

Negative Control Outcomes Model

Phuc Nguyen, David Pritchard, Alan Brookhart

December 29, 2023

1 Background

1.1 Unmeasured confounding

Under the potential outcome framework, a dichotomous counterfactual outcome variable Y^a is the random variable that would have been observed under treatment a . For the common setting of a dichotomous treatment A , each individual has two potential outcomes Y^0 , the outcome variable that would have been observed under the control treatment, and Y^1 , the outcome variable that would have been observed under the experimental treatment [6]. The average treatment effect (ATE) of a given treatment in a particular population is a widely-used causal estimand. For a dichotomous treatment, on the risk difference scale, the measure of the ATE in the population is defined as $E[Y^1] - E[Y^0]$. Similarly, the measure of the ATE on the risk ratio scale is defined as $E[Y^1]/E[Y^0]$. The measure of the average treatment effect on the treated (ATT), defined as $E[Y^1|A = 1] - E[Y^0|A = 1]$, is another estimand of interest.

In observational studies, the ATE can be identified—expressed as a function of the observed data—under identifying assumptions including consistency, no interference, positivity, and exchangeability [11]. In this chapter, we focus solely on issues related to exchangeability being violated and hereafter assume consistency, no interference, and positivity hold. (Interested readers can see [6, 11] for details on the other assumptions.) Exchangeability means that the risk of the outcome under potential treatment a is the same among the treated and untreated individuals, or $E[Y^a|A = 0] = E[Y^a|A = 1] \quad \forall a \in \{0, 1\}$:

Assumption 1 (*Exchangeability*) $Y^a \perp A$ for all a

Under exchangeability, the ATE is equivalent to the observed association, e.g. the causal risk difference is equivalent to $E[Y|A = 1] - E[Y|A = 0]$. A weaker assumption known as conditional exchangeability also enables the identification of the ATE:

Assumption 2 (*Conditional exchangeability*) $Y^a \perp A|L$ for all a

where L is a set of covariates. Under conditional exchangeability, the causal risk difference is:

$$E[Y^1] - E[Y^0] = \sum_l \left(E[Y|A = 1, L = l] - E[Y|A = 0, L = l] \right) P(L = l) \quad (1)$$

Ideal randomized experiments automatically lead to exchangeability, however, observational studies often do not because factors L used to determine treatment assignments are also likely to be associated with the outcome. The association between treatment A and outcome Y is then made up of the true causal effect and a bias induced by factors L . When factors L are common causes of the treatment and the outcome, we refer to the bias as confounding [6]. In observational studies of medical interventions, confounding is common as physicians prescribe treatments based on their assessment of a patient’s disease risk and patients seek treatment for reasons often related to prognosis [1]. Under mild technical assumptions, confounding implies that exchangeability does not hold [6]. However, conditional exchangeability can still hold if we have measured a sufficient set for confounding adjustment [6]. We can determine this set using the backdoor criterion [12] and the knowledge of the true causal directed acyclic graph (DAG) that generated the data. (See Chapter 7 in [6] for more details.)

In practice, we can never guarantee that we have adjusted for a sufficient set, for instance, because we don’t know the true causal DAG or because it’s impossible to collect data on some variables. Returning then to our previous example of healthcare studies where physicians prescribe the treatments of interest, it is likely impossible to be sure that the set of covariates being controlled for accounts for all possible risk factors for the outcome that influenced the treatment decision. Confounding that requires adjustment for unmeasured variables, which is impossible, is referred to as unmeasured confounding. Under mild technical assumptions, unmeasured confounding implies that conditional exchangeability also does not hold [6]. It creates a systemic bias in the causal estimate that persists even as the study’s sample size increases. It is because of this clear opportunity for violation of identifying assumptions that unmeasured confounding is often considered to be the main threat to the validity of observational studies [3]. Although the no unmeasured confounding assumption is untestable [16], negative controls are useful tools to detect and sometimes correct bias from unmeasured confounding [8].

