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 outcomes" (aka "negative outcome controls") or "placebo treatments". Traditional approaches employing such information typically focus on point identification and assume (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"). We show 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. Specifically, we provide a sensitivity analysis or 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. We also frame the conventional difference-in-difference approach as a special case of this, with a strict equi-confounding assumption regarding the pre-treatment outcome's relationship to the treatment group indicator, to which our approach allows several useful relaxations.

1 Introduction

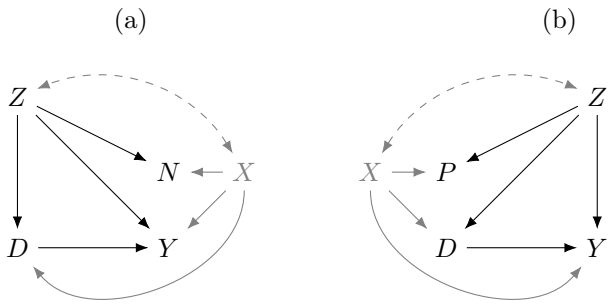
Unobserved confounding plagues many, if not most, attempts to draw causal inferences from observational data. Moreover, chance imbalances between the treatment and control groups in randomized experiments and sample selection and measurement error can further threaten the validity of causal inferences 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 placebo outcomes and treatments are shown in Figure 1. In this figure N is a placebo outcome and P is a placebo treatment. 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 outcomes and treatments. Arnold et al. (2016); Arnold and Ercumen (2016) discuss the potential to 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.

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

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

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 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. Finally, we frame the conventional difference-in-difference approach as a special case, 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 Proposal

Our proposed framework for using information from placebo outcomes and treatments relies on the omitted variable bias paradigm for regression analysis. However, we also provide relaxations to this parametric approach in Appendix C, 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 outcomes and treatments 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. Additionally, we note that, in all

the settings we discuss, one could ignore the placebos and directly reason about the strength of unobserved confounding of the main treatment-outcome relationship using the methods developed in Cinelli and Hazlett (2020) or Chernozhukov et al. (2022). But this would leave the information about the placebo unused. We start by discussing placebo outcomes, then discuss placebo treatments, and finally discuss "double placebos."[1]

2.1 Placebo outcomes

2.1.1 Placebo outcomes and the traditional omitted variables bias framework

We start our discussion of placebo outcomes by leveraging the traditional omitted variable bias framework. (Angrist and Pischke, 2008) Suppose that we want to run 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.

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

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. We will refer to these as unobserved confounders but note that they might not strictly be common causes of both D and Y. The motivation for the long regression as well as the interpretation of Z should be based on the use of causal models relating the variables of interest. So Z is a set of unobserved variables and we must run 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.

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

Traditional omitted variable bias tells us that the bias in using $\beta_{Y\sim D|X}$ to estimate $\beta_{Y\sim D|Z,X}$ can be characterised as the product of two regression coefficients. $\beta_{Y\sim Z|D,X}$ captures the relationship between Y and Z after linearly adjusting for D and X. $\beta_{Z\sim D|X}$ captures the relationship between D and Z after linearly adjusting for X.

$$\widehat{\text{bias}_Y} = \hat{\beta}_{Y\sim D|X} - \beta_{Y\sim D|Z,X} = \beta_{Y\sim Z|D,X}\beta_{Z\sim D|X} \tag{3}$$

Cinelli and Hazlett (2020) discuss this framework and extensions to it as a sensitivity analysis or partial identification strategy. But what if we also observe a placebo outcome, N? If we think that the relationship between D and N suffers from a similar level of confounding as does the relationship between D and Y, then how might be leverage N to understand $\widehat{\text{bias}_Y}$ and hence $\beta_{Y\sim D|Z,X}$? Consider a long and short regression for N as an outcome and again with D as the treatment. The long regression of N on D, X, and Z can be found in Equation 4.

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

The short regression of N on just D and X can be found in Equation 5.

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

As with Y, we are only able to estimate the short regression for N. The bias in using $\beta_{N\sim D|X}$ to estimate $\beta_{N\sim D|Z,X}$ can be characterised as in Equation 6.

$$\widehat{\text{bias}_N} = \hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X} = \beta_{N\sim Z|D,X}\beta_{Z\sim D|X} \tag{6}$$

It will often be useful to consider Z to be a rich enough set of variables to identify both the effect of D on Y and of D on N, but the mechanics of our proposed approach do not rely on this. Suppose that we go ahead and estimate both short regressions. It turns out that we can leverage information about the negative outcome fairly easily by first solving for $\beta_{N\sim Z|D,X}$, then assuming that $\beta_{Y\sim Z|D,X} = m \times \beta_{N\sim Z|D,X}$, and finally plugging into the expression for $\widehat{\text{bias}_Y}$.

[1] Any placebo could be phrased as either a placebo outcome or a placebo treatment in our framework. But 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. As we will see, this will result in our reasoning about how the actual outcome relates to the unobserved confounders relative to how the placebo outcome does. But we could consider the pre-treatment outcome as a placebo treatment on a different scale from the actual treatment and then reason about how the actual treatment relates to the unobserved confounders relative to how the placebo treatment does. The latter framing may be more difficult to reason about but it is possible.

$$\beta_{Y\sim Z|D,X} = m \times \beta_{N\sim Z|D,X} = m \times \frac{\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}}{\beta_{Z\sim D|X}} \tag{7}$$

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

Now we have an expression for $\widehat{\text{bias}_Y}$ in terms of $m = \frac{\beta_{Y\sim Z|D,X}}{\beta_{Y\sim Z|D,X}} = \frac{\beta_{Y\sim Z|D,X}\beta_{Z\sim D|X}}{\beta_{Y\sim Z|D,X}\beta_{Z\sim D|X}} = \frac{\widehat{\text{bias}_Y}}{\widehat{\text{bias}_N}}$, $\beta_{N\sim D|Z,X}$, and $\hat{\beta}_{N\sim D|X}$. We have already estimated $\hat{\beta}_{N\sim D|X}$ from the data and so we are left with a bias expression in terms of two parameters. We can use this for partial identification by establishing reasonable 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. $\beta_{N\sim D|Z,X}$ captures any direct causal relationship between D and N. However, in many cases differing scales for N and Y might lead to difficulty in interpreting or reasoning about the values for m. Different Z's might also have different scales making the using m difficult. Can we do better than this?

2.1.2 Re-expressing $\widehat{\text{bias}_Y}$

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 that we worried about in the previous section.

$$m = \frac{\beta_{Y\sim Z|D,X}}{\beta_{Y\sim Z|D,X}} = \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}} \times \frac{\left[\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(Z^{\perp D,X})}\right]}{\left[\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Z^{\perp D,X})}\right]} = \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \triangleq k \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \tag{9}$$

We can then plug Equation 9 into Equation 8 to arrive at a bias expression in terms of a ratio of scale-independent partial-correlation parameters.[2]

$$\widehat{\text{bias}_Y} = k \times \left(\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}\right) \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \tag{10}$$

Equation 10 looks similar to Equation 8, but k is not equal to m and we adjust for scale differences between N and Y. $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}$ captures the scale differences between N and Y and ameliorates the need to worry about the scale of Z by allowing us to reason about partial correlations. It is obvious that $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} = \frac{\left[\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(Z^{\perp D,X})}\right]}{\left[\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Z^{\perp D,X})}\right]}$. The numerator of the right side contains the transformation that turns $R_{Y\sim Z|D,X}$ into $\beta_{Y\sim Z|D,X}$; the denominator of the right side contains the transformation that turns $R_{N\sim Z|D,X}$ into $\beta_{N\sim Z|D,X}$. This is how we transformed m into k and, therefore, how we are able to reason about partial correlations rather than regression coefficients.

Finally, we can show that $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} = \frac{\left[\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(D^{\perp X})}\right]}{\left[\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(D^{\perp X})}\right]} = \frac{\text{se}(\hat{\beta}_{Y\sim D|X})}{\text{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\text{df}_Y}{\text{df}_N}}$, which is estimable from the observed data and is a function of commonly reported regression output. $\text{se}(\hat{\beta}_{Y\sim D|X})$ and $\text{se}(\hat{\beta}_{N\sim D|X})$ 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. Hence, $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}$ can be estimated easily from commonly reported regression output for the short regressions using available data only. Thus, we can also write $\widehat{\text{bias}_Y}$ as in Equation 11.

$$\widehat{\text{bias}_Y} = k \times (\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|X})}{\text{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\text{df}_Y}{\text{df}_N}} \tag{11}$$

Now we have an expression for $\widehat{\text{bias}_Y}$ where everything but $k = \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}$[3] and $\beta_{N\sim D|Z,X}$ can be estimated from the data. We can use this for partial identification by leveraging external information and subject matter expertise to determine plausible ranges of values for k and $\beta_{N\sim D|Z,X}$. As before, $\beta_{N\sim D|Z,X}$ captures any direct causal relationship between D and N. Since both N and D are observed variables, their relationship conditional on Z, X should be able to be specified as the

[2] See Appendix B.1 for an alternative derivation of Equation 10 that we will refer to in subsequent sections.

[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 $\beta_{N\sim D|Z,X} = 0$.

regression coefficient $\beta_{N\sim D|Z,X}$ despite scale considerations. k can be interpreted in more than one way. First, we note that k can be rewritten as in Equation 12. This means we can express bias as in Equations 13a and 13b.

$$k = \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}} = \text{sign}\left(\frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}\right) \times \frac{|R_{Y\sim Z|D,X}|}{|R_{N\sim Z|D,X}|} = \text{sign}\left(\frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}\right) \times \sqrt{\frac{R^2_{Y\sim Z|D,X}}{R^2_{N\sim Z|D,X}}} = \gamma\sqrt{\lambda} \tag{12}$$
$$\text{where } \gamma \triangleq \text{sign}\left(\frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}\right) \text{ and } \lambda \triangleq \frac{R^2_{Y\sim Z|D,X}}{R^2_{N\sim Z|D,X}}$$

$$\therefore \widehat{\text{bias}_Y} = \gamma\sqrt{\lambda} \times (\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|X})}{\text{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\text{df}_Y}{\text{df}_N}} \tag{13a}$$

$$\text{and } \widehat{|\text{bias}_Y|} = \sqrt{\lambda} \times |\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}| \times \frac{\text{se}(\hat{\beta}_{Y\sim D|X})}{\text{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\text{df}_Y}{\text{df}_N}} \tag{13b}$$

For the signed bias, $\widehat{\text{bias}_Y}$, we can reason about either k or both of γ and λ. For the unsigned bias, $\widehat{|\text{bias}_Y|}$, we only need to reason about λ. When reasoning about k, there are three equivalent interpretations. We can reason about $k = \frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}$, which is a ratio of signed partial correlations capturing the relative Y, Z and N, Z relationships. We can, however, also reason about $k = m \times \frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})} = \frac{\widehat{\text{bias}_Y}}{\widehat{\text{bias}_N}} \times \frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})}$, which is a ratio of the signed biases in terms of regression coefficients scaled by something that is estimable from the data. Finally, we could reason about $k = \frac{R_{Y\sim Z|D,X} f_{D\sim Z|X}}{R_{N\sim Z|D,X} f_{D\sim Z|X}}$, the relative level of confounding in terms of partial correlations (see Appendix B.1), which means we do not need to consider scale. $f_{D\sim Z|X}$ is Cohen's f which equals $\frac{R_{D\sim Z|X}}{\sqrt{1-R^2_{D\sim Z|X}}}$. When reasoning about λ, there are also three equivalent interpretations. We can reason about $\lambda = \frac{R^2_{Y\sim Z|D,X}}{R^2_{N\sim Z|D,X}}$, which is a ratio of partial R^2s, which lie between 0 and 1, capturing the relative Y, Z and N, Z relationships and represent the proportion of residual variation in the outcomes explained by Z. We can, however, also reason about $\lambda = \left(|m| \times \frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})}\right)^2 = \left(\frac{\widehat{|\text{bias}_Y|}}{\widehat{|\text{bias}_N|}} \times \frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})}\right)^2$, which is a ratio of the unsigned biases in terms of regression coefficients scaled by something that is estimable from the data. Finally, we could reason about $\lambda = \frac{R^2_{Y\sim Z|D,X} f^2_{D\sim Z|X}}{R^2_{N\sim Z|D,X} f^2_{D\sim Z|X}}$, the relative level of confounding in terms of partial R^2's (again see Appendix B.1), which means we do not need to consider scale. Whether we want to consider $\widehat{\text{bias}_Y}$ or $\widehat{|\text{bias}_Y|}$, we can reason either in terms of relative Y, Z and N, Z relationships or in terms of relative biases (i.e., relative levels of confounding).

The bias expression we just derived will hold for any set of variables, since it is just a mechanical property of omitting variables Z from ordinary least squares regression. However, when N is a useful placebo outcome, meaning that it suffers from comparable but not necessarily equal confounding with D as does Y and may have a direct causal relationship with D, this framework provides a simple, powerful, and flexible tool for partial identification of the causal effect of D on Y leveraging information about the placebo outcome. We can propose reasonable ranges of values for the relative level of confounding without worry about scale differences between Y and N and reasonable ranges of values for the direct causal relationships between D and N and in so doing arrive at a range of plausible estimates for $\beta_{Y\sim D|Z,X}$. These ranges of values for k (or λ) and $\beta_{N\sim D|Z,X}$ should be informed by subject matter expertise, previous studies, and a priori reasoning. See Cinelli and Hazlett (2020) for additional discussion about sensitivity analysis and reasoning in terms of partial R^2s.

