Causal progress with imperfect placebo treatments and outcomes [DRAFT]

Adam Rohde*, Chad Hazlett†

January 2023

Abstract

In the quest to make defensible causal claims from observational data, it is sometimes possible to leverage information from "placebo treatments" and "placebo outcomes" (aka "negative outcome controls") unaffected by the treatment. Traditional approaches employing such information focus largely on point identification and assume (i) "perfect placebos" (placebo treatments have precisely zero effect on the outcome; the real treatment has precisely zero effect on a placebo outcome); and (ii) that the placebo treatment/outcome suffers the same amount of confounding as does the real treatment- outcome relationship on some scale ("equi-confounding"). We take a different approach, showing how the analysis of "omitted variable bias" in regression provides a flexible and powerful way to leverage information from placebo treatments and outcomes while relaxing these assumptions. This provides an easy-to-use sensitivity analysis or partial identification framework for demonstrating the implications of any claim a user chooses to support regarding the relative strengths of confounding suffered by a placebo treatment/outcome compared to the true treatment-outcome relationship. We further permit "imperfect placebos", i.e. placebo treatments with up to some postulated non-zero effect on the outcome, or placebo outcomes experiencing effects of treatment up to some postulated strength. These tools can be easily applied with any choice of placebo treatment or outcome, observed in the same dataset or in a separate one as the main treatment-outcome relationship. Conventional difference-in-difference approaches, in both the repeated cross-section and panel settings, are a special case of this setting in which a strict equiconfounding assumption (regarding the pre-treatment outcome's relationship to the treatment group indicator) is claimed. We demonstrate several relaxations of this that are natural under our framework.

1 Introduction

Unobserved confounding plagues many attempts to draw causal inferences from observational data. Moreover, chance imbalances between the treatment and control groups in randomized experiments, sample selection, measurement error can further threaten the validity of causal inferences for both observational and experimental studies. While standard practice is to check balance on observed covariates and to statistically adjust for these variables in estimating causal effects, these attempts are often unconvincing in their ability to overcome unobserved confounding and other unmeasured threats to validity.

A still underappreciated, though increasingly discussed, practice is to leverage information from "placebo outcomes" or "placebo treatments" (also known as "negative outcome controls" and "negative exposure controls," respectively) to aid in the quest to make defensible causal claims, despite the presence of unobserved confounding (or other issues). Existing approaches using such information typically focus on simply detecting bias or on point identification (i.e., bias correction), assuming (i) that these are "perfect placebos" (treatment has precisely zero effect on a placebo outcome, or placebo treatments have precisely zero effects on the outcome of interest) and (ii) that the placebo outcome/treatment suffers the same amount of confounding as does the real treatment- outcome relationship on some scale ("equi-confounding"). Simple examples of *perfect* placebo treatments and outcomes are shown in Figure 1(a,b). In this figure N is a placebo outcome and P is a placebo treatment; however, Figure 1(b) is essentially a mirror image of Figure 1(a) with N relabelled as P. What we consider a placebo treatment versus what what we consider a placebo outcome will often have this "Necker cube" quality. A Necker cube is a simple optical illusion in which the same two-dimensional drawing of a cube can be perceived with the cube oriented in two distinct ways. See Necker (1832). Depending on the substantive context, the same causal graph might arise in two settings, one in which the placebo is viewed as a placebo treatment and one in which the placebo is viewed as a placebo outcome. We explore this more in Section 4 and discuss how the perspective taken by researchers should influence the methods they use. Figure 1(c,d) contains simple examples of *imperfect* placebos. These "imperfections" arise from adding the edge $D \to N$ to Figure 1(a) to get Figure 1(c) and from adding the edge $P \to Y$ to Figure 1(b) to get Figure 1(d).

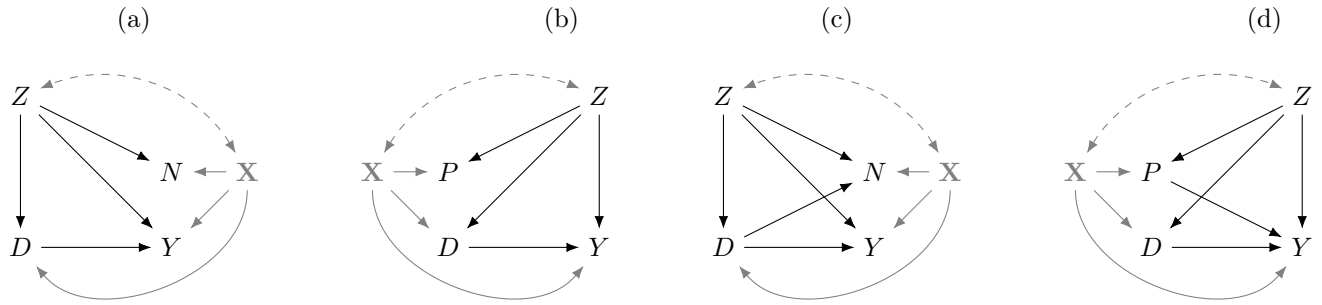
For a recent example of a placebo approach, see Ye et al. (2022). Lipsitch et al. (2010) provide an early discussion of the potential of using placebo treatments and outcomes. Arnold et al. (2016); Arnold and Ercumen (2016) discuss the potential to

*Department of Statistics, UCLA. adamrohde@ucla.edu

†Associate Professor, Departments of Statistics & Political Science, UCLA. chazlett@ucla.edu

use placebos to address sample selection and measurement bias as well as bias in randomized experiments. Shi et al. (2020) review many established procedures and recent developments in this growing literature. An important recent area of research has been dubbed "proximal" causal inference. Tchetgen et al. (2020) provides an introduction. This line of work leverages "double-negative control designs" (Shi et al., 2020) in which both a placebo outcome and a placebo treatment are present. Under certain assumptions and conditions, papers like Miao et al. (2018, 2020); Mastouri et al. (2021) show how causal effects are non-parametrically identifiable using both of these placebos. While this work is very promising, it requires two placebos, certain null effect assumptions (i.e., "perfect" placebos), and regularity conditions, potentially creating a hurdle to their easy application. In Section 2.3, we show a similar point identification result for regression when both a "perfect" placebo treatment and a "perfect" placebo outcome are available. We also relax the "double placebo" setting to allow for imperfect placebos.

Figure 1: Simple DAGs with a placebo outcome, N, and with a placebo treatment, P. D is treatment, Y is outcome, Z contains unobserved confounders, and $\mathbf{X}$ contains observed covariates.



Additionally, Sofer et al. (2016) point out that placebo outcome approaches can be seen as generalizations of the familiar difference in differences (DID) framework. DID uses a pre-treatment verison of the outcome as a placebo outcome and relies heavily on a "parallel trends" assumption that the baseline trend in the outcome is the same over time for the group that gets treated and the group that does not. (Angrist and Pischke, 2008) Parallel trends is, in fact, an assumption of no unobserved time-varying confounding; meaning that DID approaches have not circumvented the problem of unobserved confounding. Sofer et al. (2016) also show how the parallel trends assumption is equivalent to an equi-confounding assumption. Recently, there have been several attempts to relax the parallel trends assumption within the context of DID or to improve on the current practice of testing for parallel trends in the pre-treatment period. For example, see Bilinski and Hatfield (2018); Keele et al. (2019); Ryan et al. (2019); Gibson and Zimmerman (2021); Rambachan and Roth (2021). While these are all useful contributions, they may not be as easy to use as applied researchers might hope and make assumptions that researchers might be interested in relaxing. Gibson and Zimmerman (2021) points out that violations of parallel trends can be considered using an application of the omitted variable bias framework from Cinelli and Hazlett (2020) in a first difference regression. However, this approach is limited to the DID setting in which the pre-treatment outcome is directly comparable to the post-treatment outcome. This approach (and others like Rambachan and Roth (2021)) also do not allow for a relaxation of the assumption that the treatment has no direct causal relationship with the pre-treatment outcome. Additionally, this approach requires that researchers observe the pre-treatment outcome and the post-treatment outcome for the same set of units; that is, it requires panel data. Finally, the DID approach in general requires data on a pre-treatment version of the outcome, which may not always be available, making the placebo paradigm an attractive alternative.

Our contributions to this literature are simple. We show how the analysis of "omitted variable bias" in regression provides a simple yet flexible and powerful way to leverage information from placebo treatments and outcomes while relaxing assumptions of equi-confounding and perfect placebos. Specifically, we provide a partial identification framework in which users can see the implications of any claim they choose to make about the relative strengths of confounding suffered by a placebo treatment/outcome compared to the true treatment-outcome relationship, while allowing for "imperfect placebos" (e.g. placebo treatments with some non-zero effect on the outcome, placebo outcomes somewhat affected by the treatment). These tools can be used with any choice of placebo treatment or outcome, observed in the same dataset or in a separate one as the main treatment-outcome relationship. In fact, our results are simple consequences of ordinary least squares and hold for any choice of placebo. Though researchers will want to chose placebos for which the relative level of confounding is easy to consider. We also discuss how to use our framework under various assumptions about the casual graph and other practical consideration. Finally, we frame the conventional difference-in-difference approach as a special case of our framework, with a strict equi-confounding (which is equivalent to parallel trends) assumption regarding the pre-treatment outcome's relationship to the treatment group indicator, to which our approach allows several useful relaxations.

2 Leveraging placebos

Our proposed framework for using information from placebo treatments and outcomes relies on repeated application of the omitted variable bias paradigm for regression analysis. However, we also provide relaxations to this parametric approach in Appendix D, in which we develop similar methods adapted to partially linear and non-parametric estimation of treatment effects. In either case, we show how the omitted variable bias paradigm can be augmented to incorporate information from placebo treatments and outcomes to make causal progress with unequal levels of confounding for a placebo treatment/outcome compared to the true treatment-outcome relationship and with imperfect placebos. We emphasize this approach as a partial identification strategy, but it could also be seen as a sensitivity analysis for more traditional placebo approaches. We start by discussing placebo outcomes, then discuss placebo treatments, and finally discuss "double placebos."[1] In Appendix A, we show how the approaches for placebo outcomes, placebo treatments, and double placebos presented in this section can be unified under in a single framework. In section 3, we discuss how these approaches apply to the setting with multiple omitted variables. In Section 4, we discuss how these approaches can be used when considering different causal graphs.

2.1 Placebo outcomes

We start our discussion of placebo outcomes by leveraging the omitted variable bias framework. (Angrist and Pischke, 2008; Hansen, 2022) Suppose that we want to estimate the "long" regression of Y on D, X, and Z in Equation 1. In particular, we are interested in the coefficient $\beta_{Y\sim D|Z,X}$ as a measure of the causal effect of D on Y or as an approximation to the causal effect of D on Y.

$$Y = \beta_{Y\sim D|Z,X} D + \mathbf{X}\beta_{Y\sim X|D,Z} + \beta_{Y\sim Z|D,X} Z + \epsilon_l \tag{1}$$

In practice, we typically do not observe all the covariates that would allow us to estimate or approximate the causal effect of D on Y by estimating this long regression. Therefore, let us assume that Z is unobserved. We will refer to this as an unobserved confounder. Since Z is an unobserved variable and we must estimate the "short" regression of Y on D and X in Equation 2, rather than the desired long regression. Z is a variable that is "omitted" from the regression, though we wish we could have included it.

$$Y = \beta_{Y\sim D|X} D + \mathbf{X}\beta_{Y\sim X|D} + \epsilon_s \tag{2}$$

Traditional omitted variable bias analysis tells us that $\beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X}$ can be characterised as the product of two regression coefficients, as in Equation 3. $\beta_{Y\sim Z|D,X}$ captures the relationship between Y and Z after linearly partialling out D and X. $\beta_{Z\sim D|X}$ captures the relationship between D and Z after linearly partialling out X.

$$\text{bias}_{(\text{YD.X})} \triangleq \beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X} = \beta_{Y\sim Z|D,X}\beta_{Z\sim D|X} \tag{3}$$

These components alone are enough to build a powerful sensitivity analysis or partial identification framework, as discussed in Cinelli and Hazlett (2020). But can more be done when there also exists a placebo outcome, N? If we think that the relationship between D and N suffers from a similar level of unobserved confounding as does the relationship between D and Y, then how might we leverage N to understand bias_Y and, hence, $\beta_{Y\sim D|Z,X}$? Consider a long and short regression with N as the outcome and with D as the treatment. The long regression of N on D, X, and Z can be found in Equation 4. The short regression of N on just D and X can be found in Equation 5. As with Y, we are only able to consider estimating the short regression for N, as Z is unobserved. Here too $\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}$ can be characterised as in Equation 6.

$$N = \beta_{N\sim D|Z,X} D + \mathbf{X}\beta_{N\sim X|D,Z} + \beta_{N\sim Z|D,X} Z + \xi_l \tag{4}$$

$$N = \beta_{N\sim D|X} D + \mathbf{X}\beta_{N\sim X|D} + \xi_s \tag{5}$$

$$\text{bias}_{(\text{ND.X})} \triangleq \beta_{N\sim D|X} - \beta_{N\sim D|Z,X} = \beta_{N\sim Z|D,X}\beta_{Z\sim D|X} \tag{6}$$

From Equations 3 and 6, it is easy to see how to leverage information about the placebo outcome, N. First rearrange the expression for bias_N to solve for $\beta_{Z\sim D|X}$. Then plug this expression into the expression for $\text{bias}_{(\text{YD.X})}$. This can be thought of as using the placebo outcome to re-express the $Z \to D$ "arm" of the confounding path $D \leftarrow Z \to Y$ in Figure 1(a).

[1] As we mentioned, many placebos could be phrased as either a placebo outcome or a placebo treatment depending on the context. One framing may be more natural, depending on the placebo. For example, a pre-treatment version of the actual outcome will typically be best considered as a placebo outcome, since it can easily be compared to the actual outcome.

$$\beta_{Z\sim D|X} = \frac{\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}}{\beta_{N\sim Z|D,X}} \tag{7}$$

$$\implies \text{bias}_{(\text{YD.X})} = \beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X} = \beta_{Y\sim Z|D,X}\left[\frac{\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}}{\beta_{N\sim Z|D,X}}\right] = m \times \left(\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}\right) \tag{8}$$

$$\iff \beta_{Y\sim D|Z,X} = \beta_{Y\sim D|X} - m_N \times \left(\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}\right) \tag{9}$$

Equation 8 is an expression for bias_Y in terms of $m \triangleq \frac{\beta_{Y\sim Z|D,X}}{\beta_{Y\sim Z|D,X}} = \frac{\beta_{Y\sim Z|D,X}\beta_{Z\sim D|X}}{\beta_{Y\sim Z|D,X}\beta_{Z\sim D|X}} = \frac{\text{bias}_{(\text{YD.X})}}{\text{bias}_{(\text{ND.X})}}$, $\beta_{N\sim D|Z,X}$, and $\beta_{N\sim D|X}$. We have a similar expression for $\beta_{Y\sim D|Z,X}$ in Equation 9, our target parameter, that also includes $\beta_{Y\sim D|X}$. We can estimate $\beta_{N\sim D|X}$ and $\beta_{Y\sim D|X}$ from the data. The remaining terms in these expressions can be treated as parameters that can be specified by investigators. Thus, we can use Equation 9 for partial identification of $\beta_{Y\sim D|Z,X}$ by establishing plausible ranges of values for m and $\beta_{N\sim D|Z,X}$. m captures the level of relative confounding or the relative relationship between the outcomes and the unobserved confounder. This can be thought of as comparing the strength of the $Z \to Y$ "arm" of the confounding path $D \leftarrow Z \to Y$ to the relative strength of the $Z \to N$ "arm" of the confounding path $D \leftarrow Z \to N$ in Figure 1(a). Or we can consider it as comparing the strength of the entire $D \leftarrow Z \to Y$ relationship to the entire $D \leftarrow Z \to N$ relationship. $\beta_{N\sim D|Z,X}$ captures any direct causal relationship between D and N, like the $D \to N$ path in Figure 1(c). Therefore, this $\beta_{N\sim D|Z,X}$ parameter allows for "imperfect" placebo outcomes that are directly affected by the treatment. Since both N and D are observed variables, their relationship conditional on Z, X should be able to be specified as a regression coefficient despite scale considerations. In many cases differing scales for N and Y might lead to difficulty in interpreting or reasoning about the values for m. Can we do better than this?

Re-expressing m Let us consider a reparameterization of m in terms of scale-independent partial-correlation parameters, rather than regression coefficients. This will allow us to circumvent the scale issues we just mentioned.

$$m = \frac{\beta_{Y\sim Z|D,X}}{\beta_{N\sim Z|D,X}} = \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}} \times \frac{\left[\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(Z^{\perp D,X})}\right]}{\left[\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Z^{\perp D,X})}\right]} = \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} = k \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \tag{10}$$

Equation 10 provides an expression for m in terms of $k \triangleq \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}$[2][3] and $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}$. $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}$ captures the scale issues and deals only with observables. The numerator and denominator are moments of the residuals from the two short regressions. k can be interpreted in two ways. First, $k = \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}$ is a ratio of partial correlations capturing how the Y, Z relationship compares to the N, Z relationship. Second, $k = \frac{R_{Y\sim Z|D,X} f_{D\sim Z|X}}{R_{N\sim Z|D,X} f_{D\sim Z|X}}$ is a ratio of the level of confounding of the Y, D relationship to the level of confounding of the N, D relationship, where the confounding is expressed in terms of partial correlations. Quantities like $R_{Y\sim Z|D,X} f_{D\sim Z|X}$ and $R_{N\sim Z|D,X} f_{D\sim Z|X}$ are called "bias factors" in Cinelli and Hazlett (2020). $f_{D\sim Z|X}$ is Cohen's f which equals $\frac{R_{D\sim Z|X}}{\sqrt{1-R^2_{D\sim Z|X}}}$.[4][5] Equations 8 and 9 can then be rewritten as Equation 11 and 12.

$$\text{bias}_{(\text{YD.X})} = k \times \left(\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}\right) \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \tag{11}$$

$$\beta_{Y\sim D|Z,X} = \beta_{Y\sim D|X} - k \times \left(\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}\right) \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \tag{12}$$

[2] If partial R^2 are more intuitive, users can substitute in $k = \text{sign}(k) \times \sqrt{k^2}$ and consider $\text{sign}(k) = \text{sign}\left(\frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}\right)$ and $k^2 = \frac{R^2_{Y\sim Z|D,X}}{R^2_{N\sim Z|D,X}}$.

