# *University of California, Berkeley*
## U.C. Berkeley Division of Biostatistics Working Paper Series

# Challenges in Estimating the Causal Effect of an Intervention with Pre-Post Data (Part 1): Definition & Identification of the Causal Parameter

Ann M. Weber[*]        Mark J. van der Laan[†]

Maya L. Petersen[‡]

[*]Stanford University, annweber@stanford.edu

[†]University of California, Berkeley, Division of Biostatistics, laan@berkeley.edu

[‡]University of California, Berkeley, mayaliv@berkeley.edu

# Challenges in Estimating the Causal Effect of an Intervention with Pre-Post Data (Part 1): Definition & Identification of the Causal Parameter

Ann M. Weber, Mark J. van der Laan, and Maya L. Petersen

## Abstract

There is mixed evidence of the effectiveness of interventions operating on a large scale. Although the lack of consistent results is generally attributed to problems of implementation or governance of the program, the failure to find a statistically significant effect (or the success of finding one) may be due to choices made in the evaluation. To demonstrate the potential limitations and pitfalls of the usual analytic methods used for estimating causal effects, we apply the first half of a roadmap for causal inference to a pre-post evaluation of a community-level, national nutrition program. Selection into the program was non-random and strongly associated with the pre-treatment (lagged) outcome. Using structural causal models (SCM), directed acyclic graphs (DAGs) and simulated data, we demonstrate that a post treatment estimand controls for confounding by the lagged outcome but not from possible unmeasured confounders. Two separate difference-in-differences estimands have the potential to adjust for a certain type of unmeasured confounding, but introduce bias if the additional assumptions they require are not met. Our results reveal an important issue of identifiability when estimating the causal effect of a program with pre-post observational data. A careful appraisal of the assumptions underlying the causal model is imperative before committing to a statistical model and progressing to estimation.

## 1. Introduction

Interventions scaled-up to a large or national level have failed to consistently demonstrate the causal benefits anticipated by results of small-scale experimental studies (Engle et al. 2011). Challenges in evaluating a program operating at-scale are not limited to the logistical and technical constraints of surveying hundreds to thousands of households across a region or country, but also include the analytic process of determining whether the program demonstrates a benefit that actually is the result of the program. Our ability to make a causal claim from an evaluation may easily be compromised by the choices we make in the analysis, and are particularly complex in observational studies. These choices are made even more controversial with the availability of pre-treatment outcome data, as we will demonstrate in this paper. Misleading estimates of a program's benefit (in either direction) have significant policy and funding implications for the program, as well as for the people the program is intended to help.

In an introductory chapter on econometric evaluations of social programs, Nobel laureate James Heckman and co-author Edward Vytlacil point out that we often confuse the three main issues that face an evaluation: definition, identification and estimation. The authors state that "particular methods of estimation (e.g., matching or instrumental variable estimation) have become associated with 'causal inference' and even the definition of certain 'causal parameters' ..." (Heckman and Vytlacil 2007). Investigators from different disciplines will bring distinct theoretical and analytical frameworks to estimation, which can lead to differing estimates of the causal effect and contradictory conclusions, in some cases without strong theoretical justification for the approach they used. However, the selection of an estimator should happen after defining the research question and causal target parameter, and after the underlying assumptions necessary to identify the parameter are made explicit. In keeping with this logic, we divide the evaluation of an existing intervention into two papers. In this first paper, we work step-by-step through the first part of a roadmap for causal inference (van der Laan and Rose 2011), and present the second half of the roadmap, which includes the methods of estimation and inference, separately. In doing so, we hope to underscore the need to "define first, identify second, and estimate last" (quote from Judea Pearl's forward in *Targeted Learning* by van der Laan and Rose) (van der Laan and Rose 2011; Pearl 2010).

We make use of a program in Madagascar as a backdrop for exploration. The Madagascar national nutrition program presents some interesting challenges that are common in evaluations of large-scale interventions. First, the program was implemented at the community rather than the individual level. Community programs differ from individual treatment regimens in that they are typically made available to all (or most) residents of a community and sharing of information within a community is often encouraged. Examples of other community programs include national child health days (Alderman 2007), breast-feeding promotion campaigns (Popkin et al. 1991; Bhutta et al. 2008), and conditional cash transfers (Paxson and Schady 2008; Fernald, Gertler, and Neufeld 2008; Macours, Schady, and Vakis 2008). When the point of treatment is the community, the research question shifts from the more familiar individual-level treatment to what would happen to the community under a given treatment assignment. In this context, we need to reframe the question, building up by analogy from the individual to the group. We use the potential outcomes framework (also known as the counterfactual framework) popularized by the work of Rubin to help in this regard (Rubin 1973). In a counterfactual framework, we consider the

1

"ideal experiment" when posing our research question. For example, what would have happened to a given community had it received treatment (the counterfactual) when in fact it had not?

A second challenge was that the program rollout was non-random: treatment assignment was made in such a way that communities with the greatest perceived need were to receive the program first. Communities were selected for treatment if they were located within districts with a pre-program prevalence of underweight[1] that was above the national average, or if they met certain logistical criteria (e.g., a local non-profit organization was available to supervise the program). Non-random assignment of treatment makes inferences about the programs' effect susceptible to confounding if the comparison group is not exchangeable with the treated group on key determinants of the outcome. Therefore, it is imperative that we define a causal model for the system that is hypothesized to have generated the data and to examine clearly the relations and dependencies of the factors in the model (measured and unmeasured). We use a semi-parametric variant of a structural equation model for our causal model to avoid making assumptions about the underlying functional form of the data distribution (Pearl 1995, 2010). In addition, we use graphical models (directed acyclic graphs or DAGs) to make the assumptions underlying our causal models transparent. We demonstrate how DAGs can be used for locating sources of dependencies among variables.

Third, cross-sectional surveys were administered in the same communities in Madagascar pre- and post-intervention, providing multiple options for identification of a causal effect. In this paper, we contrast the definition of the outcome as either: the post treatment value or the change from pre to post treatment. We consider the long-standing controversy over the advantages and disadvantages of each (Imai 2008; Maris 1998). Using these two outcomes, we identify three statistical parameters used for interventions with pre-post data that under different assumptions are equivalent to our causal target parameter of interest. We purposefully include two difference-in-differences models commonly used with pre-post data: a change score estimand and a pooled outcome estimand (popular in the social sciences and econometrics) (Meyer 1994; Gertler et al. 2011; Lord 1956; Imai 2008). We also include a conventional approach from the epidemiology literature in which the pre-intervention outcome (or lagged outcome) is included in the conditioning set of covariates.

Finally, we present a series of data simulations, and show how the estimate of the target causal parameter diverges from the truth when the necessary assumptions for a given model fail to hold. Although the context for this paper is specific to the Madagascar study, the process is applicable to any program evaluation for which the investigator seeks to interpret an estimated effect of treatment on outcome as a causal effect.

## 2. Setting & Notation

In Madagascar, approximately 30% of children under five are estimated to be underweight (UNICEF 2011). Underweight is a near-term marker for inadequate nutrition and is estimated to be responsible for the largest proportion of the death and disease burden associated with malnutrition (Black et al. 2008). In 1999, the Madagascar National Office of Nutrition (ONN) implemented a comprehensive community-level growth-monitoring and nutrition program, incorporating multiple activities that have been found to

---

[1] Underweight is an indicator of being two standard deviations below the median weight of a reference population of well-nourished and healthy children of the same age and gender.

be associated with better child outcomes (Galasso and Umapathi 2007). The project has since expanded to include 5550 sites and covers approximately 1.1 million children (or about a third of children under 5 years of age in Madagascar) (Sharp and Kruse 2011).

The evaluation of the Madagascar program included a series of repeated cross-sectional, nationally representative, anthropometric surveys, administered pre- and post-implementation of the program, in both program participating and non-participating communities. Each survey included different children, but the same communities. The type of data collected is described in general terms in Table 1. Specifics of the data collected are given in our second paper on estimation.

Table 1: Notation used for variables, parameters, and outcomes

| Notation | Description |
|---|---|
| $V$ | Vector of time invariant community level covariates (e.g., urban location) |
| $W^c(t)$ | Vector of community level covariates that summarize individual level factors, $W_i(t)$, (i=1,…,N), for each of the N individuals sampled in the community at time $t = 0,1$ (e.g., proportion of mothers sampled in the community who are uneducated) |
| $A$ | Treatment, assigned at the community level |
| $Y^c(t) = \dfrac{1}{N}\sum\limits_{i=1}^{N} Y_i(t)$ | Community mean of individual level outcomes $Y_i(t)$ (i=1,…,N) for each of the N individuals sampled in the community at time $t = 0,1$ (e.g., weight-for-age of children under 5 years) |
| $O_j = (V_j, W^c_j(t), A_j, Y^c_j(t))$ | Observed data structure, $O_j$, for a given community j. The observed data are J independently and identically distributed copies of O. |
| $U_V, \ldots U_{Y(t)}$ | Random variation for each variable |
| $P_0$ | True data-generating distribution; $O_j \sim P_0$ |
| $Y^c_a, Y^\theta_a$ | Counterfactual outcomes; we focus on 2 outcomes: post treatment outcome, $Y^c(t = 1)$, and the change in outcome pre and post treatment, $Y^\theta = Y^c(t = 1) - Y^c(t = 0)$. For each, we define their counterfactual value under treatment level $A=a$ ($Y^c_a$ and $Y^\theta_a$, respectively). |
| $\Psi(P_0)$ | True value of the target statistical parameter (or estimand), consisting of parameter mapping $\Psi$ applied to the true data generating distribution $P_0$. We present 3 estimands labeled $\Psi^I$, $\Psi^{II}$, and $\Psi^{III}$. |
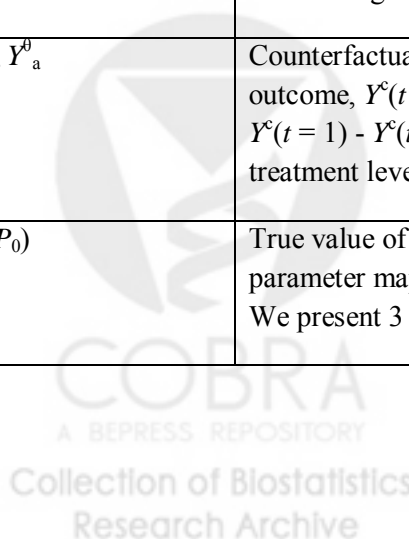
Our outcome of interest is community mean weight-for-age[2] for children under 5 years, one of the primary nutritional outcomes in children targeted by the program, and also a key determinant of the initial program implementation. In a prior analysis of the Madagascar program, the authors found that the program reduced the prevalence of child underweight in treated communities during a period of worsening malnutrition in non-treated communities (between 1997/98 and 2004) (Galasso and Umapathi 2009). We revisit these results in the context of a detailed framework for evaluation.

