Examining Dose-Response Effects

in Randomized Experiments with Partial Adherence

by

Gina L. Mazza

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved January 2018 by the
Graduate Supervisory Committee:

Kevin J. Grimm, Co-Chair
Stephen G. West, Co-Chair
David P. MacKinnon
Jenn-Yun Tein

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Understanding how adherence affects outcomes is crucial when developing and assigning interventions. However, interventions are often evaluated by conducting randomized experiments and estimating intent-to-treat effects, which ignore actual treatment received. Dose-response effects can supplement intent-to-treat effects when participants are offered the full dose but many only receive a partial dose due to nonadherence. Using these data, we can estimate the magnitude of the treatment effect at different levels of adherence, which serve as a proxy for different levels of treatment. In this dissertation, I conducted Monte Carlo simulations to evaluate when linear dose-response effects can be accurately and precisely estimated in randomized experiments comparing a no-treatment control condition to a treatment condition with partial adherence. Specifically, I evaluated the performance of confounder adjustment and instrumental variable methods when their assumptions were met (Study 1) and when their assumptions were violated (Study 2). In Study 1, the confounder adjustment and instrumental variable methods provided unbiased estimates of the dose-response effect across sample sizes (200, 500, 2,000) and adherence distributions (uniform, right skewed, left skewed). The adherence distribution affected power for the instrumental variable method. In Study 2, the confounder adjustment method provided unbiased or minimally biased estimates of the dose-response effect under no or weak (but not moderate or strong) unobserved confounding. The instrumental variable method provided extremely biased estimates of the dose-response effect under violations of the exclusion restriction (no direct effect of treatment assignment on the outcome), though less severe violations of the exclusion restriction should be investigated.

ACKNOWLEDGMENTS

I would like to thank several faculty members who mentored and inspired me throughout graduate school—Dr. Craig Enders, Dr. Kevin Grimm, Dr. David MacKinnon, Dr. Jenn-Yun Tein, and Dr. Stephen West. I truly appreciate the time, effort, and patience you devoted to furthering my education and career. I would also like to thank my graduate student friends and colleagues for their advice and reassurances. Finally, I would like to thank my parents, sisters, and husband for their encouragement and support in this endeavor and many others.

TABLE OF CONTENTS

Page

LIST OF TABLES v

LIST OF FIGURES vi

INTRODUCTION 7

Potential Outcomes Framework for Causal Inference 8

Overview of Nonadherence 11

Alternatives to Intent-to-Treat 12

Dose-Response Effect 13

Previous Research 23

Purpose of Dissertation 24

SIMULATION STUDY 1 METHOD 25

Manipulated Factors 25

Data Generation and Fitted Models 28

Evaluation Criteria 29

SIMULATION STUDY 1 RESULTS 31

Intent-to-Treat Effect 31

Dose-Response Effect 32

SIMULATION STUDY 2 METHOD 33

Manipulated Factors 34

Data Generation and Fitted Models 34

Evaluation Criteria 36

Page

SIMULATION STUDY 2 RESULTS 37

Intent-to-Treat Effect 37

Dose-Response Effect 37

DISCUSSION 40

Limitations and Future Research 42

Practical Recommendations 45

Extension to Psychological Research 46

REFERENCES 49

APPENDIX

A OMITTED VARIABLE BIAS 65

B DATA GENERATION FOR SIMULATION STUDIES 1 AND 2 67

C SAS 9.4 CODE FOR INTENT-TO-TREAT, CONFOUNDER ADJUSTMENT, AND INSTRUMENTAL VARIABLE METHODS 71

LIST OF TABLES

Table Page

1. Summary of Assumptions for Confounder Adjustment and Instrumental Variable Methods....56
2. Studies 1 and 2 Manipulated Factors....57
3. Study 1 Bias and Confidence Interval Coverage for Intent-to-Treat, Confounder Adjustment, and Instrumental Variable Methods....58
4. Study 1 Confidence Interval Width, Type I Error Rate, and Power for Intent-to-Treat, Confounder Adjustment, and Instrumental Variable Methods....60
5. Study 2 Bias and Confidence Interval Coverage for Intent-to-Treat, Confounder Adjustment, and Instrumental Variable Methods....62

LIST OF FIGURES

Figure Page

1. Study 1 Uniform, Right Skewed, and Left Skewed Adherence Distributions.....63

2. Study 2 Trellis Plot of Bias by Unobserved Confounding and Violation of the Exclusion Restriction.....64

Nonadherence is pervasive in medicine, pharmacology, and psychology and substantially impacts public health. Ignoring medical advice leads to morbidity, mortality, and avoidable medical costs, yet between 20% and 30% of prescribed medications are never filled and about half of medications for chronic diseases are not taken as prescribed (Bosworth, 2012). Adhering to recommended behaviors (e.g., dietary modifications) can be even more challenging, and such behaviors account for approximately 40% of the risk associated with preventable premature deaths in the United States (National Institutes of Health, 2015). Similarly, psychotherapy is often refused or ended prematurely, such that existing mental health problems may persist or worsen. Based on a national database of 9,173 patients with various diagnoses, 3,101 patients (33.8%) only attended one session of psychotherapy and most of the remaining 6,072 patients attended fewer than five sessions (Hansen, Lambert, & Forman, 2002). These attendance rates fall far below the 13 to 18 sessions expected to be necessary for half of patients to improve (Hansen et al., 2002), and many more referred patients never attend a session.

Understanding how adherence affects outcomes is crucial when developing and assigning interventions. However, interventions are often evaluated by conducting randomized experiments and estimating intent-to-treat effects, which ignore actual treatment received. Dose-response effects can supplement intent-to-treat effects when participants are offered the full dose but many only receive a partial dose due to nonadherence. By estimating the magnitude of the treatment effect at different levels of adherence (a proxy for different levels of treatment), dose-response effects can enhance our understanding of the treatment's efficacy and potentially improve the generalizability

of the results. In this dissertation, I conduct Monte Carlo simulations to examine when linear dose-response effects can be accurately and precisely estimated in randomized experiments comparing a no-treatment control condition to a treatment condition with partial adherence.

The remainder of this dissertation is organized as follows. I first review the potential outcomes framework for causal inference. I then introduce intent-to-treat effects and alternatives in the presence of nonadherence. Whereas many alternatives require a binary measure of adherence or dichotomization of a discrete or continuous measure of adherence, dose-response estimation relies on partial adherence. I describe two methods for estimating dose-response effects in randomized experiments with no measure of adherence in the control condition and a discrete or continuous measure of adherence in the treatment condition. Finally, I conduct two simulation studies to evaluate the performance of these methods when their assumptions are met and when their assumptions are violated.

Potential Outcomes Framework for Causal Inference

In the potential outcomes framework for causal inference (Rubin, 1974, 1977, 1978, 2005; Imbens & Rubin, 2015), the treatment effect for unit i is defined as the difference between unit i's response in the treatment condition, denoted $Y_i(Z = 1)$ or $Y_i(1)$, and its response in the control condition, denoted $Y_i(Z = 0)$ or $Y_i(0)$:

$$Y_i(1) - Y_i(0). \tag{1}$$

$Y_i(1)$ and $Y_i(0)$ are referred to as unit i's potential outcomes because only one will ultimately be realized and possibly observed. Because we cannot observe $Y_i(1)$ and

$Y_i(0)$ on the same unit (e.g., a single participant at a given time; Holland, 1986), our focus shifts from the *unit's* treatment effect to the *average* treatment effect:

$$\text{ATE} = \text{E}[Y_i(1) - Y_i(0)] = \text{E}[Y_i(1)] - \text{E}[Y_i(0)]. \quad (2)$$

To define the average treatment effect, three assumptions are necessary: consistency, stable unit treatment value assumption (SUTVA), and ignorable treatment assignment. Consistency states that a unit's potential outcome under the treatment it actually received equals its observed outcome. Letting Y_i denote unit i's observed outcome, $Y_i = Y_i(1)$ if unit i were in the treatment condition and $Y_i = Y_i(0)$ if unit i were in the control condition.

SUTVA ensures that each unit has only one potential outcome in the treatment condition and one potential outcome in the control condition. SUTVA combines the no-interference assumption that one unit's treatment assignment does not affect another unit's potential outcomes (Cox, 1958) with the assumption of no hidden variations of treatments (Rubin, 2010).[1] Interference may occur in group-based interventions, such as when a participant's engagement in the intervention depends on other members of the group. Hidden variations of treatments may exist when interventions are delivered across multiple sites or by multiple clinicians or physicians. Ensuring that the two components of SUTVA (no interference across units and no hidden variations of treatments) are met is best achieved through research design. However, if SUTVA does not hold for the

[1] VanderWeele (2009) formalized the weaker assumption of treatment-variation irrelevance, meaning variations of the treatment may exist but all result in the same potential outcome for each unit. Imbens and Rubin (2015) clarified that SUTVA does not require the treatment to be identical across all units. Rather, SUTVA states that variations of the treatment cannot alter any unit's potential outcome (p. 12).

outcome of interest, the set of represented treatments may be redefined to include previously hidden variations of treatments (Imbens & Rubin, 2015).

Ignorable treatment assignment states that treatment assignment is independent of potential outcomes. That is, treatment assignment is independent of the set of outcomes that would have been realized (though not necessarily observed) if all units had been in the treatment condition and the set of outcomes that would have been realized (though not necessarily observed) if all units had been in the control condition. Successful randomization provides strong ignorability, or independence of treatment assignment and potential outcomes and independence of treatment assignment and all baseline covariates (whether measured or unmeasured). Under strong ignorability, participants in the treatment and control conditions are equivalent, on average, at baseline and should thus only differ based on application of the treatment under investigation. Although strong ignorability is not required to define the average treatment effect, achieving ignorability is difficult without randomization.

Directly comparing the observed average outcome in the treatment condition and the observed average outcome in the control condition yields a causally valid estimate of the average treatment effect given that consistency, SUTVA, and ignorable treatment assignment hold; all units fully adhere to their assigned treatment; and all units' treatment assignment and outcome are observed (Sagarin, West, Ratnikov, Homan, Ritchie, & Hansen, 2014). However, nonadherence complicates the estimation of the average treatment effect and can compromise our ability to draw causal inferences. The co-occurrence of nonadherence and missing data is beyond the scope of this dissertation but

serves as a potential topic for future research. The remainder of this dissertation assumes complete data.

Overview of Nonadherence

Nonadherence occurs when participants' *received* treatment differs from their *assigned* treatment. For example, nonadherence would occur if a participant failed to attend all of the required sessions in a multisession intervention, did not receive a vaccination after being encouraged to do so, or only partially adhered to a prescribed drug regimen. Typically adherence is nonignorable, or related to participants' potential outcomes. Meier (1991) outlined three conditions under which adherence may be nonignorable. First, characteristics of the participants may affect both adherence and the outcome, which he termed selection effects. For example, participants' baseline risk may predict both attendance and the outcome in a multisession intervention. Second, characteristics of the treatment may lead to nonadherence, such as negative side effects from a prescribed drug regimen. Finally, the outcome may cause changes in adherence. For example, participants with substance use problems who relapse may skip sessions of an intervention to avoid reprimand (West & Sagarin, 2000). In practice, the processes leading to nonadherence are often complex, and these processes may interact, vary across conditions, or vary across participants in the same condition.

The intent-to-treat method compares the observed average outcome of participants assigned to the treatment condition to that of participants assigned to the control condition, regardless of actual treatment received:

$$\text{ITT} = \text{E}[Y_i|Z_i = 1] - \text{E}[Y_i|Z_i = 0] \quad (3)$$

where Y_i is the outcome and Z_i is treatment assignment (1 = treatment, 0 = control). The intent-to-treat effect is a causally valid estimate of the average effect of treatment *assignment*, not treatment *received*. Because treatment received is disregarded, the intent-to-treat effect conflates treatment efficacy with adherence (Sheiner & Rubin, 1995). Thus, when the magnitude of the intent-to-treat effect is lower than expected, we may question whether this results from an inefficacious treatment or from an efficacious treatment being diluted by nonadherence (Meier, 1991). Overestimation can also occur if the treatment effect is not constant across participants and those who refuse or discontinue the treatment would have experienced iatrogenic effects (West & Sagarin, 2000).

Generalization of the intent-to-treat effect assumes that the adherence pattern observed in the randomized experiment will be identical to the adherence pattern for a large-scale implementation of the intervention (Robins & Greenland, 1996), which may be implausible. Participants who are selected and agree to partake in randomized experiments may be more motivated to adhere than the typical member of the population. High levels of monitoring and greater support for the treatment regimen may also promote greater adherence in randomized experiments. Alternatively, a proven treatment or a different delivery method may elicit greater adherence (Sommer & Zeger, 1991; Robins & Greenland, 1996; Goetghebeur & Shapiro, 1996).