[3] We need $R_{N\sim Z|D,X} \neq 0$ for k to be defined. This will be the case for any useful placebo outcome. Otherwise, the placebo outcome will not have a confounded relationship with the treatment and not have any useful information that we can leverage to understand the confounding of the relationship between the treatment and the actual outcome. Whether $R_{N\sim Z|D,X} \neq 0$ is only testable under the assumption that $\beta_{N\sim D|Z,X} = 0$.

[4] A reparameterization of omitted variable bias similar to that discussed in Cinelli and Hazlett (2020) gives the following bias expressions. Using these, we see that k can be expressed as a scaled version of the ratio of $\text{bias}_{(\text{YD.X})}$ over $\text{bias}_{(\text{ND.X})}$. This is a scale independent measure of the level of confounding of the Y, D relationship relative to the level of confounding of the N, D relationship. Alternatively put, it is the level of unobserved confounding in the Y, D relationship as a percentage of the level of unobserved confounding in the N, D relationship, after resclaing to account for possible differences in the scale of Y and N. If $\text{SD}(N^{\perp D,X}) = \text{SD}(Y^{\perp D,X})$, then k is simply the ratio of biases.

$\text{bias}_{(\text{YD.X})} = \beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X} = R_{Y\sim Z|D,X} \times f_{D\sim Z|X} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(D^{\perp X})}$

$\text{bias}_{(\text{ND.X})} = \beta_{N\sim D|X} - \beta_{N\sim D|Z,X} = R_{N\sim Z|D,X} \times f_{D\sim Z|X} \times \frac{\text{SD}(N^{\perp D,X})}{\text{SD}(D^{\perp X})}$

$\therefore k = \frac{R_{Y\sim Z|D,X} f_{D\sim Z|X}}{R_{N\sim Z|D,X} f_{D\sim Z|X}} = \frac{\text{bias}_{(\text{YD.X})}}{\text{bias}_{(\text{ND.X})}} \times \frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})} = \frac{\text{bias}_{(\text{YD.X})} \times \text{SF}_Y}{\text{bias}_{(\text{ND.X})}}$ or $= \frac{\text{bias}_{(\text{YD.X})}}{\text{bias}_{(\text{ND.X})} \times \text{SF}_N}$ where $\text{SF}_Y = \frac{1}{\text{SF}_N} = \frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})}$.

[5] We can also re-write the bias-factors as follows:

Estimator and partial identification Equation 12 provides us with a moment estimator for $\beta_{Y\sim D|Z,X}$ in Equation 13.[6][7]

$$\hat{\beta}_{Y\sim D|Z,X} = \hat{\beta}_{Y\sim D|X} - k \times \left(\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}\right) \times \frac{\|Y^{\perp D,X}\|_2}{\|N^{\perp D,X}\|_2} \tag{13}$$

Given k and $\beta_{N\sim D|Z,X}$, the estimator $\hat{\beta}_{Y\sim D|Z,X}$ from Equation 13 is consistent for $\beta_{Y\sim D|Z,X}$ and asymptotically normal. We will discuss estimating standard errors in Section 2.4. Using Equation 13 in a partial identification framework, we can reason about plausible ranges of values for $\beta_{N\sim D|Z,X}$ and k to arrive at a range for plausible estimates of $\beta_{Y\sim D|Z,X}$. In this way, we can partially identify $\beta_{Y\sim D|Z,X}$.

2.2 Placebo treatments

It is possible to use similar machinery to leverage imperfect placebo treatments rather than imperfect placebo outcomes in a partial identification framework for effects of actual treatments on outcomes. However, some details change in important ways. We briefly state the problem and resulting partial identification framework here. Additional details and the complete derivation can be found in Appendix B.

Suppose that we want to run the long regression for Y on D (Equation 1) but, again, Z is an unobserved variable that we would have liked to include in the regression but cannot. Suppose also that we have a placebo treatment, P, for which we believe the confounding of P and Y from Z is similar to the confounding of D and Y from Z. We might consider running two separate regressions as we did for placebo outcomes. However, that approach is not easily adapted to imperfect placebo treatments (i.e., those for which there is a direct relationship between P and Y).[8]

Instead, suppose that P is not a descendant of D, D is not a descendant of P, P is a neutral control for the effect of D on Y, and P is a neutral control for D on Y, conditional on Z and X.[9] Further, suppose that we want to run the "long" regression of Y on D, P, X, and Z in Equation 14. In particular, we are interested in the coefficient $\beta_{Y\sim D|P,Z,X}$ as a measure of the causal effect of D on Y or as an approximation to the causal effect of D on Y. However, since Z is unobserved, we must run the "short" regression of Y on D, P and X in Equation 15, rather than the desired long regression.

$$Y = \beta_{Y\sim D|P,Z,X} D + \beta_{Y\sim P|D,Z,X} P + \mathbf{X}\beta_{Y\sim X|D,P,Z} + \beta_{Y\sim Z|D,P,X} Z + \epsilon_l \tag{14}$$

$$Y = \beta_{Y\sim D|P,X} D + \beta_{Y\sim P|D,X} P + \mathbf{X}\beta_{Y\sim X|D,P} + \epsilon_s \tag{15}$$

Following a similar omitted variables based approach as we took for placebo outcomes, we can arrive at Equations 16 and

$$\begin{aligned}
\beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X} &= R_{Y\sim Z|D,X} f_{D\sim Z|X} \frac{\mathrm{SD}(Y^{\perp D,X})}{\mathrm{SD}(D^{\perp X})} \\
\implies R_{Y\sim Z|D,X} f_{D\sim Z|X} &= (\beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X}) \frac{\mathrm{SD}(D^{\perp X})}{\mathrm{SD}(Y^{\perp D,X})} \\
&= R_{Y\sim D|X} \frac{\mathrm{SD}(Y^{\perp X})}{\mathrm{SD}(D^{\perp X})} \frac{\mathrm{SD}(D^{\perp X})}{\mathrm{SD}(Y^{\perp D,X})} - R_{Y\sim D|Z,X} \frac{\mathrm{SD}(Y^{\perp X,Z})}{\mathrm{SD}(D^{\perp X,Z})} \frac{\mathrm{SD}(D^{\perp X})}{\mathrm{SD}(Y^{\perp D,X})} \frac{\mathrm{SD}(Y^{\perp D,X,Z})}{\mathrm{SD}(Y^{\perp D,X,Z})} \\
&= f_{Y\sim D|X} - f_{Y\sim D|Z,X} \frac{\mathrm{SD}(Y^{\perp D,X,Z})}{\mathrm{SD}(D^{\perp X,Z})} \frac{\mathrm{SD}(D^{\perp X})}{\mathrm{SD}(Y^{\perp D,X})} = f_{Y\sim D|X} - f_{Y\sim D|Z,X} \sqrt{\frac{1 - R^2_{Y\sim Z|D,X}}{1 - R^2_{D\sim Z|X}}} \\
\implies 1 &= \frac{f_{Y\sim D|X}}{R_{Y\sim Z|D,X} f_{D\sim Z|X}} - \frac{f_{Y\sim D|Z,X}}{f_{Y\sim Z|D,X} R_{D\sim Z|X}}
\end{aligned}$$

[6] We follow a similar presentation of our estimators and approaches to inference as in Zhang and Ding (2022).

[7] It is also possible to show that $\frac{\widehat{\mathrm{SD}}(Y^{\perp D,X})}{\widehat{\mathrm{SD}}(N^{\perp D,X})} = \frac{\left[\frac{\widehat{\mathrm{SD}}(Y^{\perp D,X})}{\widehat{\mathrm{SD}}(D^{\perp X})}\right]}{\left[\frac{\widehat{\mathrm{SD}}(N^{\perp D,X})}{\widehat{\mathrm{SD}}(D^{\perp X})}\right]} = \frac{\mathrm{se}(\hat{\beta}_{Y\sim D|X})}{\mathrm{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\mathrm{df}_Y}{\mathrm{df}_N}}$, which is an expression containing only commonly reported regression summary statistics from the two short regressions that are estimable from the observed data. $\mathrm{se}(\hat{\beta}_{Y\sim D|X})$ and $\mathrm{se}(\hat{\beta}_{N\sim D|X})$ are the estimated standard errors from standard regression estimation of the short regressions. df_Y and df_N are the degrees of freedom from the estimated short regressions.

[8] Suppose we have a causal model like Figure 1(d). $\beta_{Y\sim Z|P,X}$ (or $R_{Y\sim Z|P,X}$) will capture the causal relationship $D \to Y$, since the path $Z \to D \to Y$ will remain open conditional on P and X alone. Expressing the bias in the regression coefficient on D from the short regression using a two regression approach like we used for placebo outcomes would, therefore, require us to reason about the causal effect of interest itself, which is not useful.

[9] We require that D and P are neutral controls for each other in that they do not change the regression coefficients for the outcome Y when we include them in the same regression. This will simplify interpretation. If a user is concerned that there could be a path opened or closed by conditioning on D or P, we suggest drawing the DAG to check. See Section 4 for further discussion.

17. All components of these expressions except for $k = \frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}} = \frac{R_{Y\sim Z|D,P,X} f_{D\sim Z|P,X}}{R_{Y\sim Z|D,P,X} f_{P\sim Z|D,X}}$[10] and $\beta_{Y\sim P|D,Z,X}$ can be estimated from the data. We can use this for partial identification by considering plausible ranges of values for k and $\beta_{Y\sim P|D,Z,X}$. $\beta_{Y\sim P|D,Z,X}$ captures any direct causal relationship between P and Y. Since both P and Y are observed variables, their relationship conditional on D, Z, X should be able to be specified as the regression coefficient $\beta_{Y\sim P|D,Z,X}$ despite scale considerations. As with placebo outcomes, we can reason about k either in terms of relative D, Z and P, Z relationships or in terms of relative levels of confounding.

$$\text{bias}_{(\text{YD.PX})} \triangleq \beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = k \times \left(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \tag{16}$$

$$\beta_{Y\sim D|P,Z,X} = \beta_{Y\sim D|P,X} - k \times \left(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \tag{17}$$

Estimator and partial identification Equation 17 provides us with a moment estimator for $\beta_{Y\sim D|P,Z,X}$ in Equation 18.[11]

$$\hat{\beta}_{Y\sim D|P,Z,X} = \hat{\beta}_{Y\sim D|P,X} - k \times \left(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\|P^{\perp D,X}\|_2}{\|D^{\perp P,X}\|_2} \tag{18}$$

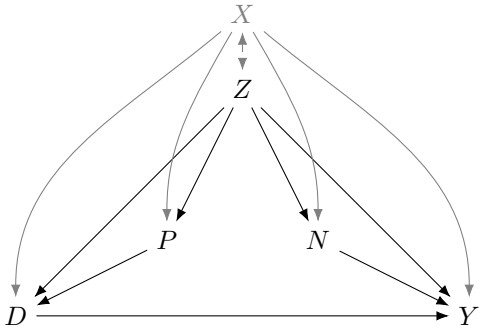
Given k and $\beta_{Y\sim P|D,Z,X}$, the estimator $\hat{\beta}_{Y\sim D|P,Z,X}$ from Equation 18 is consistent for $\beta_{Y\sim D|P,Z,X}$ and asymptotically normal. We discuss computing standard errors in Section 2.4. Using Equation 18 in a partial identification framework, we can reason about plausible ranges of values for $\beta_{Y\sim P|D,Z,X}$ and k to arrive at a range of plausible estimates of $\beta_{Y\sim D|P,Z,X}$.

2.3 Double placebos

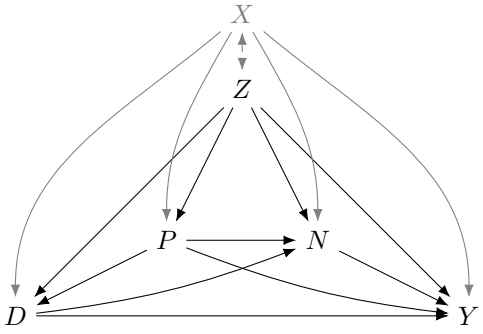
Finally, we consider the setting in which we have two placebos. Suppose we are interested in $\beta_{Y\sim D|Z,X}$ from the "long" regression in Equation 1. However, Z is unobserved and so we must only consider estimating "short" regression in Equation 2. Further, suppose we are in a setting like Figure 2 - that is, we observe both a placebo outcome and a placebo treatment. In such a "double placebo" setting, we can also develop a partial identification framework. With a "perfect" placebo treatment and outcome, meaning that the placebo treatment does not have a direct reltionship with the outcome and the placebo outcome does not have a direct relationship with the treatment as in Figure 2(a), it turns out that we can point identify $\beta_{Y\sim D|Z,X}$ from the long regression using observed data. This is similar to the proximal causal inference results (Miao et al., 2020) but where we just use OLS regression.

Figure 2: A single causal graph with both a placebo outcome, N, and with a placebo treatment, P. D is treatment, Y is outcome, and Z contains unobserved confounders. X contains observed covariates.

(a) Perfect Double Placebo



(a) Imperfect Double Placebo



[10] We need $f_{P\sim Z|D,X} \neq 0$ for k to be defined. This will be the case for any useful placebo treatment. Whether $f_{P\sim Z|D,X} \neq 0$ is only testable under the assumption that $\beta_{Y\sim P|D,Z,X} = 0$.

[11] It is also possible to show that $\frac{\widehat{\text{SD}}(P^{\perp D,X})}{\widehat{\text{SD}}(D^{\perp P,X})} = \frac{\widehat{\text{SD}}(P^{\perp D,X})}{\widehat{\text{SD}}(D^{\perp P,X})} \times \frac{\widehat{\text{SD}}(Y^{\perp D,P,X})}{\widehat{\text{SD}}(Y^{\perp D,P,X})} = \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}}$, which is an expression containing only commonly reported regression summary statistics from the short regression that is estimable from the observed data. $\text{se}(\hat{\beta}_{Y\sim D|P,X})$ and $\text{se}(\hat{\beta}_{Y\sim P|D,X})$ are the estimated standard errors from standard regression estimation of the short regression. df_Y and df_N are the degrees of freedom from the estimated short regression, these will be equal in this case.

First consider the following two sets of short and long regressions. Again, Z is an unobserved confounder. So we can only consider estimating the short regressions (i.e., Equations 19b and 19d). Traditional omitted variable bias tells us that the four expressions in Equations 20 hold.

$$Y = \beta_{Y\sim D|P,Z,X} D + \beta_{Y\sim P|D,Z,X} P + \mathbf{X}\beta_{Y\sim X|D,P,Z} + \beta_{Y\sim Z|D,P,X} Z + \epsilon_{y,l} \tag{19a}$$
$$Y = \beta_{Y\sim D|P,X} D + \beta_{Y\sim P|D,X} P + \mathbf{X}\beta_{Y\sim X|D,P} + \epsilon_{y,s} \tag{19b}$$
$$N = \beta_{N\sim D|P,Z,X} D + \beta_{N\sim P|D,Z,X} P + \mathbf{X}\beta_{N\sim X|D,P,Z} + \beta_{N\sim Z|D,P,X} Z + \epsilon_{n,l} \tag{19c}$$
$$N = \beta_{N\sim D|P,X} D + \beta_{N\sim P|D,X} P + \mathbf{X}\beta_{N\sim X|D,P} + \epsilon_{n,s} \tag{19d}$$

$$\text{bias}_{(\text{YD.PX})} \triangleq \beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = \beta_{Y\sim Z|D,P,X}\beta_{Z\sim D|P,X} \tag{20a}$$
$$\text{bias}_{(\text{YP.DX})} \triangleq \beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X} = \beta_{Y\sim Z|D,P,X}\beta_{Z\sim P|D,X} \tag{20b}$$
$$\text{bias}_{(\text{ND.PX})} \triangleq \beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X} = \beta_{N\sim Z|D,P,X}\beta_{Z\sim D|P,X} \tag{20c}$$
$$\text{bias}_{(\text{NP.DX})} \triangleq \beta_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X} = \beta_{N\sim Z|D,P,X}\beta_{Z\sim P|D,X} \tag{20d}$$

We can rewrite Equations 20b and 20c in the following way. Equation 20b becomes Equation 21a; Equation 20c becomes Equation 21b.

$$\beta_{Y\sim Z|D,P,X} = \frac{\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}}{\beta_{Z\sim P|D,X}} \tag{21a}$$
$$\beta_{Z\sim D|P,X} = \frac{\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}}{\beta_{N\sim Z|D,P,X}} \tag{21b}$$

From these, we can write bias_{YD} and, therefore, $\beta_{Y\sim D|P,Z,X}$ as in Equation 22. This is an expression containing regression coefficients from the two short regressions ($\beta_{Y\sim D|P,X}, \beta_{Y\sim P|D,X}, \beta_{N\sim D|P,X}$, and $\beta_{N\sim P|D,X}$) as well as three parameters ($\beta_{Y\sim P|D,Z,X}, \beta_{N\sim D|P,Z,X}$, and $\beta_{N\sim P|D,Z,X}$) that capture the extent to which N and P are imperfect placebos.

$$\begin{aligned}
\text{bias}_{(\text{YD.PX})} &= \beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\
&= \beta_{Y\sim Z|D,P,X}\beta_{Z\sim D|P,X} \\
&= \left[\frac{\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}}{\beta_{Z\sim P|D,X}}\right]\left[\frac{\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}}{\beta_{N\sim Z|D,P,X}}\right] \\
&= \frac{(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})(\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})}{[\beta_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X}]} \\
\therefore \beta_{Y\sim D|P,Z,X} &= \beta_{Y\sim D|P,X} - \frac{(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})(\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})}{(\beta_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X})}
\end{aligned} \tag{22}$$

Estimator and partial identification Equation 22 provides us with a moment estimator for $\beta_{Y\sim D|P,Z,X}$ in Equation 23.

$$\hat{\beta}_{Y\sim D|P,Z,X} = \hat{\beta}_{Y\sim D|P,X} - \frac{(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})}{(\hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X})} \tag{23}$$

Given $\beta_{Y\sim P|D,Z,X}, \beta_{N\sim D|P,Z,X}$, and $\beta_{N\sim P|D,Z,X}$, the estimator $\hat{\beta}_{Y\sim D|P,Z,X}$ from Equation 23 is consistent for $\beta_{Y\sim D|P,Z,X}$ and asymptotically normal. We discuss computing standard errors in Section 2.4. Using Equation 23 in a partial identification framework, we can reason about plausible ranges of values for $\beta_{Y\sim P|D,Z,X}, \beta_{N\sim D|P,Z,X}$, and $\beta_{N\sim P|D,Z,X}$ to arrive at a range of plausible estimates of $\beta_{Y\sim D|P,Z,X}$. In Figure 2(b), the partial identification parameters for double placebos capture paths as follows: $\beta_{Y\sim P|D,Z,X}$ captures $P \to Y$ and $P \to N \to Y$ (the effect of P on Y conditional on D); $\beta_{N\sim D|P,Z,X}$ captures $D \to N$; $\beta_{N\sim P|D,Z,X}$ captures $P \to N$. We can disentangle the $P \to Y$ and $P \to N \to Y$ components of $\beta_{Y\sim P|D,Z,X}$ for Figure 2(b) by using the following omitted variable expression: $\beta_{Y\sim P|D,Z,X} = \beta_{Y\sim P|N,D,Z,X} + \beta_{Y\sim N|P,D,Z,X}\beta_{N\sim P|D,Z,X}$. $\beta_{Y\sim P|N,D,Z,X}$ captures $P \to Y$; $\beta_{Y\sim N|P,D,Z,X}$ captures $N \to Y$; $\beta_{N\sim P|D,Z,X}$ captures $P \to N$ and is already a partial identification parameter. Thus, we have 4 partial identification parameters. Note that none of the partial identification parameters involve a relationship with Z; Z is only in the set of variables that is partialled out.