We use the notation shown in Table 1 based on the book on *Targeted Learning* by van der Laan and Rose (van der Laan and Rose 2011). The notation has been modified to indicate how we aggregate individual level measures up to the community level (e.g., mean maternal education), and that these are different from community variables that are considered time-independent and are measured once for the entire group (e.g., geographic location).

### 3. Causal Inference Road Map

The road map we follow links the research question to inference, making the underlying assumptions explicit for the path between the two (see van der Laan and Rose, chapters 1 and 2 for more detail) (van der Laan and Rose 2011). First, we define precisely the research question. This may seem obvious, but is often not made clear. Second, we turn the research question and relevant background knowledge into a structural causal model (SCM) (Pearl 2010), which encodes information about the relationships between the variables with a series of semi-parametric equations. Importantly, we assume that the SCM accurately represents the data generating processes that gave rise to our observed data. This is the key link from counterfactual to observed data.

Given the SCM, we specify the causal parameter of interest in the third step. The causal parameter is the parameter we would obtain under an ideal experiment and is defined using counterfactual notation. A clear specification of the causal parameter requires an understanding of the outcome; the variable(s) on which we want to intervene; the unit (or level) on which we are intervening; and the counterfactual outcome distributions (or parameters of these distributions) we want to compare. In the Madagascar case, we explore two outcomes: a post treatment value and a pre-post change score, and we want to intervene on the program availability at the community level (as opposed to program participation by an individual). We use the phrasing "intervening to set the treatment" or "setting $A=a$" to refer to the hypothetical treatment condition that we want to apply to the system when making causal contrasts. In this paper, we are interested in estimating the difference in the expectation (or mean) of counterfactual outcomes intervening to set the treatment to 1 (to receive treatment) versus intervening to set the treatment to 0 (to not receive treatment), for all communities. This contrast is known as the average treatment effect, or the ATE. We could evaluate other causal parameters of interest, such as the average treatment effect among the treated, or the ATT. The ATT contrasts the expectation of counterfactual outcomes under treatment and no treatment, but only among the treated communities. Importantly for either parameter, the contrast is made between the means of the counterfactual outcomes under each treatment regime. This is a simpler causal comparison than between the two potential outcomes for any given community, where one outcome is always unobserved.

---

[2] Weight-for-age z-score are obtained using a reference population of well- nourished and healthy children of the same age and gender.

In the fourth step, we assess identifiability, or whether the observed data, in combination with our assumptions about the data generating system, are sufficient to express the target casual parameter of interest as a parameter of the distribution of the observed data alone. This second parameter is the statistical target parameter (also referred to as the estimand; we use the terms interchangeably). In contrast to the causal parameter, the estimand is the parameter that we are able to estimate given the observed data. Because we have pre-post data, we evaluate the assumptions for three different statistical target parameters. In the first estimand, the outcome is defined as the outcome post treatment ($Y^c(t = 1)$); in the second, the outcome is defined as the change in outcome pre- vs. post-intervention ($Y^\theta$); and in the third, the outcome is pooled over time ($Y^c(t)$).

In the last steps of the roadmap, we can commit to an estimand and statistical model and proceed with the estimation. We will do so in a separate paper, presenting estimation and inference results for the observed data from Madagascar. However, in this paper, we use simulations to illustrate the different assumptions required for the three statistical parameters to be equivalent to the average treatment effect (ATE), our target parameter of interest, and the consequences when the assumptions do not hold. Clearly, the ATE is just one possible causal parameter of interest that we could have explored, but it is of interest in many health studies. We chose the ATE to demonstrate how selecting a statistical model without understanding the underlying assumptions can threaten the validity of a causal effect estimate. The choice of a different causal parameter would not eliminate this threat. We present detailed steps 1 through 4 of the roadmap next.

### 3.1 Roadmap Steps 1-3: The Research Question, Target Causal Parameter & SCM

Our causal question is: Does the intervention increase the average nutritional status of children living in the community? In this paper, we are interested in estimating a population average effect at the community level, for all communities in the target population.

The structural causal model (SCM) is characterized by a set of endogenous variables at two time points (see notation Table 1). Community variables that are not aggregates of individual factors are denoted by $V$, and are assumed to be time-invariant for the period of the study. Individual level factors aggregated up to community-level factors are denoted by a vector, $W^c(t)$, at time $t$. The community-level mean outcome for children at time t is denoted as $Y^c(t)$. The community level exposure, $A$ is assigned to zero or one as a function of $V$, $W^c(t = 0)$ and $Y^c(t = 0)$. In addition, there are unmeasured exogenous variables, $U$, that may cause random variation in each of the observed variables. Restrictions on the joint distribution of these unmeasured errors will be required for identifiability.

We pose the following SCM to explain the relationships between the variables:

$V = f_V(U_V)$

$W^c(t = 0) = f_{W(t = 0)}(V, U_{W(t = 0)})$

$Y^c(t = 0) = f_{Y(t = 0)}(V, W^c(t = 0), U_{Y(t = 0)})$

$A = f_A(V, W^c(t = 0), Y^c(t = 0), U_A)$

$W^c(t = 1) = f_{W(t = 1)}(V, W^c(t = 0), Y^c(t = 0), U_{W(t = 1)})$

$Y^c(t = 1) = f_{Y(t = 1)}(V, W^c(t = 0), Y^c(t = 0), A, W^c(t = 1), U_{Y(t = 1)})$,

5

where no assumptions are made about the shape or form of the functions. We start with a model with a minimal set of exclusion restriction assumptions about the data-generating system in order to avoid imposing restrictions that may, or may not be, supported by the data. We make a single exclusion restriction in this model: that the covariates $W^c(t = 1)$ occurring post intervention are not affected by the intervention. We impose this deliberate exclusion restriction for three reasons. First, it is a reasonable assumption in the context of the Madagascar study. Second, it is required for the estimand with the outcome pooled over time (see identifiability section for estimand III), and we apply it to the other two estimands to facilitate our comparison across estimands (although it is not required). Finally, it allows us to condition on $W^c(t = 1)$ in the models to better predict $Y^c(t = 1)$ (also not required for the first two estimands).

In the ideal experiment, we would want to know what would happen to the population mean outcome, if every community had the program, versus none of the communities had the program. We translate this into our target causal parameter as the average treatment effect (ATE) given by: $E(Y^c_1(t = 1) - Y^c_0(t = 1))$, where $Y^c_a(t)$ denotes the counterfactual community level outcome under an intervention on the SCM setting $A=a$. In the next step, we describe three identifiability results (and corresponding estimands) where we link this causal parameter to our observed data distribution.

**3.2 Roadmap Step 4: Assess Identifiability**
Causal effect estimation relies on assumptions, some of which cannot be tested. These assumptions must be made explicit when using observational data for causal inference. Specifically, the identifiability of our causal target parameter requires some form of the following two assumptions to hold: the randomization assumption (RA) and the experimental treatment assignment (ETA) assumption.

The RA (also known as the assumption of no unmeasured confounders, or of exchangeability), states that treatment, $A$, is independent of counterfactual outcome, $Y_a$, given some subset of the data. The RA is a causal assumption, and as such is not testable. However, we can draw a graphical representation of our SCM (i.e., a DAG) to check the independence assumptions given our knowledge of the underlying data generating system (Pearl 1995, 2010). By using a graphical procedure, we are able to solve the identification problem without resorting to an algebraic analysis of whether a statistical model parameter has a unique solution in terms of the parameters of the distribution of the observed variables (Pearl 2010). Detailed guidelines for reading causal diagrams are available in *An Introduction to Causal Inference* by Judea Pearl (Pearl 2010), or in *Causal Diagrams for Epidemiological Research* by Sander Greenland et. Al (Greenland, Pearl, and Robins 1999). Very briefly, the graph is drawn based on the relationships defined in the SCM, where the parents of a variable (variables on the right hand side of the equation) are connected to the child variable (variable on the left hand side of the equation) with an arrow directed towards it. A path is any sequence of lines connecting two variables. The arrow between two variables can only go in one direction, such that the paths are acyclic (i.e., the graph cannot have $A \rightarrow B \rightarrow C \rightarrow A$). Paths can either be open or blocked, depending on the direction of the arrows and whether or not a variable is conditioned on. Conditioning on a variable is represented by placing a box around it. Open paths can give rise to dependency between variables, and the absence of any open paths implies marginal independence.

