*University of Pennsylvania*
UPenn Biostatistics Working Papers

*Year* 2005 *Paper* 1

# Casual Mediation Analyses with Structural Mean Models

Thomas R. TenHave* Marshall Joffe† Kevin Lynch‡
Greg Brown** Stephen Maisto††

*ttenhave@cceb.med.upenn.edu
†mjoffe@cceb.upenn.edu
‡lynch_k@mail.trc.upenn.edu
**gregbrow@mail.med.upenn.edu
††



http://biostats.bepress.com/upennbiostat/art1

# Casual Mediation Analyses with Structural Mean Models

Thomas R. TenHave, Marshall Joffe, Kevin Lynch, Greg Brown, and Stephen Maisto

**Abstract**

We represent a linear structural mean model (SMM)approach for analyzing mediation of a randomized baseline intervention's effect on a univariate follow-up outcome. Unlike standard mediation analyses, our approach does not assume that the mediating factor is randomly assigned to individuals (i.e., sequential ignorability). Hence, a comparison of the results of the proposed and standard approaches in with respect to mediation offers a sensitivity analyses of the sequential ignorability assumption. The G-estimation procedure for the proposed SMM represents an extension of the work on direct effects of randomized treatment effects for survival outcomes by Robins and Greenland (1994) (Section 5.0 and Appendix B) and on treatment non-adherence for continuous outcomes by TenHave et al. (2004). Simulations show good estimation and confidence interval performance under unmeasured confounding relative mediation approach. Sensitivity analyses of the sequential ignorability assumption comparing the results of the two approaches are presented in the context of two suicide/depression treatment studies.

# Causal Mediation Analyses with Structural Mean Models

Thomas R. Ten Have
Marshall Joffe
Kevin Lynch
Greg Brown
Stephen Maisto

November 14, 2005

**Abstract**

We present a linear structural mean model (SMM) approach for analyzing mediation of a randomized baseline intervention's effect on a univariate follow-up outcome. Unlike standard mediation analyses, our approach does not assume that the mediating factor is randomly assigned to individuals (i.e., sequential ignorability). Hence, a comparison of the results of the proposed and standard approaches in with respect to mediation offers a sensitivity analyses of the sequential ignorability assumption. The G-estimation procedure for the proposed SMM represents an extension of the work on direct effects of randomized treatment effects for survival outcomes by Robins and Greenland (1994) (Section 5.0 and Appendix B) and on treatment non-adherence for continuous outcomes by Ten Have et al. (2004). Simulations show good estimation and confidence interval performance under unmeasured confounding relative to the standard mediation approach. Sensitivity analyses of the sequential ignorability assumption comparing the results of the two approaches are presented in the context of two suicide/depression treatment studies.

## 1.0 Introduction

We present a causal structural mean modeling (SMM; e.g., Robins (1994); Fischer-Lapp and Goetghebeur (1999); Robins (2003a)) approach for investigating whether a randomized intervention effect on a continuous outcome occurs through or around a post-randomization intermediate factor in the context of randomized behavioral health trials (i.e., mediation analysis). The proposed estimation approach relaxes the no unmeasured confounding assumption for the intermediate factor, which is required for current mediation analysis methods (Judd and Kenny (1981); Baron and Kenny (1986); MacKinnon et al. (2002)). This assumption is equivalent to randomization of the baseline intervention and of subsequent intermediate variables (i.e., "sequential ignorability"; e.g., Robins and Greenland (1992); Pearl and Robins (1995); Robins (1999)). Pearl (2001), Robins (2003b), and Robins and Rotnitzky (2004) addressed the case of

non-parametric non-identifiability of direct effects with single and multiple intermediate factors. Alternatively, Mealli et al. (2004) and Rubin (2004) applied the principal stratification approach of Frangakis and Rubin (2002) to the mediation context, where the intent-to-treat effects in certain principal strata can be interpreted as direct effects of the randomized intervention.

In an attempt to resolve the non-identifiability problem of direct effects raised by Robins and Rotnitzky (2004), we propose a parametric model analogous to the standard mediation model without an interaction between the randomized intervention effect and the mediator. However, in contrast to standard estimation methods (e.g., least squares) for fitting such a model, we propose a weighted G-estimation approach where the weights are based on regressing the intermediate variable or mediator on baseline factors, stratifying by the intervention factor. That is, with no-interaction assumptions and baseline predictors of the intermediate factor, we attempt to resolve the non-parametric non-identifiability issue raised by Robins and Rotnitzky (2004). Comparisons of the results of our approach with that of the standard approach offer sensitivity analyses of the sequential randomization assumption made by standard mediation methods.

The proposed SMM is a linear model extension of the structural failure time model for testing the direct effect of a randomized treatment by Robins and Greenland (1992). In the alternative context of assessing the causal effect of receiving treatment on outcome under treatment non-adherence, Ten Have et al. (2004) extended the weighted testing approach for direct effects on survival outcomes by Robins and Greenland (1992) to continuous outcomes. While we employ the same SMM and corresponding weighted G-estimation approach of Ten Have et al. (2004), we estimate a different parameter in a different context. That is, we focus on the direct effect of the randomized intervention rather than on the effect of the intermediate factor (e.g., effect of receiving treatment on outcome in Ten Have et al. (2004)).

Integrating the results of Joffe and Brensinger (2003), Joffe et al. (2003), Robins and Greenland (1994) (Section 5.0 and Appendix B), this modified G-estimation approach for joint estimation of the effects of the baseline treatment and mediator relies on a weight vector having a separate element for each of these two effects. The weight element for the mediator is a function of the interaction between baseline covariates and the randomized treatment with the mediator

as the dependent variable. The stronger this interaction, the more accurately we can estimate the direct effect of the baseline treatment and the effect of the mediator when the outcome is the dependent variable. We show this to be the case with simulations under the conditions of two pyschiatric behavioral intervention examples.

With data analyses and simulations, our extended SMM approach for mediation analysis will be compared to a standard mediation regression method (e.g., Judd and Kenny (1981); Baron and Kenny (1986); MacKinnon et al. (2002)). Such a method entails a linear regression of the outcome variable on randomized intervention and mediator variable, adjusting for baseline covariates, thus assuming sequential ignorability. The data analyses and simulations will focus on two behavioral intervention studies. The first is a suicide prevention study comparing collaborative care management for treating depression (and thus reducing the risk of suicide) with usual care in 293 elderly depressed primary care patients (Bruce et al. (2004)). The collaborative care management program in the intervention group was based on patient and primary care staff and physician interactions with a nurse-level behavioral health specialist (BHS). One goal of the study was to assess if a direct effect of the intervention occurred apart from use of prescriptive anti-depressant medication in treating depression at 4, 8, 12, 18, and 24 months of follow-up. Here, anti-depressant medication use is the mediator. In this investigation, we focus on estimating the 4 month direct effect of this intervention for the Hamilton depression score.