This type of expression can be used in a wide variety of settings. Though the interpretation of the partial identification parameters and of $\beta_{Y\sim D|P,Z,X}$ will depend on the underlying causal model being considered. If we are in a setting like in Figure 2(a), where "perfect" placebos are available, then $\beta_{Y\sim P|D,Z,X}$, $\beta_{N\sim D|P,Z,X}$, and $\beta_{N\sim P|D,Z,X}$ will all equal zero and we will be able to point identify $\beta_{Y\sim D|P,Z,X}$. With two imperfect placebos, users have a choice between the "single placebo" approach discussed above that considers relative levels of confounding as a partial identification parameter and the "double placebo" approach discussed here in which the presence of a second placebo means that we can account for both "arms" of the unobserved confounder but need to reason about the extent to which both placebos are imperfect.

2.4 Standard Errors

Work in progress.

3 Multiple unobserved confounders

The approaches in Sections 2.1, 2.2, and 2.3 show how to leverage placebo treatments and outcomes to learn about the effect of a treatment on an outcome when there is a single unobserved confounder. We can employ similar methods for the case where we have multiple unobserved confounders by allowing $\mathbf{Z}$ to be a matrix of unobserved variables, $\beta_{Y\sim Z|D,X}$ to be the vector of coefficients for $\mathbf{Z}$, and the long regression becomes Equation 24. We can consider $Z_{(\text{Y.DX})} = \mathbf{Z}\beta_{Y\sim Z|D,X}$ to be the linear combination of the columns of $\mathbf{Z}$ that "de-confounds" the $D \to Y$ relationship in the same way that including $\mathbf{Z}$ in Equation 24 does.[12] That is $\beta_{Y\sim D|Z,X} = \beta_{Y\sim D|Z_{(\text{Y.DX})},X}$ from Equation 24 and 25.[13]

$$Y = \beta_{Y\sim D|Z,X} D + \mathbf{X}\beta_{Y\sim X|D,Z} + \mathbf{Z}\beta_{Y\sim Z|D,X} + \epsilon_l \tag{24}$$

$$Y = \beta_{Y\sim D|Z_{(\text{Y.DX})},X} D + \mathbf{X}\beta_{Y\sim X|D,Z_{(\text{Y.DX})}} + Z_{(\text{Y.DX})} + \epsilon_l \tag{25}$$

Let's clarify two important points. First, in reality, there will almost always be multiple unobserved confounders of every relationship we want to consider. Second, even when we considering the set of unobserved variables that is rich enough to de-confound all the relationships of interest, these unobserved variables will have different relationships with the each of the observed variables of interest. Is there an alternative approach that neatly deals with these two points? Yes. In the context of placebo outcomes (Section 2.1), we discussed that m and k can be interpreted as a ratio of the level of confounding of the Y, D relationship to the level of confounding of the N, D relationship. This was only for a single unobserved variable. But why don't we start from such a comparison in the case with multiple unobserved confounders? We do this using reparameterization of the traditional omitted variable bias. (Cinelli and Hazlett, 2020)

$$\begin{aligned}
\text{bias}_{(\text{YD.X})} &= \beta_{Y\sim D|X} - \beta_{Y\sim D|Z_{(\text{Y.DX})},X} = R_{Y\sim Z_{(\text{Y.DX})}|D,X} f_{D\sim Z_{(\text{Y.DX})}|X}(\text{SD}(Y^{\perp D,X})/\text{SD}(D^{\perp X})) \\
\text{bias}_{(\text{ND.X})} &= \beta_{N\sim D|X} - \beta_{N\sim D|Z_{(\text{N.DX})},X} = R_{N\sim Z_{(\text{N.DX})}|D,X} f_{D\sim Z_{(\text{N.DX})}|X}(\text{SD}(N^{\perp D,X})/\text{SD}(D^{\perp X})) \\
\text{bias}_{(\text{YD.X})} &= m \times \text{bias}_{(\text{ND.X})} \implies \beta_{Y\sim D|Z_{(\text{Y.DX})},X} = \beta_{Y\sim D|X} - m \times (\beta_{N\sim D|X} - \beta_{N\sim D|Z_{(\text{N.DX})},X}) \\
m &= \frac{\text{bias}_{(\text{YD.X})}}{\text{bias}_{(\text{ND.X})}} = \frac{R_{Y\sim Z_{(\text{Y.DX})}|D,X} f_{D\sim Z_{(\text{Y.DX})}|X}}{R_{N\sim Z_{(\text{N.DX})}|D,X} f_{D\sim Z_{(\text{N.DX})}|X}} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} = k \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \\
\therefore \beta_{Y\sim D|Z_{(\text{Y.DX})},X} &= \beta_{Y\sim D|X} - k \times (\beta_{N\sim D|X} - \beta_{N\sim D|Z_{(\text{N.DX})},X}) \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}
\end{aligned} \tag{26}$$

In Equation 26, we start by noting that we can compare the level of confounding of the $D \to N$ relationship from $Z_{(\text{N.DX})} = \mathbf{Z}\beta_{N\sim Z|D,X}$ to the level of confounding of the $D \to Y$ relationship from $Z_{(\text{Y.DX})} = \mathbf{Z}\beta_{Y\sim Z|D,X}$. We are assuming that $\mathbf{Z}$ is a rich enough set of variables to de-confound both $D \to Y$ and $D \to N$; but we use the appropriate linear combination of the components of $\mathbf{Z}$ in each of the initial expressions for omitted variable bias. We then show that we can express $\beta_{Y\sim D|Z_{(\text{Y.DX})},X}$, our target quantity, in terms of quantities that are estimable from the data as well as k and $\beta_{N\sim D|Z_{(\text{N.DX})},X}$. $k \triangleq \frac{R_{Y\sim Z_{(\text{Y.DX})}|D,X} f_{D\sim Z_{(\text{Y.DX})}|X}}{R_{N\sim Z_{(\text{N.DX})}|D,X} f_{D\sim Z_{(\text{N.DX})}|X}}$ is the ratio of the level of total confounding of $D \to Y$ from $D \leftarrow Z \to Y$ to the level of total confounding of $D \to N$ from $D \leftarrow Z \to N$, after re-scaling one of these biases to account for scale differences between N

[12] Note that in this section and others, we will refer to including $\mathbf{Z}$ in the long regression as "de-confounding". Readers should recognize that, in reality, we are only partialling out the linear relationships with $\mathbf{Z}$. This is, therefore, a linear approximation to full non-parametric elimination of confounding from $\mathbf{Z}$.

[13] Indeed $R^2_{Y\sim Z_{(\text{Y.DX})}|D,X}$ will also equal $R^2_{Y\sim Z|D,X}$. But $R^2_{D\sim Z_{(\text{Y.DX})}|X} \leq R^2_{D\sim Z|X}$, since $Z_{(\text{Y.DX})}$ is chosen to maximize the R^2 between $\mathbf{Z}$ and Y not $\mathbf{Z}$ and D. (Cinelli and Hazlett, 2020)

and Y.[14] $\beta_{N\sim D|Z_{(\text{N.DX})},X}$ measures the causal effect of D on N, since $Z_{(\text{N.DX})}$ is the linear combination of the components of $\mathbf{Z}$ that de-confounds $D \rightarrow N$. Note that the expression at the end of Equation 26 is identical in form to the expression in Equation 12. All that changed was the interpretation of k and the Zs.

It is easy to do something similar for the placebo treatment case (Section 2.2). See Equation 27. We start by noting that we can compare the level of confounding of the $P \rightarrow Y$ relationship from $\mathbf{Z}$ to the level of confounding of the $D \rightarrow Y$ relationship from $\mathbf{Z}$ by using the same linear combination, $Z_{(\text{Y.DPX})} = \mathbf{Z}\beta_{Y\sim Z|D,P,X}$. This is because in the placebo treatment approach we run a single regression with Y as the outcome that includes both D and P on the right-hand side. Since we assume that $\mathbf{Z}$ contains a rich enough set of unobserved variables to de-confound both $P \rightarrow Y$ and $D \rightarrow Y$, $Z_{(\text{Y.DPX})}$ is also sufficient for this. We then show that we can express $\beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X}$, our target quantity, in terms of quantities that are estimable from the data as well as k and $\beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}$. $kk \triangleq \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}}{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{P\sim Z_{(\text{Y.DPX})}|D,X}}$ is the ratio of the level of total confounding of $D \rightarrow Y$ from $D \leftarrow Z \rightarrow Y$ to the level of total confounding of $P \rightarrow Y$ from $P \leftarrow Z \rightarrow Y$, after re-scaling one of these biases to account for scale differences between P and D.[15] $\beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}$ measures the causal effect of P on Y, since $Z_{(\text{Y.DPX})}$ is the linear combination of the components of $\mathbf{Z}$ that de-confounds $P \rightarrow Y$. Note that the expression at the end of Equation 27 is identical in form to the expression in Equation 17. All that changed was the interpretation of k and the Zs.

$$
\begin{aligned}
&\text{bias}_{(\text{YD.PX})} = \beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X} (\text{SD}(Y^{\perp D,P,X})/\text{SD}(D^{\perp P,X})) \\
&\text{bias}_{(\text{YP.DX})} = \beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X} = R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{P\sim Z_{(\text{Y.DPX})}|D,X} (\text{SD}(Y^{\perp D,P,X})/\text{SD}(P^{\perp D,X})) \\
&\text{bias}_{(\text{YD.PX})} = m \times \text{bias}_{(\text{YP.DX})} \implies \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim D|P,X} - m \times (\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}) \\
&m = \frac{\text{bias}_{(\text{YD.PX})}}{\text{bias}_{(\text{YP.DX})}} = \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}}{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{P\sim Z_{(\text{Y.DPX})}|D,X}} \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} = k \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \\
&\therefore \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim D|P,X} - k \times (\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}) \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})}
\end{aligned}
\tag{27}
$$

We can also do something similar for the double placebo case as well. To account for the possibility that the four different relationships we use may be de-confounded by different linear combinations of the components of $\mathbf{Z}$, we also need to scale the expression for $\beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X}$ in Equation 22 by $k_{(\text{YD / YP})} \times k_{(\text{NP / ND})} = \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{Z_{(\text{Y.DPX})}\sim D|P,X}}{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{Z_{(\text{Y.DPX})}\sim P|D,X}} \times \frac{R_{N\sim Z_{(\text{N.DPX})}|D,P,X} f_{Z_{(\text{N.DPX})}\sim P|D,X}}{R_{N\sim Z_{(\text{N.DPX})}|D,P,X} f_{Z_{(\text{N.DPX})}\sim P|D,X}}$. $k_{(\text{YD / YP})}$ is the ratio of the level of total confounding of the D,Y relationship, conditional on P, to the level of total confounding of the P,Y relationship, after re-scaling one of these biases to account for scale differences between P and D. $k_{(\text{NP / ND})}$ is the ratio of the level of total confounding of the P,N relationship to the level of total confounding of the D,N relationship, after re-scaling one of these biases to account for scale differences between P and D. See Appendix C for details.

4 A taxonomy of causal graphs

How do we know if we have a useful placebo? How do we know if it is better to treat a placebo as a placebo outcome or a placebo treatment? Are there settings in which the above approaches should not be used? Are there settings in which the proposed methods can be used for problems other than unobserved confounding? We will address these questions by developing a taxonomy of causal graphs when a single placebo is available and when two are available. For introductions to graphical causal models see Pearl (2009), Richardson and Robins (2013), Matthay and Glymour (2020), and Cinelli et al. (2022). We re-emphasize the Necker cube quality of placebos. Two causal graphs, each with a placebo, that have the same structure may invite different interpretations for whether the placebo is a placebo treatment or outcome. These choices are context specific. Sometimes the placebo may feel more like an outcome, while in other contexts the placebo feels more like a treatment.

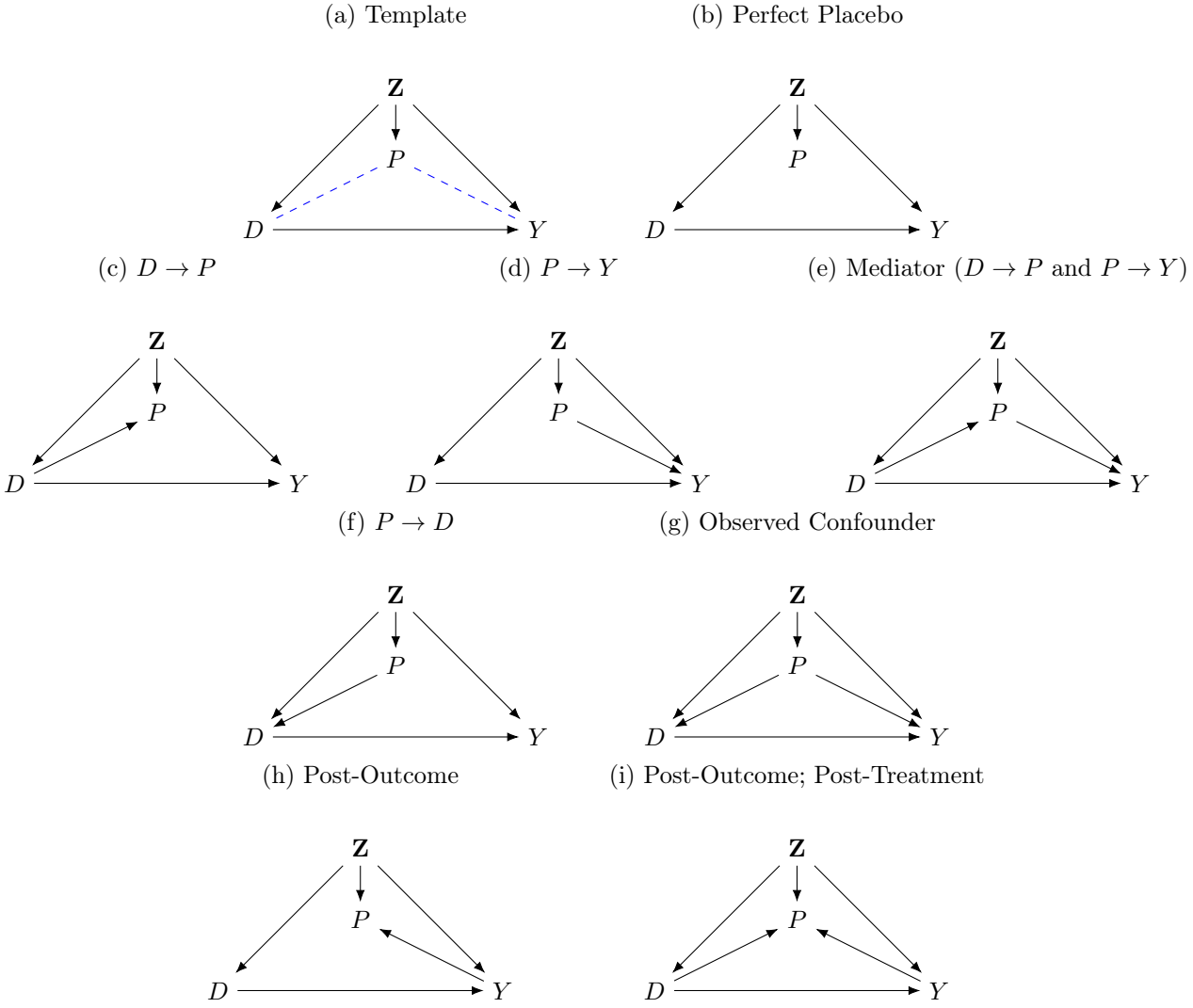
[14] The numerator and denominator in k are both "bias factors" in the language of Cinelli and Hazlett (2020). This can be interpreted as the bias in the Y, D relationship as a fraction of the bias in the N, D relationship, after re-scaling one of these biases to correct for any differences in the scales of N and Y. See footnote 4 for additional discussion of ratios of bias factors.

[15] Note that we do not cancel terms in $k = \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}}{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{P\sim Z_{(\text{Y.DPX})}|D,X}}$ since this would force us to reason about how $f_{D\sim Z_{(\text{Y.DPX})}|P,X}$ differs from $f_{D\sim \mathbf{Z}|P,X}$ and how $f_{P\sim Z_{(\text{Y.DPX})}|D,X}$ differs from $f_{P\sim \mathbf{Z}|D,X}$. It is more intuitive to reason about relative levels of bias, which does not require us to directly consider those differences.

4.1 Single Placebos

We start by listing all possible causal graphs when a single placebo is observed in Figure 3. We assume that all relevant unobserved common causes are bundled into the vector of unobserved variables $\mathbf{Z}$. Observed covariates $\mathbf{X}$ are omitted from the graphs for clarity. We refer to the placebo as P. Figure 3(a) contains a template from which the rest of the graphs in Figure 3 are derived. The two dashed edges can point in either direction or not exist.[16] Different approaches should be used when dealing with each causal structure.[17]

Figure 3: Single placebo DAGs.



Placebos Figure 3(b) is the simplest setting in which P has no direct relationship with either the treatment, D, or the outcome, Y. That is, P is a "perfect" placebo. There are two options for how we might proceed in this case. We can look at P as a placebo outcome and follow the approach laid out in Section 2.1 (and generalized in Equation 26), or we can look at P as a placebo treatment and follow the approach laid out in Section 2.2 (and generalized in Equation 27). In either case, since P is a perfect placebo, we will only have one partial identification parameter that captures the relative levels of confounding.

[16] There are $3^2 - 1 = 8$ possible graphs, excluding those with cycles.

[17] The approaches presented in Section 2.1 and 2.2 hold for any causal graph. However, the interpretation of the partial identification parameters depend on the causal graph. We, therefore, provide guidance on which approaches we suggest for each causal graph. Assuming a causal graph (or set of causal graphs) that are plausible in any application is a pre-requisite to making any credible causal claims. This section includes some additional approaches not yet discussed. Details of these can be found in the Appendix.

