Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2012 *Paper* 96

Robust Estimation of Pure/Natural Direct Effects with Mediator Measurement Error

Eric J. Tchetgen Tchetgen* Sheng Hsuan Lin†

*Harvard University, etchetge@hsph.harvard.edu

†Harvard University, b92401002@gmail.com

Robust Estimation of Pure/Natural Direct Effects with Mediator Measurement Error

Eric J. Tchetgen Tchetgen and Sheng Hsuan Lin

Abstract

Recent developments in causal mediation analysis have offered new notions of direct and indirect effects, that formalize more traditional and informal notions of mediation analysis emanating primarily from the social sciences. The pure or natural direct effect of Robins-Greenland-Pearl quantifies the causal effect of an exposure that is not mediated by a variable on the causal pathway to the outcome, and combines with the natural indirect effect to produce the total causal effect of the exposure. Sufficient conditions for identification of natural direct effects were previously given, that assume certain independencies about potential outcomes, and a rich literature on estimation of natural direct effects has since developed. A common situation in epidemiology is that the mediator is subject to measurement error, in which case, existing techniques for estimating natural direct and indirect effects could be biased and the resulting inferences could be incorrect if measurement error were ignored. In this paper, the authors consider classical measurement error of a continuous mediator. The authors propose a three-stage least-squares regression technique for estimating natural direct effects on the additive scale, that is robust to classical measurement error of the mediator under certain assumptions about the structure of confounding. The robustness property implies that no additional data such as a validation sample, nor replicate measurements of the error prone mediator are needed to recover valid mediation inferences. An important appeal of the three-stage approach is that it is easy to implement using standard software. A simulation study is provided illustrating the finite sample performance of the proposed approach as compared to the prevailing mediation technique, and the new methodology is also shown to apply under a specific form of differential additive measurement error, and to extend to multiplicative effects under a log-linear regression framework.

Original Manuscript

Robust estimation of pure/natural direct effects with mediator measurement error

Eric J. Tchetgen Tchetgen[†,‡] and Sheng Hsuan Lin[†]

Departments of Epidemiology[†]and Biostatistics[‡], Harvard University

Corresponding author: Eric J. Tchetgen Tchetgen, Department of Epidemiology, Harvard School of Public Health 677 Huntington Avenue, Boston, MA 02115.

Running head: Natural direct and indirect effects, Classical measurement error, three-stage least-squares regression

Sources of financial support:

Acknowledgements:

Abstract

Recent developments in causal mediation analysis have offered new notions of direct and indirect effects, that formalize more traditional and informal notions of mediation analysis emanating primarily from the social sciences. The pure or natural direct effect of Robins-Greenland-Pearl quantifies the causal effect of an exposure that is not mediated by a variable on the causal pathway to the outcome, and combines with the natural indirect effect to produce the total causal effect of the exposure. Sufficient conditions for identification of natural direct effects were previously given, that assume certain independencies about potential outcomes, and a rich literature on estimation of natural direct effects has since developed. A common situation in epidemiology is that the mediator is subject to measurement error, in which case, existing techniques for estimating natural direct and indirect effects could be biased and the resulting inferences could be incorrect if measurement error were ignored. In this paper, the authors consider classical measurement error of a continuous mediator. The authors propose a three-stage least-squares regression technique for estimating natural direct effects on the additive scale, that is robust to classical measurement error of the mediator under certain assumptions about the structure of confounding. The robustness property implies that no additional data such as a validation sample, nor replicate measurements of the error prone mediator are needed to recover valid mediation inferences. An important appeal of the three-stage approach is that it is easy to implement using standard software. A simulation study is provided illustrating the finite sample performance of the proposed approach as compared to the prevailing mediation technique, and the new methodology is also shown to apply under a specific form of differential additive measurement error, and to extend to multiplicative effects under a log-linear regression framework.