The second study is a suicide treatment study, evaluating the effect of a specific type of psychological therapy versus usual care in treating depression and suicide ideation in 101 patients who had recently attempted suicide. We refer to this study as the "suicide therapy study" to distinguish it from the first study, the "suicide prevention study." One mediation-oriented goal of the suicide therapy study was to assess if the effect of the randomized therapy occurred apart from use of non-study therapy in treating depression at 1, 3, 6, 12, and 18 weeks of follow-up. Here, non-study therapy is the mediator. We focus on estimating the 6 week direct effect on the Beck Depression Index (BDI) of this intervention.

The paper now proceeds to Section 2 for notation, Section 3 for models, Section 4 for assumptions, Section 5 for estimation, Section 6 for the simulation results, Section 7 for the case study analyses, and Section 8 for the discussion.

## 2.0 Notation

We define the observed and potential variables for participant $i$. However, we suppress the index $i$ to simplify the notation resulting from the addition of indices for the randomized intervention and mediators when defining the potential outcome variables. We first define the observed variables before defining the potential variables.

For the observed variables, $Y$ is the observed continuous outcome; $R$ is the observed randomized zero-one variable; $\mathbf{X}$ is the vector of observed baseline covariates other than randomization; and $M$ is the observed mediation variable. Without loss of generality, we assume $M$ is binary. The SMM approach and corresponding G-estimation equations procedure that we present can accommodate continuous $M$ in a straightforward way.

For the potential outcome, $Y_{r\,m}$ is the outcome that would be observed if participant $i$ were randomized to level $r$ of the intervention and then were to receive or exhibit level $m$ of the mediator. Accordinlgy, with $r$ and $m$ binary, we define four separate potential outcome variables: $Y_{0\,0}$, $Y_{1\,0}$, $Y_{0\,1}$, and $Y_{1\,1}$. With these four potential outcome variables, one can define the causal expectation contrasts for the direct effect of the baseline intervention and the effect of the mediator on outcome. These effects are defined more formally in Section 3.0 below.

The indices of the potential outcome, which represent levels of the baseline intervention and mediator "set" by those in control of these factors (e.g., investigators or clinicians) need to be distinguished from the observed levels of these factors for patient $i$. Given that the set levels of randomized baseline and mediators are denoted by $r$ and $m$, respectively, in the definition of $Y_{r\,m}$, we denote the observed levels of $R$ and $M$ by $\tilde{r}$ and $\tilde{m}$, respectively. To be consistent, we also denote the observed level of the baseline covariates, $\mathbf{X}$, to be $\tilde{\mathbf{x}}$. This distinction between set levels of factors and observed levels of factors is needed for the discussion of assumptions and estimation under the proposed SMM.

## 3.0 SMM Model

In our context, a SMM may be used to model jointly the causal effects of the randomized baseline intervention and the mediator. We present such a SMM in Section 3.1 and the standard model in Section 3.2.

### 3.1 Model

One specification of the SMM is:

$$Y_{r\,m} \quad = \boldsymbol{\beta}^T \tilde{\mathbf{x}} + \quad \theta_M \; m + \theta_R \; r + \epsilon_{r\,m} \tag{1}$$

for all participants regardless of what is actually observed in terms of $R$ and $M$; and where $\theta_M = E(Y_{r\,1} - Y_{r\,0} \mid \mathbf{X} = \tilde{\mathbf{x}})$; $\theta_R = E(Y_{1\,m} - Y_{0\,m} \mid \mathbf{X} = \tilde{\mathbf{x}})$; $\boldsymbol{\beta}$ is a vector of effects for baseline covariate values $\tilde{\mathbf{x}}$; and $\epsilon_{r\,m}$ is a mean zero error term with unspecified distribution with finite mean and variance. Here, $\theta_m$ represents the effect of the mediator on the outcome holding the baseline intervention fixed at $r$; and $\theta_R$ represents the direct effect of the randomized intervention on the outcome, holding the mediator fixed at $m$.

The consistency of the proposed estimators of $\theta_M$ and $\theta_R$ relies on the correct additive specification of $\theta_M \; a + \theta_R \; r$ but not on the correct specification of $\boldsymbol{\beta}^T \tilde{\mathbf{x}}$ or the distribution of $\epsilon_{r\,m}$. However, efficiency depends on how well $\boldsymbol{\beta}^T \tilde{\mathbf{x}}$ approximates the true relationship between $\mathbf{X}$ and $Y_{r\,m}$ (Robins (1994); Fischer-Lapp and Goetghebeur (1999)).

### 3.2 Standard regression model

For comparison with the SMM in (1), we present the corresponding standard linear regression model as presented by a number of authors (e.g., Judd and Kenny (1981); Baron and Kenny (1986); MacKinnon et al. (2002)). This standard linear regression model is defined as:

$$Y \quad = \quad \boldsymbol{\beta}_S^T \tilde{\mathbf{x}} + \theta_{M\;S} \; \tilde{m} + \theta_{R\,S} \; \tilde{r} + \boldsymbol{\beta}^T \mathbf{x} + \epsilon_S \tag{2}$$

for all participants; and where $\theta_{R\,S} = E(Y \mid R = 1, M = \tilde{m}, \mathbf{X} = \tilde{\mathbf{x}}) - E(Y \mid R = 0, M = \tilde{m}, \mathbf{X} = \tilde{\mathbf{x}})$; $\theta_{M\,S} = E(Y \mid R = \tilde{r}, M = 1, \mathbf{X} = \tilde{\mathbf{x}}) - E(Y \mid R = \tilde{r}, M = 0, \mathbf{X} = \tilde{\mathbf{x}})$; $\boldsymbol{\beta}_S$ is a vector of effects for baseline covariate values $\tilde{\mathbf{x}}$; and $\epsilon_S$ is a mean zero error term with a normal distribution and variance equal to $\sigma_S^2$. The parameters $\theta_{R\,S}$ and $\theta_{M\,S}$ are defined as comparisons of observed outcome expectations from different sample subgroups defined by $\tilde{r}$ and $\tilde{m}$ but not as causal contrasts of expectations under different conditions defined by $r$ and $m$ for the same individual. The comparisons of such sub-groups will only equal the causal contrasts for an individual under certain conditions listed below for the standard approach (e.g., sequential ignorability).



## 4.0 Model Assumptions

We first present assumptions for estimating the $\theta_R$ and $\theta_M$ parameters under the SMM in (1) and then the $\theta_{R\,S}$ and $\theta_{M\,S}$ parameters under the standard regression model in (2).

### 4.1 SMM Assumptions

For the SMM model, the assumptions necessary for unbiased inference are: 1) Stable Unit Treatment Value Assumption (SUTVA); 2) randomization (i.e., ignorability) of baseline intervention assignment; and 3) model assumptions including the no-interaction assumption between the baseline randomized intervention and the mediator.