The specific randomization assumption (RA) and necessary additional assumptions for our three estimands are discussed in detail below. To minimize confusion from too many arrows, we represent DAGs for each estimand using a simplified data structure that omits the observed, time-invariant, village

6

factors, *V*. We justify this simplification because *V* are exogenous to the data generating system (no arrows go into *V*) and if we condition on *V*, we do not have to worry about unblocked paths from unmeasured variables through *V*. In most cases, we also omit the exogenous variables, *U*, such that $O_j = (W^c(t), A, Y^c(t))$. The omission of the *U*'s implies that these exogenous variables are independent (discussed further with Figure 1). Paths depicted in red in the figures represent unblocked paths between the treatment and outcome variables.
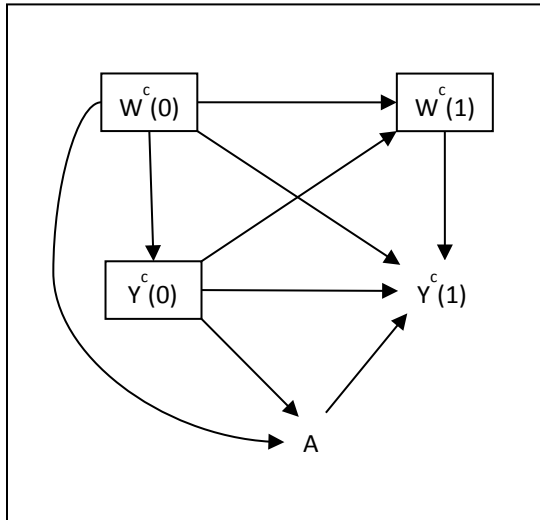


Figure 1: DAG illustrating that post-treatment outcome, $Y^c(1)$, is independent of treatment, *A*, given lagged outcome, $Y^c(0)$, pre- and post-treatment covariates, $W^c(0)$ and $W^c(1)$, and exogenous covariates, *V* (not shown). There are no unmeasured confounders.

The ETA assumption (also known as the positivity assumption) states that for the target statistical parameter to be identified there must be sufficient variation in treatment (i.e., some positive probability of both being treated and not being treated) within strata of confounders. The form of the ETA assumption depends on knowledge of the data-generating system encoded in the SCM and on the target parameter. For the average treatment effect, the strong positivity assumption states that each possible treatment level occurs with some positive probability within each stratum of the confounders (Petersen et al. 2011). But this can be weakened under additional parametric assumptions. For example, urban versus rural location is a confounder in our study in Madagascar. The strong version of the ETA assumption requires that we have both treated and untreated, urban and rural communities in our observed data. If, in fact, there were no observed treated urban communities, then we could weaken the ETA assumption by assuming (if plausible) that the treatment effect is the same among urban and rural communities. However, imposing this type of parametric assumption is risky as it requires extrapolating from an area supported by the observed data (the treatment effect among rural communities) to an area that is not (the treatment effect among urban communities) (Petersen et al. 2011). The ETA assumption will be discussed in more detail in the next paper in the context of the actual data from Madagascar. For the purposes of this paper, we accept that the ETA assumption is not violated in our study.

There are two additional assumptions that are typically invoked when investigators start from the Rubin framework of potential outcomes for causal inference: the consistency assumption and the stable unit treatment value assumption (SUTVA) (Robins, Hernán, and Brumback 2000). Both assumptions are

7

subsumed in our SCM and the implied knowledge it encodes about the underlying data generating distribution. The consistency assumption states that an individual's (or community's) potential outcome under the treatment actually received is precisely the observed outcome (Robins, Hernán, and Brumback 2000). This assumption is used to convert probabilities written in terms of counterfactuals into ordinary probabilities in terms of observed values. However, our SCM already implies the counterfactual and provides the necessary link to the observed data. In addition, the absence of hierarchical relationships between communities in our SCM implies that one community's (or individual's) outcome is unaffected by another's treatment assignment (i.e., SUTVA holds).

### 3.3 Estimand I: Outcome $Y^c(t = 1)$

For the first estimand, we define the outcome as the community specific mean post-treatment outcome, $Y^c(t = 1)$. Identifiability is based on conditioning on all baseline covariates, including the pre-treatment (or lagged) outcome (as well as the post treatment covariates $W^c(t = 1)$ assumed not to be affected by $A$, as discussed above). The randomization assumption for this estimand is:

(1) $Y_a^c(t = 1) \perp A | V, W^c(t = 0), Y^c(t = 0), W^c(t = 1)$

For the RA (1) to hold, it is sufficient that the exogenous variables for the exposure, $U_A$, be independent of the exogenous variables for the outcome, $U_{Y(t=1)}$, given $V$, $W^c(t = 0)$, $Y^c(t = 0)$, $W^c(t = 1)$. This additional independence assumption is reasonable, if we have no unmeasured common causes of $A$ and $Y^c(t = 1)$ (i.e., no confounders).

The DAG in Figure 1 encodes the information from the series of equations from the SCM in the previous section. The graphical model is particularly useful in that we can visually check that $Y^c(t = 1)$ is independent of $A$ given $W^c(t = 0)$, $Y^c(t = 0)$, and $W^c(t = 1)$. Specifically, we check that our conditioning variables block any unblocked path from $A$ to $Y^c(t = 1)$ (i.e., paths with arrows pointing into $A$), while not opening any new paths. This is referred to as the backdoor criterion (Greenland, Pearl, and Robins 1999). In Figure 1, the variables $W^c(t = 0)$, $Y^c(t = 0)$, and $V$ (not shown) are conditioned on, block the paths from $A$ to $Y^c(t = 1)$, and satisfy the backdoor criterion. Writing the graph in this way implies the independence assumptions among the exogenous variables, $U$, described previously. The RA (1) holds under this model.

We now have the following identifiability result:

$E(Y^c_a(t = 1) \mid V, W^c(t = 0), Y^c(t = 0), W^c(t = 1)) =$
$E(Y^c_a(t = 1) \mid A=a, V, W^c(t = 0), Y^c(t = 0), W^c(t = 1)) =$
$E(Y^c(t = 1) \mid A=a, V, W^c(t = 0), Y^c(t = 0), W^c(t = 1))$

where the first equality holds under the RA(1), and the second holds under our definition of the counterfactual outcomes. Note that for these conditional expectations of the outcome to be well-defined in our SCM, we need some communities with and without the treatment for each level of the conditioning variables $V$ and $W^c(t)$ (i.e., we need for the positivity assumption to hold).

A first estimand (or statistical parameter) for the average treatment effect, $\Psi^I$, follows:

$$(2) \ \Psi^I(P_0) = E_{V,W(t=0),Y(t=0),W(t=1)} \left( \begin{array}{l} E\big(Y^c(t = 1)|A = 1, V, W^c(t = 0), Y^c(t = 0), W^c(t = 1)\big) \\ -E\big(Y^c(t = 1)|A = 0, V, W^c(t = 0), Y^c(t = 0), W^c(t = 1)\big) \end{array} \right)$$

We refer to this estimand as the post treatment estimand.

### 3.4 Estimand II: Outcome $Y^{\theta}$

Next, we consider the outcome as the change in the community specific means, $Y^{\theta}$, before and after treatment. We define $Y^{\theta}$ as:

(3) $Y^{\theta} = Y^c(t = 1) - Y^c(t = 0)$

By definition of the structural equations for $Y^c(t = 1)$ and $Y^c(t = 0)$, we have the following structural equation for $Y^{\theta}$:

$Y^{\theta} = f_{Y(t=1)}(V, W^c(t = 0), Y^c(t = 0), A, W^c(t = 1), U_{Y(t=1)}) - f_{Y(t=0)}(V, W^c(t = 0), U_{Y(t=0)})$
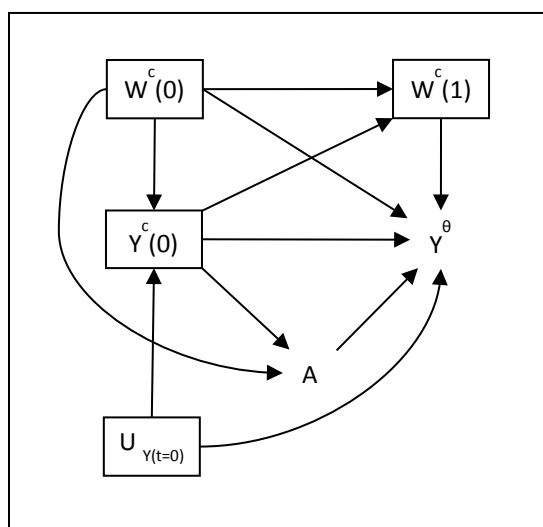


<u>Figure 2</u>: DAG illustrating that pre-post change outcome, $Y^{\theta}$, is independent of treatment, $A$, given the lagged outcome, $Y^c(0)$, pre- and post-treatment covariates, $W^c(0)$ and $W^c(1)$, and exogenous covariates, $V$ (not shown). There are no unmeasured confounders.