2.1.3 How standard errors change

We might also be interested in how the standard errors change as the partial identification parameters change. This would allow us to see when effect estimates in the partial identification framework are statistically different from zero. It turns out this is fairly straightforward to do, but it comes at the cost of adding an additional partial identification parameter. Cinelli and Hazlett (2020), show the following relationship between the standard error for $\hat{\beta}_{Y\sim D|X}$ and the standard error for $\hat{\beta}_{Y\sim D|X,Z}$.

$$\text{se}(\hat{\beta}_{Y\sim D|X,Z}) = \text{se}(\hat{\beta}_{Y\sim D|X}) \sqrt{\frac{1 - R^2_{Y\sim Z|D,X}}{1 - R^2_{D\sim Z|X}} \times \frac{\text{df}_Y}{\text{df}_Y - 1}} \tag{14}$$

In Appendix B.1, we show that $R_{N\sim Z|D,X}$ can be expressed in terms of k, $\beta_{N\sim D|Z,X}$, and $R_{D\sim Z|X}$. We can plug this into our assumption that $R^2_{Y\sim Z|D,X} = \lambda \times R^2_{N\sim Z|D,X}$, see Equation 15, and then plug into Equation 14.

$$R^2_{Y\sim Z|D,X} = \frac{\lambda}{\text{df}_N} \times \left(\frac{\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}}{\text{se}(\hat{\beta}_{N\sim D|X})} \right)^2 \times \frac{1 - R^2_{D\sim Z|X}}{R^2_{D\sim Z|X}} \tag{15}$$

$$\implies \text{se}(\hat{\beta}_{Y\sim D|X,Z}) = \text{se}(\hat{\beta}_{Y\sim D|X}) \sqrt{\frac{\text{df}_Y}{\text{df}_Y - 1} \times \left[\frac{1}{1 - R^2_{D\sim Z|X}} - \frac{1}{R^2_{D\sim Z|X}} \times \frac{\lambda}{\text{df}_N} \times \left(\frac{\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}}{\text{se}(\hat{\beta}_{N\sim D|X})} \right)^2 \right]} \tag{16}$$

As we can see, the $R^2_{D\sim Z|X}$ terms do not cancel out in Equation 16 as they do in Appendix B.1. Instead, $R^2_{D\sim Z|X}$ will need to be added as a third sensitivity parameter in addition to λ and $\beta_{N\sim D|Z,X}$. $R^2_{D\sim Z|X}$ is the proportion of residual variation in D that is explained by Z. This should not be terribly burdensome since this is a natural parameter in omitted variable bias (see Appendix B.1). Fortunately, all other terms are either already partial identification parameters or are estimable from the observed data.

2.1.4 Sensitivity analysis and robustness values

While we favor a partial identification perspective, we also recognize that many researchers will be most comfortable starting from traditional placebo outcome point identification assumptions and then considering how sensitive their effect estimates are to violations of those assumptions. An effect estimate based on point identification assumptions would have the form of Equation 17, where the specific values γ^*, λ^*, and $\beta^*_{N\sim D|Z,X}$ are implied by the point identification assumptions (or are the assumptions themselves). In Equation 17, $\hat{\beta}_{Y\sim D|Z,X}$ (note the $\hat{}$) is the researcher's estimate of $\beta_{Y\sim D|Z,X}$, which may be biased and not equal to $\beta_{Y\sim D|Z,X}$.

$$\text{Effect Estimate} = \hat{\beta}_{Y\sim D|Z,X} = \hat{\beta}_{Y\sim D|X} - \gamma^* \sqrt{\lambda^*} \times (\hat{\beta}_{N\sim D|X} - \beta^*_{N\sim D|Z,X}) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|X})}{\text{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\text{df}_Y}{\text{df}_N}} \tag{17}$$

For example, a simple selection on observables (SOO) effect estimate would be $\hat{\beta}_{Y\sim D|X}$, which corresponds to an assumption that $\lambda^* = 0$ (i.e., no unobserved confounding). A traditional difference in differences (DID) effect estimate would be $\hat{\beta}_{Y\sim D|X} - \hat{\beta}_{N\sim D|X}$, which corresponds to an assumption that $\lambda^* = \left(\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})} \right)^2 = \left(\frac{\text{se}(\hat{\beta}_{N\sim D|X})}{\text{se}(\hat{\beta}_{Y\sim D|X})} \times \sqrt{\frac{\text{df}_N}{\text{df}_Y}} \right)^2$, $\gamma^* = +1$, and $\beta^*_{N\sim D|Z,X} = 0$. Researchers might also consider a version of DID where we assume $\lambda^* = 1$. See Section 5 for more discussion of how DID is a special case of our approach.

From Equation 17, it's easy to see that we can find the value for λ or $\beta_{N\sim D|Z,X}$ that would mean the true effect is equal to $100 \times q\%$ times the estimated effect. This would happen if Equation 18 holds, since then the true effect would equal $\beta_{Y\sim D|Z,X} = \hat{\beta}_{Y\sim D|X} - [\hat{\beta}_{Y\sim D|X} - (1-q) \times \hat{\beta}_{Y\sim D|Z,X}] = (1-q) \times \hat{\beta}_{Y\sim D|Z,X}$.

$$\hat{\beta}_{Y\sim D|X} - (1-q) \times \hat{\beta}_{Y\sim D|Z,X} = \gamma \sqrt{\lambda} \times (\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|X})}{\text{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\text{df}_Y}{\text{df}_N}} \tag{18}$$

The values for λ and $\beta_{N\sim D|Z,X}$ that mean the true effect is equal to $100 \times q\%$ times the estimated effect, λ_q and β_q, while allowing the other parameters to remain at their assumed levels are shown in Equations 19.[4]

$$\lambda_q = \left(\frac{\hat{\beta}_{Y\sim D|X} - (1-q) \times \hat{\beta}_{Y\sim D|Z,X}}{\hat{\beta}_{N\sim D|X} - \beta^*_{N\sim D|Z,X}} \times \frac{\text{se}(\hat{\beta}_{N\sim D|X})}{\text{se}(\hat{\beta}_{Y\sim D|X})} \times \sqrt{\frac{\text{df}_N}{\text{df}_Y}} \right)^2 \triangleq \text{RV}_{\lambda,q} \tag{19a}$$

$$\beta_q = \hat{\beta}_{N\sim D|X} - \frac{\hat{\beta}_{Y\sim D|X} - (1-q) \times \hat{\beta}_{Y\sim D|Z,X}}{\gamma^* \sqrt{\lambda^*}} \times \frac{\text{se}(\hat{\beta}_{N\sim D|X})}{\text{se}(\hat{\beta}_{Y\sim D|X})} \times \sqrt{\frac{\text{df}_N}{\text{df}_Y}} \triangleq \text{RV}_{\beta,q} \tag{19b}$$

We call these values "robustness values," and label them $\text{RV}_{\lambda,q}$ and $\text{RV}_{\beta,q}$, since they capture how vulnerable an effect estimate is to violations of it's point identification assumptions. For example, let us consider $\text{RV}_{\lambda,0}$, which is the value for λ that would mean the true effect is zero. If the assumed value λ^* is very close to $\text{RV}_{\lambda,0}$, then the effect estimate may be vulnerable slight violations to the assumption that $\lambda = \lambda^*$, which would mean the true effect might be null. However, such judgements of closeness are inherently context specific; and there is no value for $\text{RV}_{\lambda,0}$ or set of values $\{\text{RV}_{\lambda,0}, \lambda^*\}$ from which a researcher can be confident that their effect estimate is robust to violations of their point identification assumptions,

[4] γ^* drops out of Equation 19a because it is squared.

regardless of context. Researchers can, however, compare how plausible values for the true value for λ, which we will denote as $\lambda_{\text{plausible}}$, compare to $\text{RV}_{\lambda,0}$. $\lambda_{\text{plausible}}$ captures the violations to the assumption that $\lambda = \lambda^*$ that researchers cannot rule out. If the range of $\lambda_{\text{plausible}}$ values includes $\text{RV}_{\lambda,0}$, then the effect estimate may not be robust to violations to the assumption that $\lambda = \lambda^*$ that would mean the true effect might be null. Similar discussion also applies for $\text{RV}_{\beta,0}$ and violations to the assumption that $\beta_{N\sim D|Z,X} = \beta^*$ as well as to other values for q.

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 bias expressions for placebo treatment partial identification framework here. Additional detail and derivation, as well as expressions for how standard errors change and discussion of sensitivity analysis and robustness values, can be found in Appendix A.

Suppose that we want to run the long regression for Y on D (Equation 1) but, again, Z captures the unobserved variables 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, the approach is not easily adapted to imperfect placebo treatments (i.e., those for which there is a direct causal relationship between P and Y).[5]

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.[6] Further, suppose that we want to run the "long" regression of Y on D, P, X, and Z in Equation 20. 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 21, rather than the desired long regression.

$$Y = \beta_{Y\sim D|P,Z,X} D + \beta_{Y\sim P|D,Z,X} P + X\beta_{Y\sim X|D,P,Z} + Z\beta_{Y\sim Z|D,P,X} + \epsilon_l \tag{20}$$

$$Y = \beta_{Y\sim D|P,X} D + \beta_{Y\sim P|D,X} P + X\beta_{Y\sim X|D,P} + \epsilon_s \tag{21}$$

Following a similar omitted variables based approach as for placebo outcomes, we can arrive at the expressions for the bias in using $\hat{\beta}_{Y\sim D|P,X}$ as an estimate for $\beta_{Y\sim D|P,Z,X}$ found in Equations 22a and 22b, where $\gamma \triangleq \text{sign}\left(\frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}}\right)$, $\lambda \triangleq \frac{f^2_{D\sim Z|P,X}}{f^2_{P\sim Z|D,X}}$, and $f = R/\sqrt{1-R^2}$ is Cohen's f. See Appendix A for details of the derivation.

$$\widehat{\text{bias}_D} = \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = \gamma\sqrt{\lambda} \times \left(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}} \tag{22a}$$

$$|\widehat{\text{bias}_D}| = |\hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X}| = \sqrt{\lambda} \times |\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}| \times \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}} \tag{22b}$$

These are expressions for $\widehat{\text{bias}_D} = \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X}$ where everything but $k = \gamma\sqrt{\lambda} = \frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}} = \frac{R_{Y\sim Z|D,P,X} f_{D\sim Z|P,X}}{R_{Y\sim Z|D,P,X} f_{P\sim Z|D,X}}$ [7] and $\beta_{Y\sim P|D,Z,X}$ can be estimated from the data. We can use this for partial identification by leveraging external information and subject matter expertise to determine plausible ranges of values for k (or γ and λ) 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. For the signed bias, $\widehat{\text{bias}_D}$, we can reason about either k or both of γ and λ. For the unsigned bias, $|\widehat{\text{bias}_D}|$, we only need to reason about λ. As with placebo outcomes, whether we want to consider $\widehat{\text{bias}_D}$ or $|\widehat{\text{bias}_D}|$, we can reason either in terms of relative D, Z and P, Z relationships or in terms of relative biases (i.e., relative levels of confounding). See Appendix A for expressions for how standard errors change and discussion of sensitivity analysis and robustness values.

[5] Suppose we have a causal model like Figure 3(d). $\beta_{Y\sim Z|P,X}$ (or $R^2_{Y\sim Z|P,X}$) 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.

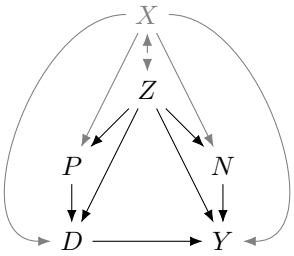
[6] We require that D and P are neutral controls for each other in that they do not change the regression coefficients for the outcome Y when we include them. This will simplify interpretation. If a user is concerned that there could be a path opened or closed by conditioning on D or P, we suggest drawing the DAG to check. See Section 4.1 for further discussion.

[7] 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 $\beta_{Y\sim P|D,Z,X} = 0$.

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 develop an additional partial identification framework in which we do not need to reason about relative levels of confounding between the placebos and the main variables but can instead just reason about the extent to which the placebos are "imperfect" (i.e., to what extent does the placebo treatment affect the main outcome of interest, as well as the placebo outcome, and to what extend is the placebo outcome affected by the main treatment of interest). With a "perfect" placebo treatment and outcome, 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 set up (Miao et al., 2020) but where we just use regression.[8]

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



First consider the following two sets of short and long regressions. Suppose we estimate both of these short regressions (i.e., Equations 23b and 23d). Then we arrive at the four omitted variable bias expressions in Equations 24. We can consider the terms on the right hand side of Equations 24 to be vectors of coefficients, when there are multiple components of Z. For example, $\beta_{Z\sim D|P,X} = \begin{bmatrix}\beta_{Z_1\sim D|P,X} & \dots & \beta_{Z_j\sim D|P,X}\end{bmatrix}$ and $\beta_{Y\sim Z|D,P,X} = \begin{bmatrix}\beta_{Y\sim Z_1|D,P,X} \\ \vdots \\ \beta_{Y\sim Z_j|D,P,X}\end{bmatrix}$, when there are j components of Z.

$$Y = \beta_{Y\sim D|P,Z,X}D + \beta_{Y\sim P|D,Z,X}P + X\beta_{Y\sim X|D,P,Z} + Z\beta_{Y\sim Z|D,P,X} + \epsilon_{y,l} \tag{23a}$$
$$Y = \beta_{Y\sim D|P,X}D + \beta_{Y\sim P|D,X}P + X\beta_{Y\sim X|D,P} + \epsilon_{y,s} \tag{23b}$$
$$N = \beta_{N\sim D|P,Z,X}D + \beta_{N\sim P|D,Z,X}P + X\beta_{N\sim X|D,P,Z} + Z\beta_{N\sim Z|D,P,X} + \epsilon_{n,l} \tag{23c}$$
$$N = \beta_{N\sim D|P,X}D + \beta_{N\sim P|D,X}P + X\beta_{N\sim X|D,P} + \epsilon_{n,s} \tag{23d}$$

$$\widehat{\text{bias}_{YD}} = \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \tag{24a}$$
$$\widehat{\text{bias}_{YP}} = \hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X} = \beta_{Z\sim P|D,X}\beta_{Y\sim Z|D,P,X} \tag{24b}$$
$$\widehat{\text{bias}_{ND}} = \hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X} = \beta_{Z\sim D|P,X}\beta_{N\sim Z|D,P,X} \tag{24c}$$
$$\widehat{\text{bias}_{NP}} = \hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X} = \beta_{Z\sim P|D,X}\beta_{N\sim Z|D,P,X} \tag{24d}$$

We can rewrite these bias expressions in the following way. Equation 24a becomes Equation 25a; Equation 24b becomes Equation 25b; Equation 24c becomes Equation 25c; Equation 24d remains the same but is repeated in Equation 25d.

$$\beta_{Y\sim D|P,Z,X} = \hat{\beta}_{Y\sim D|P,X} - \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \tag{25a}$$
$$\beta_{Y\sim Z|D,P,X} = [\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}) \tag{25b}$$
$$\beta_{Z\sim D|P,X} = (\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})[\beta_{N\sim Z|D,P,X}]^{-1} \tag{25c}$$
$$\hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X} = \beta_{Z\sim P|D,X}\beta_{N\sim Z|D,P,X} \tag{25d}$$

[8] This approach is also related to Sewall Wright's path tracing approach for linear models.