#### 4.1.1 SUTVA Assumption

SUTVA consists of two sub-assumptions. First, there is a single value for each of the potential random outcome variables ($Y_{r\,m}$) for a given patient $i$ regardless of the randomization assignment of any other patient $i'$. Notationally, this assumption implies that $Y_{r\,m}$ is defined with scalar indices for a given participant $i$, rather than vectors of indices representing baseline treatment assignments and mediator levels of all patients. This first component of SUTVA may be vulnerable when interventions require each provider to treat multiple patients so that one patient's treatment may be related to another person's treatment. This was the case with the behavioral health specialist in the suicide prevention study.

Second, there is a single value for each of the potential outcome random variables ($Y_{r\,m}$) for a given patient $i$ regardless of the method of administration of the randomized baseline intervention or the administration or occurrence of the mediator. This assumption is known as the consistency assumption as it addresses consistency of an outcome across variations of administration of treatment (Rubin (1986)). Notationally, consistency implies for patient $i$ with observed levels $\tilde{r}$ and $\tilde{m}$ for $R$ and $M$, respectively: $Y = \tilde{r}\ \tilde{m}\ Y_{\tilde{r}\,\tilde{m}} + (1-\tilde{r})\ \tilde{m}\ Y_{1-\tilde{r}\ \tilde{m}} + \tilde{r}\ (1-\tilde{m})\ Y_{\tilde{r}\ 1-\tilde{m}} + (1-\tilde{r})\ (1-\tilde{m})\ Y_{1-\tilde{r}\ 1-\tilde{m}}$. Such an identity only holds for binary $r$ and $m$, but extends in a straightforward way to continuous $m$. Hence, SUTVA allows us to relate the potential to observed outcomes and thus perform estimation under the other assumptions.

#### 4.1.2 Randomization Assumption

The randomization assumption for the SMM in (1) implies stochastic independence between the randomized baseline intervention, $R$, and potential outcomes (i.e., ignorability of $R$).

Stochastically, this means for the potential outcomes:

$$\Pr\left(Y_{1,1}, Y_{1,0}, Y_{0,1}, Y_{0,0} \mid R = \tilde{r}, \mathbf{X} = \tilde{\mathbf{x}}\right) \quad = \quad \Pr\left(Y_{1,1}, Y_{1,0}, Y_{0,1}, Y_{0,0} \mid \mathbf{X} = \tilde{\mathbf{x}}\right) . \tag{3}$$

Such an assumption implies no imbalance between randomization groups with respect to unmeasured confounders, i.e., no unmeasured confounding.

We note that for the suicide prevention study, primary care practices were randomized. However, because the within-practice design effect is so small for the outcome, the Hamilton depression scale, we ignore the clustering due to primary care practice (Bruce et al. (2004); Small et al. (2005)).

#### 4.1.3 Model Assumptions

The model assumptions for the SMM in (1) include the the additive structure of the baseline intervention and mediating factors and the no-interaction assumptions. The additive structure in (1), $\theta_M\ m + \theta_R\ r$, is the only modeling assumption needed for consistent estimation of $\theta_M$ and $\theta_R$ apart from the no-interaction assumption below. Again, consistent estimation does not depend on the correctness of $\boldsymbol{\beta}^T\tilde{\mathbf{x}}$ or of the distribution of the error term, $\epsilon_{r\,m}$, although more accurate specifications improve efficiency of estimation (see Robins (1994)).

The no-interaction assumption under the SMM in (1) consists of two components. First, in terms of "setting" the randomization and mediation levels at $r$ and $m$, $\theta_M$ does not depend on the level to which the randomized intervention, $r$, is set; nor does $\theta_R$ depend on the level to which the mediation level, $m$, is set. Second, the causal effects of the baseline intervention and mediator, $\theta_R$ and $\theta_M$, respectively, do not vary across observed sub-groups defined by different combinations of the observed variables $R$, $M$, and $X$.

Notationaly, the two components of the no-interaction assumption can be expressed as follows. For $\theta_M$, let

$$\theta_{M\,r}(\tilde{r}, \tilde{m}, \tilde{\mathbf{x}}) \quad = \quad E(Y_{r,1} \mid R = \tilde{r}; M = \tilde{m}; \mathbf{X} = \tilde{\mathbf{x}}) - E(Y_{r,0} \mid R = \tilde{r}; M = \tilde{m}; \mathbf{X} = \tilde{\mathbf{x}}) . \tag{4}$$

for set $r$ and observed levels $\tilde{r}$, $\tilde{m}$, and $\tilde{\mathbf{x}}$. Then it follows under the no-interaction assumption that $\theta_{M\,r}(\tilde{r}, \tilde{m}, \tilde{\mathbf{x}}) = \theta_M$. A similar identity holds for $\theta_R$ with respect to set level of $m$ and observed levels $\tilde{r}$, $\tilde{m}$, and $\tilde{\mathbf{x}}$. This no-interaction assumption is similar to that made for an

efficacy analysis under treatment non-compliance, where the focus is on the effect of treatment-received (Robins (1994); Fischer-Lapp and Goetghebeur (1999)).

### 4.2 Standard Regression Assumptions

The standard regression model assumptions are well known. We have: 1) sequential ignorability; 2) independence among participants; and 3) model assumptions including a no-interaction assumption. We focus on the sequential ignorability and no-interaction assumptions.

#### 4.2.1 Sequential Ignorability

While the SMM in (1) requires ignorability of $R$ under randomization of the baseline intervention assignment, the standard regression model in (2) requires sequential ignorability assumption for both the baseline intervention and mediator. The sequential ignorability assumption implies stochastic independence between these two factors and the potential outcomes, conditional on baseline covariates. Stochastically, this implies:

$$\Pr\left(Y_{1,1}, Y_{1,0}, Y_{0,1}, Y_{0,0} \mid R = \tilde{r}, M = \tilde{m}, \mathbf{X} = \mathbf{x}\right) \quad = \quad \Pr\left(Y_{1,1}, Y_{1,0}, Y_{0,1}, Y_{0,0}, \mid \mathbf{X} = \tilde{\mathbf{x}}\right). \quad (5)$$

The no confounding assumption for the mediator that is made in the literarture on standard mediation methods (e.g., Judd and Kenny (1981); Baron and Kenny (1986); MacKinnon et al. (2002)) requires the identity (5).