The DAG in Figure 2 reflects this same information. Note that $U_{Y(t=0)}$ now affects both $Y^c(t = 0)$ and $Y^{\theta}$, so we have included it in the graph. Under this model, we have a new RA for outcome, $Y^{\theta}$:

(4) $Y^{\theta}_a \perp A \mid V, W^c(t = 0), Y^c(t = 0), W^c(t = 1)$

and we can identify a statistical target parameter based on $Y^{\theta}$ that is equivalent to $\Psi^I$ (Rubin, Stuart, and Zanutto 2004). Specifically, if we define the counterfactual mean of $Y^{\theta}_a$ under an intervention on the SCM setting $A=a$ as:

(5) $E(Y^{\theta}_a) = E(Y^c_a(t = 1) - Y^c_a(t = 0)) = E(Y^c_a(t = 1)) - E(Y^c_a(t = 0))$

then we can rewrite our target causal parameter in terms of $Y^{\theta}_a$, and show that it is identical to the ATE as previously defined as $E(Y^c_1(t = 1) - Y^c_0(t = 1))$. First, the parameter is expressed as a difference in the differences of means:

(6) $E(Y^{\theta}_1 - Y^{\theta}_0) = (E(Y^c_1(t = 1)) - E(Y^c_1(t = 0))) - (E(Y^c_0(t = 1)) - E(Y^c_0(t = 0)))$

9

However, since intervening to set the treatment cannot affect the pre-treatment outcome ($Y^c_a(t = 0)= Y^c(t = 0)$), the above can be rewritten such that the mean of $Y^c(t = 0)$ cancels out to give the ATE:

(7) $(E(Y^c_1(t = 1)) - E(Y^c(t = 0))) - (E(Y^c_0(t = 1)) - E(Y^c(t = 0))) = E(Y^c_1(t = 1) - Y^c_0(t = 1))$

Under the RA (4), we can identify our statistical target parameter

$E(Y^\theta_a \mid V, W^c(t = 0), Y^c(t = 0), W^c(t = 1)) =$
$E(Y^\theta_a \mid A=a, V, W^c(t = 0), Y^c(t = 0), W^c(t = 1)) =$
$E(Y^\theta \mid A=a, V, W^c(t = 0), Y^c(t = 0), W^c(t = 1))$

and have an alternative, but equivalent, formulation of estimand $\Psi^I$:

$$(8)\ \Psi^{I^*}(P_0) = E_{V, W(t=0), Y(t=0), W(t=1)} \begin{pmatrix} E\left(Y^\theta | A = 1, V, W^C(t = 0), Y^C(t = 0), W^C(t = 1)\right) \\ -E\left(Y^\theta | A = 0, V, W^C(t = 0), Y^C(t = 0), W^C(t = 1)\right) \end{pmatrix}$$

So what is the advantage of using $Y^\theta$ over $Y^c(t = 1)$ for estimating the ATE? The main justification in the causal inference literature is that a difference method allows for both the treatment, $A$, and outcome, $Y^c(t)$, to depend on unobserved community fixed effects that are time invariant (Allison 1990; Imai 2008). To explore this advantage, we add an unmeasured confounder, $C = f_C(U_C)$, to our SCM and DAG, such that $C$ is a common cause for $A$, $Y^c(t = 0)$, and $Y^c(t = 1)$ (see Figure 3). The allowed functional forms of $f_{Y(t=0)}$ and $f_{Y(t=1)}$ in the SCM are restricted such that $C$ has a linear additive effect on $Y^c(t)$, specifically that:

$Y^c(t = 0) = f_{Y(t=0)}(V, W^c(t = 0), U_{Y(t=0)}) + C$

$Y^c(t = 1) = f_{Y(t=1)}(V, W^c(t = 0), Y^c(t = 0), A, W^c(t = 1), U_{Y(t=1)}) + C$
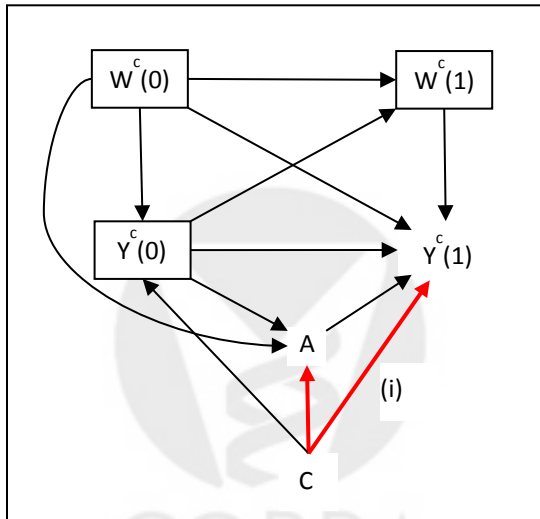


Figure 3: DAG illustrating that an unblocked path (i) is opened from treatment, $A$, to the the post-treatment outcome, $Y^c(1)$, in the presence of unmeasured confounder, $C$.

The introduction of an unmeasured confounder, $C$, opens up a backdoor path from $A$ to $Y^c(t = 1)$ (see path $A \leftarrow C \rightarrow Y^c(t = 1)$ labeled (i) and colored red in Figure 3). The RA (1) for estimand I no longer holds.

10

At first, it appears that RA(4) might hold for $Y^\theta$. If we assume $C$ has a constant additive effect on both $Y^c(t = 0)$ and $Y^c(t = 1)$, then $Y^\theta$ is not a function of $C$ when taking the difference of $Y^c$ at the two time points. The structural equation for $Y^\theta$ remains unchanged in this case.

The DAG for $Y^\theta$ in Figure 4 reflects this same information in that there is no arrow from $C$ into $Y^\theta$ (only variables on the right hand side of the equation have arrows into $Y^\theta$). Thus using $Y^\theta$ instead of $Y^c(t = 1)$ as outcome has the potential (under this specific parametric assumption) to close one backdoor pathway from $A$ to $Y^\theta$ via unmeasured confounder $C$.
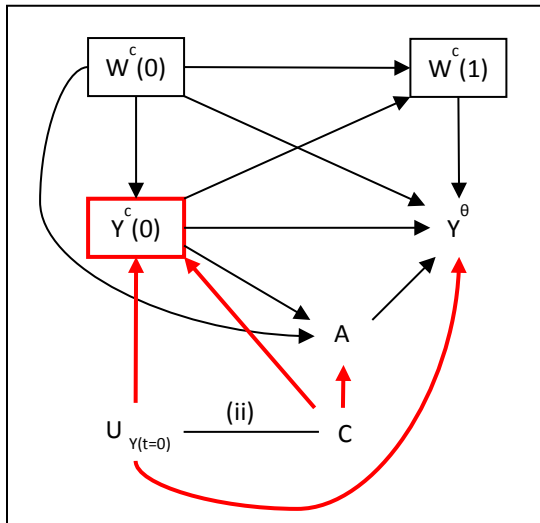


Figure 4: DAG illustrating by conditioning on pre-treatment outcome, $Y^c(0)$, in the presence of unmeasured confounder, $C$, that an unblocked path (ii) is opened from treatment, $A$, to the pre-post change outcome, $Y^\theta$, through $C$ and exogenous $U_{Y(t=0)}$.

However, on closer inspection, RA (4) does not hold under this model. Under the causal model where $C$ affects $Y^c(t = 0)$, $Y^c(t = 1)$, and $A$, conditioning on $Y^c(t = 0)$ induces new dependence between $Y^\theta$ and $A$, and opens a backdoor path through exogenous variable $U_{Y(t=0)}$ and confounder $C$. This occurs because $Y^c(t = 0)$ is a collider (two arrows go into the same variable). Conditioning on a collider opens a path that would otherwise be blocked.(Pearl 1995) This unblocked path, $A \leftarrow C - U_{Y(t=0)} \rightarrow Y^\theta$, is represented by the line between $U_{Y(t=0)}$ and $C$ (labeled (ii) in Figure 4). The path would be blocked if $Y^c(t = 0)$ is not conditioned on.

Thus, to benefit from the potential to remove unmeasured confounding from the use of $Y^\theta$ as outcome, we need a new RA (9), which is not conditional on $Y^c(t = 0)$:

(9) $Y^\theta_a \perp A \mid V, W^c(t = 0), W^c(t = 1)$

It is important to note that we have arrived at the same conclusion with DAGs that others have reached using parametric equations and analysis of covariance. In the econometrics literature, the problem is recognized as the fact that the residual on $Y^\theta$ (in a parametric equation) is necessarily correlated with the lagged outcome, $Y^c(t = 0)$, because both are a function of the random error on $Y^c(t = 0)$ (i.e., a function of $U_{Y(t=0)}$ in our SCM) (Angrist and Pischke 2009). Conditioning on $Y^c(t = 0)$ has been demonstrated to bias the treatment effect estimate under this model where the errors on $Y^c$ are serially correlated (Angrist and Pischke 2009). The method of differencing can still be applied if this correlation is thought to be

negligible (i.e., possibly when the data are from a series of cross-sections of different individuals and/or the time between cross-sections is long) (Guryan 2004). However, RA (9) still does not hold under this model without additional assumptions. We make these assumptions apparent with the use of the DAG shown in Figure 5.

By not conditioning on $Y^c(t=0)$, we open up multiple new pathways from $A$ to $Y^\theta$: directly through $Y^c(t=0)$ ($A \leftarrow Y^c(t=0) \rightarrow Y^\theta$, labeled (iii) in Figure 5); through $C$ ($A \leftarrow C \rightarrow Y^c(t=0) \rightarrow Y^\theta$, labeled (iv)); and through $U_{Y(t=0)}$ ($A \leftarrow Y^c(t=0) \leftarrow U_{Y(t=0)} \rightarrow Y^\theta$, labeled (v)). Additionally, $W^c(t=1)$ is a descendant of collider $Y^c(t=0)$, and conditioning on $W^c(t=1)$ opens up the same pathway as conditioning on $Y^c(t=0)$ (i.e., $A \leftarrow C - U_{Y(t=0)} \rightarrow Y^\theta$). However, if we do not condition on $W^c(t=1)$, then we would open up new backdoor pathways through $W^c(t=1)$ (i.e., $A \leftarrow C \rightarrow Y^c(t=0) \rightarrow W^c(t=1) \rightarrow Y^\theta$ and $A \leftarrow Y^c(t=0) \rightarrow W^c(t=1) \rightarrow Y^\theta$ labeled (vi)).
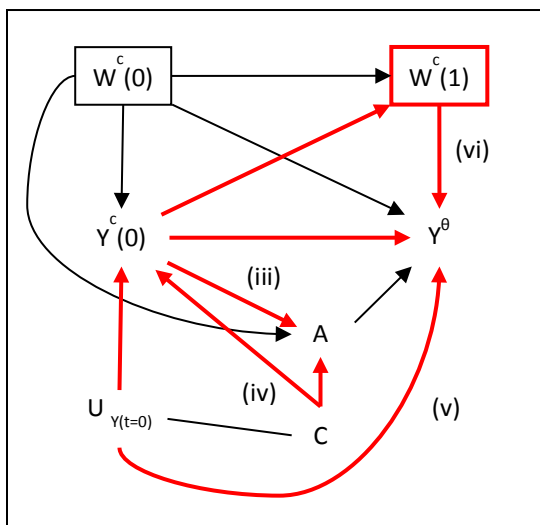


Figure 5: DAG illustrating by not conditioning on lagged outcome, $Y^c(0)$, in the presence of unmeasured confounder, $C$, that multiple unblocked paths are opened from treatment, $A$, to the pre-post change outcome, $Y^\theta$, through: (iii) lagged outcome, $Y^c(0)$; (iv) confounder, $C$; (v) exogenous $U_{Y(t=0)}$; and (vi) collider covariates $W^c(1)$.