1.2 Negative controls

Suppose that, in addition to measured variables L , there exist unmeasured common causes U of the treatment A and outcome Y . Given unmeasured confounders U , we have conditional exchangeability, that is $Y^a \perp A|U, L$ for all a . Let N be a secondary outcome variable and N^a be its potential outcome variable under treatment a . Outcome variable N is a negative control outcome (NCO) if it satisfies the following [16]:

Assumption 3 (*Negative control outcome*) $N^a = N$ and $N \perp A|U, L$ for all a

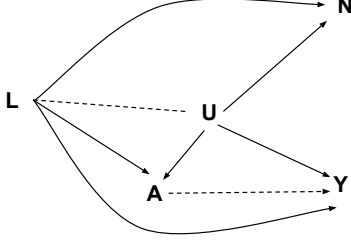


Figure 1: Causal DAG of an ideal NCO N for bias detection/correction of the effect of treatment A on outcome Y , with observed confounders L and unmeasured confounders U . Ideal NCOs have the same incoming arrows as Y , except that from A , and are U -comparable to Y [8]. A dashed line means that a relationship may exist.

A secondary exposure Z is a negative control exposure (NCE) if it satisfies the following [16]:

Assumption 4 (*Negative control exposure*) $Y^{a,z} = Y^a$ and $Y^a \perp Z|U, L$ for all a, z

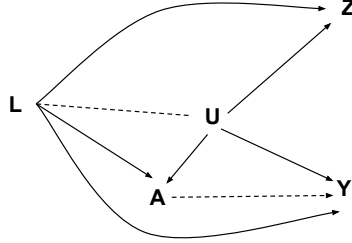


Figure 2: Causal DAG of an ideal NCE Z for bias detection/correction of the effect of treatment A on outcome Y , with observed confounders L and unmeasured confounders U . Ideal NCEs have the same incoming arrows as A , and are U -comparable to A [8].

where $Y^{a,z}$ is the potential outcome variable under primary exposure a and secondary exposure z . Informally, an NCO is a variable that is not causally affected by the treatment, and an NCE is a variable that does not causally affect the outcome. Assumptions 3 and 4 imply that conditional on (U, L) , there is no unmeasured confounding for the null causal effect of A on N , or the null causal effect of Z on Y . The ideal NCO N is U -comparable to Y , and the ideal NCE Z is U -comparable to A [16]:

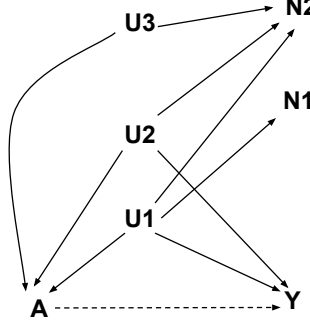


Figure 3: Causal DAG of two non-ideal NCOs $N1, N2$ for bias detection/correction of the effect of treatment A on outcome Y , with unmeasured confounders $U1, U2, U3$. Observed confounders L that cause $A, Y, N1, N2$ and may relate to $U1, U2, U3$ are not showed.

Assumption 5 (*U-comparability*) $N \not\perp U|L$ and $Z \not\perp U|L, A$

Assumption 5 means that the common causes of A and an ideal NCO N are *identical* to those of A and Y [8]. Similarly, it means that the common causes of an ideal NCE Z and Y are identical to those of A and Y [8]. Figure 1 and Figure 2 shows the causal DAGs of ideal an NCO and ideal NCE. Based on Assumptions 3, 4, and 5, empirical associations between the treatment and the NCO (or the NCE and the outcome), after controlling for all other observed variables L , are evidence of unmeasured confounding. For a comprehensive review, see Shi et al. [16]

In addition to bias detection, recent negative control techniques focus on partial or full bias correction [16]. Each technique has a unique set of assumptions and limitations. The following techniques partially correct for bias if their assumptions hold. Flanders et al. (2017) proposed adjusting for a negative control exposure to reduce bias in the estimate of the ATE using regression analyses. It works with time-series data but assumes linear relationships between the outcome and unmeasured confounders [4]. Schuemie et al. proposed empirical calibration of p-value [13] and empirical calibration of confidence intervals [14] using a set of negative controls. These calibration techniques aim to robustly control the Type I error rate and reduce the number of falsely-rejected null hypothesis of no causal effect in the presence of unmeasured confounding. They assume that the negative controls (and positive for confidence interval calibration) are to some level exchangeable with the outcome, and the bias is approximately Gaussian. The calibrated p-value and confidence interval help control the Type I error rate, but at the cost of a potentially vast increase in the Type II error rate [5].