#### 4.2.2 No-interaction assumption

The no-interaction assumption under the standard model in (2) requires invariance of the standard regression effects of the baseline intervention and mediator, $\theta_{S\,R}$ and $\theta_{M\,R}$, respectively, across observed sub-groups defined by different combinations of the observed variables $R$, $M$, and $X$. Notationally, the no-interaction assumption for the standard regression effects $\theta_{S\,R}$ and $\theta_{M\,R}$ can be expressed as follows. For $\theta_{M\,S}$, let

$$\theta_{M\,S}(\tilde{r}, \tilde{\mathbf{x}}) \quad = \quad E(Y \mid R = \tilde{r}; M = 1; \mathbf{X} = \tilde{\mathbf{x}}) - E(Y \mid R = \tilde{r}; M = 0; \mathbf{X} = \tilde{\mathbf{x}}). \quad (6)$$

for observed levels $\tilde{r}$ and $\tilde{\mathbf{x}}$. Then it follows under the no-interaction assumption that $\theta_{M\,S}(\tilde{r}, \tilde{\mathbf{x}}) = \theta_{M\,S}$. A similar identity holds for $\theta_{R\,S}$ with respect to the observed levels $\tilde{m}$ and $\tilde{\mathbf{x}}$. Note $\theta_{M\,S}(\tilde{r}, \tilde{\mathbf{x}})$ is not a function of the "set" levels $r$ and $s$, in contrast to the interaction function $\theta_{M\,r}(\tilde{r}, \tilde{m}, \tilde{\mathbf{x}})$ in (4).

## 5.0 Estimation

Under the assumptions in Section 4.0 and SMM in (1), consistent estimators of $\theta_R$ and $\theta_M$ can be obtained by solving the following weighted G-estimation equations for $\theta_M$, $\theta_R$, and $\boldsymbol{\beta}$. To obtain these equations based on observed data, we first relate the observed and potential outcome variables as follows with the potential outcome indices in (1) equal to the corresponding observed outcome indices, $m = \tilde{m}$ and $r = \tilde{r}$:

$$Y = Y_{0,0} + \tilde{m}\theta_M + \tilde{r}\theta_R + \delta_{\tilde{r},\tilde{m}} \tag{7}$$

where $\delta_{\tilde{r},\tilde{m}} = \epsilon_{\tilde{r},\tilde{m}} - \epsilon_{0,0}$ and $\epsilon_{\tilde{r},\tilde{m}}$ is a mean zero error. The $\delta_{\tilde{r},\tilde{m}}$ parameterization allows the consistency assumption to be satisfied when $\tilde{m} = \tilde{r} = 0$: $Y = Y_{0,0} = \boldsymbol{\beta}^T\tilde{\mathbf{x}} + \epsilon_{0,0}$.

For estimation of $\theta_M$ and $\theta_R$ based on (7), we obtain a candidate value for $Y_{0,0}$ for each combination of $\tilde{m}$ and $\tilde{r}$: $Y_{0,0}(\boldsymbol{\theta}^*) = Y - \tilde{m}\,\theta_M^* - \tilde{r}\,\theta_R^*$ where $\boldsymbol{\theta}^{*T} = (\theta_M^*\,\theta_R^*)$, and the elements of which are putative or candidate values for $\theta_M$ and $\theta_R$. When $\theta_R^* = \theta_R$ and $\theta_M^* = \theta_M$ under the SMM and $\hat{\boldsymbol{\beta}}$ is some estimate of $\boldsymbol{\beta}$, $Y_{0,0}(\boldsymbol{\theta}^*) - \hat{\boldsymbol{\beta}}^T\tilde{\mathbf{x}}$ and the randomized baseline intervention, $R$, are uncorrelated. Hence, we can obtain consistent estimators of $\theta_M$ and $\theta_R$ by iteratively solving the following unbiased estimating equation using a Newton-Raphson routine:

$$\sum (R - q)\mathbf{W}(\tilde{\mathbf{x}})(Y_{0,0}(\boldsymbol{\theta}) - \hat{\boldsymbol{\beta}}^T\tilde{\mathbf{x}}) \;\; = \;\; 0\,, \tag{8}$$

where $\boldsymbol{\theta}^T = (\theta_M\,\theta_R\,)$; $q \equiv Pr(R = 1)$, the proportion randomized to the baseline intervention; $\hat{\boldsymbol{\beta}}$ is obtained from a linear regression of $Y_{0,0}(\hat{\boldsymbol{\theta}})$ on $\mathbf{X}$ given an estimate of $\boldsymbol{\theta}$ from the previous iteration, $\hat{\boldsymbol{\theta}}$; and $\mathbf{W}(\tilde{\mathbf{x}})$ is a weight vector function of the observed elements of $\mathbf{X}$. Such functions can be chosen on the basis of optimal efficiency using criteria in Robins (1994). We note that the specification of $\mathbf{X}$ may differ between the functions $\boldsymbol{\beta}^T\tilde{\mathbf{x}}$ for the mean model in (1) and $\mathbf{W}(\tilde{\mathbf{x}})$. Both specficiations affect the efficiency but not the bias of the estimates arising from (8).

We specified $\mathbf{W}(\tilde{\mathbf{x}})^T = [1 \;\; \eta(\tilde{\mathbf{x}})]$ for subject $i$, with the two elements corresponding to $\theta_R$ and $\theta_M$, respectively. In the context of treatment non-adherence, the element corresponding to $\theta_M$ is the "compliance score," $\eta(\tilde{\mathbf{x}}) = \Pr(M = 1 \mid R = 1, \mathbf{X} = \tilde{\mathbf{x}}) - \Pr(M = 1 \mid R = 0, \mathbf{X} = \tilde{\mathbf{x}})$ (Joffe et al. (2003); Follmann (2000)). Given the context of mediation, we refer to this score as the "mediation score." The variability of the mediation score is a measure of the interaction between baseline covariates and the randomized intervention factor with the mediator as the dependent variable. The more baseline covariates interact with the randomized intervention (i.e.,

the more variability in the mediation score), the more precise are the estimates of $\theta_M$ and $\theta_R$. In estimating the effect of the mediator on outcome ($\theta_M$), the mediation score, $\eta(\tilde{\mathbf{x}})$, downweights participants characterized by $\mathbf{x}$ for whom the randomized intervention effect on the mediator is small and thus contributing little information to estimating the path through the mediator to outcome. Similarly, $\eta(\tilde{\mathbf{x}})$ upweights participants characterized by $\mathbf{x}$ for whom the intervention effect on the mediator is large and thus contributing more information to estimating the path through the mediator to outcome. For this paper, estimation of $\Pr(M \mid R, \mathbf{X})$ was based on the logistic model. However, the consistency of this estimating approach is not impacted by the specification of $\eta(\tilde{\mathbf{x}})$, although the efficiency of $\hat{\boldsymbol{\theta}}$ is.