Therefore, we must be willing to make three additional exclusion restrictions for our casual parameter to be identifiable in a difference model: that $Y^c(t=0)$ must not affect $A$, $W^c(t=1)$ and $Y^c(t=1)$. The semi-parametric equation for $Y^\theta$ becomes:

$$Y^\theta = Y^c(t=1) - Y^c(t=0) = f_{Y(t=1)}(V, W^c(t=0), A, W^c(t=1), U_{Y(t=1)}) - f_{Y(t=0)}(V, W^c(t=0), U_{Y(t=0)})$$

where $Y^\theta$ is no longer a function of $Y^c(t=0)$ but is still a function of $U_{Y(t=0)}$ (see Figure 6). Under this model, we can choose to either adjust for $W^c(t=1)$ or not (conditioning on $W^c(t=0)$ is sufficient and $W^c(t=1)$ is no longer a descendant of a collider).

In summary, RA (9) holds in the presence of unmeasured confounding from non-time varying factors, $C$, with a constant additive effect on $Y^c(t)$, only if $Y^c(t=0)$ does not affect $A$, $Y^c(t=1)$ and $W^c(t=1)$. The target causal parameter can now be identified as a new target parameter of the observed data distribution. The identifiability result applied to $Y^\theta$ becomes:

12

$E(Y^{\theta}_{a} \mid V, W^{c}(t = 0), W^{c}(t = 1)) =$
$E(Y^{\theta}_{a} \mid A=a, V, W^{c}(t = 0), W^{c}(t = 1)) =$
$E(Y^{\theta} \mid A=a, V, W^{c}(t = 0), W^{c}(t = 1))$

where the first equality holds under the RA (9) and the second from the definition of the counterfactual outcome $Y^{\theta}$ under our new SCM (Figure 6), giving us a new estimand for the ATE, $\Psi^{II}$:

$$(10)\ \Psi^{II}(P_0) = E_{V,W(t=0),W(t=1)} \left( \begin{array}{c} E\left(Y^{\theta} \mid A = 1, V, W^{C}(t = 0), W^{C}(t = 1)\right) \\ -E\left(Y^{\theta} \mid A = 0, V, W^{C}(t = 0), W^{C}(t = 1)\right) \end{array} \right)$$

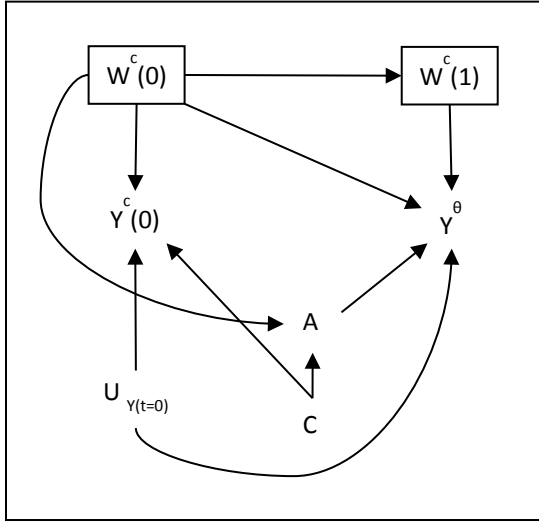which we refer to as the change score estimand.



Figure 6: DAG illustrating exclusion restrictions on lagged outcome $Y^{c}(0)$ for the pre-post change outcome, $Y^{\theta}$, in the presence of unmeasured confounder, $C$. $Y^{\theta}$, is independent of treatment, $A$, given covariates, only if $Y^{c}(0)$, does not affect treatment, $A$, post-treatment outcome, $Y^{c}(1)$, and post-treatment covariates, $W^{c}(1)$.

### 3.5 Estimand III: Outcome $Y^{c}(t)$
Finally, there is an alternate difference-in-differences estimand that pools the outcome data from both time periods together. For this approach, we need to evaluate a third causal model for identifiability. Specifically, if we are willing to make additional assumptions on the underlying causal model such that:

$(11)\ E_{V,W(t=1),W(t=0)}(Y^{c}(t) \mid A=a, V, W^{c}(t = 0), W^{c}(t = 1)) = E_{V,W(t)}(Y^{c}(t) \mid A=a, V, W^{c}(t))$, for $t = 0, 1$

then we have the following identifiability result under the new SCM:

$E(Y^{\theta}_{a} \mid V, W^{c}(t = 0), W^{c}(t = 1)) =$
$E(Y^{\theta}_{a} \mid A=a, V, W^{c}(t = 0), W^{c}(t = 1)) =$
$E(Y^{c}(t = 1) \mid A=a, V, W^{c}(t = 0), W^{c}(t = 1)) - E(Y^{c}(t = 0) \mid A=a, V, W^{c}(t = 0), W^{c} t = 1)) =$
$E(Y^{c}(t = 1) \mid A=a, V, W^{c}(t = 1)) - E(Y^{c}(t = 0) \mid A=a, V, W^{c}(t = 0))$

As with estimand II, the first equality in the identifiability result holds under the RA (9). The last equality holds under assumption (11) (i.e., by substituting $t = 1$ and $t = 0$ for $t$), giving us a third estimand for the ATE:

13

(12)

$$\Psi^{III}(P_0) = E_{V,W(t)} \begin{pmatrix} E\left(Y^C(t=1)|A=1,V,W^C(t=1)\right) - E\left(Y^C(t=0)|A=1,V,W^C(t=0)\right) \\ -E\left(Y^C(t=1)|A=0,V,W^C(t=1)\right) - E\left(Y^C(t=0)|A=0,V,W^C(t=0)\right) \end{pmatrix}$$

We refer to this final estimand as the pooled outcome estimand. However, there may be additional restrictions on the allowed data distribution for this identifiability result to hold. Starting with the SCM established for the change score estimand ($\Psi^{II}$), we work through the model separately at each time point. At time $t = 1$, assumption (11) becomes:

$E_{V,W(t=1),W(t=0)}(Y^c(t=1)\,|\,A=a,\,V,\,W^c(t=0),W^c(t=1)) = E_{V,W(t=1)}(Y^c(t=1)\,|\,A=a,\,V,\,W^c(t=1)),$

which will hold if $Y^c(t=1)$ is independent of $W^c(t=0)$ given $V$, $A$, and $W^c(t=1)$. We can use the DAG shown in Figure 7 to check whether our SCM implies this conditional independence. Under our current model, assumption (11) fails at $t = 1$ because of two unblocked paths: the direct path from $W^c(t=0)$ to $Y^c(t=1)$ (label (vii) in Figure 7); and the paths through collider $A$ (i.e., $W^c(t=0) - C \rightarrow Y^c(t=1)$ label (viii) in Figure 7). Therefore, for assumption (11) to hold at $t = 1$, we need to add two new exclusion restrictions: that $W^c(t=0)$ does not affect $Y^c(t=1)$ and does not affect $A$ (see Figure 8).
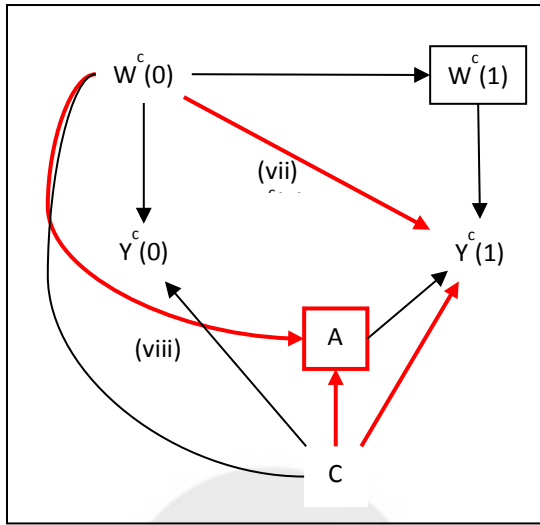


Figure 7: DAG illustrating the test of independence for the pooled outcome estimand at time t=1 in the presence of unmeasured confounder, $C$. Two unblocked paths are opened from pre-treatment covariates, $W^c(0)$, to post-treatment outcome, $Y^c(1)$: (vii) through a direct path; and (viii) the path through $C$ created by the collider treatement, $A$.

Similarly, at time $t = 0$, assumption (11) becomes:

$E_{V,W(t=1),W(t=0)}(Y^c(t=0)\,|\,A=a,\,V,\,W^c(t=0),W^c(t=1)) = E_{V,W(t=0)}(Y^c(t=0)\,|\,A=a,\,V,\,W^c(t=0))$

and we verify with a DAG that our SCM implies $Y^c(t=0)$ is independent of $W^c(t=1)$ given $V$, $A$, and $W^c(t=0)$ (Figure 9). No additional exclusion restrictions are required.