The following techniques can fully correct for bias if their assumptions hold.

Tchetgen Tchetgen (2014) proposed a control outcome calibration technique to remove the systematic bias in the estimate of the ATT using an NCO. It requires rank preservation (i.e. a constant additive individual treatment effect) [18]. Sofer et al. (2016) proposed a generalization of difference-in-differences as a negative control outcome approach to detect and correct for bias in the ATT estimate. It requires monotonic associations between unmeasured confounders and the control outcomes as well as the outcome of interest [17]. Recent techniques using a pair of NCO and NCE, called double negative controls, provide nonparametric identification of the ATE [10, 15]. For inference on the true treatment effect, they require that both categorical negative controls have as many levels as the categorical unmeasured confounders and that at least one of the working models for the NCO or the outcome is correct.

1.3 Our contribution

Most negative control techniques rely on only one NCO and/or one NCE [18, 17, 4, 10, 15], while it can be advantageous to consider multiple negative controls. Most techniques for bias correction also require that the negative control is U-comparable to the treatment and outcome of interest (i.e. Assumption 5) [18, 4, 10, 15]. However, in practice, such negative controls are likely rare to find [8]. If A and N are confounded by some, but not all, variables in set U , they might appear unassociated even though unmeasured confounding exists for A and Y (see $N1$ in Figure 3) [8]. If A and N have an unmeasured common cause not in set U , an empirical association between A and N may not prove unmeasured confounding exists for A and Y (see $N2$ in Figure 3) [8]. When there is uncertainty about the confounding structure, a set of negative controls can be chosen to detect bias across multiple domains of confounding.

Many epidemiological studies have successfully used multiple negative controls for sensitivity analysis. For example, Jackson et al. (2006) used several NCOs to detect unmeasured confounding bias in unexpectedly large estimated benefits of influenza vaccine on seniors. The NCOs were risk of influenza hospitalization before and after flu seasons, and hospitalization of outcomes not linked to influenza [7]. Dormuth et al (2009) used a set of negative controls related to health behavior to detect bias due to the healthy user effect among new users of statins [2]. In a similar example, Lousdal et al. (2020) used multiple negative controls in an observational study of mammography screening participants and non-participants to caution against unmeasured confounding. They used proxies for better health, including death from various causes other than breast cancer, as NCOs and proxies for healthier behaviors, including dental care participation, as an NCE [9]. However, the problem of how to correct bias and draw conclusions about the validity of causal estimates from studies with multiple negative controls remains open.

There are several challenges to be considered. Firstly, existing bias correction techniques are likely to provide different corrected causal estimates for different negative controls. It is then unclear how to aggregate or interpret these corrected estimates, especially if they have opposing signs. Secondly, the estimated

bias from each negative control will likely have different sizes and precisions, for example, due to differing sample sizes or incidence rates, or because the unmeasured confounding is stronger for one negative control than for another. The estimates may even have opposing signs. For example, if gender were an unmeasured confounder, then it may be promotive of prostate cancer but protective against breast cancer. The precision, direction of each negative control estimate, and variability among estimates should be taken into consideration when drawing conclusions. Though empirical calibration of p-value and confidence interval partially address these two challenges, these methods have low power to detect non-null causal effects when some bias estimates have opposing signs or are null [5]. Thirdly, it is possible that unmeasured confounding induces bias too small to affect the conclusion of the study. In that case, the bias can be safely ignored [6]. Therefore, it is important to quantify the expected magnitude of the *net* bias from all unmeasured confounders [6, 5]. Though empirical calibration of p-value and confidence interval model the distribution of the bias, their calibrated test statistics blend unmeasured confounding bias and sampling variance into a single metric [5]. It would be advantageous to estimate the magnitude of the bias separately from random error and to provide uncertainty quantification for this estimated magnitude.