Extending Joffe and Brensinger (2003), the variance-covariance for $\hat{\boldsymbol{\theta}}$ is estimated after convergence of the G-estimation algorithm with a sandwich estimator based on (8) as follows: $\text{VarCov}(\hat{\boldsymbol{\theta}}) = \mathbf{D}^{-1}\mathbf{H}^{-1}\mathbf{D}^{-1\,T}$, where $\mathbf{D}$ is a symmetric $2 \times 2$ matrix: $\mathbf{D} = \sum \frac{\partial \mathbf{S}}{\partial \boldsymbol{\theta}}$; $\mathbf{S}$ is a $2 \times 1$ column vector for patient $i$: $\mathbf{S} = (R - q)\left(Y_{0,0}(\boldsymbol{\theta}) - \boldsymbol{\beta}^T \tilde{\mathbf{x}}\right)\mathbf{W}(\tilde{\mathbf{x}})$; and $\mathbf{H}$ is a $2 \times 2$ matrix: $\mathbf{H} = \sum \mathbf{S}\mathbf{S}^T$. Instead of using the full inverse matrix of $\mathbf{H}$ above for $\text{VarCov}(\hat{\boldsymbol{\theta}})$, we use the elements of $\mathbf{H}^{-1}$ corresponding to the original elements of $\mathbf{S}$ defined with $\mathbf{W}(\tilde{\mathbf{x}})$. The resulting estimate of $\text{VarCov}(\hat{\boldsymbol{\theta}})$, evaluated at $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$, is used in Wald statistics for hypothesis testing and Wald confidence intervals for $\theta_M$ and $\theta_R$.

**6.0 Simulations**

We now present simulation results for the effects of the randomized baseline treatment and mediation factors under the conditions of the two example trials. Specifically, two sets of simulations were performed, one for each of the example datasets. Each dataset for each set of simulations was based on the corresponding characteristics of the respective example dataset and fitted SMM's: 1) sample size of the dataset (293 for the suicide prevention study and 101 for the suicide therapy study); 2) observed values of $\mathbf{X}$ and $R$ for each subject in each study; and 3) the specifications of the model parameters in $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, and $\eta(\tilde{\mathbf{x}})$. Given these specifications, we simulated $Y_{r\,m}$ and $M$.

The only unknown parameter specification not provided by estimates for the observed datasets is that for unmeasured factors related to $Y$ and $M$ (i.e., unmeasured confounders of the relationship between $Y$ and $M$). Under the SMM framework, we do not need to spec-

ify these relationships correctly. For the simulations, we assumed that the error term for $Y_{r\,m}$ in (1), $\epsilon_{r\,m}$, was decomposed into two components, one of which was related to a model for $M$. That is, we specified the following shared parameter framework: $\epsilon_{r\,m} = \epsilon^*_{r\,m} + \nu_Y \gamma$ and $\Pr(M = 1 \mid R = \tilde{r}, \mathbf{X} = \tilde{\mathbf{x}}) = \text{expit}\,(\boldsymbol{\beta}^T \tilde{\mathbf{x}} + \alpha\ r + \nu_M \gamma)$, where $\gamma$ is a normal random variable with mean zero and variance equal to one; $\nu_Y = \nu_M = 1$; $\epsilon^*_{r\,m}$ is a normal random variable with mean zero and variance equal to the variance of the oberved outcome variable in the particular example dataset; and $\boldsymbol{\beta}$ and $\alpha$ are specified to be equal to the corresponding naive estimates from the logistic regression of $M$ on $\mathbf{X}$ and $R$ without consideration of $\gamma$.

For each of the two groups of simulation specifications, we simulated 1000 sets of data for $Y$ and $M$. From the corresponding 1000 sets of fitted SMM's under (1), we computed for $\theta_R$ and $\theta_M$, the absolute and relative bias of the SMM estimate, the mean squared error (MSE), and confidence interval coverage (proportion of iterations for which the 95% confidence interval included the true value of $\theta_R$ or $\theta_M$). We computed the same simulations statistics for the standard regression model in (2). The results of the simulations for the suicide prevention and therapy studies are presented in Tables 1 and 2, respectively.

Tables 1 and 2 show that the SMM approach yields smaller bias and more accurate confidence intervals than the standard regression procedure, but at the expense of larger MSE due to greater variability. This result is consistent between the two example-based simulations and between parameters, $\theta_R$ and $\theta_M$. The reduction in bias under the SMM compared to standard regression ranges from a 257 percentage point decrease (258% to 1%) to a 5 percentage point decrease (10% to 5%). There is somewhat less consistency for the improvement in 95% confidence interval coverage under the SMM approach relative to the standard regression procedure. The improvement in coverage under the SMM ranges from the largest change of 2 to 99% for $\theta_M$ in the suicide prevention study (Table 1) to a change of 95 to 97% for $\theta_R$ in the suicide therapy study (Table 2).

The above improvements under the SMM approach relative to the standard mediation approach need to be weighed against the increased variability of the SMM relative to the standard regression approach. While it tends to yield more bias and less accurate confidence intervals than the SMM approach, the standard regression procedure is less variable in terms of the MSE

**Table 1**: Simulation results based on the suicide prevention study (N=293; $\theta_M = -1.43$ and $\theta_R = -2.58$).

| Method | Simulation Statistic | Mediation Effect ($\theta_M$) | Direct Effect ($\theta_R$) |
|---|---|---|---|
| Standard | Bias (%) | 3.68 (258%) | 0.79 (31%) |
| SMM | Bias (%) | 0.02 (1%) | 0.00 0%) |
| Standard | % Coverage | 2% | 84% |
| SMM | % Coverage | 99% | 95% |
| Standard | MSE | 14.37 | 1.46 |
| SMM | MSE | 23.43 | 1.92 |

uniformly across examples and parameters. The increase in MSE under the SMM approach ranges from a change of 1.46 to 1.92 for $\theta_R$ to a change of 25.11 to 326.95 for $\theta_M$. Finally, the simulations show that the proposed SMM estimator of $\theta_R$ can perform better than the proposed SMM estimator of $\theta_M$ uniformly across all simulation conditions.

Overall, the simulation results for the SMM approach under the suicide prevention study conditions are better than the analogous results for the suicide therapy study conditions with the smaller sample size. One can attribute this difference in results between study conditions to either the sample size difference and/or other differences between study conditions. To see if the sample size was a factor, we increased the sample size of the simulation for Table 2 under the suicide therapy study to equal the sample size for the suicide prevention study corresponding to

**Table 2**: Simulation results based on the suicide therapy study (N=101; $\theta_M = 14.59$ and $\theta_R = -3.93$).

| Method | Simulation Statistic | Mediation ($\theta_M$) | Direct Effect ($\theta_R$) |
|---|---|---|---|
| Standard | Bias (%) | -4.05 (-28%) | -0.41 (-10%) |
| SMM | Bias (%) | 0.40 (3%) | -0.18 (5%) |
| Standard | % Coverage | 73% | 95% |
| SMM | % Coverage | 90% | 97% |
| Standard | MSE | 25.11 | 7.10 |
| SMM | MSE | 326.95 | 11.60 |

Table 1. The resulting MSE's and bias for the SMM estimators of $\theta_M$ and $\theta_R$ (not displayed) are smaller than in Table 2, but still clearly larger than in Table 1.