Epidemiologic studies often aim to quantify the extent to which an exposure is mediated by an intermediate variable on the causal pathway to the outcome. Recent developments in causal mediation analysis have helped formalize notions of effect decomposition of a total effect of the exposure into its direct and indirect components. Specifically, the study of natural direct and indirect effects, also known as pure direct and indirect effects, has received considerable attention in recent causal inference literature, and formal conditions for identification of these effects are now well established[1–4]. A variety of statistical methods for estimating mediation causal effects have also developed in recent years[6–14]. However, existing mediation techniques typically rely on the key assumption that the mediator is measured without error. The assumption of an error-free mediator is sometimes inappropriate in practice, and inferences about natural direct and indirect effects may be incorrect if the mediator is subject to measurement error. This paper is primarily concerned with causal mediation analysis in the presence of classical measurement error of the mediator. In principle, if one either had access to replicate measurements of the mediator, or if one had internal or external validation data, one could possibly adapt one of several existing techniques for measurement error correction, to formally account for classical measurement error. For instance, within the context of linear models, one could use relatively straightforward regression calibration techniques[15,16] to recover unbiased inferences about natural direct and indirect effects. However, such auxiliary data are often not available in practice, and standard measurement error correction techniques may in reality only be useful to the extent that they provide a straightforward framework for obtaining simple sensitivity analyses.[16]

In this paper, the authors propose a novel three-stage least-squares (3SLS) regression approach for causal mediation analysis, that is completely robust to classical measurement error of the mediator, and that requires no additional data such as a validation sample or replicate measurements of the error prone mediator. The proposed 3SLS approach relies on an estimator of a regression

model for the mediator given the exposure and confounders of the exposure-mediator relation, and also requires an estimator of the outcome regression given exposure, mediator and confounders. The authors show that when additive effects are in view, the regression for the mediator need not necessarily be correct to obtain, using a stratified 3SLS strategy, a consistent estimator of the natural direct effect. Stratified 3SLS is also shown to confer some additional robustness against modeling error of the outcome regression. A key feature of 3SLS and stratified 3SLS, is that both methods are easy to implement using standard regression software. A simulation study is provided illustrating the finite sample performance of the proposed approach, and comparing it to the prevailing regression-based mediation technique; and the proposed methods are shown to continue to apply under a certain form of differential measurement error. Finally, in an Online Appendix, the authors also extend the approach to the multiplicative scale, and establish that, whereas one is not required to correctly specify the first stage regression in the context of additive effect decomposition, effect decomposition on the multiplicative scale requires that the first stage regression of the mediator is correctly specified.

Methodology

Notation and definitions

We introduce the notation and definitions we will be using throughout. Let E denote the exposure or treatment received by an individual, let Y denote a post-treatment outcome, and let M denote the true value of a post-treatment intermediate variable that may serve as a mediator for the treatment-outcome relationship. Let C denote the value of a set of pre-exposure confounding variables of the effects of E and M. Throughout, we assume that C can be partitioned into two sets of variables C_1 and C_2, where C_2 is known to only confound the effect of E on M, and therefore, is not a direct cause of Y, whereas C_1 may be directly related to all variables (C_2, E, M, Y). We assume independent and identically distributed sampling of C, E, M and Y. Then, the relation-

ships between these variable may be depicted as in the causal diagram in Figure 1.

Insert Figure 1

We now consider counterfactuals or potential outcomes, under possible interventions on the variables.[17,18] Let $Y(e)$ denote a subject's outcome if treatment E were set, possibly contrary to fact, to e. In the context of mediation there will also be potential outcomes for the intermediate variable. Let $M(e)$ denote a subject's counterfactual value of the intermediate M if treatment E were set to the value e. Finally, let $Y(e,m)$ denote a subject's counterfactual value for Y if E were set to e and M were set to m. Similar definitions hold for $Y(e,m,c) = Y(e,m,c_1)$ and $M(e,c)$.