Based on these considerations, we propose a Bayesian hierarchical model to estimate the magnitude of the net unmeasured confounding bias, separately from its existence, its direction, and the random error. We view this partial bias correction approach as having the following advantages. Firstly, it infers a posterior distribution of the magnitude of the bias from multiple negative control estimates, providing uncertainty quantification for the estimated magnitude. Secondly, because the negative control estimates are residual biases after all other necessary adjustments, e.g. measured confounders, weighting, matching, censoring, etc., have been made, our proposed model can work with any complex study designs and any causal estimand that satisfies its modeling assumptions. Thirdly, the proposed model can work with non-ideal NCOs, such as those in Figure 3, and have assumptions less restrictive than the U-comparability, rank-preservation, and monotonicity assumptions of existing techniques. Though it shares many assumptions with empirical p-value calibration [13], its ability to decompose the bias into its existence, direction, and magnitude makes it more suitable to model bias estimates from non-ideal negative controls. Finally, the posterior of the magnitude of the bias enables the construction of a principled decision rule for whether to proceed with the study, in situations where prior information about the effect size may be available. Our proposed decision rule can guide researchers in refining their study design to a stage where the remaining bias may be considered negligible. We will focus solely on negative control outcomes below, though we believe extensions to negative control exposures will be similar.

2 Methods

2.1 Setup

We consider the application of NCOs to assess systemic bias due to unmeasured confounding in an observational study. In this section, we refer to bias with respect to a given study design as a non-zero log risk ratio or risk difference for an outcome with a null effect, even after controlling for all observed confounding factors. A study design for our purposes also includes the proposed model of the association between the treatment and outcome. Analyses of multiple NCOs are performed and result in a set of bias estimates and their corresponding standard errors.

2.2 Assumptions

Our method relies on the following assumptions:

Assumption 6 *Consistency [11], positivity [11], no interference [11], and conditional exchangeability (Assumption 2) given all common causes (L, U) of A and Y .*

Assumption 3. *(Negative control outcomes) $N^a = N$ and $N \perp A|U, L$ for all a .*

Assumption 7 *The observed biases, for the NCOs and the outcome of interest, in terms of log risk ratio or risk difference, are approximately Gaussian, centered at the true biases, with standard deviation equaling the estimated standard error.*

Assumption 8 *The true biases, for the NCOs and the outcome of interest, are drawn from a common distribution whose parameters can be estimated.*

In this chapter, we model this common distribution with a mixture of distributions to decompose the existence, direction, and magnitude of the biases. We recommend picking a heterogeneous set of NCOs for this assumption to hold. Below, we will define a Bayesian hierarchical model for the magnitude of the confounding bias.

2.3 Model details

The following model is for a binary treatment, binary outcome, and I binary NCOs. Let \hat{S}_i be the bias estimates in terms of log risk ratio or risk difference for NCO $i = 1, \dots, I$ under a binary treatment A , and $\hat{\Sigma}$ be the bootstrap-estimated covariance of the random error in the bias estimates. We assume that the bias estimates are normally distributed around the true biases, and the estimated covariance adequately captures variation in the estimates due to random error:

$$(\hat{S}_1, \dots, \hat{S}_I)^T \sim N((S_1, \dots, S_I)^T, \hat{\Sigma}) \quad (2)$$

We also assume that the true biases are i.i.d. samples from a common bias distribution parametrized by a mixture of Gaussians and a point mass P_0 at zero:

$$S_i \sim \delta_i P_0 + (1 - \delta_i)(2M_i - 1)N(\mu, \sigma^2) \quad (3)$$

Parameter δ_i is an indicator for whether the i th bias belongs to the point mass or the Gaussian components. It models the existence of the i th bias, since it is possible for the overall bias for NCO i to be null, for example when there are multiple sources of opposing bias. Parameter M_i is an indicator for whether the bias belongs to the Gaussian component with a positive or a negative expectation. It models the direction of the bias. The motivation for this bimodal form is the following. Suppose that we have a single unmeasured confounding variable such as gender, then this variable may be promotive of prostate cancer but protective of breast cancer. The two Gaussian modes share the same variance σ^2 and an absolute value of the expectation μ . These parameters characterize the magnitude of the bias. We place hierarchical priors on μ and σ^2 :