Another factor in the discrepancy between the simulation results in Tables 1 and 2 may be the weak relationship between the randomized therapy intervention and the non-study therapy mediator (see Table 5). We altered this relationship in a number of different ways to see if it was such a factor in the discrepancy. Increasing the difference between the randomized groups with respect to the mediator (non-study therapy), reversing the sign of this difference, or increasing the overall proportion of the positive level of the mediator improved results relative to Table 2, somewhat. However, the results for MSE and bias were still not as good as in Table 1 under each of these changes. Hence, we conclude from this limited simualtion experiment that

the deterioration in the behavior of the SMM estimators (especially for $\theta_M$) under the suicide therapy study conditions in Table 2 may be due to a number of factors including the smaller sample size and the weak relationship between the baseline intervention and mediator.

Additionally, we repeated the simulations under conditions from both examples but assuming no variability in the mediation score, $\eta(\tilde{\mathbf{x}})$ (i.e., no interaction between the baseline covariates and intervention with respect to the mediator). The results clearly showed a severe deterioration in bias and MSE of $\hat{\theta}_M$ under both sets of study conditions. In contrast, the minimal bias for $\hat{\theta}_R$ observed in Tables 1 and 2 below was still observed under no covariate-baseline intervention interactions for $M$, although the MSE for $\hat{\theta}_R$ was at least twice as large than the respective MSE reported for $\hat{\theta}_R$ in Tables 1 and 2. In summary, the bias of $\hat{\theta}_R$ appears to be impacted slightly by the magnitude of the baseline covariate-intervention interaction on the mediator, although the variability of $\hat{\theta}_R$ is increased, as is the bias and variability of $\hat{\theta}_M$.

Finally, we have attempted to find the sample size at which the asymptotic results in Tables 1 and 2 deteriorated for each of the two study settings. For the SMM estimators under the suicide therapy study conditions (Table 2), halving the sample size to 50 (25 in each randomization group) did not adversely impact bias, but did increase MSE, especially for the direct effect of the baseline intervention (500% increase). Halving the sample size also substantially decreased coverage for the mediator effect on outcome to 80%. Further decreasing the sample size to 40 increased bias by 10% for the direct effect of the baseline intervention but did not increase bias for the effect of the mediator. Under the suicide prevention study conditions, halving the sample size to 150 (75 in each group) for the SMM estimators under the suicide prevention study conditions adversely impacted bias (30-40% increase) and MSE but not confidence interval coverage. Hence, study conditions appear to impact the minimal sample size necessary for inference under the proposed SMM approach. In one case, the threshold sample size for asymptotic validity is approximately 40, whereas in the other, it is greater than 150.

**7.0 Data Analyses**

In Tables 3 to 6, the two examples reveal how the SMM approach can be a check of the sensitivity of the standard mediation analysis to the sequential ignorability assumption. Tables 3 and 5 pertain to the suicide prevention study (N=293), while Tables 4 and 6 pertain to

**Table 3**: For the suicide prevention study, means (standard deviations in parentheses) and and proportions for the Hamilton outcome and proportion of patients taking anti-depressant medication, respectively, by randomized intervention group or by whether they took anti-depression medication.

| Group | Hamilton | Medication |
|---|---|---|
| Usual Care | 13.55 (8.35) | 0.45 |
| Intervention | 11.50 (7.38) | 0.85 |
| No medication | 13.14 (8.09) | |
| Medication | 12.23 (12.23) | |

the suicide therapy study (N=101). For each example, the first table (Tables 3 or 4) presents descriptive means and proportions for ITT differences in the continuous outcome and the binary mediator factor. The second table (Tables 5 and 6) for each example presents the estimated ITT effect on outcome and the SMM and standard regression estimates of the effects of the baseline intervention and the mediator factor. The analyses presented below for each of the studies assume there is no interaction between the baseline intervention and the mediator. Tests of such interactions based on the standard regression model yielded p-values of 0.53 and 0.77 for the suicide prevention and therapy studies, respectively. While one may attribute the lack of significant interactions to lack of power, the magnitude of the p-values suggests a much larger sample size will be needed to achieve significant interactions.

The descriptive statistics in Tables 3 and 4 suggest similarities between the two examples in terms of the ITT comparisons of outcome but not in terms of the ITT comparison of the mediator factor. The ITT mean contrasts for outcome are significant in both studies. Tables 5 and 6 confirm this with the corresponding model-based ITT confidence intervals. Hence, an analysis of the mediation of these significant ITT effects is justified.

**Table 4**: For the suicide therapy study, Means (standard deviations in parentheses) and and proportions for the BDI outcome and proportion of patients using non-study therapy (mediator), respectively, by randomized intervention group or by whether they used non-study therapy.

| Group | BDI | Non-study Therapy |
|---|---|---|
| Usual Care | 19.33 (12.07) | 0.25 |
| Study Therapy | 14.02 (14.77) | 0.08 |
| No Non-study Therapy | 17.08 (14.78) | |
| Non-study Therapy | 15.11 (12.07) | |

Tables 3 and 4 do indicate differences between the two examples in terms of the level of use of the mediator factor by patients and also the ITT effect on the mediator factors. These differences in descriptive statistics coincide with the different clinical meanings of these two mediators. Most of the depressed patients in the suicide prevention study used medication regardless of whether they were in the behavioral health specialist arm or not. In contrast, in the suicide therapy study, fewer of the suicidal patients used non-study therapy in either arm, although a higher proportion of the usual care group used non-study therapy than the randomized therapy group. These mediator results in both examples were expected by the respective clinical investigators. One of the goals of the BHS intervention in the suicide prevention study was to increase the use of anti-depressive medications among the depressed patients, whereas the randomized therapy intervention was not intended to increase the use of non-study therapy, although it was not discouraged. Nonetheless, the study investigators wanted to see how the effectiveness of the study therapy was due to differential non-study therapy use between patients in the two randomized arms.

**Table 5**: For the suicide prevention study, ITT, standard regression, and SMM estimates are presented for the direct effects of the randomized behavioral health specialist intervention and the mediator (anti-depressant medication). Standard errors and nominal 95% confidence intervals are in parentheses.

| Method | Direct Effect | Medication Effect |
|---|---|---|
| ITT | -3.12 (0.82) | |
| | (-4.72, -1.51) | |
| Standard | -2.67 (0.89) | -1.19 (0.94) |
| | (-4.41, -0.93) | (-3.03, 0.65) |
| SMM | -2.58 (1.27) | -1.43 (2.34) |
| | (-5.07, -0.10) | (-6.01, 3.15) |

Given the differences between the two examples with respect to the mediator results in Tables 3 and 4, we now proceed to the sensitivity analysis of the standard mediation results for each example, comparing the SMM and standard regression results in Tables 5 and 6. In Table 5 for the suicide prevention study example (N=293), we see that the SMM and standard regression approaches yield similar estimates of $\theta_M$ and $\theta_R$. In contrast, in Table 6 for the suicide therapy study (N=101), we see that the SMM and standard mediation methods lead to different results. This difference may indicate less unmeasured confounding in the suicide prevention study than in the suicide prevention study. We now examine this evidence in more detail in terms of these two studies.