From these, we can write $\widehat{\text{bias}_{YD}}$ and, therefore, $\beta_{Y\sim D|P,Z,X}$ as in Equation 26. This is an expression containing estimated regression coefficients from the two short regressions ($\hat{\beta}_{Y\sim D|P,X}, \hat{\beta}_{Y\sim P|D,X}, \hat{\beta}_{N\sim D|P,X}$, 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. Equations 25 and 26 hold mechanically for OLS regression.

$$\begin{aligned}
\widehat{\text{bias}_{YD}} &= \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\
&= \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \\
&= \left[(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})[\beta_{N\sim Z|D,P,X}]^{-1}\right]\left[[\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})\right] \\
&= (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})[\beta_{N\sim Z|D,P,X}\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}) \\
&= (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})([\hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X}]^{-1})^{\top}(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}) \\
\therefore \beta_{Y\sim D|P,Z,X} &= \hat{\beta}_{Y\sim D|P,X} - (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})([\hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X}]^{-1})^{\top}(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})
\end{aligned} \tag{26}$$

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) 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.

3 Alternate Proposal

[This is a possible alternative approach to the proposal section.]

We start our discussion of placebos by leveraging the traditional omitted variable bias framework. (Angrist and Pischke, 2008) Suppose that we want to run the "long" regression of Y on D, X, and Z in Equation 27. 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.

$$Y = \beta_{Y\sim D|Z,X}D + X\beta_{Y\sim X|D,Z} + Z\beta_{Y\sim Z|D,X} + \epsilon_l \tag{27}$$

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. We will refer to these as unobserved confounders but note that they might not strictly be common causes of both D and Y. The motivation for the long regression as well as the interpretation of Z should be based on the use of causal models relating the variables of interest. So Z is a set of unobserved variables and we must run the "short" regression of Y on D and X in Equation 28, rather than the desired long regression. Z is a variable that is "omitted" from the regression, though we wish we could have included it.

$$Y = \beta_{Y\sim D|X}D + X\beta_{Y\sim X|D} + \epsilon_s \tag{28}$$

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 29 and 30. 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 29b and 30b). Then traditional omitted variables bias analysis tells us that the four omitted variable bias expressions in Equations 31 hold.[9]

[9]We can consider the terms on the right hand side of Equations 31 to be vectors of coefficients, when there are multiple components of Z. For

$$Y = \beta_{Y\sim D|P,Z,X} D + \beta_{Y\sim P|D,Z,X} P + X\beta_{Y\sim X|D,P,Z} + Z\beta_{Y\sim Z|D,P,X} + \epsilon_{y,l} \tag{29a}$$
$$Y = \beta_{Y\sim D|P,X} D + \beta_{Y\sim P|D,X} P + X\beta_{Y\sim X|D,P} + \epsilon_{y,s} \tag{29b}$$

$$N = \beta_{N\sim D|P,Z,X} D + \beta_{N\sim P|D,Z,X} P + X\beta_{N\sim X|D,P,Z} + Z\beta_{N\sim Z|D,P,X} + \epsilon_{n,l} \tag{30a}$$
$$N = \beta_{N\sim D|P,X} D + \beta_{N\sim P|D,X} P + X\beta_{N\sim X|D,P} + \epsilon_{n,s} \tag{30b}$$

$$\widehat{\text{bias}_{YD}} = \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \tag{31a}$$
$$\widehat{\text{bias}_{YP}} = \hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X} = \beta_{Z\sim P|D,X}\beta_{Y\sim Z|D,P,X} \tag{31b}$$
$$\widehat{\text{bias}_{ND}} = \hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X} = \beta_{Z\sim D|P,X}\beta_{N\sim Z|D,P,X} \tag{31c}$$
$$\widehat{\text{bias}_{NP}} = \hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X} = \beta_{Z\sim P|D,X}\beta_{N\sim Z|D,P,X} \tag{31d}$$

We can rewrite Equations 31b and 31c in the following way. Equation 31b becomes Equation 32a; Equation 31c becomes Equation 32b. 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.

$$\beta_{Y\sim Z|D,P,X} = [\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}) \tag{32a}$$
$$\beta_{Z\sim D|P,X} = (\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})[\beta_{N\sim Z|D,P,X}]^{-1} \tag{32b}$$

We start with the case in which we have both a placebo outcome and a placebo treatment. We simply plug both Equations 32b and 32a into the expression for $\widehat{\text{bias}_{YD}}$, followed by Equation 31d. In this way, we can write $\widehat{\text{bias}_{YD}}$ as in Equation 33. 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.

$$\begin{aligned}
\widehat{\text{bias}_{YD}} &= \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\
&= \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \\
&= \left[(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})[\beta_{N\sim Z|D,P,X}]^{-1}\right]\left[[\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})\right] \\
&= (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})[\beta_{N\sim Z|D,P,X}\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}) \\
&= (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})([\hat{\beta}_{N\sim P|D,X} - \beta_{N\sim P|D,Z,X}]^{-1})^{\top}(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})
\end{aligned} \tag{33}$$

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 32b into the expression for $\widehat{\text{bias}_{YD}}$. In this way, we can also write $\widehat{\text{bias}_{YD}}$ as in Equation 34. 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.

example, $\beta_{Z\sim D|P,X} = \begin{bmatrix}\beta_{Z_1\sim D|P,X} & \cdots & \beta_{Z_j\sim D|P,X}\end{bmatrix}$ and $\beta_{Y\sim Z|D,P,X} = \begin{bmatrix}\beta_{Y\sim Z_1|D,P,X} \\ \vdots \\ \beta_{Y\sim Z_j|D,P,X}\end{bmatrix}$, when there are j components of Z.

$$\begin{aligned}\widehat{\text{bias}_{YD}} &= \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\ &= \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \\ &= \left[(\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X})[\beta_{N\sim Z|D,P,X}]^{-1}\right]\beta_{Y\sim Z|D,P,X} \\ &= (\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}) \times m_N \\ &\text{where } m_N \triangleq [\beta_{N\sim Z|D,P,X}]^{-1}\beta_{Y\sim Z|D,P,X}\end{aligned} \tag{34}$$

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 32a into the expression for $\widehat{\text{bias}_{YD}}$. In this way, we can write $\widehat{\text{bias}_{YD}}$ as in Equation 35. 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.

$$\begin{aligned}\widehat{\text{bias}_{YD}} &= \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\ &= \beta_{Z\sim D|P,X}\beta_{Y\sim Z|D,P,X} \\ &= \beta_{Z\sim D|P,X}\left[[\beta_{Z\sim P|D,X}]^{-1}(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})\right] \\ &= m_P \times (\hat{\beta}_{N\sim D|P,X} - \beta_{N\sim D|P,Z,X}) \\ &\text{where } m_P \triangleq \beta_{Z\sim D|P,X}[\beta_{Z\sim P|D,X}]^{-1}\end{aligned} \tag{35}$$

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 33, we substitute in using both 32b and 32a. In Equation 34, we substitute in using only 32b. In Equation 35, we substitute in using only 32a. Equations 33, 34 and 35 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 4.1 for further discussion.

We can use Equations 33, 34 and 35 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. Now, in many cases differing scales for N and Y (or for P and D) might lead to difficulty in interpreting or reasoning about the values for m_N (or m_P). Different Z's might also have different scales making the using m_N (or m_P) difficult. Can we do better than this in the single placebo settings?

[Then we would move to a discussion of reparametarization of m_N and m_P and their interpretation, then SEs, then RVs. All of this would remain essentially unchanged from what we already have above.]

4 Practical considerations

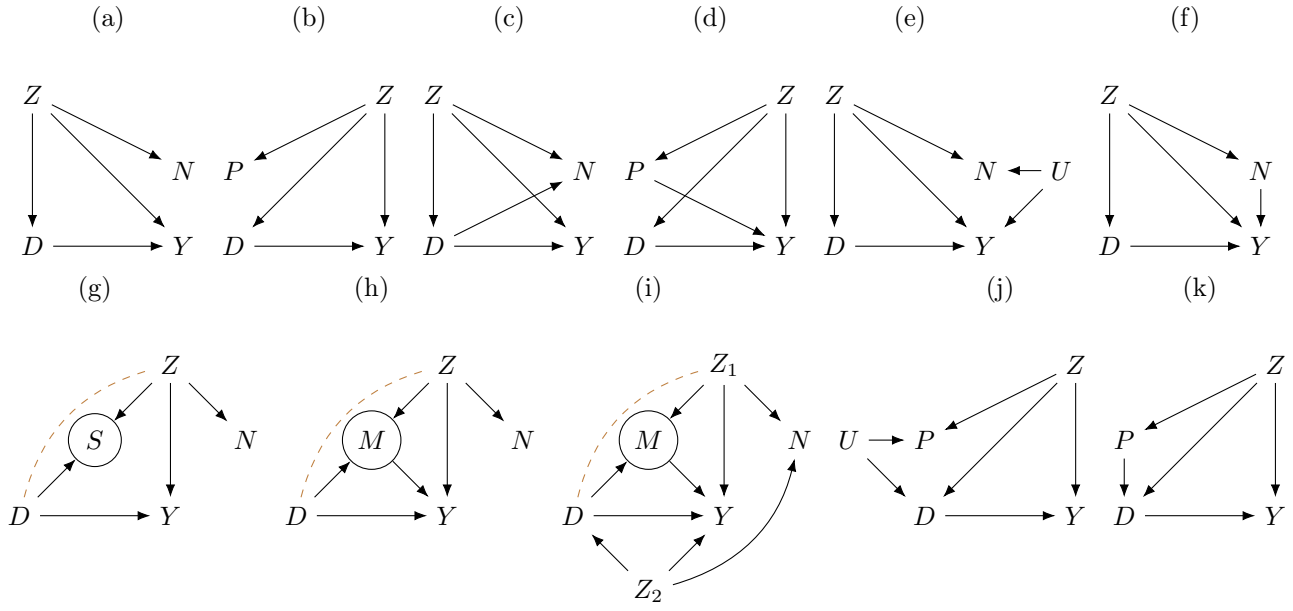
Let us turn to some practical items that researchers should consider when using the proposed methods. First, we discuss that a causal model should underlie and inform the use of our proposed method. We discuss several example graphical causal models. Such graphs will inform how we interpret the user-defined parameters of our partial identification framework. Second, we consider how we should conceptualize the set of unobserved variables Z and what this means for interpretation. Third, we discuss how various types of data can be used in the proposed framework. Finally, we discuss software under development to aid in practical applications of the methods.

4.1 Causal models are crucial to clarify interpretation

How do we know if we have useful placebos? How do we know exactly which causal relationships are captured by each parameter in the partial identification framework? The answer to both questions is to build a causal model. We will briefly discuss graphical causal models as tools for these purposes. For introductions to graphical causal models see Pearl (2009), Richardson and Robins (2013), Matthay and Glymour (2020), and Cinelli et al. (2022). Directed acyclic graphs (DAGs) are simple graphical causal models that allow researcher to formalize and visualize their knowledge of how variables in the system under study causally relate to one another. In a DAG, each node represents a variable, and the edges between nodes represent the causal relationships between those variables. Simple rules govern the conditional independencies between variables once a DAG has been constructed. While it can be burdensome to construct a causal model, doing so is necessary to making any causal inferences. Without a causal model illustrating how variables are assumed to relate, we do not have any information on which variables can be safely conditioned on, which cannot be, what the challenges for causal inference there may be, or what might be able to be done about these challenges.

In Figure 1, we saw two simple examples of DAGs in which placebos are present. Figure 3(a,b) replicate Figure 1 without the observed covariates. In Figure 3(a), we have a placebo outcome, N. The relationship between N and D, the treatment, is confounded by Z, a set of unobserved variables. So too is the relationship between Y, the outcome, and D confounded by Z. These confounding relationships consist of two "arms," $Z \to D$ and $Z \to Y$ (or $Z \to N$). Since $Z \to D$ is common to both, we can choose to focus on how the $Z \to Y$ relationship compares to the $Z \to N$ relationship. Or we can consider how the entire $D \leftarrow Z \to Y$ relationship compares to the entire $D \leftarrow Z \to N$ relationship. This is precisely what our proposed partial identification framework asks us to consider in the parameter λ (or k). Figure 3(c,d) are the same graphs but now with imperfect placebos, that is the placebo has a direct causal relationship with the actual outcome or treatment. This type of relaxation is captured in the $\beta_{N\sim D|Z,X}$ (or $\beta_{Y\sim P|D,Z,X}$) parameter in our partial identification framework.

Figure 3: Example DAGs with a placebo outcome, N, or a placebo treatment, P. D is treatment, Y is outcome, Z contains unobserved confounders, and U contains additional unobserved variables. Additional observed covariates are omitted for simplicity.