Alternatives to Intent-to-Treat

To investigate the average effect of treatment *received*, alternatives to intent-to-treat have been proposed in the presence of nonadherence (see Sagarin et al., 2014 for a review). Many alternatives require a binary measure of adherence (e.g., receiving a

seasonal influenza vaccination or not) or dichotomization of a discrete or continuous measure of adherence. However, decisions about how to dichotomize a discrete or continuous measure of adherence are often arbitrary and data-driven. Dichotomization is also "rarely justified from either a conceptual or statistical perspective" (MacCallum, Zhang, Preacher, & Rucker, 2002, p. 20). First, dichotomization distorts differences in adherence across participants. For example, suppose that adherence is defined as attending at least seven out of ten sessions. Although we may view a participant who attended six sessions as being more similar to a participant who attended seven sessions than to a participant who never attended, the participant who attended six sessions would be grouped with the participant who never attended. Second, dichotomization leads to loss of information and lower measurement reliability (MacCallum et al., 2002). Third, dichotomization attenuates bivariate associations. Finally, nonlinear associations cannot be investigated given that adherence has only two possible values. Because adherence often cannot reasonably be characterized as "all-or-none," this dissertation focuses on partial adherence in randomized experiments.

Dose-Response Effect

The ideal experimental design for understanding the impact of dose on outcomes would involve randomly assigning participants to receive different doses of the treatment and then ensuring that all participants are 100% adherent. However, this experimental design is typically infeasible—withholding the full dose may be unethical and 100% adherence may be unachievable or impractical. More commonly, participants are offered the full dose but many only receive a partial dose due to nonadherence. Using these data,

we can estimate the magnitude of the treatment effect at different levels of adherence, which serve as a proxy for different levels of treatment.

Some methods for estimating dose-response effects require measures of adherence in both the treatment and control conditions (e.g., Holland, 1988; Efron & Feldman, 1991; Jin & Rubin, 2008). With these methods, the outcomes of participants assigned to one treatment are compared to those of similarly adherent participants assigned to the other treatment. However, assumptions about how adherence to one treatment relates to adherence to the other treatment may be difficult to justify, and adherence across different forms of treatment may not be comparable (e.g., medication management versus psychotherapy for children with attention-deficit hyperactivity disorder, MTA Cooperative Group, 1999; see Cooper & Richardson, 1986 for a discussion of unfair comparisons). No-treatment and literature control conditions also limit the applicability of these methods. In this dissertation, I examine confounder adjustment and instrumental variable methods for estimating linear dose-response effects in randomized experiments comparing a no-treatment control condition to a treatment condition with partial adherence. The confounder adjustment and instrumental variable methods are much more applicable to psychological research because they do not require a measure of adherence in the control condition.

Confounder Adjustment Method

To estimate dose-response effects in randomized experiments with partial adherence, researchers commonly regress the outcome on adherence using data from only those participants in the treatment condition. Most models assume a linear association between adherence and the outcome, though attempts have been made to allow for a

nonlinear association (e.g., using regression splines; Ramsay, 1988). Regardless of the specified functional form, causal inferences about the dose-response effect may be difficult to justify due to selection effects.

To illustrate, consider the effectiveness trial of the New Beginnings Program, a preventive intervention designed to reduce mental health problems and substance use in children following their parents' divorce or separation (Sandler et al., 2017). Of the 477 parents randomized to the ten-session treatment condition, 111 (23.3%) never attended, 309 (64.8%) attended between one and nine sessions, and 57 (11.9%) attended all ten sessions. Adherence (here defined as attendance) was likely nonignorable. For example, mothers who reported greater conflict with the other parent at baseline were more likely to drop out of the intervention early than sustain attendance (Mauricio et al., 2017), and interparental conflict was related to outcomes of interest such as children's mental health problems. Observing a negative association between number of sessions attended and children's mental health problems at posttest may be due to receiving more of the treatment, but it may also be due to interparental conflict confounding the association between number of sessions attended and children's mental health problems (see Appendix A for a description of omitted variable bias in equations).

The confounder adjustment method requires the following assumptions to draw valid causal inferences about the dose-response effect (see Table 1 for a summary).

1. Consistency states that a participant's potential outcome under the dose actually received equals the observed outcome.
2. The stable unit treatment value assumption states that there is no interference across participants and no hidden variations of treatments.

3. Positivity states that all participants have a nonzero probability of receiving each dose.
4. The adherence level is measured without error.
5. The adherence level and potential outcomes are conditionally independent. That is, the adherence level must be ignorable conditional on the baseline covariates (confounders).[2] This assumption implies that all of the confounders are measured without error.
6. The covariate distributions for participants who received dose d overlap with the covariate distributions for participants who received dose d' where $d \neq d'$.
7. The functional form of the dose-response curve and the relations between the baseline covariates and outcome are correctly specified.

To satisfy Assumption 5, all baseline covariates that theoretically relate to both adherence and the outcome (confounders) must be included in the model as follows:

$$Y_i = b_0 + b_1 D_i + \sum_{j=1}^{k} b_{(j+1)} X_{ji} + \varepsilon_i \quad (4)$$

where Y_i is the outcome; D_i is treatment received (dose); X_{1i}, X_{2i},…, X_{ki} are baseline covariates; and $i = 1, 2, \ldots, n_T$ indexes the n_T participants in the treatment condition.[3] In the earlier example, children's mental health problems would be regressed on number of sessions attended, baseline interparental conflict, and any other baseline covariates believed to be related to both number of sessions attended and children's mental health problems. When all confounders are perfectly measured and included in Equation 4 (and

[2] This assumption is sometimes referred to as selection on observables.

[3] This dissertation focuses on confounders that are not influenced by the treatment.

all other assumptions outlined above are met), b_1 allows for causal inferences about the average linear effect of receiving more of the treatment. However, when some, but not all, confounders are observed, conditioning on an observed confounder can increase or decrease bias (Kenny, 2004; Clarke, Kenkel, & Rueda, 2016). For example, if an observed confounder and an unobserved confounder have countervailing effects (i.e., one induces positive bias and the other induces negative bias), then conditioning on the observed confounder can increase bias. Because the population model for Y_i is unknown, theory should guide which baseline covariates to include in Equation 4 (but see Mayer, Thoemmes, Rose, Steyer, & West, 2014).

Assumption 6 may also be difficult to justify in practice. Diagnosing overlap becomes challenging with more than one or two confounders (Schafer & Kang, 2008), yet more than one or two confounders can easily exist. Without sufficient overlap, ordinary least squares (OLS) estimation extrapolates beyond the observed data while relying heavily on the specified functional form.[4] This extrapolation can lead to biased and unstable parameter estimates.

Instrumental Variable Method

Unlike the confounder adjustment method, the instrumental variable method for estimating dose-response effects uses data from participants in both the treatment and control conditions and allows for unobserved confounding between adherence and the outcome. The instrumental variable method relies on the existence of one or more so-

[4] When adherence is discrete and has only a few possible values, propensity score adjustment can be used to reduce a large set of baseline covariates to a one-number summary, which helps diagnose overlap and avoid extrapolation beyond the observed data (Foster, 2003; McCaffrey et al., 2013). However, achieving balance across participants with different levels of adherence may be difficult without a very large sample size.

called instruments that (1) cause treatment actually received, (2) only affect the outcome via treatment actually received (referred to as the exclusion restriction), and (3) do not share common causes with the outcome (Hernán & Robins, 2006). In most randomized experiments, treatment assignment serves as an excellent instrument because it causes treatment actually received (Condition 1) and does not share common causes with the outcome with successful randomization (Condition 3); Condition 2 is usually the only condition at risk for failure.

Dunn and Bentall (2007), Maracy and Dunn (2011), and Ginestet, Emsley, and Landau (2017) described instrumental variable methods for estimating dose-response effects based on the potential outcomes framework. Letting Z_i denote treatment assignment (1 = treatment, 0 = control), $Y_i(z)$ denote participant i's potential outcome if assigned to treatment z, and $D_i(z) \geq 0$ denote participant i's potential dose if assigned to treatment z, they relied on the following assumptions (see Table 1 for a summary).[5]

1. Consistency states that $D_i = D_i(\mathrm{z})$ and $Y_i = Y_i(z)$ where D_i and Y_i denote participant i's observed dose and outcome, respectively.
2. The stable unit treatment value assumption states that there is no interference across participants and no hidden variations of treatments. This assumption ensures that each participant has only one value of $D_i(0)$, $D_i(1)$, $Y_i(0)$, and $Y_i(1)$.
3. Participants are randomly assigned to the treatment and control conditions.

[5] The notation $Y_i(z)$ is consistent with Ginestet et al. (2017), though these authors used R_i instead of Z_i to denote treatment assignment. The notation $Y_i(z, D_i(z))$ has been used elsewhere (e.g., Imai, Keele, & Tingley, 2010 for mediation analysis; Imbens & Rubin, 2015, p. 517).

4. Participants in the control condition cannot access the treatment (referred to as one-sided nonadherence), such that $D_i(0)$ is assumed to equal zero for all participants.
5. The functional form of the dose-response curve is correctly specified. Dunn and Bentall (2007) and Maracy and Dunn (2011) outlined instrumental variable methods when assuming either a linear or quadratic dose-response curve. Only a linear dose-response curve is considered here.
6. When modeling a linear dose-response curve, one plausible instrument must exist. Modeling a quadratic dose-response curve requires at least two instruments. More generally, at least as many instruments as endogenous regressors are needed.
7. Due to the exclusion restriction, the average treatment effect is zero for participants who receive a dose of zero.

Under these assumptions, we can estimate a dose-response curve for participants with $D_i(1) > 0$.

With a single binary instrument (e.g., treatment assignment), the average linear effect of receiving more of the treatment for participants with $D_i(1) > 0$ can be estimated as

$$\frac{\mathrm{E}[Y_i|Z_i = 1] - \mathrm{E}[Y_i|Z_i = 0]}{\mathrm{E}[D_i|Z_i = 1] - \mathrm{E}[D_i|Z_i = 0]} \tag{5}$$

where Y_i is the outcome, D_i is treatment received (dose), and Z_i is treatment assignment (1 = treatment, 0 = control). Under one-sided nonadherence (Assumption 4), $\mathrm{E}[D_i|Z_i = 0] = 0$ such that the denominator equals the average dose in the treatment

condition (i.e., $\mathrm{E}[D_i|Z_i = 1]$). Equation 5 is known as the Wald estimator (Wald, 1940); it is the ratio of the average effect of treatment assignment on the outcome (i.e., the intent-to-treat effect defined in Equation 3) to the average effect of treatment assignment on treatment received. Because we assume that Z_i is only associated with Y_i through D_i (Conditions 2 and 3), dividing the variation in Y_i that is generated by variation in Z_i (the numerator of Equation 5) by the variation in D_i that is generated by variation in Z_i (the denominator of Equation 5) consistently estimates the causal effect of D_i on Y_i (Morgan & Winship, 2015). Condition 1 ensures that the denominator of Equation 5 is nonzero.

The Wald estimator is restricted to a single binary instrument and no covariates. A more general procedure outlined by Dunn and Bentall (2007) and Maracy and Dunn (2011) estimates dose-response effects using two-stage least squares (TSLS).[6] For a linear dose-response curve, TSLS estimation involves the following steps.

1. Stage 1: Using data from all participants (in both the treatment and control conditions), treatment received D_i (e.g., number of sessions attended where $D_i = 0$ for all participants in the control condition) is regressed on treatment assignment and the baseline covariates. The resulting model is used to calculate predicted scores $\widehat{D}_i$ for all participants.
2. Stage 2: The outcome is regressed on $\widehat{D}_i$ and the baseline covariates. This model does not include treatment assignment due to the exclusion restriction (i.e., because we assume that treatment assignment only affects the outcome via treatment actually received). The regression coefficient for $\widehat{D}_i$ represents the

[6] Interested readers should refer to Fischer-Lapp and Goetghebeur (1999), Dunn and Bentall (2007), and Maracy and Dunn (2011) for a procedure based on structural mean modeling with G-estimation.

average linear effect of receiving more of the treatment for participants with $D_i(1) > 0$.

Unlike D_i, $\widehat{D}_i$ is not endogenous because it is a linear function of treatment assignment and the baseline covariates, which we assume are exogenous (DeMaris, 2014). The inclusion of baseline covariates is not required but increases efficiency when the baseline covariates are strongly associated with D_i (Dunn & Bentall, 2007).

When sequentially estimating the first and second stage equations, the residuals from the second stage equation are calculated using $\widehat{D}_i$, which does not account for uncertainty in the predicted scores. Instead, the observed scores D_i should be used to calculate the residuals in the second stage equation (Hanushek & Jackson, 1977, pp. 267-269; Angrist & Pischke, 2009, p. 140; Wooldridge, 2010, pp. 101-102), which is automated by specialized software routines (e.g., the SYSLIN procedure in SAS; Angrist & Pischke, 2009, p. 122). This estimate of the residual variance is then used to calculate the standard errors of the regression coefficients in the second stage equation.[7]

The instrumental variable method can be implemented using other estimators, including limited information maximum likelihood, Bayesian, generalized method of methods, and structural mean modeling (see Burgess, Small, & Thompson, 2015 for a review). With a single instrument, these estimators (excluding Bayesian) provide the same causal estimate as the TSLS estimator described in this section (Burgess et al., 2015). In this dissertation, I implement the instrumental variable method using TSLS

[7] Dunn and Bentall (2007), Maracy and Dunn (2011), and Ginestet et al. (2017) assumed homoscedasticity of the residuals in the first and second stage equations, which I also assume throughout this dissertation.

estimation. TSLS estimation is available through the SYSLIN procedure in SAS as well as several other statistical software packages (e.g., R, Stata).