Figure 3(c) has the addition of $D \to P$. Since the placebo is post-treatment in this setting, it is likely that we will want to treat the placebo as an imperfect placebo outcome. Therefore, we can follow the approach laid out in Section 2.1 (and generalized in Equation 26). [18] Figure 3(d) has $P \to Y$. Here there are also two options. The first would be to treat P as an imperfect placebo treatment and follow the approach laid out in Section 2.2 (and generalized in Equation 27). The second would be to view P as a placebo outcome, as in the case that P is a pre-treatment version of the outcome. In this setting, we can employ the approach laid out in Equation 28. The key here is the inclusion of P in the regression of Y on D.

$$
\begin{aligned}
&\text{bias}_{(\text{YD.PX})} = \beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}(\text{SD}(Y^{\perp D,P,X})/\text{SD}(D^{\perp P,X}))\\
&\text{bias}_{(\text{PD.X})} = \beta_{P\sim D|X} - \beta_{P\sim D|Z_{(\text{P.DX})},X} = R_{P\sim Z_{(\text{P.DX})}|D,X} f_{D\sim Z_{(\text{P.DX})}|X}(\text{SD}(P^{\perp D,X})/\text{SD}(D^{\perp X}))\\
&\text{bias}_{(\text{YD.PX})} = m \times \text{bias}_{(\text{PD.X})} \implies \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim D|P,X} - m \times (\beta_{P\sim D|X} - \beta_{P\sim D|Z_{(\text{P.DX})},X})\\
&m = \frac{\text{bias}_{(\text{YD.PX})}}{\text{bias}_{(\text{PD.X})}} = \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}}{R_{P\sim Z_{(\text{P.DX})}|D,X} f_{D\sim Z_{(\text{P.DX})}|X}} \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})} \times \frac{\text{SD}(D^{\perp X})}{\text{SD}(P^{\perp D,X})} = k \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})} \times \frac{\text{SD}(D^{\perp X})}{\text{SD}(P^{\perp D,X})}\\
&\therefore \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim D|P,X} - k \times (\beta_{P\sim D|X} - \beta_{P\sim D|Z_{(\text{P.DX})},X}) \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})} \times \frac{\text{SD}(D^{\perp X})}{\text{SD}(P^{\perp D,X})}
\end{aligned}
\tag{28}
$$

In Equation 28, $k \triangleq \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}}{R_{P\sim Z_{(\text{P.DX})}|D,X} f_{D\sim Z_{(\text{P.DX})}|X}}$ is the ratio of the level of confounding D and Y, conditional on P, to the level of confounding D and P, after re-scaling one of these biases to account for scale differences. Since we included P in the regression of Y on D, we see that the confounding of D and Y does not include the $Z \to P \to Y$ path, making comparison with the confounding of D and P easier. $\beta_{P\sim D|Z_{(\text{P.DX})},X}$ measures the causal effect of D on P, which based on Figure 3(d) should be zero.

Mediators Figure 3(e) has $D \to P$ and $P \to Y$, making P a mediator between D and Y. In this case there is more than one option. First, Zhang and Ding (2022) provide guidance on how to take an omitted variables bias approach to partial identification of the direct effect and indirect effect. The direct effect approach deals with the fact that conditioning on P opens the collider path $D \to P \leftarrow Z$ and involves specifying partial identification parameters for each of the $Z \to D$, $Z \to P$, and $Z \to Y$ relationships. Suppose we want to consider the total effect. We could use the approach outlined in Cinelli and Hazlett (2020), but this approach for the total effect applied to a setting with a mediator requires that we reason about $R^2_{Y\sim Z|D}$ which captures the $Z \to Y$ and $Z \to P \to Y$ relationships. A wrinkle is that $P \to Y$ is a component of the total effect we want to identify. An alternative would be to follow the approach laid out in Section 2.1 (and generalized in Equation 26), where the mediator is treated as a placebo outcome. Here there is again the wrinkle that we will be required to reason about $D \to P$ (above we refer to the placebo outcomes as N), which is a component of the total effect we want to identify. We could also follow the approach in Equation 29. This does something similar but where we treat the mediator as a placebo treatment. The wrinkle remains that we need to reason about $P \to Y$, a component of the total effect we want to identify.

$$
\begin{aligned}
&\text{bias}_{(\text{YD.X})} = \beta_{Y\sim D|X} - \beta_{Y\sim D|Z_{(\text{Y.DX})},X} = R_{Y\sim Z_{(\text{Y.DX})}|D,X} f_{D\sim Z_{(\text{Y.DX})}|X}(\text{SD}(Y^{\perp D,X})/\text{SD}(D^{\perp X}))\\
&\text{bias}_{(\text{YP.DX})} = \beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X} = R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{P\sim Z_{(\text{Y.DPX})}|D,X}(\text{SD}(Y^{\perp D,P,X})/\text{SD}(P^{\perp D,X}))\\
&\text{bias}_{(\text{YD.X})} = m \times \text{bias}_{(\text{YP.DX})} \implies \beta_{Y\sim D|Z_{(\text{Y.DX})},X} = \beta_{Y\sim D|X} - m \times (\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X})\\
&m = \frac{\text{bias}_{(\text{YD.X})}}{\text{bias}_{(\text{YP.DX})}} = \frac{R_{Y\sim Z_{(\text{Y.DX})}|D,X} f_{D\sim Z_{(\text{Y.DX})}|X}}{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{P\sim Z_{(\text{Y.DPX})}|D,X}} \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp X})} \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(Y^{\perp D,P,X})} = k \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp X})} \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(Y^{\perp D,P,X})}\\
&\therefore \beta_{Y\sim D|Z_{(\text{Y.DX})},X} = \beta_{Y\sim D|X} - k \times (\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}) \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp X})} \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(Y^{\perp D,P,X})}
\end{aligned}
\tag{29}
$$

Observed Confounders Figure 3(f) has $P \to D$. This is similar to an observed confounder. Figure 3(g) has $P \to D$ and $P \to Y$, making P an observed confounder. In both of these cases, there are two options. First, we could follow the approach laid out in Cinelli and Hazlett (2020), where P is an observed covariate. Second, we could follow the approach in Equation 30. The key here is to include P in the regression of Y on D and to also consider the regression of D on P.

[18] In Figure 3(c), the approach from Section 2.2 will not work well. The regression $Y \sim D + P$ will run into the problem that $\beta_{Y\sim D|P} = \beta_{Y\sim D|P,Z} + \beta_{Y\sim Z|D,P}\beta_{Z\sim D|P}$ and $\beta_{Z\sim D|P}$ captures both $Z \to D$ and $Z \to P \leftarrow D$. This means that a parameter that we will need to reason about will capture the effect of conditioning on a collider.

$$
\begin{aligned}
&\text{bias}_{(\text{YD.PX})} = \beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}(\text{SD}(Y^{\perp D,P,X})/\text{SD}(D^{\perp P,X}))\\
&\text{bias}_{(\text{DP.X})} = \beta_{D\sim P|X} - \beta_{D\sim P|Z_{(\text{D.PX})},X} = R_{D\sim Z_{(\text{D.PX})}|P,X} f_{P\sim Z_{(\text{D.PX})}|X}(\text{SD}(D^{\perp P,X})/\text{SD}(P^{\perp X}))\\
&\text{bias}_{(\text{YD.PX})} = m \times \text{bias}_{(\text{DP.X})} \implies \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim D|P,X} - m \times (\beta_{D\sim P|X} - \beta_{D\sim P|Z_{(\text{D.PX})},X})\\
&m = \frac{\text{bias}_{(\text{YD.PX})}}{\text{bias}_{(\text{DP.X})}} = \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}}{R_{D\sim Z_{(\text{D.PX})}|P,X} f_{P\sim Z_{(\text{D.PX})}|X}} \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})} \times \frac{\text{SD}(P^{\perp X})}{\text{SD}(D^{\perp P,X})} = k \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})} \times \frac{\text{SD}(P^{\perp X})}{\text{SD}(D^{\perp P,X})}\\
&\therefore \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim D|P,X} - k \times (\beta_{D\sim P|X} - \beta_{D\sim P|Z_{(\text{D.PX})},X}) \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})} \times \frac{\text{SD}(P^{\perp X})}{\text{SD}(D^{\perp P,X})}
\end{aligned}
\tag{30}
$$

In Equation 30, $k \triangleq \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{D\sim Z_{(\text{Y.DPX})}|P,X}}{R_{D\sim Z_{(\text{D.PX})}|P,X} f_{P\sim Z_{(\text{D.PX})}|X}}$ is the ratio of the level of confounding D and Y, conditional on P, to the level of confounding D and P, after re-scaling one of these biases to account for scale differences. Since we included P in the regression of Y on D, we see that the confounding of D and Y does not include the paths that run through P, making comparison with the confounding of D and P easier. $\beta_{D\sim P|Z_{(\text{D.PX})},X}$ measures the causal effect of P on D.

Post-outcome Figure 3(h) has $Y \to P$. Figure 3(i) has $Y \to P$ and $D \to P$. In both cases, P is a post-outcome variable. In both of these cases, there are two options. First, we could follow the approach laid out in Cinelli and Hazlett (2020), where we ignore P. Second, we could follow the approach in Equation 31. The key is to consider the regression of P on Y and D.

$$
\begin{aligned}
&\text{bias}_{(\text{YD.X})} = \beta_{Y\sim D|X} - \beta_{Y\sim D|Z_{(\text{Y.DX})},X} = R_{Y\sim Z_{(\text{Y.DX})}|D,X} f_{D\sim Z_{(\text{Y.DX})}|X}(\text{SD}(Y^{\perp D,X})/\text{SD}(D^{\perp X}))\\
&\text{bias}_{(\text{PY.DX})} = \beta_{P\sim Y|D,X} - \beta_{P\sim Y|D,Z_{(\text{P.YDX})},X} = R_{P\sim Z_{(\text{P.YDX})}|Y,D,X} f_{Y\sim Z_{(\text{P.YDX})}|D,X}(\text{SD}(P^{\perp Y,D,X})/\text{SD}(Y^{\perp D,X}))\\
&\text{bias}_{(\text{YD.X})} = m \times \text{bias}_{(\text{PY.DX})} \implies \beta_{Y\sim D|Z_{(\text{Y.DX})},X} = \beta_{Y\sim D|X} - m \times (\beta_{P\sim Y|D,X} - \beta_{P\sim Y|D,Z_{(\text{P.YDX})},X})\\
&m = \frac{\text{bias}_{(\text{YD.X})}}{\text{bias}_{(\text{PY.DX})}} = \frac{R_{Y\sim Z_{(\text{Y.DX})}|D,X} f_{D\sim Z_{(\text{Y.DX})}|X}}{R_{P\sim Z_{(\text{P.YDX})}|Y,D,X} f_{Y\sim Z_{(\text{P.YDX})}|D,X}} \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})} \times \frac{\text{SD}(P^{\perp X})}{\text{SD}(D^{\perp P,X})} = k \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(D^{\perp X})} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(P^{\perp Y,D,X})}\\
&\therefore \beta_{Y\sim D|Z_{(\text{Y.DX})},X} = \beta_{Y\sim D|X} - k \times (\beta_{P\sim Y|D,X} - \beta_{P\sim Y|D,Z_{(\text{P.YDX})},X}) \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(D^{\perp X})} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(P^{\perp Y,D,X})}
\end{aligned}
\tag{31}
$$

In Equation 31, $k \triangleq \frac{R_{Y\sim Z_{(\text{Y.DX})}|D,X} f_{D\sim Z_{(\text{Y.DX})}|X}}{R_{P\sim Z_{(\text{P.YDX})}|Y,D,X} f_{Y\sim Z_{(\text{P.YDX})}|D,X}}$ is the ratio of the level of confounding of D and Y to the level of confounding of Y and P, after re-scaling one of these biases to account for scale differences. $\beta_{P\sim Y|D,Z_{(\text{P.YDX})},X}$ measures the causal effect of Y on P.

4.2 Double Placebos

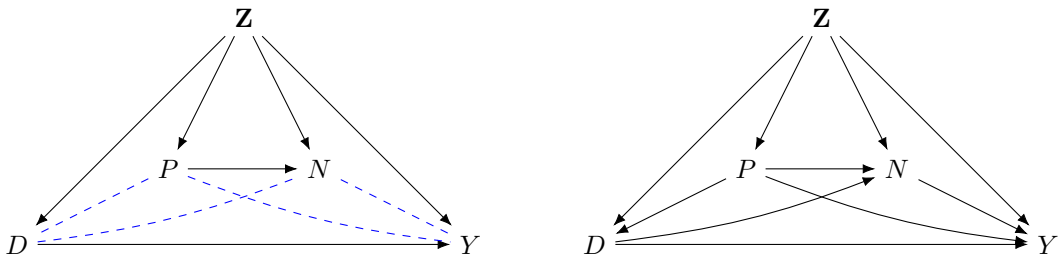
In the double placebo case, there are many more possible graphs. In this section, we discuss a single graph that we believe should be useful in many contexts. We assume a fully connected graph. We only make some slight assumptions on the direction of some edges for ease of exposition. That is, all the dashed edges in Figure 4(a) could have either direction or not exist[19] (excluding combinations of edge directions that lead to cycles). However, we work with the specific case shown in Figure 4(b) since this should cover many practical settings and researchers interested in alternative graphs can reason through the implications of their favored graph and its implications for interpreting the partial identification parameters with relative ease. Again, we assume that all relevant unobserved common causes are bundled into the vector of unobserved variables $\mathbf{Z}$. Observed covariates $\mathbf{X}$ are omitted from the graphs for clarity. See Appendix C for details on the double placebo case generalized to multiple unobserved confounders.

In Figure 4(b), the partial identification parameters for double placebos capture paths as follows: $\beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}$ captures $P \to Y$ and $P \to N \to Y$ (the effect of P on Y conditional on D); $\beta_{N\sim D|P,Z_{(\text{N.DPX})},X}$ captures $D \to N$; $\beta_{N\sim P|D,Z_{(\text{N.DPX})},X}$ captures $P \to N$. We can disentangle the $P \to Y$ and $P \to N \to Y$ components of $\beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}$ for Figure 4(b) by using the following omitted variable expression: $\beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim P|N,D,Z_{(\text{Y.DPX})},X} + \beta_{Y\sim N|P,D,Z_{(\text{Y.DPX})},X}\beta_{N\sim P|D,Z_{(\text{Y.DPX})},X}$. $\beta_{Y\sim P|N,D,Z_{(\text{Y.DPX})},X}$ captures $P \to Y$; $\beta_{Y\sim N|P,D,Z_{(\text{Y.DPX})},X}$ captures $N \to Y$; $\beta_{N\sim P|D,Z_{(\text{Y.DPX})},X}$ captures $P \to N$ and should be similar to the existing partial identification parameter $\beta_{N\sim P|D,Z_{(\text{N.DPX})},X}$. We also have $k_{(\text{YD / YP})} \times k_{(\text{NP / ND})} =$

[19] There are $3^4 = 81$ possible graphs, assuming that users know which of the two placebos is causally prior to the other; but the graphs with cycles should be excluded. Making the conservative assumption that the dashed edges in Figure 4(a) do exist, then there are $2^4 = 16$ possible fully connected graphs; again, the graphs with cycles should be excluded.

Figure 4: Double placebo causal graph.

(a) Template (b) Observed confounder and mediator



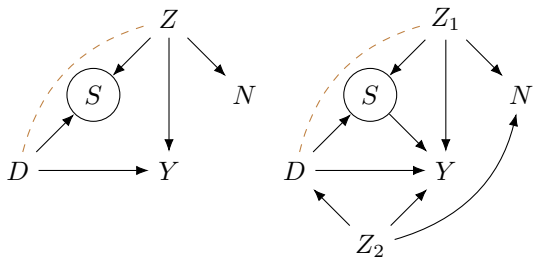
$\frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{Z_{(\text{Y.DPX})}\sim D|P,X}}{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{Z_{(\text{Y.DPX})}\sim P|D,X}} \times \frac{R_{N\sim Z_{(\text{N.DPX})}|D,P,X} f_{Z_{(\text{N.DPX})}\sim P|D,X}}{R_{N\sim Z_{(\text{N.DPX})}|D,P,X} f_{Z_{(\text{N.DPX})}\sim D|P,X}}$. $k_{(\text{YD / YP})}$ is the ratio of the level of total confounding of the D,Y relationship, conditional on P, to the level of total confounding of the P,Y relationship, after re-scaling one of these biases to account for scale differences. $k_{(\text{NP / ND})}$ is the ratio of the level of total confounding of the P,N relationship to the level of total confounding of the D,N relationship, after re-scaling one of these biases to account for scale differences.

4.3 Sample selection and other applications

Arnold et al. (2016) suggests that placebos can be used to address sample selection and measurement bias in addition to unobserved confounding. This might look like Figure 5. Take Figure 5(a) as an example. Here selection of units into the study sample is indicated by a special node S. Since we are limited to the study sample, we are implicitly conditioning on S, which opens a non-causal path from D to Y. However, if sample selection also opens a similar non-causal path from D to N, then we may be able to use information about the placebo outcome N to inform how sample selection biases the estimate of the effect of D on Y. See Rohde and Hazlett (20XX) for a discussion of causal graphs and sample selection.

Figure 5: Example causal graphs with a placebo outcome, N. D is treatment; Y is outcome; Z contains unobserved confounders; S represents sample selection. Additional observed covariates are omitted for simplicity.

(a) (b)



5 Difference in differences as special case

In this section, we explore how the approaches put forth in this paper correspond to various relaxations to the standard assumptions of difference in differences identification strategies. Let's start by recalling the typical difference in differences (DID) setup and identification strategy. Suppose we have panel data for two periods for a cohort of units. We observe the outcome of interest in period 1, which we will label N, as well as the outcome of interest in period 2, which we will call Y. In period 1, no units are treated but in period 2 some units are treated. We indicate the group of units that is treated with $G = 1$. We indicate actual treatment with $D = 1$. We therefore have that in period 1 $D = 0$ for all units and that in period 2 $G = D$. For now, we assume that we do not need to adjust for any covariates for the DID approach to work.