Note that we cannot add any arrows back that were removed for estimand II (i.e., $Y^c(t=0)$ cannot affect $A$, $W^c(t=1)$ or $Y^c(t=1)$). Under the additional restriction assumptions that $W^c(t=0)$ does not affect $A$ and

14

$Y^c(t = 1)$, our causal target parameter, the ATE, is equivalent to estimand III.   In settings where background knowledge makes it plausible to assume this more restrictive causal model, alternative estimation approaches offer some important advantages over traditional approaches, which will be discussed in the next paper.
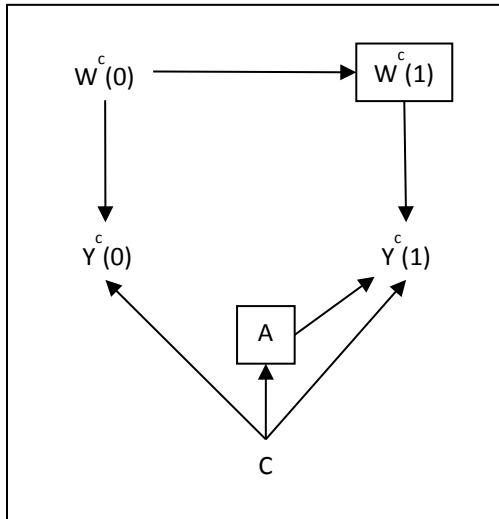


Figure 8: DAG illustrating exclusion restrictions on pre-treatment covariates $W^c(0)$ for the pooled outcome estimand at time t=1 in the presence of unmeasured confounder, $C$.  The post-treatment outcome, $Y^c(1)$, is independent of $W^c(0)$ given treatment, $A$, and post-treatment covariates, $W^c(1)$, only if $W^c(0)$ does not affect $Y^c(1)$ and does not affect $A$.
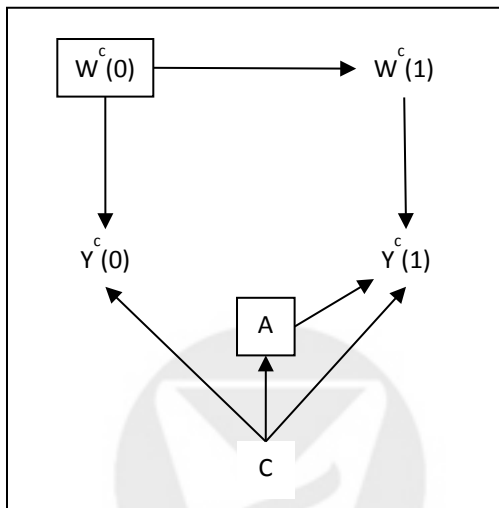


Figure 9: DAG illustrating exclusion restrictions on pre-treatment covariates $W^c(0)$ for the pooled outcome estimand at time t=0 in the presence of unmeasured confounder, $C$.  The lagged outcome, $Y^c(0)$, is independent of post-treatment covariates, $W^c(1)$, given treatment, $A$, and $W^c(0)$, if $W^c(0)$ does not affect post-treatment outcome, $Y^c(1)$, and does not affect $A$.

## 4. Illustration of Results Using Simulated Data

In this section, we present a series of simulations to demonstrate the need for the additional exclusion restrictions for the difference-in-differences estimands ($\Psi^{II}$ and $\Psi^{III}$).  The programming language R, version 2.13.1, was used for the simulations (the code is available in the Appendix).  As with the DAGs,

15

we exclude the observed village factors, $V$, from the simulations. We present eight scenarios based on different SCMs represented by the DAGs in the previous section. In all cases, $Y^c$(t), $W^c$(t) and $C$ are continuous, normally distributed and a function of additive linear terms. Treatment variable, $A$, is dichotomous and the true parameter of interest, the ATE, has a value of 1. For each scenario and estimand, linear regression with main terms was used to estimate the relevant conditional expectation from a sample of 100,000 observations. These estimates are reported in Table 2.

The first simulation is based on the starting SCM for the post treatment estimand ($\Psi^I$) represented in Figure 1. Under this model, RA (1) holds. We obtain identical estimates of the target parameter whether the outcome is defined as $Y^c(t = 1)$ or $Y^\theta$ (Figure 2 and RA (4)). The estimate is nearly equal to the target parameter value of 1 (simulation #1, Table 2). However, when we introduce an unmeasured confounder, $C$, in the second simulation, RA (1) and RA (4) no longer hold and the estimate diverges from the truth (simulation #2). This result is in keeping with a backdoor pathway being open from $A$ to outcome $Y^c(t = 1)$ through $C$ (path (i) in Figure 3) or with dependence between $Y^\theta$ and $A$ through $U_{Y(t = 0)}$ and confounder $C$ (path (ii) in Figure 4).

Table 2: Estimates[3] for the ATE under various models and sample size (true value =1)

| Sim #[4] | Figure | Assumptions | Conditional on $Y^c(t = 0)$[5] | Estimate | |
|---|---|---|---|---|---|
| 1 | 1 & 2 | RA (1) or (4), no unmeasured confounders, $A$ does not affect $W^c(t = 1)$ | Yes | $\Psi^I$ | 0.99 |
| 2 | 3 & 4 | RA (1) or (4) with unmeasured confounder $C$, $A$ does not affect $W^c(t = 1)$ | Yes | $\Psi^I$ | 3.43 |
| 3 | 5 | RA (9) with unmeasured confounder $C$, $A$ does not affect $W^c(t = 1)$, $Y^c(t = 0)$ affects $A$, $W^c(t = 1)$, and $Y^c(t = 1)$ | No | $\Psi^{II}$ | 4.78 |
| 4 | 6 | RA (9) with unmeasured confounder $C$, $A$ does not affect $W^c(t = 1)$, and $Y^c(t = 0)$ does not affect $A$, $W^c(t = 1)$, or $Y^c(t = 1)$ | No | $\Psi^{II}$ | 0.98 |
| 5 | N/A | Same as simulation 4 but $Y^c(t = 0)$ affects $A$ | No | $\Psi^{II}$ | -1.78 |
| 6 | 7 | RA (9) with unmeasured confounder $C$, $A$ does not affect $W^c(t = 1)$, $Y^c(t = 0)$ does not affect $A$, $W^c(t = 1)$, or $Y^c(t = 1)$ Assumption (11) but $W^c(t = 0)$ affects $A$ and $Y^c(t = 1)$ | No | $\Psi^{III}$ | 6.64 |
| 7 | 8 & 9 | RA (9) with unmeasured confounder $C$, $A$ does not affect $W^c(t = 1)$, $Y^c(t = 0)$ does not affect $W^c(t = 1)$, $A$, or $Y^c(t = 1)$ Assumption (11), and $W^c(t = 0)$ does not affect $A$ or $Y^c(t = 1)$ | No | $\Psi^{III}$ | 1.02 |
| 8 | N/A | Same as simulation 7 but $Y^c(t = 0)$ affects $A$ | No | $\Psi^{III}$ | -0.99 |

---

[3] Conditional expectation from a sample of 100,000 observations for a specified scenario and estimand using linear regression with main terms

[4] Reference number for the simulation scenario described in the text

[5] Indicator of whether or we conditioned on $Y^c$(t=0) in the simulation

Switching to our change score estimand ($\Psi^{II}$), we demonstrate that the estimate for the ATE diverges from 1 when not conditioning on $Y^c(t = 0)$ (simulation #3) because it opens up new pathways from $A$ to $Y^0$ (paths (iii) to (vi) in Figure 5). By adding the necessary exclusion restrictions for estimand II in the fourth simulation (i.e., Figure 6), the estimate once again nearly equals the target value (simulation #4). These results are comparable to those for the post treatment estimand with no unmeasured confounding (simulation #1). However, in simulation #5, we add that $Y^c(t = 0)$ affects $A$ into the previous scenario for estimand II. In this fifth scenario, estimand II will diverge from the truth.

The sixth simulation represents the model for our pooled outcome estimand ($\Psi^{III}$), where at time $t = 1$, $Y^c(t = 1)$ is not independent of $W^c(t = 0)$ given $V$, $A$, and $W^c(t = 1)$ (Figure 7). As expected, the estimate for estimand III diverges from 1 (simulation #6). However, when the paths from $W^c(t = 0)$ to $A$ and $Y^c(t = 1)$ are removed (Figure 8), estimand III is equal to the ATE (simulation #7). Finally, in simulation #8, we add that $Y^c(t = 0)$ affects $A$ into the previous scenario for estimand III, and the estimate once again diverges from the truth. As with simulation #5, this last simulation demonstrates that even if we can accept all the other exclusion restrictions for estimand III, we still must be willing to accept that $Y^c(t = 0)$ does not affect $A$ for the difference-in-differences estimands to equal the target parameter.

In summary, the above simulations show that when there is an unmeasured confounder, the post treatment estimand is not equal to the ATE whereas the change score and pooled outcome estimands might be, but only under additional assumptions. We demonstrate that even with an additive constant confounder $C$, we can get into trouble by using these latter two estimands ($\Psi^{II}$ and $\Psi^{III}$) if $Y^c(t = 0)$ affects $A$ (i.e., $Y^c(t = 0)$ is a confounder).