There are many more settings in which the proposed methods can be used. Figure 3 presents several more. In this Figure, we do not explicitly include the observed covariates to make exposition clearer; but these should be included in actual uses of DAGs, as conditioning on some observed covariates could change conditional independencies that users should consider.[10] In Figure 3(e), we might simply consider U, an unobserved common cause of N and Y but not D, to be contained within Z and reason about the relative strength of the Z, Y and Z, N relationships. It turns out that this is not necessary. Since we do not condition on N or Y in our analysis, U ends up not having an impact on the causal effect of interest or on the interpretation of any of our partial identification parameters.

Figure 3(f) is an apparently similar case, but one in which we should not come to the same conclusion. In this example, the parameter $\lambda = \frac{R^2_{Y\sim Z|D,X}}{R^2_{N\sim Z|D,X}}$ needs some additional consideration. The denominator, $R^2_{N\sim Z|D,X}$, will be capturing the $Z \to N$ relationship. However, the numerator, $R^2_{Y\sim Z|D,X}$, will actually be capturing both the $Z \to Y$ relationship and the $Z \to N \to Y$ relationship. This means that the $Z \to N$ relationship is appearing in both the numerator and denominator of λ. Therefore, it may be less likely that λ would equal 1.[11] But λ not being equal to some single value is exactly the sort of occasion that our approach was built for. Moreover, if Figure 3(f) is our DAG, we could also consider including N in our short regression for Y. N is not a collider and so doing this would not create any new back door paths or difficulties in interpretation. In this DAG, we can see that $\beta_{Z\sim D|X} = \beta_{Z\sim D|X,N}$ (or $R_{D\sim Z|X} = R_{D\sim Z|X,N}$) and so the cancellation of these terms would still happen in the derivation of our bias expression. We could then consider $\lambda = \frac{R^2_{Y\sim Z|D,X,N}}{R^2_{N\sim Z|D,X}}$, which would compare the $Z \to Y$ and $Z \to N$ arms. This would not be the case in all DAGs. In some settings including N in the short regression for Y could lead to newly opened non-causal paths or difficulties with interpretation. We emphasize the importance of careful consideration and use of causal models to make all such determinations.

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 3(g,h,i). Take Figure 3(g) 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.

There are settings in which users may encounter more difficulty in applying the proposed methods. Figure 3(j) appears to be similar to Figure 3(e), but it is not quite. Here we want to consider a placebo treatment, P, that has a common cause with the actual treatment, D. Since conditioning on U is not required to identify the effect of D on Y, we would typically not include it in the set of unobserved variables Z that we would like to but cannot include in our estimation strategy. However, as discussed in Appendix A, our proposal for the use of placebo treatments requires that P and D be included in the same short regression with outcome Y and for the inclusion of P to not change the coefficient on D. In Figure 3(j), however, this inclusion would mean that the numerator of λ, $R^2_{D\sim Z|P,X}$, would capture $D \leftarrow U \to P \leftarrow Z$ and $Z \to D$, while the denominator, $R^2_{P\sim Z|D,X}$ would capture $P \leftarrow U \to D \leftarrow Z$ and $Z \to P$. Whether or not $D \leftarrow U \to P \leftarrow Z$ and $P \leftarrow U \to D \leftarrow Z$ can be easily compared, in combination with $Z \to D$ and $Z \to P$, may be difficult to determine, since conditioning on colliders often creates unintuitive associations. A similar example can be found in Figure 3(k), where, conditional on D, $R^2_{P\sim Z|D,X}$ would capture $P \to D \leftarrow Z$ and $Z \to P$ but $R^2_{D\sim Z|P,X}$ would capture only $Z \to D$. Again, this puts us in a setting like Figure 3(f) where $R^2_{P\sim Z|D,X}$ and $R^2_{D\sim Z|P,X}$ capture different structural relationships. The extent to which such issues make interpretation and use of our proposed methods difficult is specific to every given problem.[12]

For additional discussion and examples of causal graphical models in which placebos are present, see Lipsitch et al. (2010); Arnold et al. (2016). We refrain providing specific graphical or counterfactual criteria that must be met for the use of our methods, like those discussed in Tchetgen et al. (2020). This is because our approach is correct regardless of the underlying structural relationships between variables; it is only in the interpretation of the partial identification parameters that structural relationships matter. But there may be many settings in which researchers are willing to make assumptions about the range of plausible values for the partial identification parameters outside the most simple settings. We hope that the discussion in this section provides a guide to how users of our proposed methods can proceed in their applications. We stress that it is crucial for researchers to carefully construct and consider their causal model when employing our proposed methods. While the bias expressions we provide will hold mechanically, the structural relationships captured by the partial identification parameters may be more complex than a naive analysis suggests.

[10] We also strongly believe that all causal graphical models should include a sample selection node. See Rohde and Hazlett (20XX) for further discussion. These are again omitted for simplicity in most cases here.

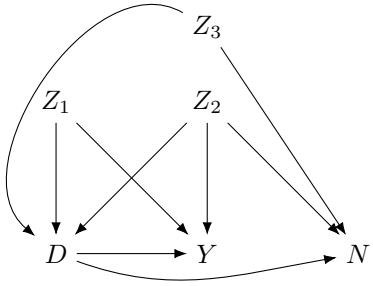
[11] Though this is not impossible, since partial R^2s represent the proportion of residual variance explained, after linearly accounting for D, X, and this could differ if N or Y has more residual variation.

[12] Some users may wonder whether multiple placebo outcomes or placebo treatments can be used in concert. We refer to XXXX for discussion of a related approach. An explicit extension of our framework to multiple placebos may be the subject of future work.

4.2 What should be contained in Z?

In this section we discuss how to think about the set of unobserved variables that are contained in the set Z. We focus on the placebo outcome case for exposition but similar discussion would apply to placebo treatments as well. We suggest one of two approaches. We will use Figure 4 to guide our discussion. In this Figure, we partition the unobserved common causes that we might consider into three separate sets: Z_1 contains common causes of D and Y only; Z_2 contains variables that are common causes of D, Y, and N; Z_3 contains common causes of D and N only.

Figure 4: DAG with Z partitioned into three sets of variables: Z_1 contains common causes of D and Y only; Z_2 contains variables that are common causes of D, Y, and N; Z_3 contains common causes of D and N only. N is a placebo outcome, D is the treatment, and Y is the actual outcome.



Clearly, researchers will need to consider at least Z_1 and Z_2 as the components of the set of unobserved common causes that we would like to condition on (i.e., include in our regression) but cannot. These variables contain all the common causes of D and Y. Note that since these variables are all unobserved, we can consider a sufficiently rich set so that we can identify the effect of D on Y (i.e., block all the non-causal paths between D and Y).[13] So, a researcher may consider $Z = \{Z_1, Z_2\}$. This is the set of unobserved confounders that is sufficient to de-confound the effect of D on Y, but not necessarily sufficient to de-confound the effect of D on N. With this choice of Z, we can consider how the confounding of D and Y (from $Z = \{Z_1, Z_2\}$) compares to the confounding of D and N. This relative level of confounding is captured by the partial identification parameter $\lambda = \frac{R^2_{Y\sim Z|D,X}}{R^2_{N\sim Z|D,X}}$. If Z_1 is empty, then λ is a comparison of the $Z_2 \to Y$ relationship and the $Z_2 \to N$ relationship. If Z_1 is not empty, then λ is a comparison of the $\{Z_1, Z_2\} \to Y$ relationship and the $\{Z_1, Z_2\} \to N$ relationship, in which we know that there is no $Z_1 \to N$ connection. This should influence the relative levels of confounding (i.e., λ values) that are deemed plausible. Here we must consider the relative level of confounding of D and Y versus of D and N from just the variables that are common causes of D and Y. With $Z = \{Z_1, Z_2\}$, the $\beta_{N\sim D|Z,X}$ partial identification parameter would capture both the $D \to N$ relationship and the $D \leftarrow Z_3 \to N$ relationship. This means that $\beta_{N\sim D|Z,X}$ needs to be interpreted correctly relative to the choice of Z. The extent of confounding of D and N from Z_3 and the strength of the $D \to N$ relationship need to be considered. If Z_3 is empty, then $\beta_{N\sim D|Z,X}$ simply captures the $D \to N$ relationship.

Now what if a researcher considers $Z = \{Z_1, Z_2, Z_3\}$? This is the set of unobserved confounders that is sufficient to de-confound the effect of D on Y and the effect of D on N. In this case, λ is a comparison of the $\{Z_1, Z_2, Z_3\} \to Y$ relationship and the $\{Z_1, Z_2, Z_3\} \to N$ relationship, where we know that there is no $Z_1 \to N$ connection or $Z_3 \to Y$ connection. Again, this choice for Z should influence the λ values that are deemed plausible. In this case, however, we can consider the total confounding of D and Y (from $\{Z_1, Z_2\}$) with the total confounding of D and N (from $\{Z_2, Z_3\}$). With $Z = \{Z_1, Z_2, Z_3\}$, $\beta_{N\sim D|Z,X}$ captures just the direct causal path from D to N. We expect that the second approach (i.e., $Z = \{Z_1, Z_2, Z_3\}$) may be more intuitive in most cases; but both are equally valid.

In the difference in differences setting where N is a pre-treatment outcome, Z should include all unobserved confounders, including time-varying confounders. In this case, λ will capture differences in confounding resulting from any time-varying confounders, which can make the confounding relationship between Z and Y different from that between Z and N. As we will see, standard difference in differences approaches assume that there are no unobserved time-varying confounders and that the level of confounding of D and Y is exactly equal to the level of confounding of D and N.

[13] Again, we emphasize that researchers ought to consider how well linear regression may do in estimating their causal effect of interest. Are assumptions of linearity and effect homogeneity defensible or is linear regression being used to obtain an approximation?

4.3 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.

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

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.

$$\begin{aligned}
\text{DIM}_Y &= \mathbb{E}[Y|G=1] - \mathbb{E}[Y|G=0] \\
&= \mathbb{E}[Y_1|G=1] - \mathbb{E}[Y_0|G=0] \text{ by consistency and since } D=G \text{ in period 2} \\
&= \underbrace{\mathbb{E}[Y_1|G=1] - \mathbb{E}[Y_0|G=1]}_{\text{ATT}_Y} + \underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[Y_0|G=0]}_{\text{Bias}_Y} \\
&= \text{ATT}_Y + \text{Bias}_Y
\end{aligned} \tag{36}$$

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).

$$\begin{aligned}
\text{Parallel Trends} &\triangleq \underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[N_0|G=1]}_{\text{Trend}_{G=1}} = \underbrace{\mathbb{E}[Y_0|G=0] - \mathbb{E}[N_0|G=0]}_{\text{Trend}_{G=0}} \\
&\iff \underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[Y_0|G=0]}_{\text{Bias}_Y} = \underbrace{\mathbb{E}[N_0|G=1] - \mathbb{E}[N_0|G=0]}_{\text{Bias}_N} \triangleq \text{Equi-Confounding}
\end{aligned} \tag{37}$$

Now, we recall our constraint that no units are treated in period 1 and so $N_0 = N$ for both the treated and control groups. So we see that $\text{Bias}_N = \text{DIM}_N$.

$$\text{Bias}_N = \mathbb{E}[N_0|G=1] - \mathbb{E}[N_0|G=0] = \mathbb{E}[N|G=1] - \mathbb{E}[N|G=0] = \text{DIM}_N \tag{38}$$

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 most 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.[14]

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,

$$\begin{aligned}
\text{DIM}_Y &= \beta_{Y\sim G} = \beta_{Y\sim G|Z} + \beta_{Y\sim Z|G}\beta_{Z\sim G} \\
\text{DIM}_N &= \beta_{N\sim G} = \beta_{N\sim G|Z} + \beta_{N\sim Z|G}\beta_{Z\sim G}
\end{aligned} \tag{39}$$

where Z contains 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. From here, we see that $\text{ATT}_Y = \text{DIM}_Y - \text{DIM}_N = \beta_{Y\sim G} - \beta_{N\sim G}$.[15] We can also see that

$$\text{ATT}_Y + \text{Bias}_Y = \text{DIM}_Y = \beta_{Y\sim G} = \beta_{Y\sim G|Z} + \beta_{Y\sim Z|G}\beta_{Z\sim G} \tag{40}$$

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 $\text{DIM}_N - \beta_{N\sim G|Z} = \beta_{N\sim Z|G}\beta_{Z\sim G}$. In the traditional DID setup, N is a pre-treatment outcome and so there is zero effect of G on N, i.e. $\beta_{N\sim G|Z} = 0$. So we

[14] When we get to relaxing the requirement that N is a pre-treatment version of the outcome Y, parallel trends type assumptions will be less intuitive than equi-confounding type assumptions, which is another reason we favor the equi-confounding view.