$$\mu \sim |N(\nu, \tau^2)| \quad (4)$$

$$\sigma^2 \sim |Cauchy(s_0)| \quad (5)$$

The reason for forcing μ to be nonnegative is to prevent identifiability issues due to the form of (2). Because the number of available NCOs is typically small, we can use a weakly informative half-Cauchy prior for σ^2 as suggested by [?], and recommend that scale s_0 be chosen based on the data application to construct an informative prior when possible. We also place priors on the bias existence and bias direction parameters:

$$\delta_i \sim \text{Bern}(\phi_\delta) \quad (6)$$

$$M_i \sim \text{Bern}(\phi_M) \quad (7)$$

$$\phi_M \sim \text{Beta}(u_0, v_0) \quad (8)$$

$$\phi_\delta \sim \text{Beta}(r_0, s_0) \quad (9)$$

When the treatment groups are balanced and there is no confounding bias, ϕ_δ will ideally be 1, and δ_i will be 1 for all i . Parameter ϕ_δ gives the posterior predictive probability that the bias of a new outcome is null. Inference on $\{M_i\}_i^I, \phi_M$ is not of interest for our model. They exist to prevent biases of opposing signs from pulling the magnitude of the bias toward zero or inflating the magnitude's variance. When $u_0 = v_0 = 1$ or $r_0 = s_0 = 1$, these priors simplify to a uniform distribution that can be used as an uninformative prior. For an uninformative prior, we can set $\nu = 0$ and $\tau^2 = 1^2$, though we recommend setting informative priors when possible.

We can extend this model to treatment with multiple treatment groups. A comparison of two treatment groups will be referred to as a treatment setting.

Additional priors on ν and τ^2 can also be specified to allow some sharing of information across the settings.

2.4 Expected bias magnitude conditional on outcomes prevalence

We observe that, on the risk difference scale, the estimated bias and its standard error tend to be greater when the NCO’s prevalence—the average probability of observing the outcome in the population—is closer to 0.5, and smaller when it’s closer to 0 or 1. On the log risk ratio scale, we also observe correlations between an NCO’s prevalence and its bias’ magnitude and variance. Figure 4 shows log risk ratios and risk differences of simulated NCOs with identical associations to U but with different prevalence rates.

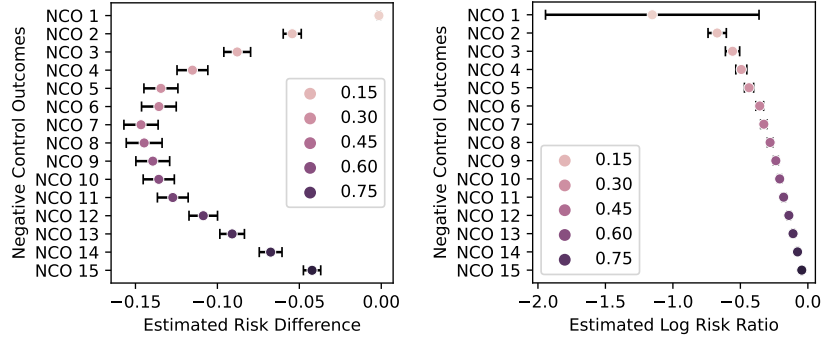


Figure 4: Log risk ratios and risk difference of simulated NCOs with identical associations to unmeasured confounder U but with different outcome prevalence rates.