## 5. Discussion

Pre-post program evaluations (with data from treatment and control groups) present investigators with multiple causal models to choose from for identifying a causal effect of the program. Causal assumptions are necessary to obtain a valid estimate of a casual effect, and each of these models relies on a different set of assumptions. However, the causal model needs to be defined before committing to a statistical model (as opposed to selecting an estimand based on the estimation procedure it allows). In this paper, we use the structure of an existing program evaluation with pre-post data as the basis for defining several commonly used causal models. First we define the outcome as the post treatment value, $Y^c(t = 1)$, and present a causal model that requires a minimal set of exclusion restriction assumptions for identification of the ATE (estimand I). Under the key assumption of no unmeasured confounding, the simple post treatment estimand ($\Psi^I$) equals our target parameter (simulation #1 in table 1). As expected, $\Psi^I$ and the ATE diverge (i.e., are no longer equal) if an unmeasured factor, $C$, is introduced that confounds the relationship between treatment and outcome (simulation #2).

Since unmeasured cofounding is a realistic scenario in observational studies, it is not surprising that a difference-in-differences approach is often favored to try to address this issue. A differencing model is advantageous in that it "subtracts out" the effect of unmeasured confounders with a constant additive effect on the outcome at the two time points. The commonly accepted identifying assumption for the difference-in-differences estimand is a randomization assumption (RA (9)) typically referred to in the econometrics literature as the parallel trend assumption. However, through a step-by-step process of checking graphical models, we show that several exclusion restrictions are necessary for the difference-in-differences estimand to equal the ATE. In order to take advantage of this approach, we must be willing

17

to assume that the lagged outcome, $Y^c(t = 0)$, does not affect treatment, $A$, the post-treatment covariates, $W^c(t = 1)$, or the post-treatment outcome, $Y^c(t = 1)$. These are very strong assumptions about the lagged outcome! Under conditions where these restrictions do not hold, $\Psi^{II}$ and $\Psi^{III}$ will generally not be equivalent to the ATE (as illustrated with simulations #3, 5, 6, and 8). In fact, a difference-in-differences estimand has the potential to diverge further from the wished for causal effect than the post treatment estimand adjusting for all baseline covariates, even in the presence of an unmeasured confounder with a constant additive effect.

The exclusion restrictions become more numerous for the model that pools the outcome from both time periods ($\Psi^{III}$). In order for the pooled outcome estimand to be equal to the causal parameter of interest, we must add to the list of assumptions for the change score estimand ($\Psi^{II}$) that $W^c(t = 0)$ does not affect $A$ or $Y^c(t = 1)$.

Although the exclusion restriction assumptions for difference-in-differences models may seem unrealistic, it is important to note that they are often applied to data from serial cross-sections of different persons from the same communities separated in time by many years (Guryan 2004). It is possible under certain conditions that the pre-intervention outcome and covariates do not directly affect the post-intervention outcome and covariates, and are associated with post intervention outcome and covariates due only to fixed community level factors that affect both. In other words, $Y^c(t = 0)$ may be predictive of $Y^c(t = 1)$, but only due to shared common causes $C$ or $V$. The advantage of controlling for unmeasured fixed effects with a difference-in-differences estimand must be weighed against what is known about the underlying data generating system and the associated model assumptions.

In this paper, we demonstrate the power (and importance) of using graphical models (DAGs) to make the assumptions underlying a causal model transparent. The graphs prove to be invaluable tools for locating sources of dependencies among variables from confounders or colliders that may result in bias. Ideally, researchers should spend the time working with DAGs prior to conducting a study in order to collect the necessary data for a valid analysis. In an *ex-post facto* evaluation, it falls to the analyst to make use of DAGs, expert opinion, and other tools at their disposal, before proceeding with estimation. For example, given that the statistical model is semi-parametric, some of the exclusion restrictions are testable (i.e., that $Y^c(t = 1)$ is independent of $W^c(t = 0)$ given A, $W^c(t = 1)$ and $V$).

## 6. Conclusions
This paper might be more accurately titled: "To ignore or not to ignore the unobservables? That is the question." Although different disciplines control for observed confounders in different ways, we all agree that we should control for them in the best way possible. However, disagreement runs deep with respect to those factors that we do not observe. In the Madagascar evaluation, we are faced with a bias trade-off between a single post-treatment estimand that conditions on the pre-treatment outcome (a measured confounder) but assumes no unmeasured confounders, and two difference-in-differences estimands that address certain types of unmeasured confounders but do not condition on the pre-treatment outcome. The following is a re-phrasing of text from an article that compares two methods (one from epidemiology, one from econometrics) in such a way that it speaks to this bias trade-off (Hogan and Lancaster 2004). An economist will view the single post-treatment estimand with suspicion because the assumption of no unmeasured confounders seems unrealistic. On the other hand, an epidemiologist or biostatistician will be wary of the difference-in-differences estimands that purport to "subtract out" a variable that has not

been (and possibly cannot be) observed. In reality, both rely on assumptions that cannot be empirically verified.

Given that the unobservables may also be unknown, how should we decide whether to ignore them or not? The authors of this same article have a take on the disciplinary differences that we find helpful:

> "… Many epidemiologic studies differ from those in social sciences in that the collection of candidate confounders is an integral part of study design. By contrast, important research questions in economics and the social sciences are usually addressed by analysing data that have been collected or are maintained by government agencies or survey organizations (e.g., Current Population Survey, Medicaid data, etc.). The databases serve as important resources for investigating a wide variety of issues, but the variables are not typically selected for a specific research agenda. Consequently, econometric methods for causal inference are predicated on the existence of at least one and possibly several unmeasured confounding variables; therefore, confounding is essentially viewed as an omitted variables problem that leads to correlation between errors and covariates (endogeneity)..."(Hogan and Lancaster 2004)

In the absence of a third alternative (to ignore, not to ignore, other?), we conclude that the answer to the question lies in part with the source of the data, and in part with what is known about the data generating system. Consider the situation where information is collected on known confounders, for a well-defined research question, based on expert knowledge and the use of SCMs and DAGs. In this case, ignoring unobserved variables may be the best choice because the measurement of confounders was integral to the study and we can do a good job of controlling for them. We also avoid imposing additional untestable restrictions on the data. In the alternate situation where the research question is defined after the data were collected (e.g., from a national government survey), using a model that averages away the unobservables may be the better choice (again based on the research question, expert knowledge and DAGs).

Epidemiologists who use survey data collected for another purpose than their own can learn an important lesson from economists in choosing a statistical model for evaluation. For example, suppose an analyst wants to evaluate the effect of the national school lunch program on obesity among children in the U.S. using publically available data. The methods used by economists to evaluate labor or other policies may be necessary to account for unmeasured factors that may have influenced selection into the school lunch program. Similarly, economists who plan detailed measurement into their survey design *a priori* can learn an important lesson from epidemiologists. By incorporating DAGs into the planning process, researchers can identify a sufficient set of observables that should be measured to control for confounding. In this way, we can avoid models that only hold under assumptions that may be implausible.

In summary, our results reveal an important issue of identifiability that is not clearly articulated in the published literature. In the context of evaluating a community level intervention with pre-post data, we are confronted with a trade-off between statistical models that require expert knowledge about the observed data before choosing one over the other. Specifically, the SCM should incorporate expert knowledge about the data generating process that gave rise to the observed data. Any assumptions necessary to obtain a valid estimate of the desired causal effect should be reflected in the SCM and supported by this knowledge. The step of evaluating the assumptions for a given study should not be

19

overlooked prior to selecting a model and proceeding with estimation. Failure to do so can result in choosing an estimand that is not equivalent to the target casual parameter.

If our knowledge is sufficient to accurately represent the underlying data generating distribution, then our casual model may help us choose between estimands (e.g., whether the post treatment estimand is closer to the ATE than the pooled outcome estimand). In many cases, however, our knowledge will be insufficient and we won't know that the SCM holds for either estimand (or know which estimand is closer to our target parameter). Importantly, if we have strong evidence that a) there is important unmeasured confounding, and that b) the data do not support any other assumptions on which our identifiability results rely, then the target parameter is not identifiable. We cannot disregard this evidence; we risk obtaining a biased estimate. Instead, it is at this juncture that we must consider redefining our research question and target parameter before proceeding. The threat to validity from selecting a statistical model without understanding the underlying assumptions transcends our work and is applicable to any evaluation of an intervention.

### Acknowledgements

# References

Alderman, Harold. 2007. "Improving Nutrition through Community Growth Promotion: Longitudinal Study of the Nutrition and Early Child Development Program in Uganda." *World Development* no. 35 (8):1376-1389.

Allison, Paul D. 1990. "Change scores as dependent variables in regression analysis." In *Sociological Methodology*, 93-114. Oxford & Cambridge, MA: Basil Blackwell Ltd.

Angrist, Joshua, and Jorn-Steffen Pischke. 2009. Mostly harmless econometrics: an empiricist's companion. Princeton University Press.

Bhutta, Z.A., T. Ahmed, R.E. Black, S. Cousens, K. Dewey, E. Giugliani, B.H. Haider, B. Kirkwood, S.S. Morris, H.P.S. Sachdev, and M. Shekar. 2008. "What works? Interventions for maternal and child undernutrition and survival." *The Lancet* no. 371:417-40.

Black, Robert E., Lindsay H. Allen, Zulfiqar A. Bhutta, Laura E. Caulfield, Mercedes de Onis, Majid Ezzati, Colin Mathers, and Juan Rivera. 2008. "Maternal and child undernutrition: global and regional exposures and health consequences." *The Lancet* no. 371 (9608):243-260.

Engle, Patrice L., Lia C. H. Fernald, Harold Alderman, Jere Behrman, Chloe O'Gara, Aisha Yousafzai, Meena Cabral de Mello, Melissa Hidrobo, Nurper Ulkuer, Ilgi Ertem, and Selim Iltus. 2011. "Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries." *The Lancet* no. 378 (9799):1339-1353.

