rho_1 U,$$

where $\boldsymbol{\rho} = (\rho_0, \rho_1)$;

$E(B \mid U;\boldsymbol{\eta}) = \eta_0 + \eta_1 U$, where $\boldsymbol{\eta} = (\eta_0, \eta_1)$, and ρ_1, η_1 and β_3 are bounded away from zero, so that *U* confounds both the null association between *A* and *Y* and the association of *B* with *Y*. In this simple setting, when *U* is unobserved, the asymptotic bias of the least-squares estimate $\widehat{\beta}_1$ of the marginal effect of *A* on *Y,* can be re-expressed as $\frac{\beta_3}{\rho_1}\frac{\rho_1^2 \,\mathrm{var}(U)}{\mathrm{var}(A)} = \frac{\beta_3}{\rho_1}(1-\lambda)$, where

$\lambda = \frac{E[\mathrm{var}(A \mid U)]}{\mathrm{var}(A)} = \frac{E[\mathrm{var}(A \mid U)]}{\mathrm{var}[E(A \mid U)] + E[\mathrm{var}(A \mid U)]}$. When A and U are continuous, the fraction $1-\lambda$ can be interpreted as the proportion of the variance of *A* due to the effect of*U* on the mean of*A*; if $1-\lambda = 0$ there is no uncontrolled confounding and thus there is no corresponding bias, whereas, as the association between *U* and the mean of *A* explains an increasing proportion of the variance of *A*, $1-\lambda$ tends to 1 and the worst-case confounding bias corresponds to the limiting value $\frac{\beta_3}{\rho_1}$. As λ directly involves the variance of *U,* there is generally no hope of estimating it from the observed data. However, as we show next, appropriate use of the negative control exposure B permits identification of the sign of the bias of $\widehat{\beta}_1$ with no further assumption required. Specifically, consider the statistic $T = \frac{\sum_i (Y_i - \bar{Y})B_i}{\sum_i (A_i - \bar{A})B_i}$. We will show below that *T* converges in probability to $\beta_1 + \frac{\beta_3}{\rho_1}$, so that $T - \hat{\beta}_1$ converges to $\lambda\frac{\beta_3}{\rho_1}$. Now, because $0 < \lambda < 1$ the sign of $\lambda\frac{\beta_3}{\rho_1}$ agrees with the sign of $\frac{\beta_3}{\rho_1}(1-\lambda)$; thus in large samples, we can expect the sign of

$T-\hat{\beta}_1$ to generally agree with the sign of the bias of the least-squares estimator of the effect of A on Y. Formally, this is true with probability tending to one. Furthermore, it is interesting to note that for an assumed value of λ, one obtains the following bias corrected least-squares estimator $\tilde{\beta}_1(\lambda)=T-\frac{T-\widehat{\beta}_1}{\lambda}$. In principle, this opens up the possibility of performinga simple sensitivity analysis by varying λ to assess the potential impact of the magnitude of uncontrolled confounding on the corrected least-squares estimator $\tilde{\beta}_1(\lambda)$.

The above discussion did not allow for the presence of observed confounders *L*. Nonetheless, the results generalize quite naturally if confounding adjustment is done by stratification. However, no similar results are currently available for situations where confounding adjustment is performed either by conditioning in the model or by inverse-probability weighting.

We finally provide technical arguments supporting our results. First, to derive the large sample bias of $\widehat{\beta}_1$, we note that

$$\widehat{\beta}_1=\frac{\sum_i (Y_i-\bar{Y})A_i}{\sum_i (A_i-\bar{A})A_i}=\beta_1+\beta_3\frac{\sum_i (U_i-\bar{U})A_i}{\sum_i (A_i-\bar{A})A_i}+\frac{\sum_i (\varepsilon_{y,i}-\overline{\varepsilon_y})A_i}{\sum_i (A_i-\bar{A})A_i}\text{, where}$$

$\varepsilon_{y,i}=Y_i-(\beta_0+\beta_1 A_i+\beta_3 U_i)$, and the overbar denotes the sample average. The third term in the above expression for $\widehat{\beta}_1$ converges to zero in probability with increasing sample size, whereas the second term, which constitutes the bias of least-squares –

$$\beta_3\frac{\sum_i (U_i-\bar{U})A_i}{\sum_i (A_i-\bar{A})A_i}=\beta_3\rho_1\frac{\sum_i (U_i-\bar{U})U_i}{\sum_i (A_i-\bar{A})A_i}+\beta_3\frac{\sum_i (U_i-\bar{U})(A_i-\rho_1 U_i)}{\sum_i (A_i-\bar{A})A_i}\text{--}$$

converges in probability to $\frac{\beta_3}{\rho_1}\frac{\rho_1^2\,\mathrm{var}(U)}{\mathrm{var}(A)}=\frac{\beta_3}{\rho_1}(1-\lambda)$, where we use the fact that $\frac{\sum_i (U_i-\bar{U})(A_i-\rho_1 U_i)}{\sum_i (A_i-\bar{A})A_i}$ converges in probability to $\frac{E([U-E(U)][A-E(A\mid U)])}{E([A-E(A)]A)}=0$.

It remains to show that *T* converges to $\beta_1+\frac{\beta_3}{\rho_1}$. This follows from

$$T=\frac{\sum_i (Y_i-\bar{Y})B_i}{\sum_i (A_i-\bar{A})B_i}=\beta_1+\beta_3\frac{\sum_i (U_i-\bar{U})B_i}{\sum_i (A_i-\bar{A})B_i}+\frac{\sum_i (\varepsilon_{y,i}-\bar{\varepsilon}_y)B_i}{\sum_i (A_i-\bar{A})A_i}\,,$$

where the third term converges to zero in probability, and the second term can be written:

$$\beta_3 \frac{\sum_i (U_i - \bar{U})B_i}{\sum_i (A_i - \bar{A})B_i} = \beta_3 \frac{\sum_i (U_i - \bar{U})(\eta_1 U_i + \epsilon_{b,i})}{\sum_i [\rho_1 (U_i - \bar{U}) + (\epsilon_{3,i} - \bar{\epsilon}_3)(\eta_1 U_i + \epsilon_{b,i})}$$

$$= \beta_3 \eta_1 \frac{\sum_i (U_i - \bar{U})U_i}{\sum_i [\eta_1 \rho_1 (U_i - \bar{U})U_i + \eta_1 (\epsilon_{3,i} - \bar{\epsilon}_3)U_i + \rho_1 (U_i - \bar{U})\epsilon_{b,i} + (\epsilon_{3,i} - \bar{\epsilon}_3)\epsilon_{b,i}]} +$$

$$\beta_3 \frac{\sum_i (U_i - \bar{U})\epsilon_{b,i}}{\sum_i [\eta_1 \rho_1 (U_i - \bar{U})U_i + \eta_1 (\epsilon_{3,i} - \bar{\epsilon}_3)U_i + \rho_1 (U_i - \bar{U})\epsilon_{b,i} + (\epsilon_{3,i} - \bar{\epsilon}_3)\epsilon_{b,i}}.$$

The second term in this last equation is easily shown to converge to zero in probability by an application of the law of large numbers, whereas the first term converges in probability to the desired quantity.