Weak instruments. Instruments that are only weakly associated with receipt of treatment present several issues. First, the instrumental variable method is biased in finite samples, particularly with weak instruments (Bound, Jaeger, & Baker, 1995; Morgan & Winship, 2015). To understand why, consider the Wald estimator in Equation 5. Suppose that the average effect of treatment assignment on treatment received is zero in the population. In a finite sample, the estimated average effect of treatment assignment on treatment received and thus the denominator of Equation 5 will likely be nonzero but small (Hernán & Robins, in press). A small denominator will inflate the estimate of the dose-response effect. Second, using too many weak instruments (i.e., specifying too many overidentifying restrictions) yields biased parameter estimates with confidence intervals that are too narrow (Bound et al., 1995; Staiger & Stock, 1997; Angrist & Pischke, 2009, pp. 205-209). With fewer overidentifying restrictions, weak instruments yield parameter estimates with (appropriately) wide confidence intervals (Angrist & Pischke, 2009, p. 209). Finally, violations of the exclusion restriction are most severe with weak instruments. Violating the exclusion restriction biases the numerator of Equation 5 because the variation in Y_i that is generated by variation in Z_i cannot be solely attributed to the association of Z_i with Y_i through D_i; a small denominator (i.e., a weak instrument) amplifies this bias.[8]

[8] Stolzenberg and Relles (1990) and Virdin (1993) conducted simulations to evaluate the sensitivity of Heckman's selection model to assumption violations, including violations of the exclusion restriction. Heckman (1976, 1979) proposed the selection model to correct for bias resulting from sample selection. Briefly, Heckman's selection model combines the model of interest with a model for selection (e.g., adherence) and correlates their residuals. These two models can share predictors, but the model for

The exclusion restriction is often considered the most problematic assumption of the instrumental variable method. Thus, violations of the exclusion restriction are investigated in the simulations that follow.

Previous Research

Confounder adjustment and instrumental variable methods have been extensively discussed and evaluated in the literature (e.g., Hanushek & Jackson, 1977; Cohen, Cohen, West, & Aiken, 2003; Angrist & Pischke, 2009; Wooldridge, 2010; Morgan & Winship, 2015; Hernán & Robins, in press). However, most existing research on the instrumental variable method, particularly in the context of nonadherence, focuses on binary endogenous regressors (e.g., Bound et al., 1995; Angrist, Imbens, & Rubin, 1996; Staiger & Stock, 1997; Jo, 2002; DeMaris, 2014). Factors commonly manipulated in simulations comparing confounder adjustment and instrumental variable methods include sample size, strength of the instrument(s), violations of the exclusion restriction, and strength of unobserved confounding. As discussed earlier, these simulations suggest that the instrumental variable method performs poorly (high bias and low precision) with small sample sizes and with weak or invalid instruments. When unobserved confounding is weak or absent, the confounder adjustment method (properly specified) outperforms the instrumental variable method (low to no bias and high precision).

Fischer-Lapp and Goetghebeur (1999), Dunn and Bentall (2007), Maracy and Dunn (2011), and Ginestet et al. (2017) conducted simulations investigating confounder adjustment and instrumental variable methods for dose-response effects. In the first three

---

selection should contain at least one predictor not in the model of interest (referred to as an exclusion restriction).

of these papers, the endogenous regressor was defined as a linear combination of normally distributed variables and ranged from zero to one to represent the proportion of dose received.[9] All four papers generated unobserved confounding, but the number of other factors manipulated was limited. In addition to unobserved confounding (absent, moderate, strong), Ginestet et al. (2017) manipulated the sample size (N = 100, 300, 500) and strength of the instruments (weak, moderate, strong). However, their objective was to compare their proposed semi-parametric Stein-like estimator to the OLS and TSLS estimators rather than to evaluate the performance of the OLS and TSLS estimators. None of the four papers investigated violations of the exclusion restriction or varied the adherence distribution.

Purpose of Dissertation

I conducted Monte Carlo simulations to examine when linear dose-response effects can be accurately and precisely estimated in randomized experiments comparing a no-treatment control condition to a treatment condition with partial adherence. Monte Carlo simulations help determine the finite sampling properties of estimators, unlike analytic derivations that establish the asymptotic properties of estimators. The first simulation study evaluated the performance of the confounder adjustment and instrumental variable methods when their assumptions were met. The second simulation study assessed the sensitivity of the confounder adjustment and instrumental variable methods to assumption violations. In addition to the confounder adjustment and instrumental variable methods, the intent-to-treat method was applied in both simulation studies.

[9] Generated values less than zero or greater than one were set to zero and one, respectively.

Simulation Study 1 Method

In Study 1, I compared the confounder adjustment and instrumental variable methods when their assumptions were met. Although the manipulated factors (sample size, adherence distribution in the treatment condition, magnitude of the dose-response effect) did not represent assumption violations, I hypothesized that some of the conditions outlined below would be more optimal than others. In particular, I expected the confounder adjustment method to provide unbiased estimates of the dose-response effect under all of the conditions investigated in Study 1 and to provide the greatest power when the sample size was large and the dose-response effect was strong. Because treatment assignment served as a strong and valid instrument across all conditions (described in more detail below), I expected the instrumental variable method to provide unbiased estimates of the dose-response effect except perhaps when the sample size was small (N = 200). Finally, I expected the instrumental variable method to provide the greatest power when the sample size was large, dose-response effect was strong, and adherence distribution in the treatment condition was either uniform or left skewed.

Manipulated Factors

I implemented a full factorial design with three factors: sample size (200, 500, 2,000), adherence distribution in the treatment condition (uniform, right skewed, left skewed), and magnitude of the dose-response effect (zero, ρ_{YD} = .00; weak, ρ_{YD} = .10; moderate, ρ_{YD} = .30; strong, ρ_{YD} = .50). The manipulated factors and their levels are summarized in the first block of Table 2. Manipulating these factors produced 36 design cells. Each design cell contained 1,000 replications, yielding a total of 36,000 samples to analyze.

Sample size was manipulated because the instrumental variable method provides consistent, but not unbiased, parameter estimates (Hanushek & Jackson, 1977; Angrist & Krueger, 2001). Although samples of 200 and 500 participants are common in psychological research, the instrumental variable method may require large samples (e.g., 2,000; Dunn & Bentall, 2007; DeMaris, 2014). The magnitude of the dose-response effect was manipulated to examine the Type I error rate and power. A zero, weak, moderate, or strong dose-response effect was represented by ρ_{YD} = .00, .10, .30, or .50, respectively, based on Cohen's (1988) effect size guidelines.

Finally, the adherence distribution in the treatment condition was manipulated to represent adherence patterns of theoretical or practical interest. Most previous simulations investigating the performance of the instrumental variable method have assumed that adherence (or some other endogenous regressor) was binary and have manipulated the proportion of cases in the two levels (e.g., Chiburis, Das, & Lokshin, 2011, 2012; DeMaris, 2014). For example, DeMaris (2014) varied the proportion of cases in each level of the endogenous regressor (50 – 50 versus 15 – 85 split) and found that the uneven split worsened the performance (mean square error and power) of the instrumental variable method under some conditions. In the present study, adherence in the treatment condition was generated as discrete with nine categories (range = 0, 1,…, 8) and with either a uniform, right skewed, or left skewed distribution (see Figure 1). These nine categories might represent attending between zero and eight sessions of an intervention. Under a uniform adherence distribution, each level of adherence consisted of $\frac{1}{9} \approx 11.1\%$ of cases, in expectation. Proportions of .17, .14, .13, .11, .10, .10, .09, .08, and .08 (for zero to eight sessions) were used to define the right skewed adherence

distribution, and proportions of .08, .08, .09, .10, .10, .11, .13, .14, and .17 (for zero to eight sessions) were used to define the left skewed adherence distribution. With these proportions, the variance of the (right or left) skewed adherence distribution (6.6731) approximately equaled the variance of the uniform adherence distribution (6.6667, difference = 0.0064). However, the means of the uniform, right skewed, and left skewed adherence distributions differed (4.0000, 3.3700, 4.6300). A right skewed adherence distribution would occur when many participants never initiate or only minimally adhere to their assigned treatment (e.g., the *effectiveness* trial of the New Beginnings Program described earlier where 23.3% of parents never attended). A left skewed adherence distribution would occur when most participants fully or almost fully adhere to their assigned treatment (e.g., the *efficacy* trial of the New Beginnings Program where mothers attended an average of 82.8% of the sessions; Wolchik et al., 2000).

Notice that the left skewed adherence distribution mirrored the right skewed adherence distribution. However, the effect of treatment assignment on treatment received was stronger with a left skewed adherence distribution than with a right skewed adherence distribution. Given previous research on instrumental variable methods, differences in performance across the right and left skewed adherence distributions may be due to differences in the magnitude of the effect of treatment assignment on treatment received but cannot be due to differences in data sparseness. However, differences in performance across the uniform and (right or left) skewed adherence distributions may be due to differences in data sparseness or differences in the magnitude of the effect of treatment assignment on treatment received. The expected correlation between treatment assignment and treatment received was .7385 with a uniform adherence distribution,

.6780 with a right skewed adherence distribution, and .7850 with a left skewed adherence distribution. Because treatment assignment served as a strong instrument across all conditions, the undesirable properties of weak instruments outlined earlier did not apply.

Data Generation and Fitted Models

All of the data were generated and analyzed in SAS 9.4. Data generation proceeded according to the following steps.

1. For all cases, I generated treatment assignment Z_i (1 = treatment, 0 = control) by randomly drawing from a Bernoulli distribution with probability of success .50.
2. For cases in the control condition, I set $D_i = 0$ and generated the outcome Y_i by randomly drawing from a standard normal distribution (mean = 0, variance = 1).
3. For cases in the treatment condition, I generated adherence D_i by randomly drawing from a standard normal distribution (mean = 0, variance = 1) and then creating nine categories (range = 0, 1,…, 8) based on thresholds of z = −1.2206, −0.7647, −0.4307, −0.1397, 0.1397, 0.4307, 0.7647, 1.2206 (for a uniform adherence distribution), z = −0.9542, −0.4959, −0.1510, 0.1257, 0.3853, 0.6745, 0.9945, 1.4051 (for a right skewed adherence distribution), or z = −1.4051, −0.9945, −0.6745, −0.3853, −0.1257, 0.1510, 0.4959, 0.9542 (for a left skewed adherence distribution). These thresholds correspond to the proportions listed in the previous section.
4. For cases in the treatment condition, I generated the outcome Y_i according to the following equation:

$$Y_i = b_1 D_i + \varepsilon_i \quad (6)$$

where D_i is adherence. I generated ε_i by randomly drawing from a normal distribution (mean = 0, variance = σ_ε^2). As described in Appendix B, the specified values of the (unstandardized) regression coefficient and residual variance in Equation 6 were $b_1 = 0.0000$ and $\sigma_\varepsilon^2 = 1.0000$, $b_1 = 0.0387$ and $\sigma_\varepsilon^2 = 0.9900$, $b_1 = 0.1162$ and $\sigma_\varepsilon^2 = 0.9100$, or $b_1 = 0.1936$ and $\sigma_\varepsilon^2 = 0.7500$ to achieve a zero (ρ_{YD} = .00), weak (ρ_{YD} = .10), moderate (ρ_{YD} = .30), or strong (ρ_{YD} = .50) dose-response effect. Equation 6 did not include an intercept such that $\mathrm{E}[Y_i|Z_i = 1, D_i = 0] = \mathrm{E}[Y_i|Z_i = 0] = 0$. That is, the expected value of Y_i for cases in the treatment condition with $D_i = 0$ equaled the expected value of Y_i for cases in the control condition, which was zero.

For each of the 36,000 samples, I applied the intent-to-treat and confounder adjustment methods using the REG procedure in SAS and the instrumental variable method using the SYSLIN procedure in SAS (see Appendix C for example code). When performing TSLS estimation via the SYSLIN procedure, the second stage equation did not include an intercept (NOINT option) due to the exclusion restriction.