Fernald, L. C. H., Paul J. Gertler, and Lynnette M. Neufeld. 2008. "Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades." *The Lancet* no. 371 (9615):828-837.

Galasso, Emanuela, and Nithin Umapathi. 2009. "Improving nutritional status through behavioural change: lessons from Madagascar." *Journal of Developmental Effectiveness* no. 1 (1):60-85.

Galasso, Emanuela, and Nithin Umapathi. 2007. Improving nutritional status through behavioral change : lessons from Madagascar In *World Bank Policy Research Working Paper Series No. 4424*. Washington DC.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2011. "Impact evaluation in practice." In. Washington DC: The International Bank for Reconstruction and Development / The World Bank.

Greenland, Sander, Judea Pearl, and James M. Robins. 1999. "Causal Diagrams for Epidemiologic Research." *Epidemiology* no. 10 (1):37-48.

Guryan, Jonathan. 2004. "Desegregation and Black Dropout Rates." *The American Economic Review* no. 94 (4):919-943.

Heckman, James J., and Edward J. Vytlacil. 2007. "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics*, edited by J. Heckman James and E. Leamer Edward, 4779-4874. Elsevier.

Hogan, Joseph W, and Tony Lancaster. 2004. "Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies." *Statistical Methods in Medical Research* no. 13 (1):17-48. doi: 10.1191/0962280204sm351ra.

Imai, Kosuke. 2008. Causal Inference Lecture Notes: Causal Inference with Repeated Measures in Observational Studies. Department of Politics, Princeton University.

Lord, Frederic M. 1956. "The Measurement of Growth." *Educational and Psychological Measurement* no. 16 (4):421-437. doi: 10.1177/001316445601600401.

Macours, Karen, Norbert Schady, and Renos Vakis. 2008. Cash Transfers, Behavioral Changes, and the Cognitive Development of Young Children: Evidence from a Randomized Experiment. Washington, DC: World Bank.

Maris, Eric. 1998. "Covariance adjustment versus gain scores—revisited." *Psychological Methods* no. 3 (3):309-327. doi: 10.1037/1082-989x.3.3.309.

Meyer, Bruce D. 1994. "Natural and Quasi-Experiments in Economics." *National Bureau of Economic Research Technical Working Paper Series* no. 170.

Paxson, Christina, and Norbert Schady. 2008. "Does money matter? The effects of cash transfers on child health and development in rural Ecuador." *Economic Development and Cultural Change* no. 59 (1):187-229.

Pearl, Judea. 1995. "Causal diagrams for empirical research." *Biometrika* no. 82 (4):669-688. doi: 10.1093/biomet/82.4.669.

———. 2010. "An Introduction to Causal Inference." *The International Journal of Biostatistics* no. 6 (2).

Petersen, Maya L., Kristen E. Porter, Susan Gruber, Yue Wang, and Mark J. van der Laan. 2011. "Positivity." In *Targeted Learning: Causal Inference for Observational and Experimental Data*, edited by M.J. van der Laan and Sherri Rose. Berlin Heidelberg New York: Springer.

Popkin, Barry M., Judy Canahuati, Patricia E. Bailey, and Chloe O'Gara. 1991. "An evaluation of a national breast-feeding promotion programme in Honduras." *Journal of Biosocial Science* no. 23 (01):5-21. doi: doi:10.1017/S0021932000019027.

Robins, James M., Miguel Ángel Hernán, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* no. 11 (5):550-560.

Rubin, D.B. 1973. "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies." *Journal of Educational Psychology* no. 66:688-701.

Rubin, Donald B., Elizabeth A. Stuart, and Elaine L. Zanutto. 2004. "A Potential Outcomes View of Value-Added Assessment in Education." *Journal of Educational and Behavioral Statistics* no. 29 (1):103-116.

Sharp, Maryanne, and Ioana Kruse. 2011. Health, nutrition, and population in Madagascar 2000-09. In *Africa Human Development Series*. Washington D.C.: The World Bank.

UNICEF. 2011. *Info by Country: Madagascar Statistics* 2011 [cited February 9 2011]. Available from http://www.unicef.org/infobycountry/madagascar_statistics.html.

van der Laan, Mark J., and Sherri Rose. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Berlin Heidelberg New York: Springer.

## Appendix: R Code for Simulations

R Code for simulations 1-6:

```r
#-----------------------
# In all, I assume W1 is not affected by A, and exclude observed exogenous variables, V
#-----------------------
set.seed(100)
n <- 100000
C<-rnorm(n,0,4)
W0<-rnorm(n,0,4)
#-----------------------
# Run 1: Example for figure 3.1: estimand I, controlling for Y0
# No unmeasured confounding C
Y0<-rnorm(n,0.5*W0,4)
A<-rbinom(n,1,1/(1+exp(-0.5*W0-0.5*Y0)))
W1<-rnorm(n,W0+Y0,4)
Y1<-rnorm(n,W0+2*Y0+A+W1,4)
est1 <- glm(Y1~A+W0+W1+Y0)
#-------------------------------------------
# Run 2: Example for figure 3.3: estimand I
# Introduce unmeasured confounder C that affects Y(0), Y(1) and A
Y0<-rnorm(n,0.5*W0+C,4)
A<-rbinom(n,1,1/(1+exp(-0.5*W0-0.5*Y0-0.5*C)))
W1<-rnorm(n,W0+Y0,4)
Y1<-rnorm(n,W0+Y0+A+W1+C,4)
est2 <- glm(Y1~A+W0+W1+Y0)
#-------------------------------------------
# Run 3: Example for figure 3.5: estimand II, not controlling for Y(0)
# Unmeasured confounder C
Y0<-rnorm(n,0.5*W0+C,4)
A<-rbinom(n,1,1/(1+exp(-0.5*W0-0.5*Y0-0.5*C)))
W1<-rnorm(n,W0+Y0,4)
Y1<-rnorm(n,W0+Y0+A+W1+C,4)
Yd<-Y1-Y0
est3 <- glm(Yd~A+W0+W1)
#-------------------------------------------
# Run 4: Example for figure 3.6: estimand II, not controlling for Y(0)
# Confounder C, assume Y(0) does not affect A, W(1), or Y(1); i.e., no confounding by Y(0)
Y0<-rnorm(n,0.5*W0+C,4)
A<-rbinom(n,1,1/(1+exp(-0.5*W0-0.5*C)))
W1<-rnorm(n,W0,4)
Y1<-rnorm(n,W0+A+W1+C,4)
Yd<-Y1-Y0
est4 <- glm(Yd~A+W0+W1)
#-------------------------------------------
# Run 5: Example for figure 3.7: estimand III
# Confounder C, assume Y(0) does not affect A, W(1), or Y(1); i.e., no confounding by Y(0)
# Assumption (11) but W(0) affects A and Y(1)
Y0<-rnorm(n,0.5*W0+C,4)
A<-rbinom(n,1,1/(1+exp(-0.5*W0-0.5*C)))
W1<-rnorm(n,W0,4)
```

```
Y1<-rnorm(n,W0+A+W1+C,4)
# Reshape wide to long
id <- paste("id", 1:n, sep="")
data_wide <- data.frame(id,C,A,W0,Y0,W1,Y1)
data_long <- reshape(data_wide,
        varying = 4:7,
        idvar = "id",
        direction = "long",
        timevar = "T",
        new.row.names = NULL,
        sep = "")
est5 <- glm(Y~A+W+T+A*T,data=data_long)
#-------------------------------------------
# Run 6: Example for figure 3.8: estimand III
# Confounder C, assume Y(0) does not affect A, W(1), or Y(1); i.e., no confounding by Y(0)
# Assumption (11) and W(0) does not affect A or Y(1)
Y0<-rnorm(n,0.5*W0+C,4)
A<-rbinom(n,1,1/(1+exp(-0.5*C)))
W1<-rnorm(n,W0,4)
Y1<-rnorm(n,A+W1+C,4)
# Reshape wide to long
id <- paste("id", 1:n, sep="")
data_wide <- data.frame(id,C,A,W0,Y0,W1,Y1)
data_long <- reshape(data_wide,
        varying = 4:7,
        idvar = "id",
        direction = "long",
        timevar = "T",
        new.row.names = NULL,
        sep = "")
est6 <- glm(Y~A+W+T+A*T,data=data_long)
#-------------------------------------------
# Run 7: Example adding Y(0) affects A into run 4
Y0<-rnorm(n,0.5*W0+C,4)
A<-rbinom(n,1,1/(1+exp(-0.5*W0-0.5*Y0-0.5*C)))
W1<-rnorm(n,W0,4)
Y1<-rnorm(n,W0+A+W1+C,4)
Yd<-Y1-Y0
est7 <- glm(Yd~A+W0+W1)
#-------------------------------------------
# Run 8: Example adding Y(0) affects A into run 6
Y0<-rnorm(n,0.5*W0+C,4)
A<-rbinom(n,1,1/(1+exp(-0.5*Y0-0.5*C)))
W1<-rnorm(n,W0,4)
Y1<-rnorm(n,A+W1+C,4)
# Reshape wide to long
id <- paste("id", 1:n, sep="")
data_wide <- data.frame(id,C,A,W0,Y0,W1,Y1)
data_long <- reshape(data_wide,
        varying = 4:7,
        idvar = "id",
```

24

```
        direction = "long",
        timevar = "T",
        new.row.names = NULL,
        sep = "")
est8 <- glm(Y~A+W+T+A*T,data=data_long)
#------
est_all <-rbind(est1$coeff["A"],est2$coeff["A"],est3$coeff["A"],est4$coeff["A"],
est5$coeff["A:T"],est6$coeff["A:T"],est7$coeff["A"],est8$coeff["A:T"])
est_all
```