The standard DID identification strategy identifies the ATT for Y (ATT_Y) in the following way. Note that Y_d and N_d are the potential outcomes when D is set, perhaps counterfactually, to $D = d$ with $d \in \{0, 1\}$. We start by revisiting a standard decomposition of the simple difference in means (DIM) between the treated and control groups for Y.

$$\begin{aligned}
\text{DIM}_Y &= \mathbb{E}[Y|G=1] - \mathbb{E}[Y|G=0] \\
&= \mathbb{E}[Y_1|G=1] - \mathbb{E}[Y_0|G=0] \text{ by consistency and since } D = G \text{ in period 2} \\
&= \underbrace{\mathbb{E}[Y_1|G=1] - \mathbb{E}[Y_0|G=1]}_{\text{ATT}_Y} + \underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[Y_0|G=0]}_{\text{Bias}_Y} \\
&= \text{ATT}_Y + \text{Bias}_Y
\end{aligned} \tag{32}$$

We make a "parallel trends" assumption that the trend in the mean control potential outcomes for the treated group is the same as the trend for the control group (i.e., $\mathbb{E}[Y_0|G=1] - \mathbb{E}[N_0|G=1] = \mathbb{E}[Y_0|G=0] - \mathbb{E}[N_0|G=0]$). The parallel trends assumption can be re-ordered into an "equi-confounding" assumption that the mean baseline difference between the treated and control groups in period 2 is the same as the mean baseline difference between the treated and control groups in period 1. The equivalence of parallel trends and equi-confounding is discussed in Sofer et al. (2016).

$$\begin{aligned}
\text{Parallel Trends} &\triangleq \underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[N_0|G=1]}_{\text{Trend}_{G=1}} = \underbrace{\mathbb{E}[Y_0|G=0] - \mathbb{E}[N_0|G=0]}_{\text{Trend}_{G=0}} \\
&\iff \underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[Y_0|G=0]}_{\text{Bias}_Y} = \underbrace{\mathbb{E}[N_0|G=1] - \mathbb{E}[N_0|G=0]}_{\text{Bias}_N} \triangleq \text{Equi-Confounding}
\end{aligned} \tag{33}$$

Now, we recall our constraint that N is a pre-treatment version of the outcome. This means that the treatment cannot have an effect on the pre-treatment outcome (and that no units are treated in period 1). So $N_1 = N_0 = N$ for both the treated and control groups. So we see that $\text{DIM}_N = \text{ATT}_N + \text{Bias}_N = \text{Bias}_N$.

$$\begin{aligned}
\text{DIM}_N &= \mathbb{E}[N|G=1] - \mathbb{E}[N|G=0] \\
&= \mathbb{E}[N_1|G=1] - \mathbb{E}[N_0|G=0] \\
&= \mathbb{E}[N_1|G=1] - \mathbb{E}[N_0|G=1] + \mathbb{E}[N_0|G=1] - \mathbb{E}[N_0|G=0] \\
&= \underbrace{\mathbb{E}[N_1 - N_0|G=1]}_{\text{ATT}_N} + \underbrace{\mathbb{E}[N_0|G=1] - \mathbb{E}[N_0|G=0]}_{\text{Bias}_N} \\
&= \mathbb{E}[N - N|G=1] + \text{Bias}_N = \text{Bias}_N
\end{aligned} \tag{34}$$

Therefore, we have that $\text{Bias}_Y = \text{Bias}_N = \text{DIM}_N$ and that $\text{DIM}_Y = \text{ATT}_Y + \text{Bias}_Y$. Plugging into the later, we have $\text{DIM}_Y = \text{ATT}_Y + \text{DIM}_N$, which can be rearranged to yield the standard difference in differences identification result, $\text{ATT}_Y = \text{DIM}_Y - \text{DIM}_N$, which makes clear where the DID name comes from.

Parallel Trends and Equi-Confounding While it is common to make parallel trends assumptions, equi-confounding assumptions are mathematically equivalent and really are not any less intuitive to consider. Equi-confounding assumes that the difference in mean control potential outcomes for the treated and control groups in period 2 is the same as the difference for the treated and control groups in period 1. This means that, on average, the way that the treatment and control groups differ for the outcome in period 1 can be used as a stand in for the way that the treatment and control groups differ for the outcome in period 2. This is a somewhat different assumption but researchers should be able to consider this and possible violations to it just as easily as the parallel trends assumptions and its violations. In what follows, we will stick to equi-confounding type assumptions, rather than parallel trends assumptions since they have a more direct connection to the bias that we are trying to overcome, namely Bias_Y. But every equi-confounding type assumption we might consider has analogous parallel trends assumption, as we will show.

Relaxing Assumptions Using what we've already seen, it is easy to see that we can relax the assumption that there is no effect of D on N by allowing ATT_N to be non-zero. Further, we can relax the assumption that Bias_Y exactly equals Bias_N by assuming that $\text{Bias}_Y = m_N \times \text{Bias}_N$. Together, these relaxations give us the expression for ATT_Y in Equation 35.

$$\begin{aligned}
\text{ATT}_Y &= \text{DIM}_Y - \text{Bias}_Y \\
&= \text{DIM}_Y - m_N \times \text{Bias}_N \\
&= \text{DIM}_Y - m_N \times (\text{DIM}_N - \text{ATT}_N)
\end{aligned} \tag{35}$$

That is, we can express ATT_Y as a difference between DIM_Y and an "adjusted" version of DIM_N, where the adjustment is parameterized by $m_N = \frac{\text{Bias}_Y}{\text{Bias}_N}$ and ATT_N. Our assumption that $\text{Bias}_Y = m_N \times \text{Bias}_N$ is similar to the equi-confounding assumption from the standard DID approach. And the expression $\text{ATT}_Y = \text{DIM}_Y - m_N \times (\text{DIM}_N - \text{ATT}_N)$ is also similar to a difference in differences type identification result. So we are not too far from familiar results but this approach would apply to more general placebo outcomes and settings outside the standard DID setting. Equation 35 could be used for partial identification in cases with binary treatments and a placebo outcome. Researchers would estimate DIM_Y and DIM_N and then reason about plausible values for m_N and ATT_N to partially identify ATT_Y.[20] But how does this all relate to the regression-based results we discuss in Section 2?

Connecting to Regression Since G is binary, a regression of Y or N on G will give the difference in means DIM_Y or DIM_N, respectively. Additionally, we can use traditional omitted variable bias analysis to show that,

$$\begin{aligned} \text{DIM}_Y &= \beta_{Y\sim G} = \beta_{Y\sim G|Z} + \beta_{Y\sim Z|G}\beta_{Z\sim G} \\ \text{DIM}_N &= \beta_{N\sim G} = \beta_{N\sim G|Z} + \beta_{N\sim Z|G}\beta_{Z\sim G} \end{aligned} \tag{36}$$

where Z is a linear combination of a vector of all the time-varying and non-time-varying confounders of the G, Y and G, N relationships. $\beta_{Y\sim G}$ is the regression coefficient on G from the regression $Y = \beta_{Y\sim G}G + \epsilon_1$. $\beta_{Y\sim G|Z}$ is the regression coefficient on G from the regression $Y = \beta_{Y\sim G|Z}G + Z\beta_{Y\sim Z|G} + \epsilon_2$. All the other regression coefficients are defined similarly. We can also see that Equation 37 holds.

$$\text{ATT}_Y + \text{Bias}_Y = \text{DIM}_Y = \beta_{Y\sim G} = \beta_{Y\sim G|Z} + \beta_{Y\sim Z|G}\beta_{Z\sim G} \tag{37}$$

Under the standard DID assumptions, we could write $\text{ATT}_Y = \beta_{Y\sim G} - \beta_{N\sim G}$, but we want to relax these assumptions. Under an assumption of effect homogeneity, or using OLS regression as an approximation, we have that $\text{ATT}_Y = \beta_{Y\sim G|Z}$, which means that $\text{Bias}_Y = \beta_{Y\sim Z|G}\beta_{Z\sim G}$. We can also see that, $\text{DIM}_N = \beta_{N\sim G}$ and so $\text{DIM}_N - \beta_{N\sim G|Z} = \beta_{N\sim G} - \beta_{N\sim G|Z} = \beta_{N\sim Z|G}\beta_{Z\sim G}$. Now we can see that $\beta_{Z\sim G} = \frac{\beta_{N\sim G} - \beta_{N\sim G|Z}}{\beta_{N\sim Z|G}}$. Plugging this into the expression for ATT_Y, we get Equation 38.

$$\begin{aligned} \text{ATT}_Y = \beta_{Y\sim G|Z} &= \beta_{Y\sim G} - \beta_{Y\sim Z|G}\beta_{Z\sim G} \\ &= \beta_{Y\sim G} - \beta_{Y\sim Z|G}\left[\frac{\beta_{N\sim G} - \beta_{N\sim G|Z}}{\beta_{N\sim Z|G}}\right] \\ &= \beta_{Y\sim G} - m_N \times (\beta_{N\sim G} - \beta_{N\sim G|Z}) \text{ where } m_N \triangleq \frac{\beta_{Y\sim Z|G}}{\beta_{N\sim Z|G}} \end{aligned} \tag{38}$$

Under the assumption of effect homogeneity for the effect of D on N or using OLS as an approximation, we see that $\beta_{N\sim G|Z} = \text{ATT}_N$, $m_N = \frac{\beta_{Y\sim Z|G}}{\beta_{N\sim Z|G}} = \frac{\beta_{Y\sim Z|G}\beta_{Z\sim G}}{\beta_{N\sim Z|G}\beta_{Z\sim G}} = \frac{\text{Bias}_Y}{\text{Bias}_N}$, and $\beta_{N\sim G} = \text{DIM}_N$. Thus, Equations 35 and 38 are equivalent. Note that we don't need to make the effect homogeneity for the effect of D on N for Equation 38 to be useful, however. We can simply parameterize the expression for ATT_Y with $m_N = \frac{\beta_{Y\sim Z|G}}{\beta_{N\sim Z|G}}$ and $\beta_{N\sim G|Z}$. Addtionally, if we do not want to make such an assumption for the effect of D on Y, we can make $\beta_{Y\sim G|Z}$ our inferential target and still make use of Equation 38. This framing connects us back to the framework laid out in Section 2.

We can also define $k_N \triangleq \frac{R_{Y\sim Z|G}}{R_{N\sim Z|G}}$ and can show that $m_N = \frac{\beta_{Y\sim Z|G}}{\beta_{N\sim Z|G}} = k_N \times \frac{\text{SD}(Y^{\perp G})}{\text{SD}(N^{\perp G})}$. If the scale of N and Y differ, it will be easier to reason about the value of k_N than to reason about the value of m_N. So we now can see how we can relax both the assumption that N is a pre-treatment outcome with no treatment effect from G and the assumption that we have perfect equi-confounding. We also have seen how the latter relaxation could be parameterized in terms of $\beta_{N\sim G|Z}$ and m_N or k_N, where k_N will often be easier to interpret. This all can be done while maintaining an identification strategy with roots that connect easily to the standard difference in differences assumptions and identification strategy.

Moreover, we can see standard difference in differences as the special case where we assume that $\beta_{Y\sim G|Z} = 0$ and $m_N = 1$ or equivalently $k_N = \frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}$. This value of k_N arising could be the result of simply a fortunate balancing coincidence

[20] We suspect there are more settings in which multiple applications of the decomposition DIM = ATT + Bias could yield useful expressions for causal effects with binary treatments, similar to how Section 2 involves multiple applications of the omitted variables bias formula. For example, another possibility is for a binary placebo treatment P with actual binary treatment D and outcome Y. We may have that

$$\underbrace{\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]}_{\text{DIM}_{Y\sim D}} = \underbrace{\mathbb{E}[Y_{D=1}|D=1] - \mathbb{E}[Y_{D=0}|D=1]}_{\text{ATT}_{Y\sim D}} + \underbrace{\mathbb{E}[Y_{D=0}|D=1] - \mathbb{E}[Y_{D=0}|D=0]}_{\text{Bias}_{Y\sim D}}$$

$$\underbrace{\mathbb{E}[Y|P=1] - \mathbb{E}[Y|P=0]}_{\text{DIM}_{Y\sim P}} = \underbrace{\mathbb{E}[Y_{P=1}|P=1] - \mathbb{E}[Y_{P=0}|P=1]}_{\text{ATT}_{Y\sim P}} + \underbrace{\mathbb{E}[Y_{P=0}|P=1] - \mathbb{E}[Y_{P=0}|P=0]}_{\text{Bias}_{Y\sim P}}$$

If we assume that $\text{Bias}_{Y\sim D} = m_P \times \text{Bias}_{Y\sim P}$ then we can write $\text{ATT}_{Y\sim D} = \text{DIM}_{Y\sim D} - m_P \times (\text{DIM}_{Y\sim P} - \text{ATT}_{Y\sim P})$. We could then partially identify $\text{ATT}_{Y\sim D}$ by estimating $\text{DIM}_{Y\sim D}$ and $\text{DIM}_{Y\sim P}$ and reasoning about plausible values for $m_P \triangleq \frac{\text{Bias}_{Y\sim D}}{\text{Bias}_{Y\sim P}}$ and $\text{ATT}_{Y\sim P}$.

for the ratios $\frac{R_{Y\sim Z|G}}{R_{N\sim Z|G}}$ and $\frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}$. But it is more likely to be interpreted as the scale of N and Y being the same (i.e., $\text{SD}(N^{\perp G}) = \text{SD}(Y^{\perp G})$) in addition to $R_{Y\sim Z|G} = R_{N\sim Z|G}$. In this way, the assumption that $k_N = 1$ can be thought of as less demanding than the assumption that $m_N = 1$. So we could consider a "new" form of difference in difference where we assume $k_N = 1$ but do not assume that $m_N = 1$.[21] However, the above framework need not be limited to any specific choice for k_N. Rather, we can consider a range of plausible values for k_N, perhaps including the value that would bring the true effect to zero under the assumption that $\beta_{N\sim G|Z} = 0$, $k_N = \frac{\text{DIM}_Y}{\text{DIM}_N} \times \frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}$, as well as $k_N = 1$ and $k_N = \frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}$.

Transforming unequal confounding assumptions into unequal trend assumptions Finally, we show that any assumption for m_N corresponds to an assumption on the trend in control potential outcomes. We will show that any assumption for m_N can be turned into an assumption about how the trends in control potential outcomes compare. An assumption about a difference in confounding would be something like Equation 39. An assumption about a difference in trend would be something like Equation 40.

$$\underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[Y_0|G=0]}_{\text{Bias}_Y} = m_N \times \underbrace{(\underbrace{\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0]}_{\text{DIM}_N} - \beta_{N\sim G|Z})}_{\text{Bias}_N} \tag{39}$$

$$\iff \mathbb{E}[Y_0|G=1] = \mathbb{E}[Y_0|G=0] + m_N \times (\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0] - \beta_{N\sim G|Z})$$

$$\underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[N|G=1]}_{\text{Trend}_{G=1}} = w_N \times \underbrace{(\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0])}_{\text{Trend}_{G=0}} \tag{40}$$

$$\iff \mathbb{E}[Y_0|G=1] = \mathbb{E}[N|G=1] + w_N \times (\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0])$$

For this exposition, we assume $\beta_{N\sim G|Z} = 0$ for simplicity. Both Equations 39 and 40 provide an expression for $\mathbb{E}[Y_0|G=1]$, which, in the DID setting, is unobserved while all the other components of the expressions are estimable from the data. We can set these two expressions for $\mathbb{E}[Y_0|G=1]$ equal to each other and then solve for w_N to get the trend assumption that corresponds to our assumption on m_N.

$$\mathbb{E}[Y_0|G=0] + m_N \times (\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0]) = \mathbb{E}[N|G=1] + w_N \times (\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0]) \tag{41}$$

$$\begin{aligned} \iff w_N &= \frac{\mathbb{E}[Y_0|G=0] + m_N \times (\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0]) - \mathbb{E}[N|G=1]}{\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0]} \\ \iff m_N &= \frac{\mathbb{E}[N|G=1] + w_N \times (\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0]) - \mathbb{E}[Y_0|G=0]}{\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0]} \end{aligned} \tag{42}$$

We could similarly write m_N in terms of w_N, if we preferred to make an assumption on w_N, the difference in the trend, and wanted to know what this implies about the unequal confounding. Additionally, we know that $m_N = k_N \times \frac{\text{SD}(Y^{\perp G})}{\text{SD}(N^{\perp G})}$ so that we can make an assumption on k_N and see what this implies about w_N, the trend, or make a trend assumption on w_N and see what this implies about k_N.

Additional Relaxations N could be some placebo outcome other than a pre-treatment outcome. It turns out that the framework above does not require that N is a pre-treatment outcome, just that it is some other placebo outcome that we expect to have somewhat similar confounding from Z as does Y. In this setting, while we want to understand the causal effect of G on Y, we leverage N, some outcome we do not directly care about. In general, G may have an effect on or some additional relationship with N and the confounding of G and N need not be the same as the confounding of G and Y. These relaxations are already covered above. Recall that G captures the group of units that is treated in period 2.

We also do not need to restrict ourselves to the panel setting in which we have data on the same units over time. We can have data on the same units with two different outcomes N and Y for the same time period or we can have data on completely separate sets of units, so long as we are able to estimate the regressions $Y = \hat{\beta}_{Y\sim D}D + \epsilon$ and $N = \hat{\beta}_{N\sim D}D + \xi$ (or obtain standard regression summary statistics from them). We can just think about the variable D as indicating whether or not a unit is treated at the relevant time period; we do not need to track the group of units that will be treated in a future time period.

Additionally, it turns out that, in our framework, we don't need to restrict ourselves to binary treatments so long as our inferential target is $\beta_{Y\sim D|Z}$. We can replace the difference in means in the above discussion with $\beta_{Y\sim D}$ and $\beta_{N\sim D}$. In the most relaxed case, all of the our discussion actually boils down to mechanical results for OLS regressions that will hold without any assumptions. We also allow for observed covariates to be included as well.

[21] This approach avoids the scale issues discussed in Sofer et al. (2016).

6 Application and software

Work in progress.

In this section, we introduce an `R` package that facilitates partial identification using imperfect placebos based on the approach discussed in this paper. The package is called `placeboLM` and documentation can be found at XXXX. We also demonstrate usage of this package in a simple application.