[15] The difference in differences can also be estimated in a single regression, but we stick to this setting with two regressions so that we can relax more of the standard DID assumptions, as we will see shortly.

will call $\text{DIM}_N - \beta_{N\sim G|Z}$ "Bias_N," as above. We now know that $\text{Bias}_N = \beta_{N\sim Z|G}\beta_{Z\sim G}$. Therefore, we can see that the equi-confounding assumption that $\text{Bias}_Y = \text{Bias}_N$ is equivalent to $\beta_{Y\sim Z|G}\beta_{Z\sim G} = \beta_{N\sim Z|G}\beta_{Z\sim G} \iff \beta_{Y\sim Z|G} = \beta_{N\sim Z|G}$. So an assumption of equi-confounding (or parallel trends), as well as effect homogeniety, is the same as assuming the regression coefficient capturing the Y, Z relationship is the same as that capturing the N, Z relationship, where both regressions also include D. Slight variations on the omitted variables bias expression from Cinelli and Hazlett (2020) give us expressions for Bias_Y and Bias_N as in Equation 41, where $f_{G\sim Z} = R_{G\sim Z}/\sqrt{1-R^2_{G\sim Z}}$ is Cohen's f.

$$\begin{aligned}\text{Bias}_Y = \text{DIM}_Y - \text{ATT}_Y = \text{DIM}_Y - \beta_{Y\sim G|Z} = \beta_{Y\sim G} - \beta_{Y\sim G|Z} = \frac{\text{SD}(Y^{\perp G})}{\text{SD}(G)} R_{Y\sim Z|G} f_{G\sim Z} \\ \text{Bias}_N = \text{DIM}_N - \beta_{N\sim G|Z} = \beta_{N\sim G} - \beta_{N\sim G|Z} = \frac{\text{SD}(N^{\perp G})}{\text{SD}(G)} R_{N\sim Z|G} f_{G\sim Z}\end{aligned} \tag{41}$$

The equi-confounding assumption can therefore also be seen as in Equation 42. Equi-confounding, in regression, can be seen as an assumption about how $R^2_{Y\sim Z|G}$ compares to $R^2_{N\sim Z|G}$.

$$\begin{aligned}&\text{Bias}_Y = \text{Bias}_N \\ \iff &\frac{\text{SD}(Y^{\perp G})}{\text{SD}(G)} R_{Y\sim Z|G} f_{G\sim Z} = \frac{\text{SD}(N^{\perp G})}{\text{SD}(G)} R_{N\sim Z|G} f_{G\sim Z} \\ \iff &\frac{R_{Y\sim Z|G}}{R_{N\sim Z|G}} = \frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})} \iff \frac{R^2_{Y\sim Z|G}}{R^2_{N\sim Z|G}} = \left(\frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}\right)^2\end{aligned} \tag{42}$$

Assumption Relaxation 1: G could have an effect on N In the traditional DID setup, N is a pre-treatment outcome and so there is zero effect of G on N, i.e. $\beta_{N\sim G|Z} = 0$. However, we could relax this assumption to the case where G does have a Z conditional effect on N. So we relax the assumption that $\beta_{N\sim G|Z} = 0$ and allow $\beta_{N\sim G|Z}$ to take any value. Since we observe N as well as G, we know the units of these variables and hence will be able to reason about their relationship in terms of a regression coefficient. In particular, we can reason about how $\beta_{N\sim G|Z}$ might compare to $\beta_{N\sim G}$, which we are able to estimate. Relaxing this assumption of the standard DID framework leads to the equi-confounding type assumption in Equation 43, where $\beta_{N\sim G|Z}$ is a sensitivity parameter. This can then be used to explore how different values for the sensitivity parameter changes the estimate of ATT_Y or $\beta_{Y\sim G|Z}$.

$$\begin{aligned}&\text{Bias}_Y = \text{Bias}_N = \text{DIM}_N - \beta_{N\sim G|Z} \\ \implies &\text{ATT}_Y = \text{DIM}_Y - \text{Bias}_Y = \text{DIM}_Y - \text{DIM}_N - \beta_{N\sim G|Z}\end{aligned} \tag{43}$$

We can, thus, allow for G to have some non-zero effect on N or some other association with N other than through Z and still maintain assumptions and an identification strategy that closely resemble those of the standard differences in differences framework.

Assumption Relaxation 2: Unequal confounding We will now consider an additional relaxation of the standard difference in difference assumptions. As we saw above, the equi-confounding (or parallel trends) assumption is an assumption that $\text{Bias}_Y = \text{Bias}_N$. We also just saw that this is equivalent to an assumption that $\beta_{Y\sim Z|G} = \beta_{N\sim Z|G}$, which is in turn equivalent to $\frac{R^2_{Y\sim Z|G}}{R^2_{N\sim Z|G}} = \left(\frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}\right)^2$. We can relax this assumption to be an assumption on the relative amount of confounding, rather than perfect equality in confounding. That is, for some scalar m, we get the relationships in Equation 44.

$$\begin{aligned}&\text{Bias}_Y = m \times \text{Bias}_N \\ \iff &(\text{DIM}_Y - \beta_{Y\sim G|Z}) = m \times (\text{DIM}_N - \beta_{N\sim G|Z}) \\ \iff &\beta_{Y\sim Z|G}\beta_{Z\sim G} = m \times \beta_{N\sim Z|G}\beta_{Z\sim G} \qquad \text{or } \frac{\text{SD}(Y^{\perp G})}{\text{SD}(G)} R_{Y\sim Z|G} f_{G\sim Z} = m \times \frac{\text{SD}(N^{\perp G})}{\text{SD}(G)} R_{N\sim Z|G} f_{G\sim Z} \\ \iff &\beta_{Y\sim Z|G} = m \times \beta_{N\sim Z|G} \qquad \text{or } \frac{R^2_{Y\sim Z|G}}{R^2_{N\sim Z|G}} = \left(m \times \frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}\right)^2\end{aligned} \tag{44}$$

If we define $\lambda \triangleq \frac{R^2_{Y\sim Z|G}}{R^2_{N\sim Z|G}}$ and $\gamma \triangleq \text{sign}\left(\frac{R_{Y\sim Z|D,X}}{R_{N\sim Z|D,X}}\right)$, then we can see that under this assumption, we get the relationship between m, λ and γ in Equation 45.

$$m = \frac{\beta_{Y\sim Z|G}}{\beta_{N\sim Z|G}} = \gamma\sqrt{\lambda} \times \frac{\text{SD}(Y^{\perp G})}{\text{SD}(N^{\perp G})} \tag{45}$$

If the scale of N and Y differ, it will be easier to reason about the value of λ than to reason about the value of m, but the two have a simple relationship that simply accounts for any scale differences. An assumption about the value of λ (and $\beta_{N\sim G|Z}, \gamma$) is equivalent to an assuming an equi-confounding type assumption. This can then be used to explore how different values for the parameters λ, $\beta_{N\sim G|Z}$, and γ change the estimate of ATT_Y or $\beta_{Y\sim G|Z}$ in a difference in difference type expression.

$$\begin{aligned} \text{Bias}_Y = m \times \text{Bias}_N = \gamma\sqrt{\lambda} \times \frac{\text{SD}(Y^{\perp G})}{\text{SD}(N^{\perp G})} \times \text{Bias}_N = \gamma\sqrt{\lambda} \times \frac{\text{SD}(Y^{\perp G})}{\text{SD}(N^{\perp G})} \times (\text{DIM}_N - \beta_{N\sim G|Z}) \\ \implies \text{ATT}_Y = \beta_{Y\sim G|Z} = \text{DIM}_Y - \gamma\sqrt{\lambda} \times \frac{\text{SD}(Y^{\perp G})}{\text{SD}(N^{\perp G})} \times (\text{DIM}_N - \beta_{N\sim G|Z}) \end{aligned} \tag{46}$$

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 or λ (and γ), where λ 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 the standard difference in difference assumption as assuming that $\beta_{Y\sim G|Z} = 0$ and $m = 1$ or equivalently $\lambda = \left(\frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}\right)^2$. This value of λ arising could be the result of simply a fortunate balancing coincidence for the ratios $\frac{R^2_{Y\sim Z|G}}{R^2_{N\sim Z|G}}$ 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 $R^2_{Y\sim Z|G} = R^2_{N\sim Z|G}$. In this way, the assumption that $\lambda = 1$ can be thought of as less demanding than the assumption that $m = 1$. So we could consider a "new" form of difference in difference where we assume $\lambda = 1$ not that $m = 1$. This approach also avoids the scale issues discussed in Sofer et al. (2016). However, the above frame work need not be limited to any specific choice for λ. Rather, we can consider a range of plausible values for λ, perhaps including the value that would bring the true effect to zero under the assumption that $\beta_{N\sim G|Z} = 0$, $\lambda = \left(\frac{\text{DIM}_Y}{\text{DIM}_N} \times \frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}\right)^2$, as well as $\lambda = 1$ and $\lambda = \left(\frac{\text{SD}(N^{\perp G})}{\text{SD}(Y^{\perp G})}\right)^2$.

Transforming unequal confounding assumptions into unequal trend assumptions Finally, we show that any assumption for k corresponds to an assumption on the trend in control potential outcomes. We will show that any assumption for m or k 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 47. An assumption about a difference in trend would be something like Equation 48.

$$\begin{aligned} &\underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[Y_0|G=0]}_{\text{Bias}_Y} = m \times \underbrace{(\underbrace{\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0]}_{\text{DIM}_N} - \beta_{N\sim G|Z})}_{\text{Bias}_N} \\ &\iff \mathbb{E}[Y_0|G=1] = \mathbb{E}[Y_0|G=0] + m \times (\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0] - \beta_{N\sim G|Z}) \end{aligned} \tag{47}$$

$$\begin{aligned} &\underbrace{\mathbb{E}[Y_0|G=1] - \mathbb{E}[N|G=1]}_{\text{Trend}_{G=1}} = w \times \underbrace{(\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0])}_{\text{Trend}_{G=0}} \\ &\iff \mathbb{E}[Y_0|G=1] = \mathbb{E}[N|G=1] + w \times (\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0]) \end{aligned} \tag{48}$$

For this exposition, we assume $\beta_{N\sim G|Z} = 0$ for simplicity. Both Equations 47 and 48 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 to get the trend assumption that corresponds to our assumption on m.

$$\mathbb{E}[Y_0|G=0] + m \times (\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0]) = \mathbb{E}[N|G=1] + w \times (\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0]) \tag{49}$$

$$\begin{aligned} \iff w &= \frac{\mathbb{E}[Y_0|G=0] + m \times (\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0]) - \mathbb{E}[N|G=1]}{\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0]} \\ \iff m &= \frac{\mathbb{E}[N|G=1] + w \times (\mathbb{E}[Y_0|G=0] - \mathbb{E}[N|G=0]) - \mathbb{E}[Y_0|G=0]}{\mathbb{E}[N|G=1] - \mathbb{E}[N|G=0]} \end{aligned} \tag{50}$$

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

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.

6 Example

[update to do example from the proximal causal inference intro paper - Tchetgen et al. (2020); Connors et al. (1996) - https://www.icpsr.umich.edu/web/HMCA/studies/2957
or something else from the health and medical data archive - https://www.icpsr.umich.edu/web/pages/HMCA/index.html]
[also need to update the plots]

Let us consider a 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 library. We use the non-experimental dataset to illustrate the methods discussed in this paper and then compare these results with the estimates from the experimental data.

Draw from R code (and MAS 420 slides) to write this up. Describe data and program. We use 1974 earnings (N) as a placebo outcome for 1978 earnings (Y). The treatment (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. The ideal regression we would like to estimate is

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

We want to understand how omitting these variables has biased our estimate of the effect of the training program on 1978 earnings. To aid in this, we leverage the placebo outcome of 1974 earnings. That is, we estimate the following regressions and use the results of these regressions in the placebo outcome partial identification framework from above.

$$Y = \beta_{Y\sim D|X} D + X\beta_{Y\sim X|D} + \epsilon_s$$
$$N = \beta_{N\sim D|X} D + X\beta_{N\sim X|D} + \xi_s$$

Since the placebo outcome is pre-treatment earnings (specifically 1974 earnings), the treatment has no causal effect on the the placebo outcome. It might be possible for 1974 earnings to be associated with being in the experimental sample since the experimental sample contains "disadvantaged" individuals. Hence, there could be a connection between 1974 earnings and treatment other than through Z and X (i.e., $\beta_{N\sim D|Z,X} \neq 0$).

Is this right? Could $N \to D$? Low earners are more likely to be in the experimental study and hence more likely to be treated? The experimental study was in fact limited to disadvantaged workers lacking basic job skills. In this case, N might be a confounder of D and Y

COuld it be that the DID approach works well because it assume that bias Y = DID N, which would capture the confounding and the level of $\beta_{N\sim D|Z,X}$?

With the results from the estimated regressions, we can use the bias expression to see how different values for k might change our estimate and can consider what range of values for k might be reasonable.

$$|\widehat{\text{bias}_Y}| = \sqrt{k} \times |\hat{\beta}_{N\sim D|X} - \underbrace{\beta_{N\sim D|Z,X}}_{=0}| \times \frac{\text{se}(\hat{\beta}_{Y\sim D|X})}{\text{se}(\hat{\beta}_{N\sim D|X})} \times \sqrt{\frac{\text{df}_Y}{\text{df}_N}}$$

Since 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. We can also consider how well a standard DID approach and the $k = 1$ DID approach would perform in this example. See Figure 5 for a detailed picture of the sensitivity analysis.

The naive estimate (estimated by regressing 1978 earnings on the treatment indicator with some covariates) is negative. This is not the sign we would intuitively expect. The experimental estimate is positive, which makes sense because training should probably boost earnings. The RV is saying that the confounding in the post-treatment period would only need to be 0.43 times the confounding in the pre-treatment period to bring the negative naive estimate to zero. This uses the R^2 parameterization of relative confounding. 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. Since this is larger than the RV, we see that the "true" effect has the opposite sign relative to the naive effect.

The standard DID estimate implicitly makes an assumption that the confounding in the post-treatment period is actually 0.715 times the confounding in the pre-treatment period (again using the R^2 parameterization), which gets us close to the

experimental estimate. The fact that these two are close is something of a coincidence. We discuss this in detail in the following paragraphs.

The overriding takeaway from this analysis is that a fairly low level of relative confounding is enough to reverse the sign of our naive estimated effect. Regardless of the specific level we might favor, it is easy to argue that the confounding in the post-treatment period would be greater than 0.43 times the confounding in the pre-treatment period. While an assumption that the level of confounding is equal in terms of R^2's overestimates the true or experimental effect, it provides a much better estimate than our naive estimate that does not use the placebo outcome. Moreover, using our framework, we are not limited to considering a single level of relative confounding. We might choose to report both the $m = 1$ and $k = 1$ DID estimates, the RV, and the range of estimates corresponding to the plausible range for relative confounding. Surely this range excludes values for k close to zero, since the placebo outcome is a pre-treatment version of the actual outcome. As such, we can conclude that there is likely a positive effect of the treatment on the outcome.