Evaluation Criteria

The results were evaluated based on bias, standardized bias, confidence interval coverage, confidence interval width, power, and Type I error rate. The parameters of interest were the intent-to-treat effect and the dose-response effect. Bias refers to the difference between the average parameter estimate across the 1,000 replications within a given design cell and the corresponding population parameter.[10] Standardized bias was

[10] To assess bias introduced by categorizing adherence, I generated a sample with 25,000,000 cases within each design cell and obtained parameter estimates from a correctly specified model. Differences between these parameter estimates and the corresponding population parameters specified in SAS ranged from

calculated by dividing bias by the standard deviation of the parameter estimate across the 1,000 replications within a given design cell (i.e., the empirical standard error). Collins, Schafer, and Kam (2001) suggested that standardized bias noticeably and adversely affects Type I and Type II error rates, efficiency, and confidence interval coverage when standardized bias exceeds |0.40| or |0.50|. These standardized bias values indicate that the parameter estimate on average fell more than 0.40 or 0.50 standard errors above or below the population parameter. I deemed standardized bias values greater than |0.40| problematic. The confidence interval coverage rate refers to the proportion of replications where the 95% confidence interval contained the population parameter. Following Collins et al. (2001), I deemed coverage rates below 90% problematic. The confidence interval width measures precision and refers to the difference between the upper and lower bounds. Power and Type I error rate were evaluated based on the proportion of replications where the 95% confidence interval did not contain zero. This proportion should equal the nominal significance level α = .05 when the population parameter was zero, though Type I error rates within [.0365, .0635] were deemed acceptable. When the population parameter was zero and the null hypothesis was rejected, I also examined the proportion of replications where zero was above (below) the confidence interval. This proportion should equal $\frac{\alpha}{2}$ = .025, though right (left) tail rejection rates within [.0153, .0347] were deemed acceptable.[11]

−0.00018 to −0.00007 for the intent-to-treat method, 0.00003 to 0.00006 for the confounder adjustment method, and −0.00003 to −0.00002 for the instrumental variable method. The population parameters specified in SAS were used in the calculations described in this section.

[11] The 95% confidence intervals for the Type I error rate and right (left) tail rejection rate were calculated as $p \pm 1.96\sqrt{p(1-p)/n}$ where p is the proportion of interest (.05 or .025) and n is the number of replications. With 1,000 replications, the 95% confidence intervals for the Type I error rate and right (left)

Simulation Study 1 Results

Table 3 reports bias and confidence interval coverage for all 36 conditions investigated in Study 1. Table 4 reports confidence interval width, Type I error rate, and power.

Intent-to-Treat Effect

As shown in Table 3, estimates of the intent-to-treat effect were unbiased on average across all conditions (bias range = −0.0026 to 0.0005, standardized bias range = −0.0258 to 0.0056). Coverage rates ranged from .9400 to .9570. For the nine conditions where the population parameter was zero, Type I error rates ranged from .0440 to .0570, right tail rejection rates ranged from .0220 to .0310, and left tail rejection rates ranged from .0210 to .0260. For the 27 conditions where the population parameter was nonzero, power ranged from .1470 to 1.0000 depending on the sample size and the magnitude of the dose-response effect. As shown in Table 4, power increased as the sample size and dose-response effect increased. Conditions with a left skewed adherence distribution provided the greatest power, followed by conditions with a uniform adherence distribution and then conditions with a right skewed adherence distribution. Power varied by adherence distribution because the means of the left skewed, uniform, and right skewed adherence distributions differed (4.6300, 4.0000, 3.3700), which affected the magnitude of the intent-to-treat effect. However, as shown in Table 4, the confidence interval width did not vary by adherence distribution (range = 0.5562 to

---

tail rejection rate were $.05 \pm 1.96\sqrt{.05(1-.05)/1{,}000} = [.0365, .0635]$ and $.025 \pm 1.96\sqrt{.025(1-.025)/1{,}000} = [.0153, .0347]$, respectively.

0.5568, 0.3513 to 0.3515, and 0.1754 to 0.1755 for conditions with N = 200, 500, or 2,000, respectively).

Dose-Response Effect

As shown in Table 3, estimates of the dose-response effect were unbiased on average across all conditions (confounder adjustment: bias range = −0.0015 to −0.0002, standardized bias range = −0.0375 to −0.0173; instrumental variable: bias range = −0.0003 to 0.0002, standardized bias range = −0.0331 to 0.0131). Coverage rates ranged from .9440 to .9560 for the confounder adjustment method and from .9470 to .9690 for the instrumental variable method.

For the nine conditions where the population parameter was zero, Type I error rates, right tail rejection rates, and left tail rejection rates ranged from .0440 to .0560, .0230 to .0350, and .0200 to .0220, respectively, for the confounder adjustment method and from .0470 to .0530, .0210 to .0240, and .0240 to .0290, respectively, for the instrumental variable method. For the 27 conditions where the population parameter was nonzero, power ranged from .1670 to 1.0000 for the confounder adjustment method and from .2590 to 1.0000 for the instrumental variable method. As shown in Table 4, power increased as the sample size and dose-response effect increased. The adherence distribution did not affect power for the confounder adjustment method. However, for the instrumental variable method, conditions with a left skewed adherence distribution provided the greatest power, followed by conditions with a uniform adherence distribution and then conditions with a right skewed adherence distribution. Differences in the strength of the instrument may explain this pattern of results. Recall that the expected correlation between treatment assignment and treatment received was .7850

with a left skewed adherence distribution, .7385 with a uniform adherence distribution, and .6780 with a right skewed adherence distribution. However, as discussed earlier, differences in power across the uniform and (right or left) skewed adherence may also be due to differences in data sparseness. Finally, the instrumental variable method consistently provided greater power than the confounder adjustment method (until power approached its upper asymptote of 1.0000), and the confounder adjustment method provided confidence intervals that were between 1.2079 and 1.7998 times wider than those from the instrumental variable method (confidence interval width range = 0.0416 to 0.1531 and 0.0251 to 0.1171, respectively). This is because the confounder adjustment method used data from only those cases in the treatment condition, whereas the instrumental variable method used data from cases in both the treatment and control conditions.

Simulation Study 2 Method

In Study 2, I compared the confounder adjustment and instrumental variable methods when their assumptions were violated. Specifically, I investigated sensitivity to unobserved confounding (an assumption of the confounder adjustment method but not the instrumental variable method) and to violations of the exclusion restriction (an assumption of the instrumental variable method but not the confounder adjustment method). I hypothesized that the confounder adjustment and instrumental variable methods would perform poorly (high bias) when their assumptions were violated. However, the severity of these assumption violations was unknown.

Manipulated Factors

I implemented a full factorial design with three factors: magnitude of the dose-response effect (zero, ρ_{YD} = .00; weak, ρ_{YD} = .10; moderate, ρ_{YD} = .30; strong, ρ_{YD} = .50), unobserved confounding (absent, $\rho_{DU} = \rho_{YU}$ = .00; weak, $\rho_{DU} = \rho_{YU}$ = .10; moderate, $\rho_{DU} = \rho_{YU}$ = .30; strong, $\rho_{DU} = \rho_{YU}$ = .50), and magnitude of the direct effect of treatment assignment on the outcome (zero, δ = 0.00; weak, δ = 0.20; moderate, δ = 0.50). The magnitude of the dose-response effect was manipulated to examine the Type I error rate and power (see Study 1). The effect of the unobserved confounder on adherence in the treatment condition and the effect of the unobserved confounder on the outcome were equal. No, weak, moderate, or strong unobserved confounding was represented by $\rho_{DU} = \rho_{YU}$ = .00, .10, .30, or .50 respectively, based on Cohen's (1988) effect size guidelines for product-moment correlations. The magnitude of the direct effect of treatment assignment on the outcome (exclusion restriction violation) was manipulated by adding a constant difference between the treatment and control conditions on the outcome. No, weak, or moderate exclusion restriction violations were represented by δ = 0.00, 0.20, or 0.50, respectively, based on Cohen's (1988) effect size guidelines for standardized mean differences. The sample size was set to 2,000 across all conditions. The manipulated factors and their levels are summarized in the second block of Table 2. Manipulating these factors produced 48 design cells. Each design cell contained 1,000 replications, yielding a total of 48,000 samples to analyze.

Data Generation and Fitted Models

All of the data were generated and analyzed in SAS 9.4. Data generation proceeded according to the following steps.

1. For all cases, I generated treatment assignment Z_i (1 = treatment, 0 = control) by randomly drawing from a Bernoulli distribution with probability of success .50.
2. For cases in the control condition, I set $D_i = 0$ and generated the unobserved confounder U_i and outcome Y_i by randomly drawing from a bivariate normal distribution with mean vector $\begin{bmatrix}0\\0\end{bmatrix}$ and covariance matrix $\begin{bmatrix}1 & \rho_{YU}\\ \rho_{YU} & 1\end{bmatrix}$. ρ_{YU} equaled .00, .10, .30, or .50 for conditions with no, weak, moderate, or strong unobserved confounding, respectively.
3. For cases in the treatment condition, I generated the unobserved confounder U_i and adherence D_i by randomly drawing from a bivariate normal distribution with mean vector $\begin{bmatrix}0\\0\end{bmatrix}$ and covariance matrix $\begin{bmatrix}1 & \rho_{DU}\\ \rho_{DU} & 1\end{bmatrix}$ where ρ_{DU} equaled .00, .10, .30, or .50. I then created nine categories for D_i (range = 0, 1,…, 8) based on thresholds of z = −1.2206, −0.7647, −0.4307, −0.1397, 0.1397, 0.4307, 0.7647, 1.2206 (for a uniform adherence distribution with $\frac{1}{9} \approx 11.1\%$ of cases in each level of adherence). Only the uniform adherence distribution was considered in Study 2.
4. For cases in the treatment condition, I generated the outcome Y_i according to the following equation:

$$Y_i = b_0 + b_1 D_i + b_2 U_i + \varepsilon_i \quad (7)$$

where D_i is adherence and U_i is the unobserved confounder. I generated ε_i by randomly drawing from a normal distribution (mean = 0, variance = σ_ε^2). Appendix B provides the specified values of the (unstandardized) regression

coefficients and residual variance to manipulate the dose-response effect, strength of unobserved confounding, and violation of the exclusion restriction.

Conditions with no unobserved confounding and no violation of the exclusion restriction overlapped with conditions from Study 1 with a uniform adherence distribution.[12]

For each of the 48,000 samples, I applied the intent-to-treat and confounder adjustment methods using the REG procedure in SAS and the instrumental variable method using the SYSLIN procedure in SAS (see Appendix C for example code). When performing TSLS estimation via the SYSLIN procedure, the second stage equation did not include an intercept (NOINT option) due to the exclusion restriction. However, the exclusion restriction was violated under some of the conditions investigated in Study 2. Similarly, U_i was omitted when applying the confounder adjustment method, which was an assumption violation under some of the conditions investigated in Study 2.

Evaluation Criteria

As in Study 1, the results were evaluated based on bias, standardized bias, confidence interval coverage, confidence interval width, power, and Type I error rate.[13] When the population parameter was zero and the null hypothesis was rejected, I also examined the proportion of replications where zero was above (below) the confidence

[12] New data were generated and analyzed for Study 2 conditions that overlapped with Study 1 conditions. The results were consistent across Studies 1 and 2 for these conditions.

[13] To assess bias introduced by categorizing adherence, I generated a sample with 25,000,000 cases within each design cell and obtained parameter estimates from a correctly specified model. Differences between these parameter estimates and the corresponding population parameters specified in SAS ranged from −0.00030 to −0.00023 for the intent-to-treat method, 0.00005 to 0.00009 for the confounder adjustment method, and 0.00000 to 0.00002 for the instrumental variable method. The population parameters specified in SAS were used in the calculations described in this section. Relative to the analysis models in Appendix C, the correctly specified models also included the unobserved confounder. For design cells where the exclusion restriction was violated, parameter estimates from the corresponding design cells with no exclusion restriction violation were compared to the population parameters specified in SAS.

interval. The parameters of interest were the intent-to-treat effect and the dose-response effect.

Simulation Study 2 Results

Figure 2 plots bias for all 48 conditions investigated in Study 2. Table 5 reports bias and confidence interval coverage by unobserved confounding and exclusion restriction violation; bias and confidence interval coverage were averaged across strength of the dose-response effect (zero, weak, moderate, strong) in Table 5.

Intent-to-Treat Effect

As shown in Figure 2, estimates of the intent-to-treat effect were unbiased on average across all conditions (bias range = −0.0009 to −0.0006, standardized bias range = −0.0215 to −0.0141). Coverage rates ranged from .9540 to .9620. For the condition where the population parameter was zero, the Type I error rate, right tail rejection rate, and left tail rejection rate fell within the acceptable ranges. For the 47 conditions where the population parameter was nonzero, power increased as the magnitude of the population parameter increased, whereas the confidence interval width remained fairly constant across conditions (range = 0.1752 to 0.1757).

Dose-Response Effect

Confounder adjustment method. For conditions with no unobserved confounding (ρ_{DU} = ρ_{YU} = .00), estimates of the dose-response effect were unbiased on average (bias range = −0.0004 to −0.0003, standardized bias = −0.0327 for all conditions) and coverage rates equaled .9570 (for all conditions). For conditions with weak unobserved confounding (ρ_{DU} = ρ_{YU} = .10), estimates of the dose-response effect were more biased on average (bias range = 0.0016 to 0.0034), though standardized bias and

coverage rates still fell within the acceptable ranges (standardized bias range = 0.1544 to 0.2884, coverage rates range = .9530 to .9600). However, estimates of the dose-response effect were biased on average and coverage rates were low for conditions with moderate unobserved confounding ($\rho_{DU} = \rho_{YU} = .30$; bias range = 0.0182 to 0.0367, standardized bias range = 1.7477 to 3.0415, coverage rates range = .1580 to .5840) and conditions with strong unobserved confounding ($\rho_{DU} = \rho_{YU} = .50$; bias range = 0.0620 to 0.1243, standardized bias range = 5.8863 to 10.1253, coverage rate = .0000 for all conditions).[14] As shown in Figure 2, bias was unaffected by violations of the exclusion restriction (an assumption of the instrumental variable method but not the confounder adjustment method).