6.1 National Supported Work Demonstration

Let us first consider a familiar substantive example. The National Supported Work Demonstration (NSW) was a job training program aimed at helping disadvantaged worker build basic skills in the 1970s. The program randomly assigned participants to training positions. These consisted of a group that received the benefits of the program (the treatment group) and a group that did not receive any benefits (the control group). (LaLonde, 1986) The data also include a non-experimental dataset in which individuals not from the NSW program but from the Panel Study of Income Dynamics (PSID) are added as additional control units. The data that we work with is "a subset that consists of males with three (2 pre-training and 1 post-training) years of earnings data (the outcome of interest). This subset has been considered in (Dehejia and Wahba 1999; Smith and Todd 2005; Firpo 2007)." (Callaway, 2019) The data can be found in the "`qte`" `R` package. We use the non-experimental dataset (called `lalonde.psid`) to illustrate the methods discussed in this paper and then compare these results with the estimates from the experimental data (called `lalonde.exp`).[22]

We use 1974 earnings (`re74`, i.e. N) as a placebo outcome for 1978 earnings (`re78`, i.e. Y). The treatment (`treat`, i.e. D) is experiencing the job training program. We also use the set of covariates (X) in the data including age, education (years and degree or not), marital status, and demographic indicators. We recognize that there are unobserved covariates (Z) that we wish we could condition on but cannot. Ideally, we would like to estimate the regression `lm(re78 ~ treat + X + Z)`, which corresponds to a regression like $Y = \beta_{Y\sim D|Z,X} D + X\beta_{Y\sim X|D,Z} + Z\beta_{Y\sim Z|D,X} + \epsilon_l$. Due to the unobserved confounders, we will not be able to point identify $\beta_{Y\sim D|Z,X}$. Instead, take the partial identification approach proposed by the present paper and leverage information on 1974 earnings as a placebo outcome. Equation **??**, provides us with the following expression for $\beta_{Y\sim D|Z,X}$. We can get $\hat{\beta}_{Y\sim D|X}$, $\hat{\beta}_{N\sim D|X}$, $\text{se}(\hat{\beta}_{Y\sim D|X})$, $\text{se}(\hat{\beta}_{N\sim D|X})$, df_Y, and df_N from running these short regressions: `lm(re78 ~ treat + X)` and `lm(re74 ~ treat + X)`, which correspond to $Y = \beta_{Y\sim D|X} D + X\beta_{Y\sim X|D} + \epsilon_s$ and $N = \beta_{N\sim D|X} D + X\beta_{N\sim X|D} + \xi_s$, respectively.

$$\beta_{Y\sim D|Z,X} = \hat{\beta}_{Y\sim D|X} - k_N \times (\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|X})}{\text{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\text{df}_Y}{\text{df}_N}}$$

Since the placebo outcome is pre-treatment earnings (specifically 1974 earnings), it is likely that the treatment has no causal effect on the the placebo outcome. Therefore, we may want to consider that $\beta_{N\sim D|Z,X} = 0$. With the results from the estimated regressions, we can use the above expression to see how different values for k_N (and $\beta_{N\sim D|Z,X}$) translate to estimates for $\beta_{Y\sim D|Z,X}$. `placeboLM` does all of this for us, however. Basic use of `placeboLM` takes the following form.

```
# install.packages("devtools")
devtools::install_github("Adam-Rohde/placeboLM")
library(placeboLM)

data(lalonde,package = "qte")

plm = placeboLM(
    outcome = "re78",
    treatment = "treat",
    placebo_outcome = "re78",
    X = c("age", "education", "black", "hispanic", "married", "nodegree"),
    data = lalonde.psid)

placeboLM_summary(plm)
placeboLM_contour_plot(plm)
placeboLM_line_plot(plm, m_effect_D_on_N = 0)
```

[22] Since this data includes a randomized experiment, we can also estimate the effect using the experimental data to get an unbiased estimate of the true treatment effect. This experimental estimate can be used to back out the true value for k, assuming that $\beta_{N\sim D|Z,X} = 0$.

The last three commands result in the following table and the two plots in Figure 6. The table contains "perfect placebo" effect estimates for $\beta_{Y\sim D|X,Z}$. These include the selection on observables (SOO) estimate, which is equal to $\hat{\beta}_{Y\sim D|X}$. This estimate corresponds to an assumption that $k_N = 0$. The table also includes two DID estimates (one assuming $m = 1$ and one assuming $k = 1$; both assume that `m_effect_D_on_N` $= \beta_{N\sim D|Z,X} = 0$). The $m = 1$ DID estimate corresponds to standard DID estimation.

	SOO	Perfect Placebo DID m=1	Perfect Placebo DID k=1
Effect Estimate	-5928.11	1718	3115.306

Figure 6 contains two plots: a contour plot and a line plot. The contour plot shows how the effect estimate for $\beta_{Y\sim D|X,Z}$ changes as k_N and $\beta_{N\sim D|Z,X}$ change. We've noted that it is likely that $\beta_{N\sim D|Z,X} = 0$. The line plot shows us how the effect estimate for $\beta_{Y\sim D|X,Z}$ changes as k_N changes, under the assumption that $\beta_{N\sim D|Z,X} = 0$. The SOO estimate is negative. This is not the sign we would intuitively expect, since training should probably boost earnings. The $m = 1$ DID estimate, however, is positive and implicitly makes an assumption that the confounding in the post-treatment period is actually 0.715 times the confounding in the pre-treatment period, which gets us close to the (positive) experimental estimate of 1671.13. The $k = 1$ DID estimate is also positive and somewhat overestimates the experimental estimate. The experimental estimate implies that the "true" confounding in the post-treatment period is actually 0.706 times the confounding in the pre-treatment period.

Without knowledge of the experimental estimate, however, we might consider the line plot and make an argument that the the relationship between Z and `re78` probably takes the same sign as the relationship between Z and `re74`, since both outcomes are measures of earnings. This would put us on the right half of the line plot. Further, we might expect that the relative level of confounding is pretty similar, i.e. k_N is somewhere near 1; again reasoning that the outcomes are fairly similar. This analysis would lead us to consider effect estimates similar to the two DID estimates, which it turns out would get us close to the experimental estimate. Most importantly, we would likely hone in on a range of plausible effect estimates in which most or all effect estimate are positive, which is different from the naive SOO estimate. Moreover, we are not tied to a strict parallel trends or equi-confounding assumption and can consider a range of plausible levels of relative confounding.

7 Discussion

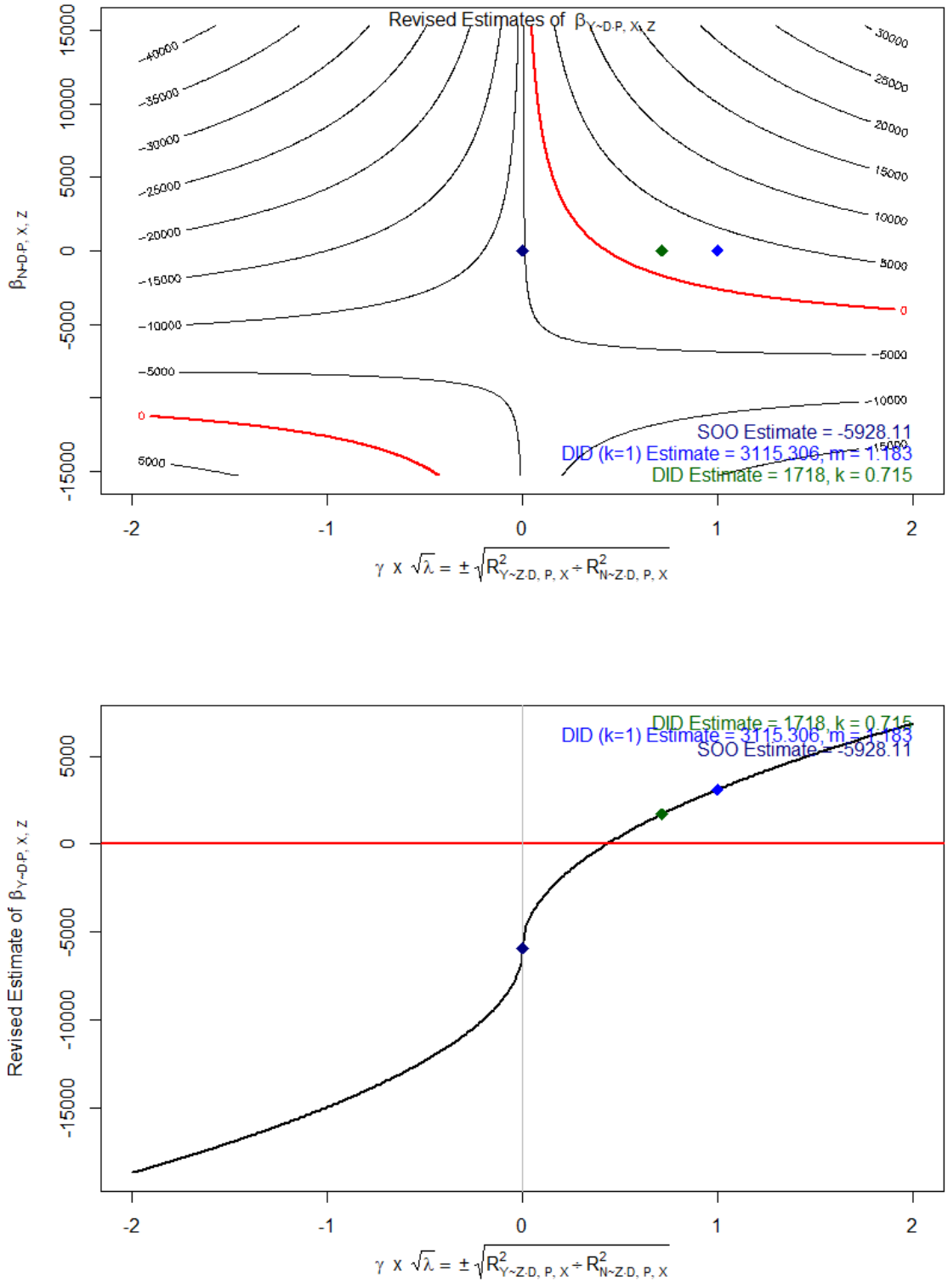
Work in progress.

Discussion of changing the philosophical approach to causal inference to phrase things as realistic assumptions first - ie start out assuming there are unobserved confounders (a realistic causal model); use all the data available (whether instruments, placebo outcomes or treatments, covariates, knowledge of discontinuities, and other knowledge based parametric assumptions); use all of these to hone in on what partial identification strategy might best suit your needs based on the information available; use this to bound or explore the range of plausible effects. But don't assume perfection and then just check how sensitive things are - stick to what is scientific / grounded first and then build in clearly assumptions or simplifications as you move toward more refined ranges of effect estimates.

7.1 Data Structures

It turns out that there are multiple data structures that are compatible with use of our proposed methods. Table 1 provides a simple summary. This table is meant to suggest various settings in which our framework can apply. It is possible that there are other settings in which it could be used. We wish to emphasize that placebos need not be pre-treatment measures of the main outcome of interest (i.e., the "actual outcome" as we say in the table). We might have placebo treatments or placebo outcomes that are different from the main outcome of interest but that suffer from comparable levels of confounding. We may have panel data or cross-sectional data. We may have a single sample of data that contains two time periods, two samples for two different time periods, a single sample for a single time period, or even just typical regression summary statistics. The key is that we have access to the inputs to our bias expressions. That is, we need $\hat{\beta}_{Y\sim D|X}$, $\hat{\beta}_{N\sim D|X}$, $\text{se}(\hat{\beta}_{Y\sim D|X})$, $\text{se}(\hat{\beta}_{N\sim D|X})$, df_Y, and df_N (or the analogous placebo treatment inputs). Whether these come from one sample, two samples, or just from regression output from existing studies, it does not matter, so long as we are able to reason about the plausible values for λ and $\beta_{N\sim D|Z,X}$. Reasoning about these may vary in difficulty depending on the data structure; but many data structures are compatible with our approach in principle.

Figure 6: Partial identification plots for National Supported Work example with 1974 earnings as placebo outcome.



References

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Table 1: Data structures where the proposed framework could apply

Data Structure	**Placebo**	**Panel vs Cross-Section**	**One vs Two Samples**	**Description**
Panel DID	pre-treatment actual outcome	panel	one	observe pre and post treatment actual outcome for same set of units
Cross-Sectional DID	pre-treatment actual outcome	cross-section	two (or regression summary statistics)	observe pre and post treatment actual outcome for different sets of units; could also simply observe typical regression summaries output rather than full set of observations
Panel Placebo	general placebo treatment or outcome	panel	one	observe post-treatment actual outcome and pre-treatment placebo for same set of units
Cross-Sectional Placebo	general placebo treatment or outcome	cross-section	one	observe actual outcome and placebo for same set of units
Two-Sample Placebo	general placebo treatment or outcome	cross-section	two (or regression summary statistics)	observe actual outcome and placebo for different sets of units; could also simply observe typical regression summaries output rather than full set of observations

Arnold, B. F. and Ercumen, A. (2016). Negative Control Outcomes: A Tool to Detect Bias in Randomized Trials. *JAMA*, 316(24):2597–2598.

Arnold, B. F., Ercumen, A., Benjamin-Chung, J., and John M. Colford, J. (2016). Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies. *Epidemiology*, 27(5):637–641.

Bilinski, A. and Hatfield, L. A. (2018). Nothing to see here? non-inferiority approaches to parallel trends and other model assumptions.

Callaway, B. (2019). *qte: Quantile Treatment Effects*. R package version 1.3.0.

Chernozhukov, V., Cinelli, C., Newey, W., Sharma, A., and Syrgkanis, V. (2022). Long story short: Omitted variable bias in causal machine learning. Working Paper 30302, National Bureau of Economic Research.

Cinelli, C., Forney, A., and Pearl, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research*, 0(0):00491241221099552.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):39–67.

Gibson, L. and Zimmerman, F. (2021). Measuring the sensitivity of difference-in-difference estimates to the parallel trends assumption. *Research Methods in Medicine & Health Sciences*, 2(4):148–156.

Hansen, B. (2022). *Econometrics*. Princeton University Press.

Keele, L. J., Small, D. S., Hsu, J. Y., and Fogarty, C. B. (2019). Patterns of effects and sensitivity analysis for differences-in-differences.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, 76(4):604–620.

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

Mastouri, A., Zhu, Y., Gultchin, L., Korba, A., Silva, R., Kusner, M., Gretton, A., and Muandet, K. (2021). Proximal causal learning with kernels: Two-stage estimation and moment restriction. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7512–7523. PMLR.

Matthay, E. C. and Glymour, M. M. (2020). A graphical catalog of threats to validity. *Epidemiology*, 31(3):376–384.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993.

Miao, W., Shi, X., and Tchetgen, E. T. (2020). A confounding bridge approach for double negative control inference on causal effects.

Necker, L. A. (1832). LXI. Observations on some remarkable optical phænomena seen in Switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid.

Pearl, J. (2009). *Causality*. Cambridge University Press, Cambridge.

Rambachan, A. and Roth, J. (2021). An honest approach to parallel trends.

Richardson, T. and Robins, J. (2013). Single world intervention graphs: a primer. *Working Paper, University of Washington, Seattle*.

Rohde, A. and Hazlett, C. (20XX). Revisiting sample selection as a threat to internal validity: New lessons, examples, and tools. *XXXX*, XX(XX):XXX–XXX.

Ryan, A. M., Kontopantelis, E., Linden, A., and James F Burgess, J. (2019). Now trending: Coping with non-parallel trends in difference-in-differences analysis. *Statistical Methods in Medical Research*, 28(12):3697–3711. PMID: 30474484.

Shi, X., Miao, W., and Tchetgen, E. (2020). A selective review of negative control methods in epidemiology. *Current Epidemiology Reports*, 7:190–202.

Sofer, T., Richardson, D. B., Colicino, E., Schwartz, J., and Tchetgen, E. J. T. (2016). On Negative Outcome Control of Unobserved Confounding as a Generalization of Difference-in-Differences. *Statistical Science*, 31(3):348 – 361.

Tchetgen, E. J. T., Ying, A., Cui, Y., Shi, X., and Miao, W. (2020). An introduction to proximal causal learning.

Ye, T., Chen, S., and Zhang, B. (2022). The role of placebo samples in observational studies.

Zhang, M. and Ding, P. (2022). Interpretable sensitivity analysis for the baron-kenny approach to mediation with unmeasured confounding.

A Unified placebo framework

[Work in progress.]

In this section, we show how the three approaches discussed in the main text for placebo outcomes, placebo treatments, and double placebos can be unified under in a single framework. To the careful reader, these connections should be expected. Suppose that we want to run the "long" regression of Y on D, X, and Z in Equation 43. In particular, we are interested in the coefficient $\beta_{Y\sim D|Z,X}$ as a measure of the causal effect of D on Y or as an approximation to the causal effect of D on Y.

$$Y = \beta_{Y\sim D|Z,X} D + X\beta_{Y\sim X|D,Z} + Z\beta_{Y\sim Z|D,X} + \epsilon_l \tag{43}$$

Sadly, we typically do not observe all the covariates that would allow us to estimate or approximate the causal effect of D on Y by running this long regression. Specifically, let us assume that Z captures the unobserved variables that we would have liked to include in the regression but cannot. So Z is a set of unobserved variables and we must run the "short" regression of Y on D and X in Equation 44, rather than the desired long regression.

$$Y = \beta_{Y\sim D|X} D + X\beta_{Y\sim X|D} + \epsilon_s \tag{44}$$

Now suppose we observe a placebo outcome, a placebo treatment, or both a placebo outcome and a placebo treatment that we believe suffer from similar levels of unobserved confounding as does our main treatment-outcome relationship of interest. How can we leverage the information we have about these placebos to make progress in addressing the unobserved confounding? We will start by assuming that we have both a placebo outcome and a placebo treatment. We will develop a partial identification framework for this "double placebo" setting. We will then show that when only a placebo outcome or only a placebo treatment is available, the framework will "reduce" to a "single placebo" partial identification framework.