As a result, we want to model μ and σ^2 as smooth functions of the outcome prevalence x_i . Therefore, we replace equation (3.3):

$$S_i \sim \delta_i P_0 + (1 - \delta_i)(2M_i - 1)N(\mu(x_i), \sigma^2(x_i)) \quad (10)$$

One option is to fit a linear regression of polynomials of the prevalence. Here, we present another option to learn the functions by placing Gaussian process priors on the function $\mu(\cdot)$ and $\sigma^2(\cdot)$, replacing equations (3.4) and (3.5) with:

$$\mu \sim |GP(f_1, c_\mu)| \quad (11)$$

$$\log(\sigma^2) \sim GP(f_2, c_{\sigma^2}) \quad (12)$$

where f is a mean function and c_μ, c_{σ^2} is a covariance kernel. We choose to use the Gaussian kernel to model a smooth function μ and σ^2 :

$$c_\mu(x_l, x_h) = d_1 e^{-\frac{(x_l - x_h)^2}{2\kappa_1^2}} \quad (13)$$

$$c_{\sigma^2}(x_l, x_h) = d_2 e^{-\frac{(x_l - x_h)^2}{2\kappa_2^2}} \quad (14)$$

where d_1, d_2 are amplitude parameters and κ_1, κ_2 are length scales, which determine the wiggleness of the functions. We place informative priors on the length scales so that they are between 0 and 1 with 99% probability, because input x (outcome prevalence) ranges from 0 to 1. We recommend setting informative priors for $d_1 \sim |N(0, s_1)|$, $d_2 \sim |Normal(0, s_2)|$ by choosing s_1, s_2 based on the data. We set the mean function $f_1 = f_2 = 0$.

To predict the magnitude of the bias of a new outcome with proportion x^* , we use the property that $\mu(x_1), \dots, \mu(x_I), \mu(x^*)$ are jointly normal:

$$\begin{bmatrix} \mu \\ \mu(x^*) \end{bmatrix} \sim |N(\mathbf{0}, \begin{bmatrix} K & K^* \\ K^{*T} & c(x^*, x^*) \end{bmatrix})| \quad (15)$$

where K^* is the covariance between observed μ and new $\mu(x^*)$. Using the conditional normal formula, we can get a posterior predictive for $\mu(x^*)$:

$$p(\mu(x^*)|\mu) \sim |N(A, V)| \quad (16)$$

$$A = K^{*T} K^{-1} \mu \quad (17)$$

$$V = c(x^*, x^*) - K^{*T} K^{-1} K^* \quad (18)$$

2.5 Posterior computation

We use PyMC for inference using Hamiltonian Monte Carlo.

2.6 Decision rule

We propose the construction of a decision rule based on a bias magnitude threshold parameter and a probability bound parameter. The probability bound parameter specifies how certain we want to be that the bias in the treatment-outcome association is less than the bias magnitude threshold parameter. For example, we might create an a priori decision rule stating that if the magnitude of the bias in terms of risk difference is less than 0.1 with at least 95% probability, we will proceed with the study. In situations where they are available, the bias magnitude threshold parameter can be selected based on previous results. For example, if a previous study has shown an effect size difference of 0.3 between two populations that are comparable to those in the new study of interest, then a bias of 0.1 might be considered acceptable. If the estimated probability is less than the pre-specified probability bound parameter, the study may be abandoned since the results would potentially be invalid.

Procedurally, we specify a priori a decision rule that if the magnitude of the bias is less than θ with at least $\rho\%$ probability, we will proceed with the study. Here, θ is the bias magnitude threshold parameter, and ρ is the probability bound parameter, chosen based on domain expert knowledge. Using the posterior predictive distribution of the magnitude of the bias for a new outcome $p(S_{I+1}|\hat{S}_1, \dots, \hat{S}_I)$, we want to calculate $\pi = Pr(S_{I+1} < \theta|\hat{S}_1, \dots, \hat{S}_I)$ —the posterior probability that the magnitude is less than θ . If $\pi \geq \rho$, we may proceed with the study of interest.