6.1 Why does the standard DID estimate do so well recovering the experimental estimate?

It turns out that the standard DID estimate is doing well in recovering the experimental estimate here by coincidence. In general, we can show that

$$k_{\text{true}} = \left(\frac{|\beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X}|}{|\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}|}\right)^2 \times \left(\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})}\right)^2$$

$$k_{\text{DID}} = \left(\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})}\right)^2$$

So k_{true} will only equal k_{DID} when $|\beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X}|$ equals $|\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}|$. This is something of a truism, since $|\beta_{Y\sim D|X} - \beta_{Y\sim D|Z,X}| = |\beta_{N\sim D|X} - \beta_{N\sim D|Z,X}|$ is actually just the parallel trends or equi-confounding assumption.

In our present example, we know that $\beta_{N\sim D|Z,X} = 0$ since N is a pre-treatment outcome. Additionally, we estimate $\beta_{Y\sim D|Z,X}$ from the experimental data, while estimating $\beta_{Y\sim D|X}$ and $\beta_{N\sim D|X}$ from the non-experimental data. The fact that we observe that $0.706 = \hat{k}_{\text{true}} \approx \hat{k}_{\text{DID}} = 0.715$ means that $|\hat{\beta}_{Y\sim D|X} - \hat{\beta}_{Y\sim D|Z,X}| \approx |\hat{\beta}_{N\sim D|X}|$:

$$\hat{k}_{\text{true}} \approx \hat{k}_{\text{DID}} \implies |\hat{\beta}_{Y\sim D|X} - \hat{\beta}_{Y\sim D|Z,X}| \approx |\hat{\beta}_{N\sim D|X}|$$

But since $\hat{\beta}_{Y\sim D|Z,X} = 1671.13$ and $(\beta_{Y\sim D|X} = -5928.11, \beta_{N\sim D|X} = -7646.11)$ are estimated on separate sets of observations, the fact that $\beta_{Y\sim D|X} - \beta_{N\sim D|X} = 1718.00$ is close to $\hat{\beta}_{Y\sim D|Z,X} = 1671.13$ is a coincidence. There is nothing mechanical here that is leading this to be the case. The estimate of $\hat{\beta}_{Y\sim D|Z,X}$ could have been something different, if the treatment effect were different. Or $\beta_{Y\sim D|X} - \beta_{N\sim D|X}$ could have equaled something else if the relative level of confounding were different.

We can also think about this in terms of m. The standard DID approach assumes that $m = 1$. Recall that the true value for m is defined as follows.

$$m = \frac{\beta_{Y\sim Z|D,X}}{\beta_{N\sim Z|D,X}} = \frac{\text{SD}(Y^{\perp D,X})/\text{SD}(Z^{\perp D,X})}{\text{SD}(N^{\perp D,X})/\text{SD}(Z^{\perp D,X})} \times \frac{\text{Cor}(Y^{\perp D,X}, Z^{\perp D,X})}{\text{Cor}(N^{\perp D,X}, Z^{\perp D,X})}$$

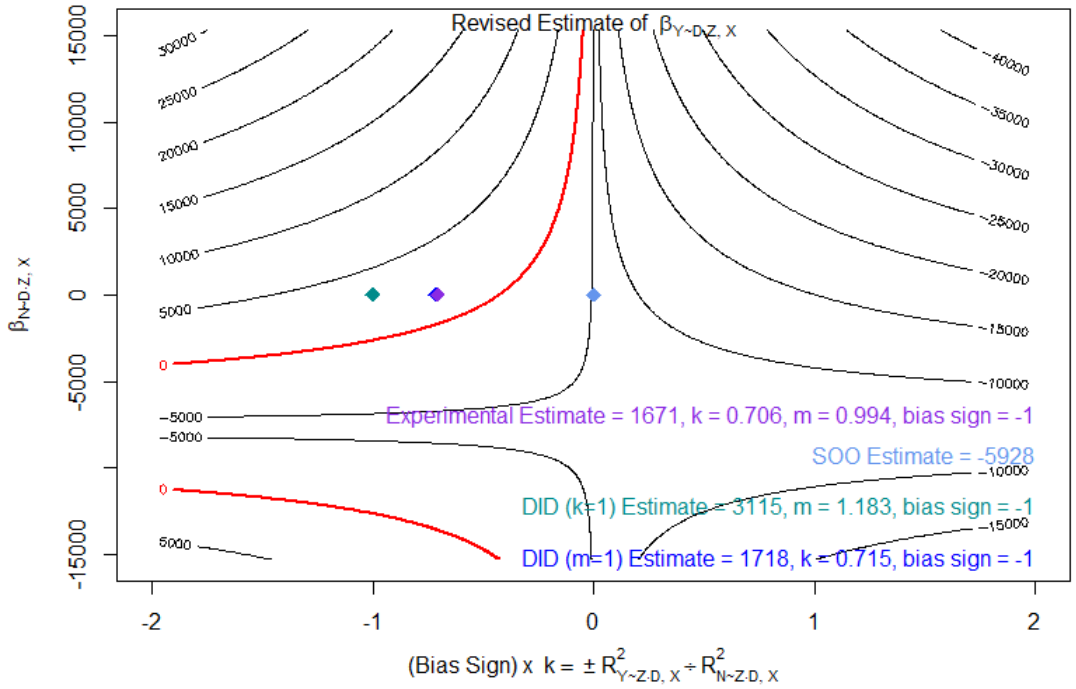
This means that, m can equal 1 (or approximately 1) only when $\frac{\text{Cor}(Y^{\perp D,X}, Z^{\perp D,X})}{\text{Cor}(N^{\perp D,X}, Z^{\perp D,X})}$ and $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}$ are (approximate) reciprocals of each other. But there is nothing that makes this mechanically true. Indeed, if something did, then we would always be justified in assuming the standard version of parallel trends or equi-confounding. But this is clearly not the case. The fact that the true value for m in this example is close to 1 does not mean that it had to be.

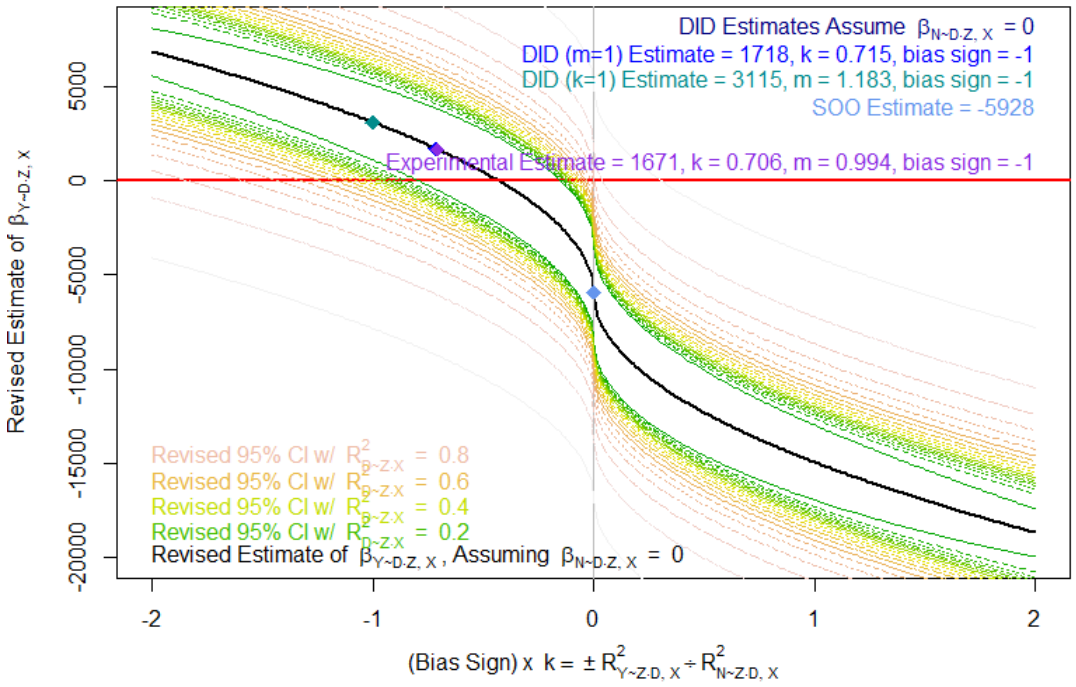
[We can create a scenario where the DID estimate does not recover the experimental estimate. For example, by limiting the observational data to 1978 earnings greater than 1000. This would create another observational dataset from which we could conduct our analysis. Or we could limit the observational data to 1974 earnings greater than 1000. In this case, the k=1 DID would actually do better than the standard. These versions might be useful to show.]

6.2 Why do the $k = 1$ DID estimate and standard DID estimate differ?

The $k = 1$ DID estimate will only equal the standard DID estimate when $\text{SD}(N^{\perp D,X}) = \text{SD}(Y^{\perp D,X})$. We can show that

Figure 5: Partial identification plots for National Supported Work example with 1974 earnings as placebo outcome.





$$\text{DID Bias}_{\text{standard}} = \sqrt{\left(\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})}\right)^2} \times (\hat{\beta}_{N\sim D|X} - 0) \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \quad = \hat{\beta}_{N\sim D|X}$$

$$\text{DID Bias}_{k=1} = \sqrt{1} \times (\hat{\beta}_{N\sim D|X} - 0) \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} \quad = \hat{\beta}_{N\sim D|X} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}$$

Which then means that the two estimates are

$$\text{DID Estimate}_{\text{standard}} = \hat{\beta}_{Y\sim D|X} - \hat{\beta}_{N\sim D|X}$$
$$\text{DID Estimate}_{k=1} = \hat{\beta}_{Y\sim D|X} - \hat{\beta}_{N\sim D|X} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}$$

So it is easy to see that these estimates are only equal when $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} = 1$ or $\text{SD}(N^{\perp D,X}) = \text{SD}(Y^{\perp D,X})$. In the present example, we see that $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})} = 1.183$. So the two estimates are not equal.

When $\text{SD}(Y^{\perp D,X}) > \text{SD}(N^{\perp D,X})$, the $k = 1$ DID estimate will subtract a larger value from $\hat{\beta}_{Y\sim D|X}$, the naive estimate of the treatment effect, than does the standard ($m = 1$) DID estimate. This is because the standard DID estimate assumes that $k = \left(\frac{\text{SD}(N^{\perp D,X})}{\text{SD}(Y^{\perp D,X})}\right)^2$, which will be less than 1 and therefore reduces the estimate for bias. When $\text{SD}(Y^{\perp D,X}) > \text{SD}(N^{\perp D,X})$, the standard DID approach assumes less bias than does the $k = 1$ version. When $\text{SD}(Y^{\perp D,X}) < \text{SD}(N^{\perp D,X})$, the $k = 1$ DID estimate will subtract a smaller value from $\hat{\beta}_{Y\sim D|X}$ than does the standard ($m = 1$) DID estimate. Similar logic applies. When $\text{SD}(Y^{\perp D,X}) < \text{SD}(N^{\perp D,X})$, the standard DID approach assumes more bias than does the $k = 1$ version. There are many factors that determine $\text{SD}(Y^{\perp D,X})$ and $\text{SD}(N^{\perp D,X})$ and these are not limited to the relative levels of confounding. There is no reason why these need to equate or have a relationship with the true value of k. The difference in scale could be a result of the treatment effect itself altering the scale of Y or different relationships with Z altering the scales differently or the presence of other unequal causes of N and Y that are not unobserved confounders. In the end, since $\frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}$ and $\beta_{Y\sim D|X}$ and $\beta_{N\sim D|X}$ are estimable from the data, we only need to worry about reasoning about the sensitivity parameters k, which captures the relative levels of confounding as described in partial R^2's, and $\beta_{N\sim D|Z,X}$, which in this case we know is equal to zero.

6.3 Unemployment in 1975 as placebo outcome

We can also use some other variable as our placebo outcome, like 1975 unemployment. This is a binary variable and so has a completely different scale from 1978 earnings. The 1975 unemployment indicator is equivalent to an indicator for 1975 earnings being zero. Despite this, unemployment may have a different level of confounding, since it is not a measure of earnings and is a measure from 1975 rather than 1978. So this might not be the best placebo outcome to use; but it is illustrative of the point that any placebo outcome could be used. See Figure 6 for a detailed picture of the sensitivity analysis. In this example, we still know that $\beta_{N\sim D|Z,X} = 0$, since 1975 unemployment is a pre-treatment measure. The naive estimate and experimental estimate are both the same as before. The RV is says that the confounding of earnings in the post-treatment period would only need to be 0.068 times the confounding of unemployment in the pre-treatment period to bring the negative naive estimate to zero. This uses the R^2 parameterization of relative confounding so the scale difference is not an issue. The experimental estimate implies that the "true" confounding of earnings in the post-treatment period is actually 0.112 times the confounding of unemployment in the pre-treatment period. Since this is larger than the RV, we see that the "true" effect has the opposite sign relative to the naive effect. The standard DID estimate implicitly makes an assumption that the confounding of earnings in the post-treatment period is actually essentially 0 times the confounding of unemployment in the pre-treatment period (again using the R^2 parameterization). This does not get us close to the experimental estimate, since it underestimates the relative level of confounding. The $k = 1$ version of DID also does not do a good job of recovering the experimental estimate, since it dramatically overestimates the relative level of confounding. Note that the experimental estimate corresponds to a m value of almost 15,000; interpreting this value is difficult since the meaning of $\beta_{Y\sim Z|D,X}$ is different from $\beta_{N\sim Z|D,X}$. Again, the overriding takeaway from this analysis is that a low level of relative confounding is enough to reverse the sign of our naive estimated effect. Using our framework, we might consider the range of estimates corresponding to the plausible range for relative confounding. Surely this range excludes values for k close to zero, since the placebo outcome has a clear relationship with the actual outcome. As such, we can conclude that there is likely a positive effect of the treatment on the outcome.