**Suicide Prevention Study**

**Table 6**: For the suicide treatment study, ITT, standard regression, and SMM estimates of effects are presented for the randomized therapy intervention and the mediator (non-study therapy). Standard errors and nominal 95% confidence intervals are in parentheses.

| Method | Direct Effect | Non-Study Therapy Effect |
|---|---|---|
| ITT | -6.35 (2.53)<br>(-11.37, -1.33) | |
| Standard | -6.86 (2.60)<br>(-12.01, -1.70) | -3.05 (3.46)<br>(-9.92, 3.82) |
| SMM | -3.93 (3.09)<br>(-9.98, 2.12) | 14.59 (15.87)<br>(-16.52, 45.69) |

The SMM and standard regression estimates for the suicide intervention study in Table 5 are in agreement in estimating a significant direct effect of the behavioral health specialist intervention apart from increasing anti-depressant use among the depressed patients. The estimated direct effect of this intervention under both the SMM and standard regression approaches is an approximate reduction of 2.5 Hamilton units. However, the SMM confidence interval is wider than the standard regression confidence intervals, as one would expect from the MSE results in the simulations. Nonetheless, both the SMM and standard regression intervals do not surround the null value of zero. The significant direct effects of the presence of BHS on reducing depression could be the result of the impact of this specialist on the staff and physicians of the practices. That is, one would expect that the presence of BHL would raise the sensitivity of the staff and providers in treating depression. We also see that both the SMM and standard

regression approaches indicate a non-significant result for the mediator (medication use) effect on outcome.

Estimating the direct effect of the BHS intervention under the SMM approach required covariates that interact with the randomized intervention factor in terms of the mediator. One strategy for identifying such predictors is to perform logistic regression of medication use on baseline covariates stratified by randomization arm. For the group not randomized to the BHS, we did not find any signficant predictors of taking medication ($p > .50$), except for baseline anti-depressant medication status ($p = .03$). For the group randomized to the BHS, site, past medication history, and baseline medication status are strongly predictive of the mediator medication factor ($p < .001$). The difference in predictive power of the baseline factors between the randomized groups is reflected in the difference in the Area-Under-the-Curve (AUC), which is a measure of fit of the logistic model. The AUC for the randomized to BHS group is 0.92. compared to the AUC of 0.67 for the randomized to usual care group. The distribution of the estimated mediation scores based on these predictive factors, $\hat{\eta}(\tilde{\mathbf{x}})$, appears to be sufficient, as evidenced by the range of scores (-0.08 to 0.72) and quartiles (-0.06, 0.55, and 0.70).

**Suicide Therapy Study**

In contrast to the suicide prevention study in Table 5, the SMM and standard regression estimates for the suicide therapy study in Table 6 are not in agreement, indicating possible unmeasured confounding of the standard regression results. Specifically, for the suicide therapy study, G-estimation under the proposed SMM leads to a reduced estimate of $\theta_R$ relative to the standard regression estimate. Hence, under the standard approach there is a significant direct effect of the study therapy apart from any impact on the use of non-study therapy, whereas the SMM approach indicates that there is not sufficient evidence for such inference. This contrast in inference for the direct effect of the study therapy between the two approaches may indicate a departure from the sequential ignorability assumption made by the standard approach. Such a departure may arise because of unmeasured variables related to non-study therapy use and the outcome, depression at 6 weeks. One such variable may be stigma, which may reduce the use of non-study therapy and other treatments, thus increasing depression. Adjusting for such confounding with baseline randomization as implemented in the G-estimation approach would

possibly lead to a reduced estimate of the direct effect, as is the case in Table 6.

Inferentially, the SMM and standard approaches agree with respect to the effect of non-study therapy on the outcome ($\theta_M$), in that both approaches yield confidence intervals surrounding one. However, the SMM-based estimate of $\theta_M$ is large, as is its standard error. This result conforms to the large simulation-based MSE for $\theta_M$ in Table 2. Nonetheless, Table 2 indicates such variability in the $\theta_M$ estimate does not preclude more accurate inference from the G-estimation estimate of $\theta_R$.

In assessing the amount of information for more efficient SMM estimation of the direct effect of randomized therapy, we again evaluate the predictors of the the mediator, taking non-study therapy, stratifed by randomization arms. For the group not randomized to the study therapy, we did not find any significant predictors of non-study therapy before 6 weeks ($p > .45$), except for baseline suicide ideation status ($p = .03$). For the group randomized to the study therapy, we did not find any significant predictors ($p > .30$) of non-study therapy use before 6 months. The AUC's for the logistic models including baseline suicide ideation status and apriori-chosen but non-signficant factors (race, gender, and baseline BDI) are very similar for the two randomized groups: the AUC for the randomized to therapy group is 0.77 compared to the AUC of 0.74 for the randomized to usual care group. Hence, the distribution of the estimated mediation scores based on these predictive factors, $\hat{\eta}(\tilde{\mathbf{x}})$, may not have been as sufficient as for the suicide prevention study. Nonetheless, the suicide therapy study appeared to have a wider range of estimated mediation scores (-0.99 to 0.21) than did the suicide prevention study (-0.72 to 0.08). The spread of the quartiles for the suicide therapy study mediation scores (-0.20, -0.10, and 0.01) indicates some skewness but with higher mass toward zero.

**8.0 Discussion**

We have proposed a new approach to analyzing direct effects of randomized baseline interventions in the presence of a post-randomization mediaton factor. This approach is based on a linear model extension of a weighted test-based approach by Robins and Greenland (1992) for testing direct and mediator effects effects with respect to survival outcomes. A similar approach was implemented by Ten Have et al. (2004) but in a different context, that of treatment non-compliance with treatment-received as the post-randomization factor. In contrast to Ten Have

et al. (2004), who investigated estimation of the intermediate factor (adherence to randomized intervention), we focus this approach on estimating the direct effect of the randomization factor when the post-randomization factor is a mediator. The modified G-estimation approach is based on separate weights for the direct baseline and mediator effects. We show through simulations that the strength of the baseline covariate-randomized intervention interaction on the mediator is crucial in estimating the effect of the mediator, but somewhat less so for estimating the direct effect of the randomized baseline intervention.

In this paper, we related the proposed SMM approach to the standard regression approach to mediation analysis in two ways. To examine estimation properties under unmeasured confounding, we have compared these two approaches through simulations. As sensitivity analyses of the randomized mediatior assumption under the standard regression approach, we have compared these approaches in data analyses for the two suicide studies. The simulation results show that the SMM approach yields less bias and improved confidence interval coverage than the standard regression approach under departures from the sequential ignorability assumption, especially for the direct effect of the baseline randomized factor. However, this improvement comes at the expense of increased variability on the part of the SMM estimators. Weighing variability against bias, the SMM approach is compatible with the clinical trial strategy of protecting against bias (e.g., setting the Type I error), while trying to minimize variability with sample size. The increased variability of the SMM estimatators notwithstanding, the resulting sensitivity analyses for the two studies indicate that there may be unmeasured confounding biasing the standard regression results in one study but not other.