We begin by considering some simple OLS regression results for two "short regressions." Consider the two sets of short and long regressions in Equations 45 and 46. These regress our two outcomes (actual Y and placebo N) on our two treatments (actual D and placebo P) as well as observed covariates, X, and the unobserved confounders Z. The short regressions omit Z. Suppose we estimate both of these short regressions (i.e., Equations 45b and 46b). Then traditional omitted variables bias analysis tells us that the four omitted variable bias expressions in Equations 47 hold.

$$Y = \beta_{Y\sim D|P,Z,X} D + \beta_{Y\sim P|D,Z,X} P + X\beta_{Y\sim X|D,P,Z} + Z\beta_{Y\sim Z|D,P,X} + \epsilon_{y,l} \tag{45a}$$
$$Y = \beta_{Y\sim D|P,X} D + \beta_{Y\sim P|D,X} P + X\beta_{Y\sim X|D,P} + \epsilon_{y,s} \tag{45b}$$

$$N = \beta_{N\sim D|P,Z,X} D + \beta_{N\sim P|D,Z,X} P + X\beta_{N\sim X|D,P,Z} + Z\beta_{N\sim Z|D,P,X} + \epsilon_{n,l} \tag{46a}$$
$$N = \beta_{N\sim D|P,X} D + \beta_{N\sim P|D,X} P + X\beta_{N\sim X|D,P} + \epsilon_{n,s} \tag{46b}$$

$$\widehat{\text{bias}_{YD}} = \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \tag{47a}$$
$$\widehat{\text{bias}_{YP}} = \hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X} = \beta_{Z\sim P|D,X}\beta_{Y\sim Z|D,P,X} \tag{47b}$$
$$\widehat{\text{bias}_{ND}} = \hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X} = \beta_{Z\sim D|P,X}\beta_{N\sim Z|D,P,X} \tag{47c}$$
$$\widehat{\text{bias}_{NP}} = \hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X} = \beta_{Z\sim P|D,X}\beta_{N\sim Z|D,P,X} \tag{47d}$$

We can rewrite Equations 47b and 47c in the following way. Equation 47b becomes Equation 48a; Equation 47c becomes Equation 48b. Using these, we can easily arrive at three expressions for $\widehat{\text{bias}_{YD}}$, one for the setting in which we have both a placebo outcome and a placebo treatment, one for the setting in which we only have a placebo outcome, and one for the setting in which we only have a placebo treatment.

$$\beta_{Y\sim Z|D,P,X} = [\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}) \tag{48a}$$
$$\beta_{Z\sim D|P,X} = (\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})[\beta_{N\sim Z|D,P,X}]^{-1} \tag{48b}$$

We start with the case in which we have both a placebo outcome and a placebo treatment. We simply plug both Equations 48b and 48a into the expression for $\widehat{\text{bias}_{YD}}$, followed by Equation 47d. In this way, we can write $\widehat{\text{bias}_{YD}}$ as in Equation 49. This is an expression containing estimated regression coefficients from the two short regressions ($\hat{\beta}_{Y\sim P|D,X}, \hat{\beta}_{N\sim D|P,X}$, and

$\hat{\beta}_{N\sim P|D,X}$) as well as three "partial identification parameters" ($\beta_{Y\sim P|D,Z,X}, \beta_{N\sim D|P,Z,X}$, and $\beta_{N\sim P|D,Z,X}$) that capture the extent to which N and P are imperfect placebos.

$$\begin{aligned}
\widehat{\text{bias}_{YD}} &= \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\
&= \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \\
&= \left[(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})[\beta_{N\sim Z|D,P,X}]^{-1}\right]\left[[\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})\right] \\
&= (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})[\beta_{N\sim Z|D,P,X}\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}) \\
&= (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})([\hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X}]^{-1})^{\top}(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})
\end{aligned} \tag{49}$$

This type of expression can be used in a wide variety of settings. Though the interpretation of the partial identification parameters and of $\beta_{Y\sim D|P,Z,X}$ will depend on the underlying causal model being considered. If we are in a setting like in Figure 2, where "perfect" placebos are available, then $\beta_{Y\sim P|D,Z,X}, \beta_{N\sim D|P,Z,X}$, and $\beta_{N\sim P|D,Z,X}$ will all equal zero and we will be able to point identify $\beta_{Y\sim D|P,Z,X}$. Moreover, $\beta_{Y\sim D|P,Z,X}$ would equal $\beta_{Y\sim D|Z,X}$, which was the coefficient of interest. However, if users believe they have perfect placebos, then non-parametric point identification results are available from the proximal causal inference literature. (Miao et al., 2020) But what can we do when we only have one placebo?

Let's consider the case where we only have a placebo outcome, N. In this case $P = \emptyset$. We plug only Equation 48b into the expression for $\widehat{\text{bias}_{YD}}$. In this way, we can also write $\widehat{\text{bias}_{YD}}$ as in Equation 50. This is an expression containing an estimated regression coefficient from the short regressions ($\hat{\beta}_{N\sim D|P,X}$) as well as two partial identification parameters. $\beta_{N\sim D|P,Z,X}$ captures the extent to which N is an imperfect placebo outcome, meaning it has a relationship with D conditional on X and Z. m_N compares the level of confounding of the Y, D relationship to the level of confounding of the N, D relationship; we will discuss m_N in depth shortly. If we are in a setting like Figure 1(a), where a perfect placebo outcome is available, then $\beta_{N\sim D|P,Z,X} = 0$ and we need only consider m_N as a partial identification parameter.

$$\begin{aligned}
\widehat{\text{bias}_{YD}} &= \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\
&= \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \\
&= \left[(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})[\beta_{N\sim Z|D,P,X}]^{-1}\right]\beta_{Y\sim Z|D,P,X} \\
&= (\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}) \times m_N \\
&\text{where } m_N \triangleq [\beta_{N\sim Z|D,P,X}]^{-1}\beta_{Y\sim Z|D,P,X}
\end{aligned} \tag{50}$$

Finally, let's consider the case where we only have a placebo treatment, P. In this case we simply cannot run the short regression with N as the outcome, since we have no placebo outcome N. We plug only Equation 48a into the expression for $\widehat{\text{bias}_{YD}}$. In this way, we can write $\widehat{\text{bias}_{YD}}$ as in Equation 51. This is an expression containing an estimated regression coefficient from the short regression ($\hat{\beta}_{Y\sim P|D,X}$) as well as two partial identification parameters. $\beta_{Y\sim P|D,Z,X}$ captures the extent to which N is an imperfect placebo outcome, meaning it has a relationship with D conditional on X and Z. m_P compares the level of confounding of the Y, D relationship to the level of confounding of the Y, P relationship; we will discuss m_P in depth shortly. If we are in a setting like Figure 1(b), where a perfect placebo treatment is available, then $\beta_{Y\sim P|D,Z,X} = 0$ and we need only consider m_P as a partial identification parameter.

$$\begin{aligned}
\widehat{\text{bias}_{YD}} &= \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\
&= \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \\
&= \beta_{Z\sim D|P,X}\left[[\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})\right] \\
&= m_P \times (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}) \\
&\text{where } m_P \triangleq \beta_{Z\sim D|P,X}[\beta_{Z\sim P|D,X}]^{-1}
\end{aligned} \tag{51}$$

As we have seen, the relationship between these three expressions for $\widehat{\text{bias}_{YD}}$ is simple. First, when we do not have P, the placebo treatment, we simply do not include it in the short regressions. When we do not have N, the placebo outcome, we simply do not estimate the short regression with N as the outcome. Second, in Equation 49, we substitute in using both 48b and 48a. In Equation 50, we substitute in using only 48b. In Equation 51, we substitute in using only 48a. Equations 49, 50 and 51 hold mechanically for two OLS regressions like our two short regressions.

When using placebo treatments, we may want to limit ourselves to settings in which $\beta_{Y\sim D|P,Z,X} = \beta_{Y\sim D|Z,X}$, where the latter was our original coefficient of interest. That is, it will be most productive to consider placebo treatments when P is not a descendant of D, D is not a descendant of P, P is a neutral control for the effect of D on Y, and P is a neutral control

for D on Y. By neutral controls, we mean that they do not change the interpretation of the regression coefficients when we include them. If a user is concerned that there could be a path opened or closed by conditioning on D or P, we suggest consulting your causal model to check. See Section **??** for further discussion.

We can use Equations 49, 50 and 51 for partial identification by establishing reasonable ranges of values for the "partial identification parameters." That is, we can partially identify $\beta_{Y\sim D|P,Z,X}$ (via our expression for $\widehat{\text{bias}_{YD}}$) by specifying plausible ranges of values for $\beta_{Y\sim P|D,Z,X}, \beta_{N\sim D|P,Z,X}$, and $\beta_{N\sim P|D,Z,X}$ based on subject matter expertise and knowledge external to the data, when both a placebo outcome and treatment are available. We can partially identify $\beta_{Y\sim D|P,Z,X}$ by specifying plausible ranges of values for $\beta_{N\sim D|P,Z,X}$ and m_N, when only a placebo outcome is available. Finally, we can partially identify $\beta_{Y\sim D|P,Z,X}$ by specifying plausible ranges of values for $\beta_{Y\sim P|D,Z,X}$ and m_P, when only a placebo treatment is available. $\beta_{Y\sim P|D,Z,X}, \beta_{N\sim D|P,Z,X}$, and $\beta_{N\sim P|D,Z,X}$ capture direct relationships between the placebo treatment and the actual outcome, between the placebo outcome and the actual treatment, and between the placebo treatment and the placebo outcome, conditional on Z and X. m_N compares the level of confounding of the Y, D relationship to the level of confounding of the N, D relationship. m_P compares the level of confounding of the Y, D relationship to the level of confounding of the Y, P relationship. With two imperfect placebos, users have a choice between the "single placebo" approaches that consider relative levels of confounding as a partial identification parameter and the "double placebo" approach discussed here in which the presence of a second placebo means that we can account for both "arms" of the unobserved confounder (i.e., $Z \to D$ and $Z \to Y$) using our placebos but still need to reason about the extent to which both placebos are imperfect.

B Placebo treatments

We now consider placebo treatments in more depth. Suppose that we want to run the long regression for Y on D (Equation 1) but, again, Z is an unobserved variable that we would have liked to include in the regression but cannot. Suppose also that we have a placebo treatment, P, for which we believe the confounding of P and Y from Z is similar to the confounding of D and Y from Z.

Suppose that P is not a descendant of D, D is not a descendant of P, P is a neutral control for the effect of D on Y, and P is a neutral control for D on Y.[23] Further, suppose that we want to run the "long" regression of Y on D, P, X, and Z in Equation 52. In particular, we are interested in the coefficient $\beta_{Y\sim D|P,Z,X}$ as a measure of the causal effect of D on Y or as an approximation to the causal effect of D on Y.

$$Y = \beta_{Y\sim D|P,Z,X} D + \beta_{Y\sim P|D,Z,X} P + X\beta_{Y\sim X|D,P,Z} + \beta_{Y\sim Z|D,P,X} Z + \epsilon_l \tag{52}$$

Since Z is unobserved, we must consider estimating the "short" regression of Y on D, P and X in Equation 53, rather than the desired long regression. Z is a variable that is "omitted" from the regression, though we wish we could have included it.

$$Y = \beta_{Y\sim D|P,X} D + \beta_{Y\sim P|D,X} P + X\beta_{Y\sim X|D,P} + \epsilon_s \tag{53}$$

Traditional omitted variable bias tells us that $\beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X}$ can be characterised as the product of two regression coefficients. $\beta_{Y\sim Z|D,P,X}$ captures the relationship between Y and Z after linearly adjusting for D, P and X. $\beta_{Z\sim D|P,X}$ captures the relationship between D and Z after linearly adjusting for P and X. But how can we leverage information about our placebo treatment P? The omitted variable bias framework also tells us that $\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}$ can similarly be characterised as the product of two regression coefficients.

$$\text{bias}_{(\text{YD.PX})} = \beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = \beta_{Y\sim Z|D,P,X}\beta_{Z\sim D|P,X} \tag{54}$$

$$\text{bias}_{(\text{YP.DX})} = \beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X} = \beta_{Y\sim Z|D,P,X}\beta_{Z\sim P|D,X} \tag{55}$$

It turns out that we can leverage information about the placebo treatment fairly easily by first rearranging the expression for $\text{bias}_{(\text{YP.DX})}$ to solve for $\beta_{Y\sim Z|D,P,X}$ and then plugging this into the expression for $\text{bias}_{(\text{YD.PX})}$. This can be thought of as using the placebo treatment to re-express the $Z \to Y$ "arm" of the confounding path $D \leftarrow Z \to Y$ in Figure 1(b).

$$\beta_{Y\sim Z|D,P,X} = \frac{\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}}{\beta_{Z\sim P|D,X}} \tag{56}$$

$$\implies \text{bias}_{(\text{YD.PX})} = \beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = \left[\frac{\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}}{\beta_{Z\sim P|D,X}}\right] \times \beta_{Z\sim D|P,X} = m_P \times \left(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \tag{57}$$

$$\iff \beta_{Y\sim D|P,Z,X} = \beta_{Y\sim D|P,X} - m_P \times \left(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \tag{58}$$

[23] See Section 4 for further discussion.

Now we have an expression for $\text{bias}_{(\text{YD.PX})}$ in terms of $m_P \triangleq \frac{\beta_{Z\sim D|P,X}}{\beta_{Z\sim P|D,X}} = \frac{\beta_{Y\sim Z|D,P,X}\beta_{Z\sim D|P,X}}{\beta_{Y\sim Z|D,P,X}\beta_{Z\sim P|D,X}} = \frac{\text{bias}_{(\text{YD.PX})}}{\text{bias}_{(\text{YP.DX})}}$, $\beta_{Y\sim P|D,Z,X}$, and $\beta_{Y\sim P|D,X}$. We have a similar expression for $\beta_{Y\sim D|P,Z,X}$, out target parameter, that also includes $\beta_{Y\sim D|P,X}$. We can estimate $\beta_{Y\sim P|D,X}$ and $\beta_{Y\sim D|P,X}$ from the data. The remaining parameters can be treated as partial identification parameters. We can use this for partial identification by establishing reasonable ranges of values for m_P and $\beta_{Y\sim P|D,Z,X}$. m_P captures the level of relative confounding or the relative relationship between the treatments and the unobserved confounder. This can be thought of as comparing the strength of the $Z \to D$ "arm" of the confounding path $D \leftarrow Z \to Y$ to the strength of the $Z \to P$ "arm" of the confounding path $P \leftarrow Z \to Y$ in Figure 1(b). Or we can consider it as comparing the strength of the entire $D \leftarrow Z \to Y$ relationship to the strength of the entire $P \leftarrow Z \to Y$ relationship. $\beta_{Y\sim P|D,Z,X}$ captures any direct causal relationship between P and Y. However, in many cases differing scales for D and P might lead to difficulty in interpreting or reasoning about the values for m_P. As in the case of placebo outcomes, we ask: can we do better than this?

Re-expressing m_P As before, let us consider a reparameterization of m_P in terms of scale-independent partial-correlation parameters, rather than regression coefficients. This will allow us to circumvent the scale issues that we worried about in the previous section.

$$
\begin{aligned}
m_P = \frac{\beta_{Z\sim D|P,X}}{\beta_{Z\sim P|D,X}} &= \frac{R_{D\sim Z|P,X}}{R_{P\sim Z|D,X}} \times \frac{\left[\frac{\text{SD}(Z^{\perp P,X})}{\text{SD}(D^{\perp P,X})}\right]}{\left[\frac{\text{SD}(Z^{\perp D,X})}{\text{SD}(P^{\perp D,X})}\right]} \times \frac{\text{SD}(Z^{\perp D,P,X})}{\text{SD}(Z^{\perp D,P,X}} \\
&= \frac{R_{D\sim Z|P,X}/\sqrt{1-R^2_{D\sim Z|P,X}}}{R_{P\sim Z|D,X}/\sqrt{1-R^2_{P\sim Z|D,X}}} \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \quad = \frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}} \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = k_P \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})}
\end{aligned}
\tag{59}
$$

Equation 59 provides an expression for m_P in terms of $k_P \triangleq \frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}}$[24][25] and $\frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})}$. The numerator and denominator of k_P are monotonic increasing functions of $R_{D\sim Z|P,X}$ and $R_{P\sim Z|D,X}$, respectively. These functions are non-linear and run to infinity as R^2 goes to 1. But their ratio should still be a fairly intuitive partial identification parameter, especially when we have rough ideas of the values for the partial correlations. $\frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})}$ captures the scale issues. The numerator and denominator are moments of the residuals from the two short regressions. k_P can be interpreted in two ways. As with placebo outcomes, we can reason either in terms of relative D, Z and P, Z relationships or in terms of relative levels of confounding.[26] Equations 57 and 58 can then be rewritten as Equation 60 and 61.

$$
\text{bias}_{(\text{YD.PX})} = k_P \times \left(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \tag{60}
$$

$$
\beta_{Y\sim D|P,Z,X} = \beta_{Y\sim D|P,X} - k_P \times \left(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \tag{61}
$$

Equation 61 provides us with a moment estimator for $\beta_{Y\sim D|Z,X}$ in Equation 62.[27]

$$
\hat{\beta}_{Y\sim D|P,Z,X} = \hat{\beta}_{Y\sim D|P,X} - k_P \times \left(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\|P^{\perp D,X}\|_2}{\|D^{\perp P,X}\|_2} \tag{62}
$$

Given k_P and $\beta_{Y\sim P|D,Z,X}$, the estimator $\hat{\beta}_{Y\sim D|P,Z,X}$ from Equation 62 is consistent for $\beta_{Y\sim D|P,Z,X}$ and asymptotically normal. We discuss computing standard errors in Section 2.4. Using Equation 62 in a partial identification framework, we can reason about plausible ranges of values for $\beta_{Y\sim P|D,Z,X}$ and k_P to arrive at different estimates of $\beta_{Y\sim D|P,Z,X}$.

[24] If f^2 is more intuitive, users can substitute in $k_P = \text{sign}(k_P) \times \sqrt{k_P^2}$ and then consider $\text{sign}(\frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}}) = \text{sign}(k_P)$ and $k_P^2 = \frac{f^2_{D\sim Z|P,X}}{f^2_{P\sim Z|D,X}}$.

[25] We need $f_{P\sim Z|D,X} \neq 0$ for k to be defined. This will be the case for any useful placebo treatment. Whether $f_{P\sim Z|D,X} \neq 0$ is only testable under the assumption that $\beta_{Y\sim P|D,Z,X} = 0$.

[26] See Appendix **??** for an alternative derivation.

[27] It is also possible to show that $\frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} = \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(Y^{\perp D,P,X})} = \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}}$, which is an expression containing only commonly reported regression summary statistics from the short regression that is estimable from the observed data. $\text{se}(\hat{\beta}_{Y\sim D|P,X})$ and $\text{se}(\hat{\beta}_{Y\sim P|D,X})$ are the estimated standard errors from standard regression estimation of the short regression. df_Y and df_N are the degrees of freedom from the estimated short regression, these will be equal in this case.