We can approximate π using Monte Carlo samples. Fixing $M_{I+1} = 1$ or predicting M_{I+1} using a signed causal diagram as discussed in [6] (since we are mostly interested in the magnitude), for each $t \in 1, \dots, T$:

1. sample $\delta_{I+1}^{(t)} \sim \text{Bern}(\phi_\delta^{(t)})$
2. sample $S_{I+1}^{(t)} \sim \delta_{I+1}^{(t)} P_0 + (1 - \delta_{I+1}^{(t)})(2M_{I+1} - 1)N(\mu^{(t)}, \sigma^{2(t)})$
3. compute $\pi^{(t)} = \mathbb{1}(S_{I+1}^{(t)} < \theta)$

where T is the number of MCMC iterations; $\phi_\delta^{(t)}, \mu^{(t)}, \sigma^{2(t)}$ are posterior samples from the t th MCMC iteration. Finally, compute $\pi \approx \hat{\pi} = \frac{\sum_t \pi^{(t)}}{T}$.

todo: Add interpretation for GP: mean of new outcome more similar to the mean of observed outcomes with similar proportions.

3 Simulations

3.1 One confounding source, similar outcome prevalence rates

Data for 10000 subjects were generated according to the following scheme:

$$\begin{aligned}
U &\sim N(0, 1) \\
\text{logit}[P(N_i = 1|U)] &= 1 + \alpha_i U \\
\text{logit}[P(A = 1|U)] &= 0.3U \\
\text{logit}[P(Y = 1|A, U)] &= 1 + \alpha^{(Y)}U + \beta A \\
\alpha_i &\sim 0.3P_0 + 0.35\text{Unif}(0.5, 1) + 0.35\text{Unif}(-1, -0.5) \quad \text{for } i = 1, \dots, I \\
\alpha^{(Y)} &\sim \text{Unif}(1, 0.5)
\end{aligned}$$

where $I = 30$ is the number of NCOs used to estimate the posterior of the magnitude of bias. This data-generating scheme satisfies all assumptions required for our method. Here, some NCOs have the same confounding structure as Y , but some do not. Since there are no observed confounders, the bias in terms of the log risk ratio was estimated with $\hat{S}_i = \log(\frac{N_i^{(1)}}{N_i^{(0)}})$, where

$\overline{N_i^{(a)}} = \frac{\sum_n^{5000} N_{i,n} \mathbb{1}(A_n=a)}{\sum_n^{5000} \mathbb{1}(A_n=a)}$. We set $\beta = 0$. We use the mean model described in subsection 3.2.3 to infer a posterior predictive distribution for the bias magnitude.

todo: Show results

3.2 Multiple confounding sources, differing outcome prevalence rates

Now we consider a simulation setting with two sources of confounding and a set of NCOs that include 1) NCOs that have the same confounding structure as A and Y ; 2) NCOs that are caused by some, but not all, common causes of A and Y ; and 3) NCOs that are caused by a variable not a common cause of A and Y . Data for 10000 subjects were generated according to the following scheme:

$$\begin{aligned} U_1 &\sim N(0, 1), \quad U_2 \sim N(0, 1) \\ \text{logit}[P(A = 1|U_1, U_2)] &= 0.3U_1 + 0.4U_2 \\ \text{logit}[P(N_i = 1|U_1, U_2)] &= \nu_i + \alpha_{1,i}U_1 + \alpha_{2,i}U_2 \\ \text{logit}[P(Y_l = 1|A, U_1, U_2)] &= \nu^{(Y)} + \alpha_{1,l}^{(Y)}U_1 + \alpha_{2,l}^{(Y)}U_2 + \beta A \\ \alpha_{j,i} &\sim 0.3P_0 + 0.49\text{Unif}(0.5, 1) + 0.21\text{Unif}(-1, -0.5) \\ \alpha_{j,l}^{(Y)} &\sim 0.3P_0 + 0.49\text{Unif}(0.5, 1) + 0.21\text{Unif}(-1, -0.5) \end{aligned}$$