For the three conditions where the population parameter was zero, the Type I error rate, right tail rejection rate, and left tail rejection rate fell within the acceptable ranges. For the 45 conditions where the population parameter was nonzero, power increased as the magnitude of the population parameter increased, whereas the confidence interval width remained fairly constant across conditions (range = 0.0417 to 0.0483).

Instrumental variable method. For conditions where the exclusion restriction was met ($\delta = 0.00$), estimates of the dose-response effect were unbiased on average (bias range = −0.0001 to 0.0000, standardized bias range = −0.0098 to 0.0012) and coverage rates ranged from .9450 to .9680. However, estimates of the dose-response effect were

[14] Standardized bias increases as sample size increases because parameters are more precisely estimated. Thus, I investigated standardized bias when $N = 200$. Although standardized bias was lower in magnitude when the sample size was 200 instead of 2,000, the conclusions remained the same. Standardized bias was less than |0.40| for conditions with no or weak unobserved confounding but greater than |0.40| for conditions with moderate or strong unobserved confounding.

extremely biased on average and coverage rates dropped to zero for conditions with a weak exclusion restriction violation (δ = 0.20; bias range = .04995 to .05004, standardized bias range = 6.1170 to 7.3416) and conditions with a moderate exclusion restriction violation (δ = 0.50; bias range = 0.1250 to 0.1251, standardized bias range = 15.1761 to 18.0134).[15,16] As shown in Figure 2, bias was unaffected by unobserved confounding (an assumption of the confounder adjustment method but not the instrumental variable method).

For the condition where the population parameter was zero and the exclusion restriction was met, the Type I error rate, right tail rejection rate, and left tail rejection rate fell within the acceptable ranges. However, for the two conditions where the population parameter was zero and the exclusion restriction was violated, the Type I error rate equaled 1.0000 and zero always fell below the confidence interval. That is, the positive direct effect of treatment assignment on the outcome was mistakenly considered a positive dose-response effect. For the 45 conditions where the population parameter was nonzero, power increased as the magnitude of the population parameter increased, whereas the confidence interval width remained fairly constant across conditions (range = 0.0290 to 0.0319). As in Study 1, the instrumental variable method provided

[15] When the sample size was 200 instead of 2,000, standardized bias was less than |0.40| for conditions where the exclusion restriction was met but greater than |0.40| for conditions where the exclusion restriction was violated.

[16] Even under less severe violations of the exclusion restriction (represented by standardized mean differences of δ = 0.05 and 0.10), estimates of the dose-response effect were biased on average and coverage rates were low (bias range = 0.0124 to 0.0125, standardized bias range = 1.4875 to 1.8550, coverage rates range = .6230 to .6600 for δ = 0.05; bias range = 0.0249 to 0.0250, standardized bias range = 3.0102 to 3.6984, coverage rates range = .0590 to .1440 for δ = 0.10).

narrower confidence intervals than did the confounder adjustment method (confidence interval width range = 0.0290 to 0.0319 and 0.0417 to 0.0483, respectively).

Discussion

In randomized experiments, nonadherence occurs when participants' *received* treatment differs from their *assigned* treatment. In the presence of nonadherence, reporting intent-to-treat effects is widely recommended because intent-to-treat effects maintain the integrity of randomization (Sagarin et al., 2014; Gottfredson et al., 2015). However, intent-to-treat effects support causal inferences about the average effects of treatment *assignment*, not treatment *received*. Dose-response effects can supplement intent-to-treat effects when participants are offered the full dose but many only receive a partial dose due to nonadherence. Using these data, we can estimate the magnitude of the treatment effect at different levels of adherence, which serve as a proxy for different levels of treatment. In this dissertation, I conducted Monte Carlo simulations to examine when linear dose-response effects can be accurately and precisely estimated in randomized experiments comparing a no-treatment control condition to a treatment condition with partial adherence. Specifically, I evaluated the performance of confounder adjustment and instrumental variable methods when their assumptions were met (Study 1) and when their assumptions were violated (Study 2).

In Study 1, the confounder adjustment and instrumental variable methods provided unbiased estimates of the dose-response effect and acceptable coverage rates across all conditions (sample size of 200, 500, or 2,000; uniform, right skewed, or left skewed adherence distribution; zero, weak, moderate, or strong dose-response effect). These results were mostly consistent with my hypotheses. However, I expected the

instrumental variable method to produce some bias when the sample size was small (N = 200) because the instrumental variable method provides consistent, but not unbiased, parameter estimates (Hanushek & Jackson, 1977; Angrist & Krueger, 2001). Nevertheless, this result from Study 1 was consistent with simulation results presented by Maracy and Dunn (2011) where the instrumental variable method provided unbiased estimates of the dose-response effect under a sample size of 200. This lack of bias was likely due to treatment assignment serving as a very strong instrument (ρ_{DZ} = .7385, .6780, or .7850 with a uniform, right skewed, or left skewed adherence distribution). By using data from cases in both the treatment and control conditions, the instrumental variable method consistently provided narrower confidence intervals and greater power (until power approached its upper asymptote of 1.0000) than the confounder adjustment method. This power difference would be reversed if the confounder adjustment and instrumental variable methods were based on the same effective sample size (Ginestet et al., 2017). Overall, the results from Study 1 suggested that when their assumptions are met, the confounder adjustment and instrumental variable methods can provide accurate and precise estimates of dose-response effects in randomized experiments typically conducted in psychological research (e.g., N = 200 participants, common adherence distributions).

In Study 2, the confounder adjustment method provided unbiased or minimally biased estimates of the dose-response effect under no or weak unobserved confounding ($\rho_{DU} = \rho_{YU}$ = .00 or .10) but provided biased estimates of the dose-response effect under moderate or strong unobserved confounding ($\rho_{DU} = \rho_{YU}$ = .30 or .50). Bias was unaffected by violations of the exclusion restriction (an assumption of the instrumental

variable method but not the confounder adjustment method). The instrumental variable method provided unbiased estimates of the dose-response effect when the exclusion restriction was met (δ = 0.00) but provided extremely biased estimates of the dose-response effect when the exclusion restriction was violated (δ = 0.20 or 0.50). Bias was unaffected by unobserved confounding (an assumption of the confounder adjustment method but not the instrumental variable method). These results were consistent with my hypotheses, though the severity of the exclusion restriction violations was unexpected. In a post hoc expansion of Study 2, the instrumental variable method provided biased estimates of the dose-response effect even under less severe exclusion restriction violations (δ = 0.05 or 0.10). When the exclusion restriction was violated, the magnitude of the dose-response effect was overestimated because the direct effect of treatment assignment on the outcome and the dose-response effect had the same sign (i.e., were both positive) in Study 2. The magnitude of the dose-response effect would have been underestimated if the direct effect of treatment assignment on the outcome and the dose-response effect had opposite signs. In practice, researchers should use theory to consider the likely direction of bias and its implications for causal inferences.

Limitations and Future Research

Although conditions for the simulation studies were chosen to represent published research, the generalizability of all simulation studies is limited. First, in Study 1, three adherence distributions were investigated—uniform, right skewed, and left skewed. Another common adherence distribution is U-shaped with heavy right and left tails. This adherence pattern occurs when many participants never initiate their assigned treatment, but those who initiate their assigned treatment fully or almost fully adhere to the

treatment regimen. For example, consider the JOBS II intervention for preventing mental health problems and promoting high quality reemployment among unemployed workers (Vinokur, Price, & Schul, 1995). Of the 1,249 unemployed workers randomly assigned to the JOBS II intervention, 578 (46.3%) never attended a session whereas 567 (45.4%) attended at least four of the five sessions. Dichotomization might be defensible for a U-shaped adherence distribution with extremely heavy right and left tails, a topic for future research.

Second, in Studies 1 and 2, cases were evenly split between the treatment and control conditions (as in Dunn & Bentall, 2007; Maracy & Dunn, 2011; Ginestet et al., 2017). Although an even split provides the greatest power to detect the intent-to-treat effect, uneven splits could be investigated to assess power to detect the dose-response effect (e.g., one-third of cases in the control condition and two-thirds of cases in the treatment condition). For the confounder adjustment method, power will increase as the proportion of cases in the treatment condition increases because the confounder adjustment method uses data from only those cases in the treatment condition. It is less clear how the proportion of cases in the treatment condition affects power for the instrumental variable method. Assigning a higher proportion of cases to the treatment condition provides more information about the expected value of Y_i when $D_i > 0$ and the expected change in Y_i from a one-point increase in treatment received (e.g., from attending one more session). At the same time, cases in the treatment condition that received a dose of zero and cases in the control condition each provide information about the expected value of Y_i when $D_i = 0$ due to the exclusion restriction. In a post hoc expansion of Study 1, assigning two-thirds rather than half of cases to the treatment

condition provided greater power to detect the dose-response effect (until power approached its upper asymptote of 1.0000) across all conditions. However, the optimal split on treatment assignment may depend on the adherence distribution as well as other factors not considered in Study 1.

Third, the exclusion restriction violations considered in Study 2 may have been too severe or implausible, though less severe exclusion restriction violations considered in a post hoc expansion of Study 2 still produced substantial bias. In Study 2, the exclusion restriction violations could be specified as an average difference between the treatment and control conditions on the outcome. Creating an average difference rather than a constant difference between the treatment and control conditions on the outcome is more representative of empirical data.

Fourth, only a linear dose-response curve was considered in this dissertation. Studies 1 and 2 should be expanded to include nonlinear dose-response curves. Dunn and Bentall (2007) and Maracy and Dunn (2011) outlined instrumental variable methods when assuming either a linear or quadratic dose-response curve, though other functional forms may be of interest. In future research, bias resulting from misspecifications of the functional form of the dose-response curve—an assumption violation for both the confounder adjustment and instrumental variable methods—could be investigated.

Finally, in Studies 1 and 2, the expected treatment benefit for participants who attended all eight sessions could be computed. The standard error of this predicted score can be calculated algebraically for OLS estimation (Cohen et al., 2003, p. 45) or via bootstrapping for both OLS and TSLS estimation. The expected treatment benefit under other levels of adherence may also be of interest (e.g., attendance at four out of eight

sessions). The accuracy and precision with which these predicted scores can be estimated may depend on the adherence distribution. For example, the expected treatment benefit for participants who attended all eight sessions may be more accurately and precisely estimated under a left skewed adherence distribution than under a right skewed adherence distribution.

Practical Recommendations

In practice, the confounder adjustment method's no-unobserved-confounding assumption and the instrumental variable method's exclusion restriction are strong and untestable assumptions (see Morgan & Winship, 2015, pp. 301-302).[17] For the confounder adjustment method, collecting baseline covariates that theoretically relate to both adherence and the outcome (e.g., motivation) can reduce bias and increase power. Researchers should use theory to consider possible unobserved confounders and their relations to observed confounders. Researchers should also assess the sensitivity of the results to unobserved confounding. For example, Mauro (1990) proposed a sensitivity analysis based on the correlation between the unobserved confounder and focal predictor (treatment received) ρ_{DU} and the correlation between the unobserved confounder and outcome ρ_{YU}. Researchers can assess the impact of several plausible values of ρ_{DU} and ρ_{YU} on the estimated dose-response effect. The robustness of the estimated dose-response effect to unobserved confounding has important implications for the interpretability and utility of the results.

[17] Researchers may mistakenly believe that the exclusion restriction implies that Z_i and Y_i are conditionally independent given D_i and thus that the exclusion restriction is testable. Morgan and Winship (2015, pp. 301-302) clarified that if Z_i is an invalid instrument, then Z_i will be associated with Y_i conditional on D_i. However, the converse is not true. That is, Z_i may be associated with Y_i conditional on D_i even if Z_i is a valid instrument.

For the instrumental variable method, researchers should use theory to consider why a direct effect of treatment assignment on the outcome might exist. Returning to the New Beginnings Program (Sandler et al., 2017), being randomized to the control condition may have led parents to engage in other behaviors that prevented child mental health problems and substance use (e.g., seeking out other services). Alternatively, being randomized to the control condition may have led to feelings of demoralization that exacerbated child mental health problems and substance use. Other considerations include the likely direction of bias and the strength of the instrument. Outlining a sensitivity analysis for violations of the exclusion restriction is a topic for future research (but see Hong, 2015, p. 265). Although the instrumental variable method allows for unobserved confounding, collecting baseline covariates that theoretically relate to adherence can increase power (Dunn & Bentall, 2007). Finally, researchers should consider all other assumptions listed in Table 1 before applying the confounder adjustment and instrumental variable methods.

Extension to Psychological Research

Because dose-response estimation was primarily developed within medical and pharmacological research, additional considerations are necessary to extend dose-response estimation to psychological research. As one example, dose may be more difficult to quantify in psychological research. In the New Beginnings Program, the first five sessions targeted parent-child relationship quality, the next session addressed children's exposure to interparental conflict, and the remaining four sessions covered discipline. Attendance at the first two sessions or attendance at the sixth and eighth sessions each conceivably represent a dose of two sessions, yet we may expect different

outcomes for these two families. Weighting sessions by their importance in targeting a specific outcome may be advisable when each session does not represent an equal dose. That is, the last four sessions of the New Beginnings Program may be weighted more heavily than the first six sessions when investigating improvements in discipline, but a different set of weights may be used when investigating improvements in parent-child communication. Dose may also be difficult to quantify when adherence to multiple components of an intervention is of interest, particularly when these components are interdependent (see West & Aiken, 1997 for a discussion). In the New Beginnings Program, attendance and completion of home practice were interdependent such that skipping a session would preclude a parent from being assigned and thus completing the home practice before the next session.