A number of limitations of the proposed SMM approach involve model specification and estimation. First, the SMM approach does yield more variable estimates than does the standard regression approach with variability depending on the strength of the baseline covariate-randomized intervention interaction on the mediator. Despite increased variability, the proposed SMM approach is useful in providing a sensitivity analysis of the randomized mediator assumption. Another limitation is the assumption of no interaction between the baseline intervention and mediator factors, although in each of the example studies, there was little evidence in the data of such an interation.

There are a number of potential extensions of our proposed SMM approach. First, the proposed methodology can be extended to binary outcomes (e.g., change in Hamilton greater than 50%) based on adding a direct effect of the baseline randomization to the logistic models of Ten Have et al. (2004) and Vansteelandt and Goetghebeur (2003). Second, analyzing sequences of multiple mediators is of interest (e.g., when considering medication or therapy use in each of the stages of acute, continuation, and maintenance treatment). Such an approach requires an element in $\mathbf{W}(\tilde{\mathbf{x}})$ for the mediator at each stage. Third, the presented approach for univariate outcomes may be extended to longitudinal binary and continuous outcomes, using generalized estimating equations with working correlation structures (e.g., multiple visits corresponding to the acute, continuation, and maintenance stages of treatment of depression). Fourth, development of a causal approach to assessing the interaction between the baseline intervention and mediator factors is underway (Joffe and Small (2005)).

Finally, to fully understand the mediating mechanism of a baseline intervention, the indirect effect of the intervention through the mediator may be of interest. While this paper focuses on estimation of the direct effect of the baseline intervention on outcome under the SMM, future work will use this model to estimate an indirect effect following the strategy proposed by Pearl (2001) and Robins (2003a). Such a strategy entails transforming the direct effects of the baseline intervention and mediator under the SMM to "natural" indirect effects. By "natural," Pearl (2001) and Robins (2003a) defined such an effect as the effect of changing the intermediate factor behavior of a patient if the baseline intervention assignment were hypothetically switched, but actually held to a constant level (randomized comparison group or randomized intervention group). Allowing the mediator level to vary this way is more appropriate, because it is causally impossible to assess the effect of the baseline intervention on outcome through the mediator if the mediator is held fixed (e.g., Pearl (2001). Future extensions of the SMM approach to estimate indirect effects will follow this strategy.

## References



Baron, R. M. and Kenny, D. A. (1986), "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations," *Journal of*

*Personality and Social Psychology*, 51, 1173–1182.

Bruce, M., Ten Have, T., Reynolds, C., and et al. (2004), "A Randomized Trial to Reduce Suicidal Ideation and Depressive Symptoms in Depressed Older Primary Care Patients: The PROSPECT Study," *Journal of the American Medical Association*, 291, 1081–1091.

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

Follmann, D. (2000), "On the effect of treatment among treatment compliers: An analysis of the Multiple Risk Factor Intervention Trial," *Journal of the American Statistical Association*, 95, 1101–1109.

Frangakis, C. and Rubin, D. (2002), "Principal stratification in causal inference," *Biometrics*, 58, 21–29.

Joffe, M. and Brensinger, C. (2003), "Weighting in instrumental variables and G-estimation," *Statistics in Medicine*, 22, 1285–1303.

Joffe, M. and Small, D. (2005), "Defining and estimating intervention effects for groups who will develop an auxiliary outcome," , Manuscript.

Joffe, M., Ten Have, T., and Brensinger, C. (2003), "The compliance score as a regressor in randomized trials," *Biostatistics*, 4, 327–340.

Judd, C. M. and Kenny, D. A. (1981), "Process analysis: Estimating mediation in treatment evaluations," *Evaluation Review*, 5, 602–619.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., and Sheets, V. (2002), "A comparison of methods to test mediation and other intervening variable effects," *Psychological Methods*, 7, 83–104.

Mealli, F., Imbens, G., Ferro, S., and Biggeri, A. (2004), "Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes," *Biostatistics*, 5, 207–222.

Pearl, J. (2001), "Direct and Indirect Effects," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, eds. Besnard, P. and Hanks, S., San Francisco: Morgan Kaufmann, pp. 411–420

(http://bayes.cs.ucla.edu/jp_home.html)

.

Pearl, J. and Robins, J. (1995), "Probabilistic Evaluation of Sequential Plans from Causal Models with Hidden Variables," in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, eds. Besnard, P. and Hanks, S., San Francisco: Morgan Kaufmann, pp. 443–453

(http://www.biostat.harvard.edu/~robins/research.html)

.

Robins, J. (1994), "Correcting for non-compliance in randomized trials using structural nested mean models," *Communications in Statistics, Theory and Methods*, 23, 2379–2412.

— (1999), "Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models," , 349–405

(http://www.biostat.harvard.edu/~robins/research.html)

.

— (2003a), "Semantics of causal DAG models and the identification of direct and indirect effects," in *In Highly Structured Stochastic Systems*, eds. Green, P., Hjort, N., and Richardson, S., New York: Oxford University Press, pp. 70–81

(http://www.biostat.harvard.edu/~robins/research.html)

.

— (2003b), "Semantics of causal DAG models and the identification of direct and indirect effects," , 70–81

(http://www.biostat.harvard.edu/~robins/research.html)

.

Robins, J. and Greenland, S. (1992), "Identifiability and exchangeability for direct and indirect effects," *Epidemiology*, 3, 142–155.

— (1994), "Adjusting for differential rates of PCP prophylaxis in high- versus low-dose AZT treatment arms in an AIDS randomized trial." *Journal of the American Statistical Association*, 89, 737–749.

Robins, J. and Rotnitzky, A. (2004), "Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models," *Biometrika*, 91, 763–783.

Rubin, D. (1986), "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers," *Journal of the American Statistical Association*, 81, 961–962.

— (2004), "Direct and indirect causal effects via potential outcomes," *Scandinavian Journal of Statistics*, 31, 161–170.

Small, D., Ten Have, T., and Cheng, J. (2005), "Random effects logistic models for analyzing efficacy of a longitudinal randomized treatment with non-adherence," *Statistics in Medicine*, In press.

Ten Have, T., Elliott, M., M.M., J., Zanutto, E., and Datto, C. (2004), "Causal models for randomized physician encouragement trials in treatin g primary care depression," *Journal of the American Statistical Association*, 99, 8–16.

Vansteelandt, S. and Goetghebeur, E. (2003), "Causal inference with generalized structural mean models," *Journal of the Royal Statistical Society, Series B*, accepted for publication.