6.4 Another application?

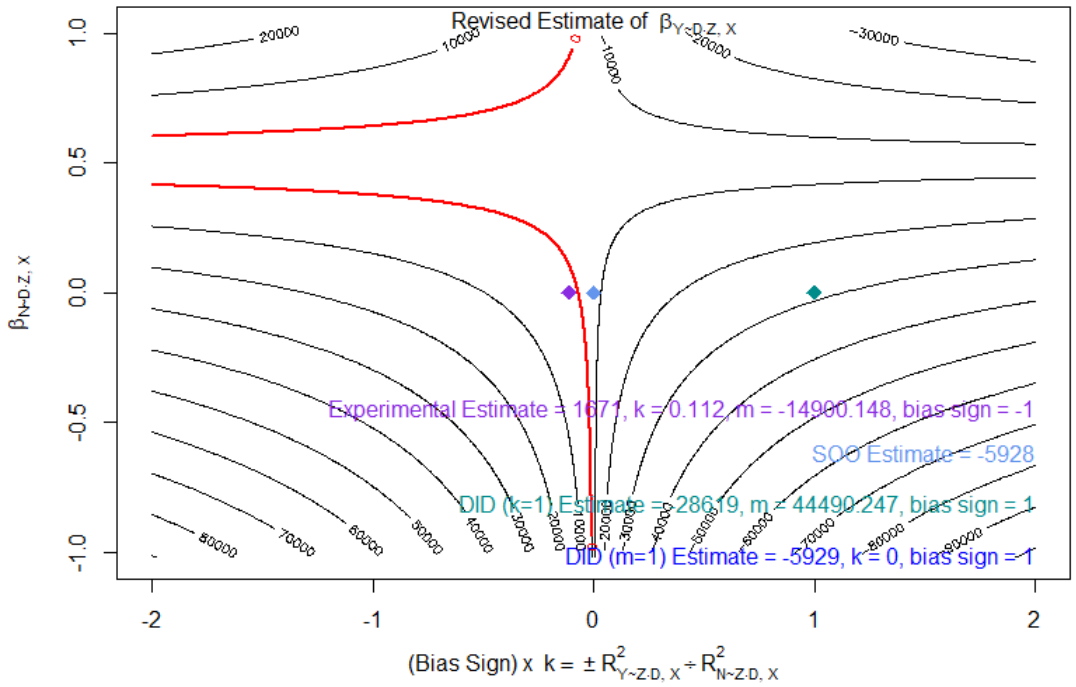
Maybe something with a non-pre-treatment placebo outcome

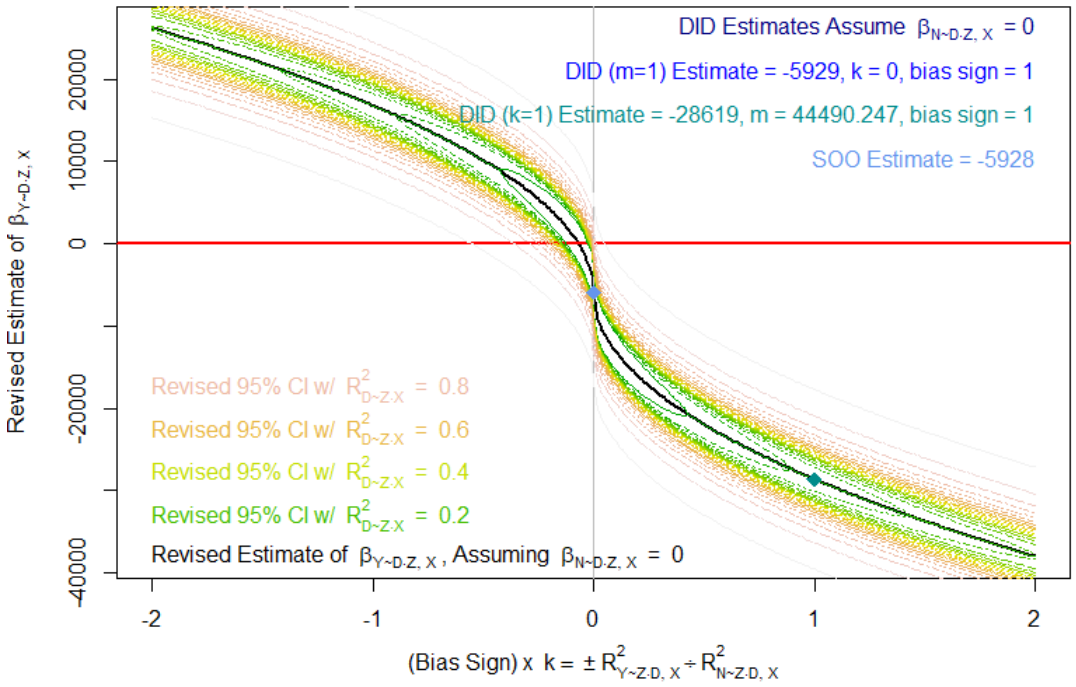
7 Software - placeboPartialID

[do we want to include a discussion of software as a section here or not?]

1. update to use γ and λ
2. add placebo treatment stuff - build in switches so it's the same funciton with a different arg

Figure 6: Partial identification plots for National Supported Work example with 1975 unemployment as placebo outcome.





3. add version where you can just input beta, se, dfs

4. find a better name: "placeboPartialID"
5. add summary feature - that reports the three simple estimates SOO, DID - figure out how to get se's for this, DID1, and RVs
6. figure out how to just say "plot(x)"
7. add complete documentation - for overall package and for each function
8. add more error handling and messages about incorrect inputs and unit testing
9. clean up and optimize code
10. add flexibility to alter charts
11. What esle would be good to include???
12. eventually publish to github and build a website
13. after building the R version add a Python version

8 Discussion

Discussion of changing the philosphical 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 insturments, 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 (does this make sense here or would this be an interesting note / review paper somewhere? can list the OLS OVB sensitivity analyses that we currently have - sensemakr, NP sensemakr, IV, placebo outcome and treatment, sample selection, DID first difference, mediation, others?)

8.1 Relation to other approaches

[do we need this section? we're trying to be more general than just DID and so far as I have seen, there are not really partial ID or sensitivity for NOCs]

DID can be seen as a special case of this approach – more on this in the next section

discuss First Difference DID Regression Sensitivity Analysis and just standard OVB Sensitivity Analysis - both could be used with the same data

discuss other unparallel trends approaches

discuss proximal CI and other NOC approaches

How does this relate to sensemakr for first difference regression?? it uses only one sensitivity parameter, though a different one.. here you're reasoning about the relative strength of the confounding relationship, not about the relationship strength directly - so this will be useful when we know most about the relationship between the unobserved confounders and the outcomes both placebo and actual; as ever, it shifts what sensitivity parameters we need to consider but doesn't eliminate the need to discuss some sensitivity parameters; no sensitivity analysis is better than the others, it's really what is easiest to reason about and what you have most information about in your particular context

include that DID and hence NOC can be thought of as a "detrended" IV and hence there is a connection to IV with violation in exclusion?

References

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

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.

Connors, Alfred F., J., Speroff, T., Dawson, N. V., Thomas, C., Harrell, Frank E., J., Wagner, D., Desbiens, N., Goldman, L., Wu, A. W., Califf, R. M., Fulkerson, William J., J., Vidaillet, H., Broste, S., Bellamy, P., Lynn, J., and Knaus, W. A. (1996). The Effectiveness of Right Heart Catheterization in the Initial Care of Critically Ill Patients. *JAMA*, 276(11):889–897.

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.

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.

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.

A Placebo treatments

A.1 Placebo treatments and the traditional omitted variables bias framework

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 captures the unobserved variables 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.[16] Further, suppose that we want to run the "long" regression of Y on D, P, X, and Z in Equation 51. 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.

$$Y = \beta_{Y\sim D|P,Z,X} D + \beta_{Y\sim P|D,Z,X} P + X\beta_{Y\sim X|D,P,Z} + Z\beta_{Y\sim Z|D,P,X} + \epsilon_l \tag{51}$$

As we mentioned, Z captures the unobserved variables that we would have liked to include in the regression but cannot. So we must run the "short" regression of Y on D, P and X in Equation 52, rather than the desired long regression. Z is a variable that is "omitted" from the regression, though we wish we could have included it.

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

Traditional omitted variable bias tells us that the bias in using $\beta_{Y\sim D|P,X}$ to estimate $\beta_{Y\sim D|P,Z,X}$ 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 the bias in using $\beta_{Y\sim P|D,X}$ to estimate $\beta_{Y\sim P|D,Z,X}$ can similarly be characterised as the product of two regression coefficients.

$$\widehat{\text{bias}_D} = \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = \beta_{Y\sim Z|D,P,X}\beta_{Z\sim D|P,X} \tag{53}$$

$$\widehat{\text{bias}_P} = \hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X} = \beta_{Y\sim Z|D,P,X}\beta_{Z\sim P|D,X} \tag{54}$$

As with placebo outcomes, it will often be useful to consider Z to be a rich enough set of variables to identify both the effect of D on Y and of P on Y, but the mechanics of our proposed approach do not rely on this. Suppose that we go ahead and estimate the short regression from Equation 52. It turns out that we can leverage information about the placebo treatment fairly easily by first solving for $\beta_{Z\sim P|D,X}$, then assuming that $\beta_{Z\sim D|P,X} = m \times \beta_{Z\sim P|D,X}$, and finally plugging into the expression for $\widehat{\text{bias}_D}$.

$$\beta_{Z\sim D|P,X} = m \times \beta_{Z\sim P|D,X} = m \times \frac{\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}}{\beta_{Y\sim Z|D,P,X}} \tag{55}$$

$$\implies \widehat{\text{bias}_D} = \beta_{Y\sim Z|D,P,X}\beta_{Z\sim D|P,X} = \beta_{Y\sim Z|D,P,X} \times \left[m \times \left(\frac{\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}}{\beta_{Y\sim Z|D,P,X}} \right) \right] = m \times \left(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X} \right) \tag{56}$$

Now we have an expression for $\widehat{\text{bias}_D}$ in terms of $m = \frac{\beta_{Z\sim D|P,X}}{\beta_{Z\sim P|D,X}} = \frac{\beta_{Y\sim Z|D,P,X}\beta_{Z\sim D|P,X}}{\beta_{Y\sim Z|D,P,X}\beta_{Z\sim P|D,X}} = \frac{\widehat{\text{bias}_D}}{\widehat{\text{bias}_P}}$, $\beta_{Y\sim P|D,Z,X}$, and $\hat{\beta}_{Y\sim P|D,X}$. We have already estimated $\hat{\beta}_{Y\sim P|D,X}$ from the data and so we are left with a bias expression in terms of two parameters. We can use this for partial identification by establishing reasonable ranges of values for m and $\beta_{Y\sim P|D,Z,X}$. m captures the level of relative confounding or the relative relationship between the treatments and the unobserved confounder. $\beta_{Y\sim P|D,Z,X}$ 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. Different Z's might also have different scales making the using m difficult. As in the case of placebo outcomes, we ask: can we do better than this?

[16] See Section 4.1 for further discussion.

A.2 Re-expressing $\widehat{\text{bias}_D}$

As before, 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 that we worried about in the previous section.

$$\begin{aligned} m = \frac{\beta_{Z\sim D|P,X}}{\beta_{Z\sim P|D,X}} &= \frac{R_{D\sim Z|P,X}}{R_{P\sim Z|D,X}} \times \frac{\left[\frac{\text{SD}(Z^{\perp P,X})}{\text{SD}(D^{\perp P,X})}\right]}{\left[\frac{\text{SD}(Z^{\perp D,X})}{\text{SD}(P^{\perp D,X})}\right]} \times \frac{\text{SD}(Z^{\perp D,P,X})}{\text{SD}(Z^{\perp D,P,X}} \\ &= \frac{R_{D\sim Z|P,X}/\sqrt{1-R^2_{D\sim Z|P,X}}}{R_{P\sim Z|D,X}/\sqrt{1-R^2_{P\sim Z|D,X}}} \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \quad = \frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}} \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \\ &\triangleq k \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \end{aligned} \tag{57}$$

Note that f is Cohen's f and equals $\frac{R}{\sqrt{1-R^2}}$. We can then plug Equation 57 into Equation 56 to arrive at an expression for $\widehat{\text{bias}_D}$ in terms of a ratio of scale-independent partial-correlation parameters.[17]

$$\widehat{\text{bias}_D} = k \times \left(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})} \tag{58}$$

Equation 58 looks similar to Equation 56, but k is not equal to m and we adjust for scale differences between D and P. The numerator and denominator of k 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 sensitivity parameter, especially when we have rough ideas of the values for the partial correlations.[18] Finally, we can 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 estimable from the observed data and is a function of commonly reported regression output. $\text{se}(\hat{\beta}_{Y\sim D|P,X})$ 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. Hence, $\frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})}$ can be estimated easily from commonly reported regression output for the short regression using available data only. Thus, we can also write $\widehat{\text{bias}_D}$ as in Equation 59.

$$\widehat{\text{bias}_D} = k \times \left(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}} \tag{59}$$

Now we have an expression for $\widehat{\text{bias}_D}$ where everything but $k = \frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}} = \frac{R_{Y\sim Z|D,P,X} f_{D\sim Z|P,X}}{R_{Y\sim Z|D,P,X} f_{P\sim Z|D,X}}$ [19] and $\beta_{Y\sim P|D,Z,X}$ can be estimated from the data. We can use this for partial identification by leveraging external information and subject matter expertise to determine plausible ranges of values for k and $\beta_{Y\sim P|D,Z,X}$. As before, $\beta_{Y\sim P|D,Z,X}$ 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. k can be interpreted in more than one way. First, we note that k can be rewritten as in Equation 60. This means we can express bias as in Equations 61a and 61b.

$$\begin{aligned} k = \frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}} &= \text{sign}\left(\frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}}\right) \times \frac{|f_{D\sim Z|P,X}|}{|f_{P\sim Z|D,X}|} = \text{sign}\left(\frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}}\right) \times \sqrt{\frac{f^2_{D\sim Z|P,X}}{f^2_{P\sim Z|D,X}}} = \gamma\sqrt{\lambda} \\ \text{where } \gamma \triangleq \text{sign}&\left(\frac{f_{D\sim Z|P,X}}{f_{P\sim Z|D,X}}\right) \text{ and } \lambda \triangleq \frac{f^2_{D\sim Z|P,X}}{f^2_{P\sim Z|D,X}} \end{aligned} \tag{60}$$

[17] See Appendix B.2 for an alternative derivation of Equation 58 that we will refer to in subsequent sections.