As another example, measuring adherence without error may be more difficult depending on how adherence is defined and where adherence occurs. In the New Beginnings Program, adherence may be defined as attendance, engagement during the sessions, or practicing the targeted skills following each session. Attendance can be easily measured with little to no error. However, engagement during the sessions is more ambiguous, and practicing the targeted skills is not directly observed by the interventionist. Although measurement error is also problematic in medical and pharmacological research, addressing this issue may differ in psychological research. For example, adherence to a drug regimen may be monitored by electronic vial caps that record when the vial was opened, electronic records of prescription refills, or chemical markers. However, other innovative methods for promoting and monitoring adherence may apply to psychological research (e.g., reminders sent via mobile apps). With random

measurement error on adherence, the instrumental variable method provides consistent estimates of dose-response effects (Foster, 2003; Goetghebeur & Vansteelandt, 2005; Dunn & Bentall, 2007; Maracy & Dunn, 2011) whereas the confounder adjustment method provides attenuated estimates of dose-response effects (but see Fritz, Kenny, & MacKinnon, 2016 for the combined effect of measurement error and unobserved confounding).

When the assumptions of the confounder adjustment and instrumental variable methods are met, supplementing intent-to-treat effects with dose-response effects can enhance our understanding of the treatment's efficacy and potentially improve the generalizability of the results. The confounder adjustment and instrumental variable methods described and evaluated in this dissertation allow for more widespread reporting of dose-response effects in psychological research, which may ultimately lead to more informed treatment decisions.

REFERENCES

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444-455. http://dx.doi.org/10.2307/2291629

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*(4), 69-85. http://dx.doi.org/10.1257/jep.15.4.69

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Bosworth, H. (2012). How can innovative uses of technology be harnessed to improve medication adherence? *Expert Review of Pharmacoeconomics & Outcomes Research, 12*(2), 133-135. http://dx.doi.org/10.1586/ERP.12.6

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association, 90*(430), 443-450. http://dx.doi.org/10.2307/2291055

Burgess, S., Small, D. S., & Thompson, S. G. (2015). A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research*. Advance online publication. http://dx.doi.org/10.1177/0962280215597579

Chiburis, R. C., Das, J., & Lokshin, M. (2011). A practical comparison of the bivariate probit and linear IV estimators (Working Paper #5601). Washington, DC: The World Bank.

Chiburis, R. C., Das, J., & Lokshin, M. (2012). A practical comparison of the bivariate probit and linear IV estimators. *Economics Letters, 117*(3), 762-766. http://dx.doi.org/10.1016/j.econlet.2012.08.037

Clarke, K. A., Kenkel, B., & Rueda, M. R. (2016). Omitted variables, countervailing effects, and the possibility of overadjustment. *Political Science Research and Methods*. Advance online publication. http://dx.doi.org/10.1017/psrm.2016.46

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351. http://dx.doi.org/10.1037//1082-989X.6.4.330

Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology, 71*(2), 179-184. http://dx.doi.org/10.1037/0021-9010.71.2.179

Cox, D. R. (1958). *Planning of experiments*. New York: Wiley.

DeMaris, A. (2014). Combating unmeasured confounding in cross-sectional studies: Evaluating instrumental-variable and Heckman selection models. *Psychological Methods, 19*(3), 380-397. http://dx.doi.org/10.1037/a0037416

Dunn, G., & Bentall, R. (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine, 26*(26), 4719-4745. http://dx.doi.org/10.1002/sim.2891

Efron, B., & Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association, 86*(413), 9-17. http://dx.doi.org/10.1080/01621459.1991.10474996

Fischer-Lapp, K., & Goetghebeur, E. (1999). Practical properties of some structural mean analyses of the effect of compliance in randomized trials. *Controlled Clinical Trials, 20*(6), 531-546.

Foster, E. M. (2003). Propensity score matching: An illustrative analysis of dose response. *Medical Care, 41*(10), 1183-1192. http://dx.doi.org/10.1097/01.MLR.0000089629.62884.22

Fritz, M. S., Kenny, D. A., & MacKinnon, D. P. (2016). The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate Behavioral Research, 51*(5), 681-697. http://dx.doi.org/10.1080/00273171.2016.1224154

Ginestet, C. E., Emsley, R., & Landau, S. (2017). Dose-response modeling in mental health using Stein-like estimators with instrumental variables. *Statistics in Medicine, 36*(11), 1696-1714. http://dx.doi.org/10.1002/sim.7265

Goetghebeur, E. J. T., & Shapiro, S. H. (1996). Analysing non-compliance in clinical trials: Ethical imperative or mission impossible? *Statistics in Medicine, 15*(24), 2813-2826. http://dx.doi.org/10.1002/(SICI)1097-0258(19961230)15:24<2813::AID-SIM366>3.0.CO;2-U

Goetghebeur, E., & Vansteelandt, S. (2005). Structural mean models for compliance analysis in randomized clinical trials and the impact of errors on measures of

exposure. *Statistical Methods in Medical Research, 14*(4), 395-415. http://dx.doi.org/10.1191/0962280205sm407oa

Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16*(7), 893-926. http://dx.doi.org/10.1007/s11121-015-0555-x

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*(3), 329-343. http://dx.doi.org/10.1093/clipsy.9.3.329

Hanushek, E. A., & Jackson, J. E. (1977). *Statistical methods for social scientists*. New York: Academic Press.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement, 5*(4), 475-492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153-161. http://dx.doi.org/10.2307/1912352

Hernán, M. A., & Robins, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology, 17*(4), 360-372. http://dx.doi.org/10.1097/01.ede.0000222409.00878.37

Hernán, M. A., & Robins, J. M. (in press). *Causal inference*. Boca Raton, FL: Chapman & Hall/CRC.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945-960. http://dx.doi.org/10.1080/01621459.1986.10478354

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology, 18*, 449-484. http://dx.doi.org/10.2307/271055

Hong, G. (2015). *Causality in a social world: Moderation, mediation, and spill-over*. West Sussex, UK: Wiley.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*(4), 309-334. http://dx.doi.org/10.1037/a0020761

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences.* New York: Cambridge University Press.

Jin, H., & Rubin, D. B. (2008). Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association, 103*(481), 101-111. http://dx.doi.org/10.1198/016214507000000347

Jo, B. (2002). Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. *Statistics in Medicine, 21*(21), 3161-3181. http://dx.doi.org/10.1002/sim.1267

Kenny, D. A. (2004). *Correlation and causality* (2nd ed.). Retrieved from http://davidakenny.net/books.htm

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19-40. http://dx.doi.org/10.1037/1082-989X.7.1.19

Maracy, M., & Dunn, G. (2011). Estimating dose-response effects in psychological treatment trials: The role of instrumental variables. *Statistical Methods in Medical Research, 20*(3), 191-215. http://dx.doi.org/10.1177/0962280208097243

Mauricio, A. M., Mazza, G. L., Berkel, C., Tein, J.-Y., Sandler, I. N., Wolchik, S. A., & Winslow, E. (2017). Attendance trajectory profiles among divorced and separated mothers and fathers in the New Beginnings Program. *Prevention Science*. Advance online publication. http://dx.doi.org/10.1007/s11121-017-0783-3

Mauro, R. (1990). Understanding L.O.V.E. (left out variables error): A method for estimating the effects of omitted variables. *Psychological Bulletin, 108*(2), 314-329. http://dx.doi.org/10.1037/0033-2909.108.2.314

Mayer, A., Thoemmes, F., Rose, N., Steyer, R., & West, S. G. (2014). Theory and analysis of total, direct, and indirect causal effects. *Multivariate Behavioral Research, 49*(5), 425-442. http://dx.doi.org/10.1080/00273171.2014.931797

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine, 32*(19), 3388-3414. http://dx.doi.org/10.1002/sim.5753

McCallum, B. T. (1972). Relative asymptotic bias from errors of omission and measurement. *Econometrica, 40*(4), 757-758. http://dx.doi.org/10.2307/1912970

Meier, P. (1991). Comment on "Compliance as an explanatory variable in clinical trials" by B. Efron & D. Feldman. *Journal of the American Statistical Association, 86*(413), 19-22. http://dx.doi.org/10.2307/2289709

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). New York: Cambridge University Press.

MTA Cooperative Group (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 56*(12), 1073-1086. http://dx.doi.org/10.1001/archpsyc.56.12.1073

National Institutes of Health (2015, January 8). *Assay development and validation for self-regulation targets*. Retrieved from https://grants.nih.gov/grants/guide/rfa-files/RFA-RM-14-020.html

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science, 3*(4), 425-441. http://dx.doi.org/10.1214/ss/1177012761

Robins, J. M., & Greenland, S. (1996). Comment on "Identification of causal effects using instrumental variables" by J. D. Angrist, G. W. Imbens, & D. B. Rubin. *Journal of the American Statistical Association, 91*(434), 456-458. http://dx.doi.org/10.2307/2291630

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688-701. http://dx.doi.org/10.1037/h0037350

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics, 2*(1), 1-26. http://dx.doi.org/10.2307/1164933

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics, 6*(1), 34-58. http://dx.doi.org/10.1214/aos/1176344064

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association, 100*(469), 322-331. http://dx.doi.org/10.1198/016214504000001880

Rubin, D. B. (2010). Reflections stimulated by the comments of Shadish (2010) and West and Thoemmes (2010). *Psychological Methods, 15*(1), 38-46. http://dx.doi.org/10.1037/a0018537

Sagarin, B. J., West, S. G., Ratnikov, A., Homan, W. K., Ritchie, T. D., & Hansen, E. J. (2014). Treatment noncompliance in randomized experiments: Statistical

approaches and design issues. *Psychological Methods, 19*(3), 317-333. http://dx.doi.org/10.1037/met0000013

Sandler, I. N., Gunn, H., Mazza, G. L., Tein, J.-Y., Wolchik, S., Berkel, C., Jones, S., & Porter, M. (2017). Effects of a program to promote high quality parenting by divorced and separated fathers. *Prevention Science*. Advance online publication. http://dx.doi.org/10.1007/s11121-017-0841-x

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods, 13*(4), 279-313. http://dx.doi.org/10.1037/a0014268

Sheiner, L. B., & Rubin, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology & Therapeutics, 57*(1), 6-15. http://dx.doi.org/10.1016/0009-9236(95)90260-0

Sommer, A., & Zeger, S. L. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine, 10*(1), 45-52. http://dx.doi.org/10.1002/sim.4780100110

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557-586. http://dx.doi.org/10.2307/2171753

Stolzenberg, R. M., & Relles, D. A. (1990). Theory testing in a world of constrained research design: The significance of Heckman's censored sampling bias correction for nonexperimental research. *Sociological Methods and Research, 18*(4), 395-415. http://dx.doi.org/10.1177/0049124190018004001

VanderWeele, T. J. (2009). Concerning the consistency assumption in causal inference. *Epidemiology, 20*(6), 880-883. http://dx.doi.org/10.1097/EDE.0b013e3181bd5638

Vinokur, A. D., Price, R. H., & Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology, 23*(1), 39-74. http://dx.doi.org/10.1007/BF02506922

Virdin, L. M. (1993). *A test of the robustness of estimators that model selection in the nonequivalent control group design* (Doctoral dissertation). Retrieved from ProQuest. (Order Number 9411023)

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics, 11*(3), 284-300. http://dx.doi.org/10.1214/aoms/1177731868

West, S. G., & Aiken, L. S. (1997). Toward understanding individual effects in multicomponent prevention programs: Design and analysis strategies. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention:*

Methodological advances from alcohol and substance abuse research (pp. 167-209). Washington, DC: American Psychological Association.

West, S. G., & Sagarin, B. J. (2000). Participant selection and loss in randomized experiments. In L. Bickman (Ed.), *Research design: Donald Campbell's legacy* (pp. 117-154). Thousand Oaks, CA: SAGE.

Wolchik, S. A., West, S. G., Sandler, I. N., Tein, J.-Y., Coatsworth, D., Lengua, L., Weiss, L., Anderson, E. R., Greene, S. M., & Griffin, W. A. (2000). An experimental evaluation of theory-based mother and mother-child programs for children of divorce. *Journal of Consulting and Clinical Psychology, 68*(5), 843-856. http://dx.doi.org/10.1037/0022-006X.68.5.843

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: The MIT Press.