for $i = 1, \dots, I$; $l = 1, \dots, 1000$; $j = 1, 2$, and NCO estimates in terms of log risk ratios are calculated as described in the last simulation. We set $\nu_i = \text{logit}(0.02 + (i-1)\frac{0.9-0.02}{I})$ so that the outcome proportions of NCOs vary from 2% to 90%. This will demonstrate the utility of having a flexible mean and variance function that accounts for outcome prevalence. We evaluate the ability of the mean model in subsection 3.2.3 and the Gaussian process model in subsection 3.2.4 to construct good posterior predictive distribution for bias in terms of log risk ratio conditional on the prevalence of unseen outcomes. We use our proposed decision rule to evaluate these simulations. A priori, we say that the bias is negligible if the bias in terms of log risk ratio is less than a bias threshold $\theta = 0.095$ with 95% probability. In each of 100 simulation replicates, we calculate π , the probability that true bias is less than θ , using 1000 pairs of $A - Y$. We also estimate $\hat{\pi}_{mean}$ and $\hat{\pi}_{gp}$ from the posterior predictive of the models as outlined in subsection 3.2.6. We repeat everything for different $I = 5, 15, 30, 50$ and different outcome prevalence rates $\nu^{(Y)} = \text{logit}(0.5), \text{logit}(0.05), \text{logit}(0.85)$.

todo: Show results

4 Analysis of real data

4.1 Data description

todo: Write

4.2 Results

5 Discussion

References

- [1] M Alan Brookhart, Til Stürmer, Robert J Glynn, Jeremy Rassen, and Sebastian Schneeweiss. Confounding control in healthcare database research: challenges and potential approaches. *Medical care*, 48(6 0):S114, 2010.
- [2] Colin R Dormuth, Amanda R Patrick, William H Shrank, James M Wright, Robert J Glynn, Jenny Sutherland, and M Alan Brookhart. Statin adherence and risk of accidents: a cautionary tale. *Circulation*, 119(15):2051–2057, 2009.
- [3] Zoe Fewell, George Davey Smith, and Jonathan AC Sterne. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *American journal of epidemiology*, 166(6):646–655, 2007.
- [4] W Dana Flanders, Matthew J Strickland, and Mitchel Klein. A new method for partial correction of residual confounding in time-series and other observational studies. *American journal of epidemiology*, 185(10):941–949, 2017.
- [5] Susan Gruber and Eric Tchetgen Tchetgen. Limitations of empirical calibration of p-values using observational data. *Statistics in medicine*, 35(22):3869–3882, 2016.
- [6] Miguel A Hernán and James M Robins. Causal inference, 2010.
- [7] Lisa A Jackson, Michael L Jackson, Jennifer C Nelson, Kathleen M Neuzil, and Noel S Weiss. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *International journal of epidemiology*, 35(2):337–344, 2006.
- [8] Marc Lipsitch, Eric Tchetgen Tchetgen, and Ted Cohen. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.)*, 21(3):383, 2010.
- [9] Mette Lise Lousdal, Timothy L Lash, W Dana Flanders, M Alan Brookhart, Ivar Sønbo Kristiansen, Mette Kalager, and Henrik Støvring. Negative controls to detect uncontrolled confounding in observational studies of mammographic screening comparing participants and non-participants. *International Journal of Epidemiology*, 49(3):1032–1042, 2020.
- [10] Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.

- [11] Ashley I Naimi and Brian W Whitcomb. Defining and identifying average treatment effects. *American Journal of Epidemiology*, 2023.
- [12] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.
- [13] Martijn J Schuemie, George Hripcsak, Patrick B Ryan, David Madigan, and Marc A Suchard. Robust empirical calibration of p-values using observational data. *Statistics in medicine*, 35(22):3883, 2016.
- [14] Martijn J Schuemie, George Hripcsak, Patrick B Ryan, David Madigan, and Marc A Suchard. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proceedings of the National Academy of Sciences*, 115(11):2571–2577, 2018.
- [15] Xu Shi, Wang Miao, Jennifer C Nelson, and Eric J Tchetgen Tchetgen. Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(2):521–540, 2020.
- [16] Xu Shi, Wang Miao, and Eric Tchetgen Tchetgen. A selective review of negative control methods in epidemiology. *Current epidemiology reports*, 7:190–202, 2020.
- [17] Tamar Sofer, David B Richardson, Elena Colicino, Joel Schwartz, and Eric J Tchetgen Tchetgen. On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 31(3):348, 2016.
- [18] Eric Tchetgen Tchetgen. The control outcome calibration approach for causal inference with unobserved confounding. *American journal of epidemiology*, 179(5):633–640, 2014.