[18] We could also assume $R_{D\sim Z|P,X} = k \times R_{P\sim Z|D,X}$; this would result in a somewhat more complicated bias expression that contains $R^2_{Y\sim Z|D,P,X}$ as an additional sensitivity parameter.

[19] 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 $\beta_{Y\sim P|D,Z,X} = 0$.

$$\therefore \widehat{\text{bias}_D} = \gamma\sqrt{\lambda} \times \left(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}} \tag{61a}$$

$$\text{and } \widehat{|\text{bias}_D|} = \sqrt{\lambda} \times |\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}| \times \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}} \tag{61b}$$

For the signed bias, $\widehat{\text{bias}_D}$, we can reason about either k or both of γ and λ. For the unsigned bias, $\widehat{|\text{bias}_D|}$, we only need to reason about λ. As with placebo outcomes, whether we want to consider $\widehat{\text{bias}_D}$ or $\widehat{|\text{bias}_D|}$, we can reason either in terms of relative D, Z and P, Z relationships or in terms of relative biases (i.e., relative levels of confounding).

A.3 How standard errors change

As with placebo outcomes, we might also be interested in how the standard errors change as the partial identification parameters change. Again, it turns out this is fairly straightforward to do, but it comes at the cost of adding an additional partial identification parameter. We will take an approach similar to that in Appendix B.2. However, we will make an assumption that $R^2_{D\sim Z|P,X} = \lambda \times R^2_{P\sim Z|D,X}$, rather than $f^2_{D\sim Z|P,X} = \lambda \times f^2_{P\sim Z|D,X}$. This will alter our expression for $\widehat{|\text{bias}_D|}$, which will now include $R^2_{Y\sim Z|D,P,X}$ as an additional sensitivity parameter. Cinelli and Hazlett (2020) show the relationship between the standard error for $\hat{\beta}_{Y\sim D|P,X}$ and the standard error for $\hat{\beta}_{Y\sim D|P,X,Z}$. Our expression will also include $R^2_{Y\sim Z|D,P,X}$ as an additional sensitivity parameter. First, in Equation 62, we show the expression we get for $R^2_{D\sim Z|P,X}$; this comes from an expression for $\widehat{|\text{bias}_P|}$ with the same form as Equation 63.

$$R^2_{D\sim Z|P,X} = \lambda \times R^2_{P\sim Z|D,X} = \lambda \times \left[\frac{(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})^2}{(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X})^2 + R^2_{Y\sim Z|D,P,X} \times \left(\text{se}(\hat{\beta}_{Y\sim P|D,X}) \times \sqrt{\text{df}_P}\right)^2}\right] \tag{62}$$

Equation 62 gives us an expression for $R^2_{D\sim Z|P,X}$ in terms of quantities that are estimable from the data, λ, $\beta_{Y\sim P|D,Z,X}$, and $R^2_{Y\sim Z|D,P,X}$. These last three will be our partial identification parameters, which are similar to those we have been considering, with one addition. Importantly λ has a slightly different interpretation. Equation 62 can then be plugged in for $R^2_{D\sim Z|P,X}$ in the expressions for bias and for standard errors from Cinelli and Hazlett (2020).

$$\widehat{|\text{bias}_D|} = \text{se}(\hat{\beta}_{Y\sim D|P,X}) \times \sqrt{\text{df}_D \times R^2_{Y\sim Z|D,P,X} \times \frac{R^2_{D\sim Z|P,X}}{1 - R^2_{D\sim Z|P,X}}} \tag{63}$$

$$\text{se}(\hat{\beta}_{Y\sim D|P,X}) = \text{se}(\hat{\beta}_{Y\sim D|P,X,Z}) \sqrt{\frac{1 - R^2_{Y\sim Z|D,P,X}}{1 - R^2_{D\sim Z|P,X}} \times \frac{\text{df}_D}{\text{df}_D - 1}} \tag{64}$$

We do not show this substitution explicitly as it does not aid in understanding. Suffice to say that we then have both a bias expression and an expression for how standard errors change. The additional sensitivity parameter is necessary to determine how standard errors change. Since this parameter is native to the type of omitted variables analysis discussed in Cinelli and Hazlett (2020), we do not believe its addition to be burdensome. We alter the meaning of λ so that we can get both bias and standard error expressions that contain it.

A.4 Sensitivity analysis and robustness values

It is again possible for researchers to start from traditional placebo treatment point identification assumptions and then consider how sensitive their effect estimates are to violations of those assumptions. An effect estimate based on point identification assumptions would have the form of Equation 65, where the specific values γ^*, λ^*, and $\beta^*_{Y\sim P|D,Z,X}$ are implied by the point identification assumptions (or are the assumptions themselves). In Equation 65, $\hat{\beta}_{Y\sim D|P,Z,X}$ (note the $\hat{}$) is the researcher's estimate of $\beta_{Y\sim D|P,Z,X}$, which may be biased and not equal to $\beta_{Y\sim D|P,Z,X}$.

$$\text{Effect Estimate} = \hat{\beta}_{Y\sim D|P,Z,X} = \hat{\beta}_{Y\sim D|P,X} - \gamma^*\sqrt{\lambda^*} \times (\hat{\beta}_{Y\sim P|D,X} - \beta^*_{Y\sim P|D,Z,X}) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}} \tag{65}$$

For example, a simple selection on observables (SOO) effect estimate would be $\hat{\beta}_{Y\sim D|P,X}$, which corresponds to an assumption that $\lambda^* = 0$ (i.e., no unobserved confounding). A traditional difference in differences (DID) effect estimate would be $\hat{\beta}_{Y\sim D|P,X} - \hat{\beta}_{Y\sim P|D,X}$, which corresponds to an assumption that $\lambda^* = \left(\frac{\text{se}(\hat{\beta}_{Y\sim P|D,X})}{\text{se}(\hat{\beta}_{Y\sim D|P,X})} \times \sqrt{\frac{\text{df}_P}{\text{df}_D}}\right)^2$, $\gamma^* = +1$, and $\beta^*_{Y\sim P|D,Z,X} = 0$. Researchers might also consider a version of DID where we assume $\lambda^* = 1$.

From Equation 65, it's easy to see that we can find the value for λ or $\beta_{Y\sim P|D,Z,X}$ that would mean the true effect is equal to $100 \times q\%$ times the estimated effect. This would happen if Equation 66 holds, since then the true effect would equal $\beta_{Y\sim D|P,Z,X} = \hat{\beta}_{Y\sim D|P,X} - [\hat{\beta}_{Y\sim D|P,X} - (1-q) \times \hat{\beta}_{Y\sim D|P,Z,X}] = (1-q) \times \hat{\beta}_{Y\sim D|P,Z,X}$.

$$\hat{\beta}_{Y\sim D|P,X} - (1-q) \times \hat{\beta}_{Y\sim D|P,Z,X} = \gamma\sqrt{\lambda} \times (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}) \times \frac{\text{se}(\hat{\beta}_{Y\sim D|P,X})}{\text{se}(\hat{\beta}_{Y\sim P|D,X})} \times \sqrt{\frac{\text{df}_D}{\text{df}_P}} \tag{66}$$

The values for λ and $\beta_{Y\sim P|D,Z,X}$ that mean the true effect is equal to $100 \times q\%$ times the estimated effect, λ_q and β_q, while allowing the other parameters to remain at their assumed levels are shown in Equations 67.[20]

$$\lambda_q = \left(\frac{\hat{\beta}_{Y\sim D|P,X} - (1-q) \times \hat{\beta}_{Y\sim D|P,Z,X}}{\hat{\beta}_{Y\sim P|D,X} - \beta^*_{Y\sim P|D,Z,X}} \times \frac{\text{se}(\hat{\beta}_{Y\sim P|D,X})}{\text{se}(\hat{\beta}_{Y\sim D|P,X})} \times \sqrt{\frac{\text{df}_P}{\text{df}_D}}\right)^2 \triangleq \text{RV}_{\lambda,q} \tag{67a}$$

$$\beta_q = \hat{\beta}_{Y\sim P|D,X} - \frac{\hat{\beta}_{Y\sim D|P,X} - (1-q) \times \hat{\beta}_{Y\sim D|P,Z,X}}{\gamma^*\sqrt{\lambda^*}} \times \frac{\text{se}(\hat{\beta}_{Y\sim P|D,X})}{\text{se}(\hat{\beta}_{Y\sim D|P,X})} \times \sqrt{\frac{\text{df}_P}{\text{df}_D}} \triangleq \text{RV}_{\beta,q} \tag{67b}$$

The robustness values $\text{RV}_{\lambda,q}$ and $\text{RV}_{\beta,q}$ can be interpreted and used as discussed for placebo outcomes.

B Alternative derivation of bias reparameterizations

B.1 Placebo outcomes

We could also leverage a reparameterization of the overall traditional omitted variable bias framework that is similar to that discussed in Cinelli and Hazlett (2020) to arrive at our expression for $\widehat{\text{bias}_Y}$. A simple re-expression yields an expression for $\widehat{\text{bias}_Y}$ in terms of scale-independent partial-correlation parameters, $R_{Y\sim Z|D,X}$ and $f_{D\sim Z|X} = R_{D\sim Z|X}/\sqrt{1 - R^2_{D\sim Z|X}}$, rather than regression coefficients. Similarly, we can write $\widehat{\text{bias}_N}$ in terms of $R_{N\sim Z|D,X}$ and $f_{D\sim Z|X} = R_{D\sim Z|X}/\sqrt{1 - R^2_{D\sim Z|X}}$. As in the traditional omitted variable bias setting, we can leverage information about the negative outcome by first solving for $R_{N\sim Z|D,X}$, then assuming that $R_{Y\sim Z|D,X} = k \times R_{N\sim Z|D,X}$, and finally plugging into the expression for $\widehat{\text{bias}_Y}$. This yields exactly the same expression for $\widehat{\text{bias}_Y}$ as in Section 2.1.2.

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

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

$$\text{Assume } R_{Y\sim Z|D,X} = k \times R_{N\sim Z|D,X} = k \times (\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}) \times \frac{1}{f_{D\sim Z|X}} \times \frac{\text{SD}(D^{\perp X})}{\text{SD}(N^{\perp D,X})}.$$

$$\begin{aligned}
\implies \widehat{\text{bias}_Y} &= \hat{\beta}_{Y\sim D|X} - \beta_{Y\sim D|Z,X} \\
&= \left[k \times (\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}) \times \frac{1}{f_{D\sim Z|X}} \times \frac{\text{SD}(D^{\perp X})}{\text{SD}(N^{\perp D,X})}\right] \times f_{D\sim Z|X} \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(D^{\perp X})} \\
&= k \times (\hat{\beta}_{N\sim D|X} - \beta_{N\sim D|Z,X}) \times \frac{\text{SD}(Y^{\perp D,X})}{\text{SD}(N^{\perp D,X})}
\end{aligned}$$

B.2 Placebo treatments

Again, we could also leverage a reparameterization of the overall traditional omitted variable bias framework that is similar to that discussed in Cinelli and Hazlett (2020) to arrive at our expression for $\widehat{\text{bias}_D}$. A simple re-expression

[20] γ^* drops out of Equation 67a because it is squared.

yields an expression for $\widehat{\text{bias}_D}$ in terms of scale-independent partial-correlation parameters, $R_{Y\sim Z|D,P,X}$ and $f_{D\sim Z|P,X} = R_{D\sim Z|P,X}/\sqrt{1-R^2_{D\sim Z|P,X}}$, rather than regression coefficients. Similarly, we can write $\widehat{\text{bias}_P}$ in terms of $R_{Y\sim Z|D,P,X}$ and $f_{P\sim Z|D,X} = R_{P\sim Z|D,X}/\sqrt{1-R^2_{P\sim Z|D,X}}$. As in the traditional omitted variable bias setting, we can leverage information about the placebo treatment by first solving for $f_{P\sim Z|D,X}$, then assuming that $f_{D\sim Z|P,X} = k \times f_{P\sim Z|D,X}$, and finally plugging into the expression for $\widehat{\text{bias}_D}$. This yields exactly the same expression for $\widehat{\text{bias}_D}$ as in Section A.2.

$$\widehat{\text{bias}_D} = \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} = R_{Y\sim Z|D,P,X} \times f_{D\sim Z|P,X} \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})}$$

$$\widehat{\text{bias}_P} = \hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X} = R_{Y\sim Z|D,P,X} \times f_{P\sim Z|D,X} \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(P^{\perp D,X})}$$

$$\text{Assume } f_{D\sim Z|P,X} = k \times f_{P\sim Z|D,X} = k \times (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}) \times \frac{1}{R_{Y\sim Z|D,P,X}} \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(Y^{\perp D,P,X})}.$$

$$\begin{aligned}
\implies \widehat{\text{bias}_D} &= \hat{\beta}_{Y\sim D|P,X} - \beta_{Y\sim D|P,Z,X} \\
&= R_{Y\sim Z|D,P,X} \times \left[k \times (\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}) \times \frac{1}{R_{Y\sim Z|D,P,X}} \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(Y^{\perp D,P,X})}\right] \times \frac{\text{SD}(Y^{\perp D,P,X})}{\text{SD}(D^{\perp P,X})} \\
&= k \times \left(\hat{\beta}_{Y\sim P|D,X} - \beta_{Y\sim P|D,Z,X}\right) \times \frac{\text{SD}(P^{\perp D,X})}{\text{SD}(D^{\perp P,X})}
\end{aligned}$$

C Partially linear and non-parametric framework

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.

C.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.

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

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

S^2_Y and S^2_N 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).[21] $\eta^2_{Y\sim Z|D,X}$ 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^2_Y and S^2_N.

We can see that

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

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

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

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

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^*.

C.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^*.