Table 1

Summary of Assumptions for Confounder Adjustment and Instrumental Variable Methods

Confounder Adjustment Method	Instrumental Variable Method
Consistency	Consistency
Stable unit treatment value assumption	Stable unit treatment value assumption
Positivity	Randomized treatment assignment
Adherence level measured without error	One-sided nonadherence
Conditional independence of adherence level	1+ strong instruments
Overlap of covariate distributions	Exclusion restriction
Correctly specified functional form	Correctly specified functional form

Table 2

Studies 1 and 2 Manipulated Factors

Study	Factor	Levels	Values			
1	Sample Size	3	200	500	2,000	
	Adherence Distribution	3	Uniform	Right Skewed	Left Skewed	
	Dose-Response Effect	4	Zero ρ_{YD} = .00	Weak ρ_{YD} = .10	Moderate ρ_{YD} = .30	Strong ρ_{YD} = .50
2	Dose-Response Effect	4	Zero ρ_{YD} = .00	Weak ρ_{YD} = .10	Moderate ρ_{YD} = .30	Strong ρ_{YD} = .50
	Unobserved Confounding	4	Absent $\rho_{DU} = \rho_{YU}$ = .00	Weak $\rho_{DU} = \rho_{YU}$ = .10	Moderate $\rho_{DU} = \rho_{YU}$ = .30	Strong $\rho_{DU} = \rho_{YU}$ = .50
	Exclusion Restriction Violation	3	Absent δ = 0.00	Weak δ = 0.20	Moderate δ = 0.50	

Note. The dose-response effect (Studies 1 and 2) and unobserved confounding (Study 2) were varied based on the product-moment correlation ρ, whereas violation of the exclusion restriction (Study 2) was varied based on Cohen's δ.

Table 3

Study 1 Bias and Confidence Interval Coverage for Intent-to-Treat, Confounder Adjustment, and Instrumental Variable Methods

Condition			Bias			Confidence Interval Coverage		
Dose-Response Effect	Sample Size	Adherence Distribution	Intent-to-Treat	Confounder Adjustment	Instrumental Variable	Intent-to-Treat	Confounder Adjustment	Instrumental Variable
Zero	200	Uniform	−0.0026	−0.0013	0.0001	.9550	.9450	.9500
		Right Skewed	−0.0026	−0.0014	0.0001	.9550	.9440	.9500
		Left Skewed	−0.0026	−0.0015	0.0001	.9550	.9440	.9500
	500	Uniform	−0.0003	−0.0007	0.0002	.9560	.9500	.9470
		Right Skewed	−0.0003	−0.0008	0.0002	.9560	.9550	.9470
		Left Skewed	−0.0003	−0.0008	0.0002	.9560	.9520	.9470
	2,000	Uniform	−0.0012	−0.0002	−0.0003	.9430	.9530	.9530
		Right Skewed	−0.0012	−0.0003	−0.0003	.9430	.9560	.9530
		Left Skewed	−0.0012	−0.0003	−0.0002	.9430	.9500	.9530
Weak	200	Uniform	−0.0022	−0.0013	0.0001	.9520	.9450	.9500
		Right Skewed	−0.0021	−0.0013	0.0001	.9510	.9440	.9500
		Left Skewed	−0.0023	−0.0015	0.0001	.9520	.9440	.9500
	500	Uniform	−0.0002	−0.0007	0.0002	.9520	.9500	.9470
		Right Skewed	−0.0001	−0.0008	0.0002	.9560	.9550	.9470
		Left Skewed	−0.0002	−0.0007	0.0002	.9520	.9520	.9470
	2,000	Uniform	−0.0011	−0.0002	−0.0003	.9410	.9530	.9540
		Right Skewed	−0.0011	−0.0002	−0.0003	.9410	.9560	.9550
		Left Skewed	−0.0012	−0.0003	−0.0002	.9410	.9500	.9540
Moderate	200	Uniform	−0.0013	−0.0012	0.0001	.9480	.9450	.9530
		Right Skewed	−0.0011	−0.0013	0.0001	.9490	.9440	.9530
		Left Skewed	−0.0016	−0.0014	0.0001	.9480	.9440	.9520

	500	Uniform	0.0000	−0.0006	0.0002	.9520	.9500	.9540
		Right Skewed	0.0002	−0.0008	0.0002	.9520	.9550	.9540
		Left Skewed	−0.0001	−0.0007	0.0002	.9540	.9520	.9520
	2,000	Uniform	−0.0011	−0.0002	−0.0003	.9400	.9530	.9600
		Right Skewed	−0.0010	−0.0002	−0.0003	.9410	.9560	.9600
		Left Skewed	−0.0012	−0.0003	−0.0002	.9410	.9500	.9600
Strong	200	Uniform	−0.0004	−0.0011	0.0001	.9510	.9450	.9680
		Right Skewed	−0.0001	−0.0012	0.0001	.9530	.9440	.9680
		Left Skewed	−0.0009	−0.0013	0.0001	.9480	.9440	.9670
	500	Uniform	0.0002	−0.0006	0.0002	.9570	.9500	.9670
		Right Skewed	0.0005	−0.0007	0.0002	.9570	.9550	.9680
		Left Skewed	0.0000	−0.0007	0.0001	.9550	.9520	.9660
	2,000	Uniform	−0.0010	−0.0002	−0.0002	.9440	.9530	.9690
		Right Skewed	−0.0008	−0.0002	−0.0003	.9440	.9560	.9690
		Left Skewed	−0.0011	−0.0003	−0.0002	.9440	.9500	.9690

Note. Whereas the intent-to-treat effect was the estimand for the intent-to-treat method, the dose-response effect was the estimand for the confounder adjustment and instrumental variable methods. Standardized bias never exceeded |0.40| (results not shown in table).

Table 4

Study 1 Confidence Interval Width, Type I Error Rate, and Power for Intent-to-Treat, Confounder Adjustment, and Instrumental Variable Methods

Condition			Confidence Interval Width			Type I Error Rate or Power		
Dose-Response Effect	Sample Size	Adherence Distribution	Intent-to-Treat	Confounder Adjustment	Instrumental Variable	Intent-to-Treat	Confounder Adjustment	Instrumental Variable
Zero	200	Uniform	0.5568	0.1529	0.0985	.0450	.0550	.0500
		Right Skewed	0.5568	0.1529	0.1171	.0450	.0560	.0500
		Left Skewed	0.5568	0.1531	0.0851	.0450	.0560	.0500
	500	Uniform	0.3515	0.0964	0.0622	.0440	.0500	.0530
		Right Skewed	0.3515	0.0964	0.0738	.0440	.0450	.0530
		Left Skewed	0.3515	0.0964	0.0537	.0440	.0480	.0530
	2,000	Uniform	0.1755	0.0480	0.0310	.0570	.0470	.0470
		Right Skewed	0.1755	0.0480	0.0368	.0570	.0440	.0470
		Left Skewed	0.1755	0.0480	0.0268	.0570	.0500	.0470
Weak	200	Uniform	0.5567	0.1522	0.0982	.1770	.1830	.3440
		Right Skewed	0.5567	0.1521	0.1168	.1470	.1670	.2590
		Left Skewed	0.5567	0.1523	0.0848	.2290	.1720	.4340
	500	Uniform	0.3515	0.0959	0.0620	.4130	.3470	.7050
		Right Skewed	0.3515	0.0959	0.0736	.3130	.3280	.5540
		Left Skewed	0.3515	0.0959	0.0536	.5170	.3460	.8230
	2,000	Uniform	0.1755	0.0478	0.0310	.9280	.8840	.9990
		Right Skewed	0.1755	0.0478	0.0367	.8230	.8790	.9840
		Left Skewed	0.1755	0.0478	0.0267	.9730	.8900	1.0000
Moderate	200	Uniform	0.5565	0.1459	0.0963	.8990	.8640	.9990
		Right Skewed	0.5565	0.1458	0.1144	.7840	.8700	.9710
		Left Skewed	0.5564	0.1460	0.0831	.9680	.8730	1.0000

	500	Uniform	0.3514	0.0920	0.0608	1.0000	.9990	1.0000
		Right Skewed	0.3514	0.0919	0.0721	.9930	.9990	1.0000
		Left Skewed	0.3514	0.0920	0.0525	1.0000	.9990	1.0000
	2,000	Uniform	0.1754	0.0458	0.0303	1.0000	1.0000	1.0000
		Right Skewed	0.1754	0.0458	0.0360	1.0000	1.0000	1.0000
		Left Skewed	0.1754	0.0458	0.0262	1.0000	1.0000	1.0000
Strong	200	Uniform	0.5564	0.1325	0.0922	1.0000	1.0000	1.0000
		Right Skewed	0.5564	0.1324	0.1095	.9950	1.0000	1.0000
		Left Skewed	0.5562	0.1326	0.0796	1.0000	1.0000	1.0000
	500	Uniform	0.3513	0.0835	0.0582	1.0000	1.0000	1.0000
		Right Skewed	0.3513	0.0835	0.0690	1.0000	1.0000	1.0000
		Left Skewed	0.3513	0.0835	0.0502	1.0000	1.0000	1.0000
	2,000	Uniform	0.1754	0.0416	0.0290	1.0000	1.0000	1.0000
		Right Skewed	0.1754	0.0416	0.0344	1.0000	1.0000	1.0000
		Left Skewed	0.1754	0.0416	0.0251	1.0000	1.0000	1.0000

Note. Whereas the intent-to-treat effect was the estimand for the intent-to-treat method, the dose-response effect was the estimand for the confounder adjustment and instrumental variable methods.

Table 5

Study 2 Bias and Confidence Interval Coverage for Intent-to-Treat, Confounder Adjustment, and Instrumental Variable Methods

Condition		Bias			Confidence Interval Coverage		
Unobserved Confounding	Exclusion Restriction Violation	Intent-to-Treat	Confounder Adjustment	Instrumental Variable	Intent-to-Treat	Confounder Adjustment	Instrumental Variable
Absent	Absent	−0.0007	−0.0004	0.0000	.9583	.9570	.9615
	Weak	−0.0007	−0.0004	0.0500	.9583	.9570	.0000
	Moderate	−0.0007	−0.0004	0.1251	.9583	.9570	.0000
Weak	Absent	−0.0007	0.0026	0.0000	.9563	.9570	.9595
	Weak	−0.0007	0.0026	0.0500	.9563	.9570	.0000
	Moderate	−0.0007	0.0026	0.1251	.9563	.9570	.0000
Moderate	Absent	−0.0008	0.0283	0.0000	.9578	.3435	.9578
	Weak	−0.0008	0.0283	0.0500	.9578	.3435	.0000
	Moderate	−0.0008	0.0283	0.1251	.9578	.3435	.0000
Strong	Absent	−0.0008	0.0962	−0.0001	.9575	.0000	.9520
	Weak	−0.0008	0.0962	0.0500	.9575	.0000	.0000
	Moderate	−0.0008	0.0962	0.1250	.9575	.0000	.0000

Note. The results in this table were averaged across strength of the dose-response effect (zero, weak, moderate, strong). Standardized bias exceeded |0.40| for the shaded cells in the "Bias" columns.

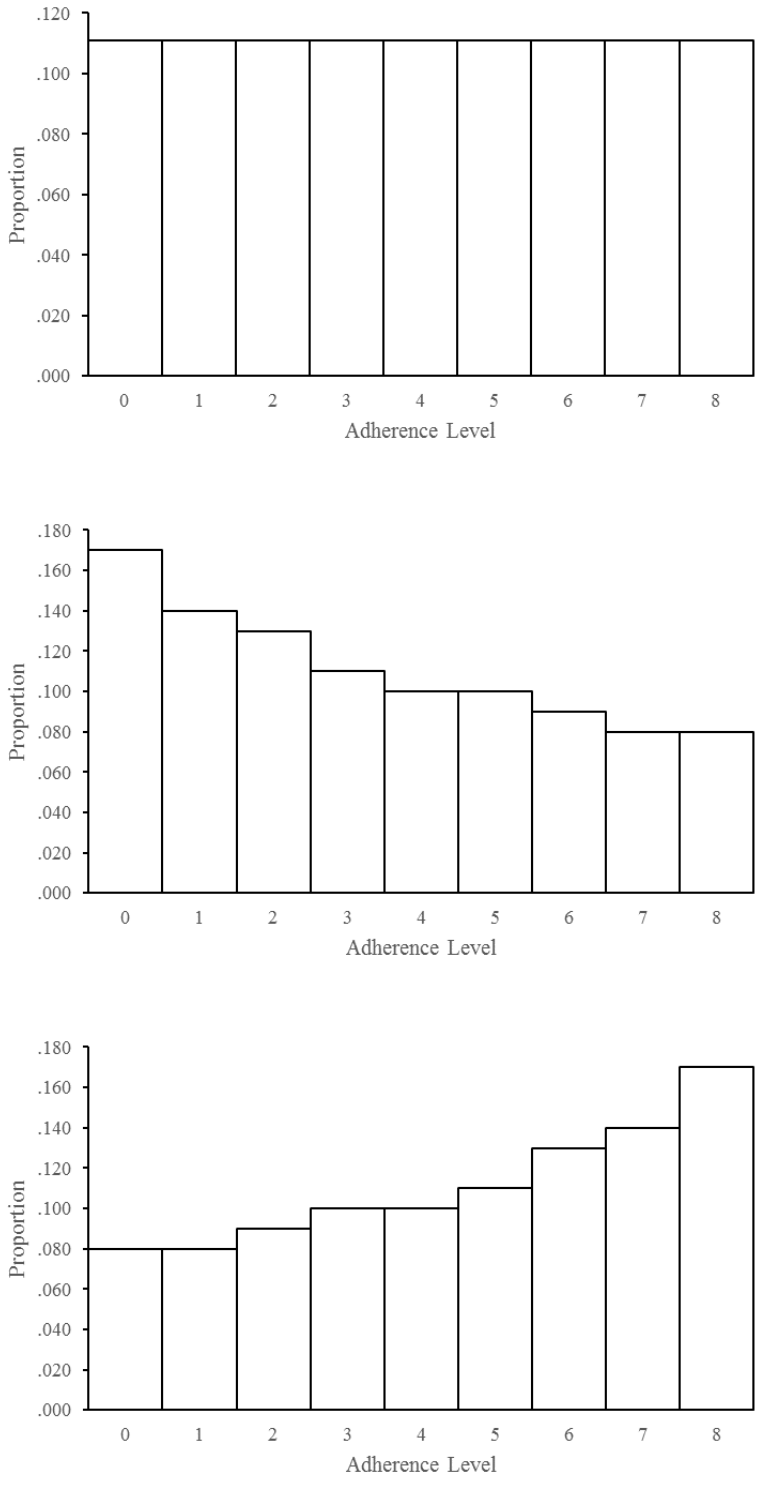

Figure 1. Study 1 uniform, right skewed, and left skewed adherence distributions.