C Double placebo with multiple unobserved confounders

Let us assume we have the causal graph in Figure 4(b). We now want to assume that there are multiple unobserved confounders captured by the vector of variables $\mathbf{Z}$. We do not assume that the same linear combination of the components of $\mathbf{Z}$ will de-confound all the relationships we care about. $Z_{(\text{Y.DPX})} = \mathbf{Z}\beta_{Y\sim Z|D,P,X}$ is the linear combination that de-confounds the Y, D and the Y, P relationship. $Z_{(\text{N.DPX})} = \mathbf{Z}\beta_{N\sim Z|D,P,X}$ is the linear combination that de-confounds the N, D and the N, P relationship. Consider the following OVB expressions, where we use the relevant linear combination for each relationship.

$$\beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim Z_{(\text{Y.DPX})}|D,P,X}\beta_{Z_{(\text{Y.DPX})}\sim D|P,X} \tag{63a}$$
$$\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim Z_{(\text{Y.DPX})}|D,P,X}\beta_{Z_{(\text{Y.DPX})}\sim P|D,X} \tag{63b}$$
$$\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z_{(\text{N.DPX})},X} = \beta_{N\sim Z_{(\text{N.DPX})}|D,P,X}\beta_{Z_{(\text{N.DPX})}\sim D|P,X} \tag{63c}$$
$$\beta_{N\sim P|D,X} - \beta_{N\sim P|D,Z_{(\text{N.DPX})},X} = \beta_{N\sim Z_{(\text{N.DPX})}|D,P,X}\beta_{Z_{(\text{N.DPX})}\sim P|D,X} \tag{63d}$$

We can re-write the middle two of these as follows.

$$\beta_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} = \frac{\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}}{\beta_{Z_{(\text{Y.DPX})}\sim P|D,X}} \tag{64a}$$
$$\beta_{Z_{(\text{N.DPX})}\sim D|P,X} = \frac{\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z_{(\text{N.DPX})},X}}{\beta_{N\sim Z_{(\text{N.DPX})}|D,P,X}} \tag{64b}$$

Now we note that there are m_1, m_2, m_3, and m_4 that satisfy the following.

$$\beta_{Z_{(\text{Y.DPX})}\sim D|P,X} = m_{\text{A}} \times \beta_{Z_{(\text{N.DPX})}\sim D|P,X} \tag{65a}$$
$$\beta_{Z_{(\text{Y.DPX})}\sim P|D,X} = m_{\text{B}} \times \beta_{Z_{(\text{N.DPX})}\sim P|D,X} \tag{65b}$$

Then we can get the following expression for $\beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X}$.

$$\begin{aligned}
\beta_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} &= \beta_{Y\sim Z_{(\text{Y.DPX})}|D,P,X}\beta_{Z_{(\text{Y.DPX})}\sim D|P,X} \\
&= m_{\text{A}} \times \beta_{Y\sim Z_{(\text{Y.DPX})}|D,P,X}\beta_{Z_{(\text{N.DPX})}\sim D|P,X} \\
&= m_{\text{A}} \times \left[\frac{\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}}{\beta_{Z_{(\text{Y.DPX})}\sim P|D,X}}\right]\left[\frac{\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z_{(\text{N.DPX})},X}}{\beta_{N\sim Z_{(\text{N.DPX})}|D,P,X}}\right] \\
&= \frac{m_{\text{A}}}{m_{\text{B}}} \times \frac{(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X})(\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z_{(\text{N.DPX})},X})}{[\beta_{N\sim P|D,X} - \beta_{N\sim P|D,Z_{(\text{N.DPX})},X}]} \\
\therefore \beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} &= \beta_{Y\sim D|P,X} - \frac{m_{\text{A}}}{m_{\text{B}}} \times \frac{(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X})(\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z_{(\text{N.DPX})},X})}{(\beta_{N\sim P|D,X} - \beta_{N\sim P|D,Z_{(\text{N.DPX})},X})}
\end{aligned} \tag{66}$$

$$\begin{aligned}
\text{Now, } \frac{m_{\text{A}}}{m_{\text{B}}} &= \frac{\beta_{Z_{(\text{Y.DPX})}\sim D|P,X}}{\beta_{Z_{(\text{N.DPX})}\sim D|P,X}}\frac{\beta_{Z_{(\text{Y.DPX})}\sim P|D,X}}{\beta_{Z_{(\text{N.DPX})}\sim P|D,X}} \\
&= \frac{R_{Z_{(\text{Y.DPX})}\sim D|P,X}R_{Z_{(\text{N.DPX})}\sim P|D,X}}{R_{Z_{(\text{N.DPX})}\sim D|P,X}R_{Z_{(\text{Y.DPX})}\sim P|D,X}} \times \frac{\text{SD}(Z^{\perp P,X}_{(\text{Y.DPX})})}{\text{SD}(Z^{\perp P,X}_{(\text{N.DPX})})}\frac{\text{SD}(Z^{\perp D,X}_{(\text{N.DPX})})}{\text{SD}(Z^{\perp D,X}_{(\text{Y.DPX})})} \\
&= \frac{R_{Z_{(\text{Y.DPX})}\sim D|P,X}R_{Z_{(\text{N.DPX})}\sim P|D,X}}{R_{Z_{(\text{N.DPX})}\sim D|P,X}R_{Z_{(\text{Y.DPX})}\sim P|D,X}} \times \frac{\text{SD}(Z^{\perp P,X}_{(\text{Y.DPX})})}{\text{SD}(Z^{\perp D,P,X}_{(\text{Y.DPX})})}\frac{\text{SD}(Z^{\perp D,P,X}_{(\text{N.DPX})})}{\text{SD}(Z^{\perp P,X}_{(\text{N.DPX})})}\frac{\text{SD}(Z^{\perp D,X}_{(\text{N.DPX})})}{\text{SD}(Z^{\perp D,P,X}_{(\text{N.DPX})})}\frac{\text{SD}(Z^{\perp D,P,X}_{(\text{Y.DPX})})}{\text{SD}(Z^{\perp D,X}_{(\text{Y.DPX})})} \\
&= \frac{R_{Z_{(\text{Y.DPX})}\sim D|P,X}R_{Z_{(\text{N.DPX})}\sim P|D,X}}{R_{Z_{(\text{N.DPX})}\sim D|P,X}R_{Z_{(\text{Y.DPX})}\sim P|D,X}} \times \frac{\sqrt{1-R^2_{Z_{(\text{N.DPX})}\sim D|P,X}}\sqrt{1-R^2_{Z_{(\text{Y.DPX})}\sim P|D,X}}}{\sqrt{1-R^2_{Z_{(\text{Y.DPX})}\sim D|P,X}}\sqrt{1-R^2_{Z_{(\text{N.DPX})}\sim P|D,X}}} \\
&= \frac{f_{Z_{(\text{Y.DPX})}\sim D|P,X}f_{Z_{(\text{N.DPX})}\sim P|D,X}}{f_{Z_{(\text{N.DPX})}\sim D|P,X}f_{Z_{(\text{Y.DPX})}\sim P|D,X}} \\
\therefore \frac{m_{\text{A}}}{m_{\text{B}}} &= \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X}f_{Z_{(\text{Y.DPX})}\sim P|D,X}}{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X}f_{Z_{(\text{Y.DPX})}\sim P|D,X}} \times \frac{R_{N\sim Z_{(\text{N.DPX})}|D,P,X}f_{Z_{(\text{N.DPX})}\sim P|D,X}}{R_{N\sim Z_{(\text{N.DPX})}|D,P,X}f_{Z_{(\text{N.DPX})}\sim D|P,X}} \triangleq k_{(\text{YD / YP})} \times k_{(\text{NP / ND})}
\end{aligned} \tag{67}$$

Thus, finally, we arrive at the expression for $\beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X}$ below.

$$\beta_{Y\sim D|P,Z_{(\text{Y.DPX})},X} = \beta_{Y\sim D|P,X} - k_{(\text{YD / YP})} \times k_{(\text{NP / ND})} \times \frac{(\beta_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X})(\beta_{N\sim D|P,X} - \beta_{N\sim D|P,Z_{(\text{N.DPX})},X})}{(\beta_{N\sim P|D,X} - \beta_{N\sim P|D,Z_{(\text{N.DPX})},X})} \tag{68}$$

Some observations:

- When $Z_{(\text{Y.DPX})} = Z_{(\text{N.DPX})}$ or both $R_{Z_{(\text{Y.DPX})}\sim D|P,X} = R_{Z_{(\text{N.DPX})}\sim D|P,X}$ and $R_{Z_{(\text{Y.DPX})}\sim P|D,X} = R_{Z_{(\text{N.DPX})}\sim P|D,X}$, then $k_{(\text{YD / YP})} \times k_{(\text{NP / ND})} = 1$.
- $k_{(\text{YD / YP})}$ captures the scaled ratio of the level of confounding of the Y, D relationship (conditional on P) to the level of confounding of the Y, P relationship (conditional on D), after re-scaling one of these biases, or the ratio of bias factors.
- $k_{(\text{NP / ND})}$ captures the scaled ratio of the level of confounding of the N, P relationship (conditional on D) to the level of confounding of the N, D relationship (conditional on P), after re-scaling one of these biases, or the ratio of bias factors.
- $\beta_{Y\sim P|D,Z_{(\text{Y.DPX})},X}$ measures the causal of P on Y (conditional on D).
- $\beta_{N\sim D|P,Z_{(\text{N.DPX})},X}$ measures the causal of D on N (conditional on P).
- $\beta_{N\sim P|D,Z_{(\text{N.DPX})},X}$ measures the causal of P on N (conditional on D).
- We could also use $k_{(\text{YD / ND})} \times k_{(\text{NP / YP})} = \frac{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{Z_{(\text{Y.DPX})}\sim D|P,X}}{R_{N\sim Z_{(\text{N.DPX})}|D,P,X} f_{Z_{(\text{N.DPX})}\sim D|P,X}} \times \frac{R_{N\sim Z_{(\text{N.DPX})}|D,P,X} f_{Z_{(\text{N.DPX})}\sim P|D,X}}{R_{Y\sim Z_{(\text{Y.DPX})}|D,P,X} f_{Z_{(\text{Y.DPX})}\sim P|D,X}}$ and have similar interpretation.

D Partially linear and non-parametric framework

[Work in progress.]

If investigators are interested in partially linear of fully non-parametric approaches to estimating treatment effects, we can use the omitted variables frameworks from Chernozhukov et al. (2022) to arrive at partial identification frameworks similar to those presented in the main text for settings in which placebo outcomes are available. We do not demonstrate the partially linear or non-parametric case for placebo treatments, but in principle this is also possible.

D.1 Partially linear framework

An investigator may be interested in estimating a partially linear model.

Actual Outcome Y:

- Long Regression: $Y = \theta_{l,Y} D + f_l(X, Z) + \epsilon_l$
- Short Regression: $Y = \theta_{s,Y} D + f_s(X) + \epsilon_s$

Placebo Outcome N:

- Long Regression: $N = \theta_{l,N} D + g_l(X, Z) + \xi_l$
- Short Regression: $N = \theta_{s,N} D + g_s(X) + \xi_s$

Chernozhukov et al. (2022) show that omitted variable bias in partially linear setting can be bounded in the following way.

$$|\hat{\theta}_{s,Y} - \theta_{l,Y}|^2 = |\widehat{\text{bias}_Y}|^2 \leq S_Y^2 \times \eta^2_{Y\sim Z|D,X} \times \frac{\eta^2_{D\sim Z|X}}{1 - \eta^2_{D\sim Z|X}}$$

$$|\hat{\theta}_{s,N} - \theta_{l,N}|^2 = |\widehat{\text{bias}_N}|^2 \leq S_N^2 \times \eta^2_{N\sim Z|D,X} \times \frac{\eta^2_{D\sim Z|X}}{1 - \eta^2_{D\sim Z|X}}$$

S_Y^2 and S_N^2 are identified and therefore estimable from the data. $\eta^2_{Y\sim Z|D,X}$ and $\eta^2_{D\sim Z|X}$ are Pearson's correlation ratios (or the non-parametric R^2s).[28] $\eta^2_{Y\sim Z|D,X}$ is the proportion of residual variation in Y explained by Z. $\eta^2_{D\sim Z|X}$ is the proportion of residual variation in D explained by Z. See Chernozhukov et al. (2022) for further discussion of how to interpret partial η^2s and for more on S_Y^2 and S_N^2.

We can see that

$$\eta^2_{N\sim Z|D,X} \geq |\hat{\theta}_{s,N} - \theta_{l,N}|^2 \times \frac{1 - \eta^2_{D\sim Z|X}}{\eta^2_{D\sim Z|X}} \times \frac{1}{S_N^2}$$

If we then assume that $\eta^2_{Y\sim Z|D,X} = k \times |\hat{\theta}_{s,N} - \theta_{l,N}|^2 \times \frac{1-\eta^2_{D\sim Z|X}}{\eta^2_{D\sim Z|X}} \times \frac{1}{S_N^2}$ and plug into the expression for $|\widehat{\text{bias}_Y}|^2$, we get

[28] $\eta^2_{D\sim Z|X} = \frac{\text{Var}(\mathbb{E}[D|Z,X]) - \text{Var}(\mathbb{E}[D|X])}{\text{Var}(D) - \text{Var}(\mathbb{E}[D|X])} = \frac{\eta^2_{D\sim Z,X} - \eta^2_{D\sim X}}{1 - \eta^2_{D\sim X}}$. $\eta^2_{Y\sim Z|D,X}$ can be similarly interpreted.

$$\widehat{|\text{bias}_Y|}^2 \leq \left[k \times |\hat{\theta}_{s,N} - \theta_{l,N}|^2 \times \frac{1-\eta^2_{D\sim Z|X}}{\eta^2_{D\sim Z|X}} \times \frac{1}{S^2_N}\right] \times \frac{\eta^2_{D\sim Z|X}}{1-\eta^2_{D\sim Z|X}} \times S^2_Y$$
$$\iff \widehat{|\text{bias}_Y|} \leq \sqrt{k} \times |\hat{\theta}_{s,N} - \theta_{l,N}| \times \frac{S_Y}{S_N}$$

This is similar to our bias expression for the fully linear version. We have a bound on bias in terms of quantities that can be estimated from the data, k, and $\theta_{l,N}$. We know that $k = \frac{\eta^2_{Y\sim Z|D,X}}{|\hat{\theta}_{s,N}-\theta_{l,N}|^2 \times \frac{1-\eta^2_{D\sim Z|X}}{\eta^2_{D\sim Z|X}} \times \frac{1}{S^2_N}} \geq \frac{\eta^2_{Y\sim Z|D,X}}{\eta^2_{N\sim Z|D,X}} \triangleq k^*$. If we assume that $\widehat{|\text{bias}_N|}^2 = |\hat{\theta}_{s,N} - \theta_{l,N}|^2 = S^2_N \times \eta^2_{N\sim Z|D,X} \times \frac{\eta^2_{D\sim Z|X}}{1-\eta^2_{D\sim Z|X}}$, the worst-case level of bias for N, then we can reason in terms of k^*.

D.2 Non-parametric framework

An investigator may also be interested in estimating a linear functional of the conditional expectation function of the outcome in a fully non-parametric setting, like $\theta_{l,Y} = \mathbb{E}[Y_1 - Y_0] = \mathbb{E}[f_Y(1, X, W) - f_Y(0, X, W)]$ for a binary treatment D, where $Y_d = f_Y(d, X, W, U_Y)$ is the equation for Y in the structural causal model under intervention to set $D = d$. Again, the investigator is only able to estimate $\theta_{s,Y} = \mathbb{E}[f^*_Y(1, X) - f^*_Y(0, X)]$, where $f^*_Y(D, X) \triangleq \mathbb{E}[Y|D, X] = \mathbb{E}[f_Y(D, X, W)|D, X]$. We define $\theta_{l,N}$ and $\theta_{s,N}$ similarly for a placebo outcome.

Chernozhukov et al. (2022) also show that omitted variable bias in fully non-parametric setting can be bounded by an expression in terms of $\eta^2_{Y\sim Z|D,X}$ and a second term that, in the case of targeting $\theta_{l,Y} = \mathbb{E}[Y_1 - Y_0]$ with a binary treatment D, is the "average gain in the conditional precision with which we predict D by using Z in addition to X," which is somewhat similar to $\eta^2_{D\sim Z|X}$. Specifically, we have that

$$|\hat{\theta}_{s,Y} - \theta_{l,Y}|^2 = \widehat{|\text{bias}_Y|}^2 \leq S^2_Y \times \eta^2_{Y\sim Z|D,X} \times \frac{1-R^2_{\alpha\sim\alpha_s}}{R^2_{\alpha\sim\alpha_s}}$$
$$|\hat{\theta}_{s,N} - \theta_{l,N}|^2 = \widehat{|\text{bias}_N|}^2 \leq S^2_N \times \eta^2_{N\sim Z|D,X} \times \frac{1-R^2_{\alpha\sim\alpha_s}}{R^2_{\alpha\sim\alpha_s}}$$

S^2_Y and S^2_N are identified and therefore estimable from the data. See Chernozhukov et al. (2022) for discussion of S^2_Y, S^2_N, and $R^2_{\alpha\sim\alpha_s}$.

We can see that

$$\eta^2_{N\sim Z|D,X} \geq |\hat{\theta}_{s,N} - \theta_{l,N}|^2 \times \frac{R^2_{\alpha\sim\alpha_s}}{1-R^2_{\alpha\sim\alpha_s}} \times \frac{1}{S^2_N}$$

If we then assume that $\eta^2_{Y\sim Z|D,X} = k \times |\hat{\theta}_{s,N} - \theta_{l,N}|^2 \times \frac{R^2_{\alpha\sim\alpha_s}}{1-R^2_{\alpha\sim\alpha_s}} \times \frac{1}{S^2_N}$ and plug into the expression for $\widehat{|\text{bias}_Y|}^2$, we get

$$\widehat{|\text{bias}_Y|}^2 \leq \left[k \times |\hat{\theta}_{s,N} - \theta_{l,N}|^2 \times \frac{R^2_{\alpha\sim\alpha_s}}{1-R^2_{\alpha\sim\alpha_s}} \times \frac{1}{S^2_N}\right] \times \frac{1-R^2_{\alpha\sim\alpha_s}}{R^2_{\alpha\sim\alpha_s}} \times S^2_Y$$
$$\iff \widehat{|\text{bias}_Y|} \leq \sqrt{k} \times |\hat{\theta}_{s,N} - \theta_{l,N}| \times \frac{S_Y}{S_N}$$

This is again a similar bias expression. We have a bound on bias in terms of quantities that can be estimated from the data, k, and $\theta_{l,N}$. We know that $k = \frac{\eta^2_{Y\sim Z|D,X}}{|\hat{\theta}_{s,N}-\theta_{l,N}|^2 \times \frac{R^2_{\alpha\sim\alpha_s}}{1-R^2_{\alpha\sim\alpha_s}} \times \frac{1}{S^2_N}} \geq \frac{\eta^2_{Y\sim Z|D,X}}{\eta^2_{N\sim Z|D,X}} \triangleq k^*$. If we assume that $\widehat{|\text{bias}_N|}^2 = |\hat{\theta}_{s,N} - \theta_{l,N}|^2 = S^2_N \times \eta^2_{N\sim Z|D,X} \times \frac{1-R^2_{\alpha\sim\alpha_s}}{R^2_{\alpha\sim\alpha_s}}$, the worst-case level of bias for N, then we can reason in terms of k^*.