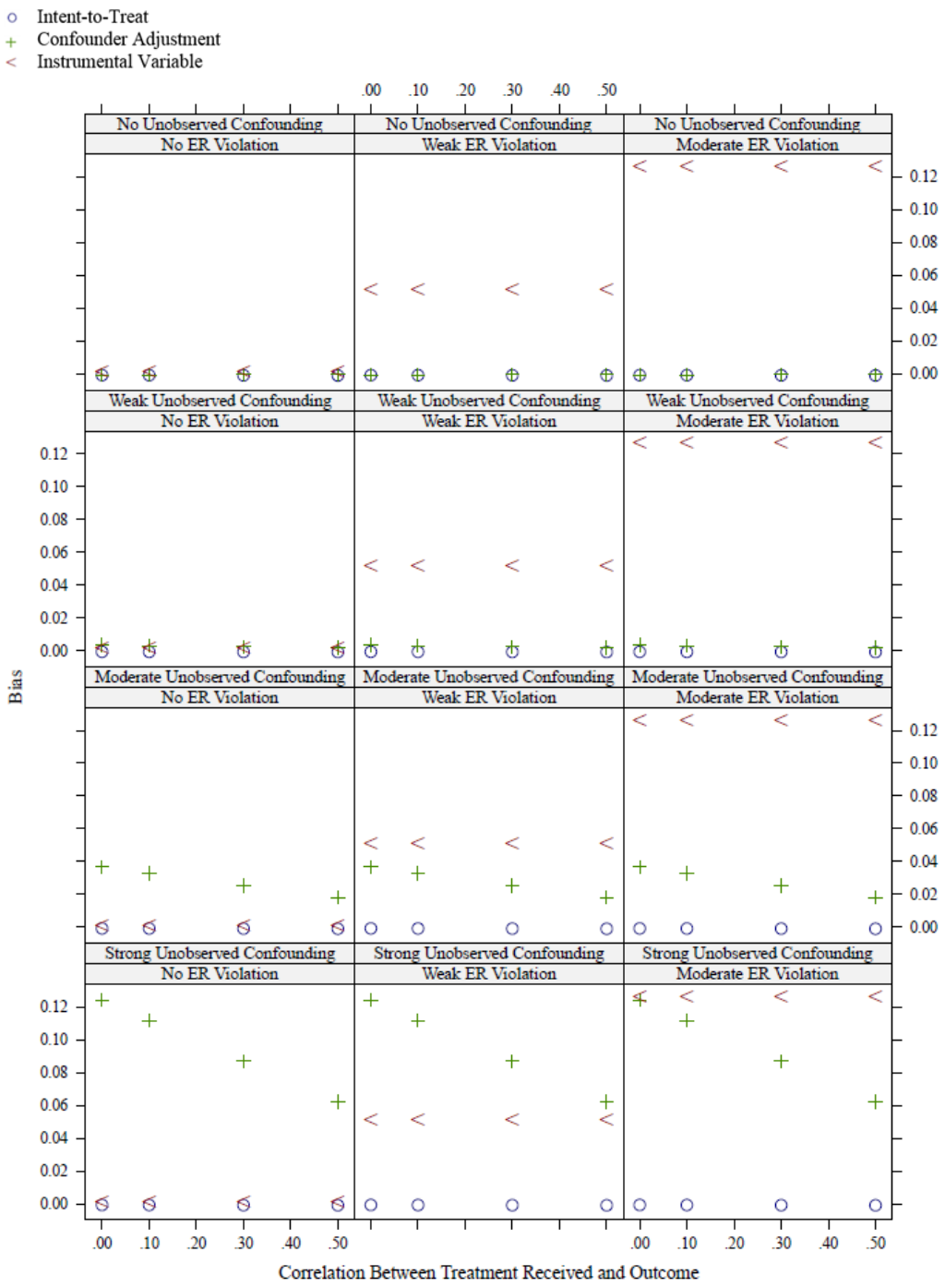

Figure 2. Study 2 trellis plot of bias by unobserved confounding and violation of the exclusion restriction (abbreviated "ER"). Whereas the intent-to-treat effect was the estimand for the intent-to-treat method, the dose-response effect was the estimand for the confounder adjustment and instrumental variable methods.

APPENDIX A

OMITTED VARIABLE BIAS

In this appendix, I describe omitted variable bias in equations. Suppose that the population model for some outcome Y_i is

$$Y_i = b_0 + b_1 D_i + b_2 X_i + \varepsilon_i \quad \text{(A1)}$$

and that b_1 is the regression coefficient of interest. Further suppose that D_i and X_i are linearly related as follows:

$$X_i = g_0 + g_1 D_i + \delta_i \quad \text{(A2)}$$

and that data on X_i are unavailable. Substituting Equation A2 into Equation A1 yields the following:

$$Y_i = b_0 + b_1 D_i + b_2 (g_0 + g_1 D_i + \delta_i) + \varepsilon_i. \quad \text{(A3)}$$

Collecting like terms in Equation A3 yields

$$Y_i = (b_0 + b_2 g_0) + (b_1 + b_2 g_1) D_i + (\varepsilon_i + b_2 \delta_i) \quad \text{(A4)}$$

where $(b_0 + b_2 g_0)$ is the intercept, $(b_1 + b_2 g_1)$ is the regression coefficient for D_i, and $(\varepsilon_i + b_2 \delta_i)$ is the residual.

When omitting X_i from Equation A1, the regression coefficient estimated for D_i equals $(b_1 + b_2 g_1)$ as in Equation A4 instead of b_1 as in Equation A1. The term $b_2 g_1$ represents bias due to omitting X_i (see Equation 3 in McCallum, 1972). In this example, the estimated regression coefficient will be unbiased if X_i either is unrelated to Y_i (i.e., $b_2 = 0$) or is unrelated to D_i (i.e., $g_1 = 0$) such that $b_2 g_1 = 0$. The estimated regression coefficient will be positively biased when b_2 and g_1 have the same sign (i.e., are both positive or both negative) and will be negatively biased when b_2 and g_1 have opposite signs.

APPENDIX B

DATA GENERATION FOR SIMULATION STUDIES 1 AND 2

Simulation Study 1

To generate the outcome in the treatment condition, values of the unstandardized regression coefficient b_1 in Equation 6 were calculated based on the following equation:

$$b_1 = \rho_{YD}\left(\frac{\sigma_Y}{\sigma_D}\right) \tag{B1}$$

where ρ_{YD} is the correlation between D_i and Y_i, $\sigma_Y = 1.0000$ is the standard deviation of Y_i, and σ_D is the standard deviation of D_i. The standard deviation of D_i equaled 2.5820 for conditions with a uniform adherence distribution and 2.5832 for conditions with a (right or left) skewed adherence distribution. Values of the residual variance σ_ε^2 were calculated based on the following equation:

$$\sigma_\varepsilon^2 = (1 - R^2)\sigma_Y^2 = (1 - \rho_{YD}^2)\sigma_Y^2 \tag{B2}$$

where $\sigma_Y = 1.0000$ is the standard deviation of Y_i and ρ_{YD} is the correlation between D_i and Y_i. Based on Equations B1 and B2, the specified values of the unstandardized regression coefficient and residual variance in Equation 6 were $b_1 = 0.0000$ and $\sigma_\varepsilon^2 = 1.0000$, $b_1 = 0.0387$ and $\sigma_\varepsilon^2 = 0.9900$, $b_1 = 0.1162$ and $\sigma_\varepsilon^2 = 0.9100$, or $b_1 = 0.1936$ and $\sigma_\varepsilon^2 = 0.7500$ to achieve a zero, weak, moderate, or strong dose-response effect.

Simulation Study 2

To generate the outcome in the treatment condition, values of the unstandardized regression coefficients b_1 and b_2 in Equation 7 were calculated based on the following equations:

$$b_1 = \beta_1\left(\frac{\sigma_Y}{\sigma_D}\right) = \left(\frac{\rho_{YD} - \rho_{YU}\rho_{DU}}{1 - \rho_{DU}^2}\right)\left(\frac{\sigma_Y}{\sigma_D}\right) \tag{B3}$$

$$b_2 = \beta_2\left(\frac{\sigma_Y}{\sigma_U}\right) = \left(\frac{\rho_{YU} - \rho_{YD}\rho_{DU}}{1 - \rho_{DU}^2}\right)\left(\frac{\sigma_Y}{\sigma_U}\right) \quad \text{(B4)}$$

where β_1 and β_2 are the standardized regression coefficients for D_i and U_i, respectively; $\sigma_Y = 1.0000$, $\sigma_D = 2.5820$, and $\sigma_U = 1.0000$ are the standard deviations of Y_i, D_i, and U_i, respectively; ρ_{YD} is the correlation between D_i and Y_i; ρ_{YU} is the correlation between U_i and Y_i; and ρ_{DU} is the correlation between D_i and U_i (see Equation 3.2.4 in Cohen, Cohen, West, & Aiken, 2003). The intercept b_0 in Equation 7 was set to 0.00, 0.20, or 0.50, where nonzero values represent violations of the exclusion restriction. These values correspond to standardized mean differences of δ = 0.00, 0.20, or 0.50, respectively, because the variance of Y_i was set to one in both the treatment and control conditions. Values of the residual variance σ_ε^2 were calculated based on the following equation:

$$\sigma_\varepsilon^2 = (1 - R^2)\sigma_Y^2 = [1 - (\beta_1\rho_{YD} + \beta_2\rho_{YU})]\sigma_Y^2 \quad \text{(B5)}$$

where $\sigma_Y = 1.0000$ is the standard deviation of Y_i; β_1 and β_2 are the standardized regression coefficients for D_i and U_i, respectively (see Equations B3 and B4); ρ_{YD} is the correlation between D_i and Y_i; and ρ_{YU} is the correlation between U_i and Y_i (see Equation 3.5.3 for R^2 in Cohen et al., 2003). The specified values of the unstandardized regression coefficients and residual variance under each condition are summarized in Table B1.

Table B1

Study 2 Specified Values of Unstandardized Regression Coefficients and Residual Variance to Generate Y_i in Treatment Condition

Manipulated Factors		Specified Values in Equation 7		
Dose-Response Effect	Unobserved Confounding	Regression Coefficient for D_i	Regression Coefficient for U_i	Residual Variance
Zero	Absent	0.0000	0.0000	1.0000
	Weak	−0.0039	0.1010	0.9899
	Moderate	−0.0383	0.3297	0.9011
	Strong	−0.1291	0.6667	0.6667
Weak	Absent	0.0387	0.0000	0.9900
	Weak	0.0352	0.0909	0.9818
	Moderate	0.0043	0.2967	0.9099
	Strong	−0.0775	0.6000	0.7200
Moderate	Absent	0.1162	0.0000	0.9100
	Weak	0.1135	0.0707	0.9051
	Moderate	0.0894	0.2308	0.8615
	Strong	0.0258	0.4667	0.7467
Strong	Absent	0.1936	0.0000	0.7500
	Weak	0.1917	0.0505	0.7475
	Moderate	0.1745	0.1648	0.7253
	Strong	0.1291	0.3333	0.6667

APPENDIX C

SAS 9.4 CODE FOR INTENT-TO-TREAT, CONFOUNDER ADJUSTMENT, AND INSTRUMENTAL VARIABLE METHODS

```
/* Intent-to-Treat Method */

PROC REG DATA = example_data;
      MODEL Y = Z;
RUN;


/* Confounder Adjustment Method */

PROC REG DATA = example_data;
      MODEL Y = D;
      WHERE Z = 1; /* Select cases in the treatment condition. */
RUN;


/* Instrumental Variable Method */

PROC SYSLIN DATA = example_data FIRST 2SLS;
      ENDOGENOUS D;
      INSTRUMENTS Z;
      equation: MODEL Y = D / NOINT;
RUN;
```