Nonparametric structural equations models and natural direct effects

The exposition is framed around a nonparametric structural equation theory of causal inference, described by Judea Pearl.[19] Structural equations provide a nonparametric algebraic interpretation of the diagram of Figure 1 corresponding to four functions, one for each variable on the causal graph:

$$C = g_C(\varepsilon_C) \tag{1}$$

$$E = g_E(C, \varepsilon_E) \tag{2}$$

$$M = g_M(C, E, \varepsilon_M) \tag{3}$$

$$Y = g_Y(C_1, E, M, \varepsilon_Y) \tag{4}$$

Each of the nonparametric functions $\{g_C, g_E, g_M, g_Y\}$ represents a causal mechanism that deter-

mines the value of the left-hand-side variable, known as the output, from variables on the right, known as the inputs. The errors $(\varepsilon_C, \varepsilon_E, \varepsilon_M, \varepsilon_Y)$ stand for all factors not included on the graph that could possibly affect their corresponding outputs when all other inputs are held constant. To be consistent with the causal graph presented in Figure 1, we require that these errors be mutually independent, but we allow their distribution to remain arbitrary. If they were not independent we would include an additional unmeasured variable U on the diagram with arrows into the relevant variables to induce independence. Lack of a causal effect of a given variable on an output is encoded by an absence of the variable from the right-hand side. For instance, the absence of a direct effect of C_2 on Y is encoded by the absence of C_2 in g_Y of equation (4), which encodes the assumption that variations in C_2 will leave Y unchanged, as long as variables E, C_1, M and ε_Y remain constant, which is also consistent with the assumption that there are no unmeasured common causes of C_2 and Y.

As stated by Pearl[19], the invariance of structural equations permits their use as a basis for modeling causal effects and potential outcomes. In fact, to emulate the intervention in which one sets $\{E = e\}$ for all individuals simply amounts to replacing the equation for E with $E = e$, producing the following set of modified equations:

$$
\begin{aligned}
C &= g_C(\varepsilon_C) \\
E &= e \\
M(e) &= g_M(C, e, \varepsilon_M) \\
Y(e) &= g_Y(C_1, e, M(e), \varepsilon_Y)
\end{aligned}
$$

with $\{M(e), Y(e) = Y(e, M(e))\}$ denoting the potential outcomes had the exposure been set to e.

Under the above NPSEM, the independence of errors $\varepsilon_M \perp\!\!\!\perp \varepsilon_Y$ implies independence of potential outcomes for different exposure values:

$$Y(e, m, c_1) \perp\!\!\!\perp M(e^*, c) \tag{5}$$

where $M(e^*, c) = g_M(c, e^*, \varepsilon_M)$ and $Y(e, m, c_1) = g_Y(c_1, e, m, \varepsilon_Y)$ are obtained upon intervening on (E, C) and (E, M, C_1) respectively, and e, e^* take values in $\{0, 1\}$.

Robins and Greenland[1] and Pearl[2] considered the following decomposition of individual total effect of exposure:

$$\begin{aligned} Y(e) - Y(e^*) &= Y(e, M(e)) - Y(e^*, M(e^*)) \\ &= \underbrace{Y(e, M(e^*)) - Y(e^*, M(e^*))}_{\text{Natural direct effect}} + \underbrace{Y(e, M(e)) - Y(e, M(e^*))}_{\text{Natural indirect effect}} \end{aligned}$$

where e^* indicates a reference or baseline value of E; for instance it is common to chose $e^* = 0$ for binary E, and e represents an active value of treatment. The first contrast on the right hand side of the second line displayed above defines individual natural direct effect of treatment E on outcome Y. The potential outcome $Y(e^*, M(e^*))$ captures the behavior of Y under the baseline treatment value, while $Y(e, M(e^*))$ describes the behavior of Y under the active treatment value, in a hypothetical situation where the mediator behaves as if treatment were set to baseline. The second contrast on the right-hand side of the expression in the display above corresponds to the natural indirect effect of treatment E on outcome Y. The potential outcome $Y(e, M(e))$ describes the behavior of Y under the active treatment value, while the second "subtracts off" the behavior of Y under the active treatment value, in a hypothetical situation where the mediator behaves as if treatment were set to baseline. In graphical terms, the individual natural indirect effect quantifies

for the individual, the effect of E on Y along the indirect causal pathway $E \rightarrow M \rightarrow Y$, but not along the direct arrow from E to Y. Because potential outcomes under conflicting exposure status are never jointly observed, individual causal effects are generally not identified. However, one can hope that under certain assumptions, population average causal effects would become identified. It is well known that the average total effect of E on Y is identified given data on (C, E, Y) in the causal diagram of Figure 1, and is given by the g-formula of Robins[20] :

$$TE\left(e, e^*, c\right) \equiv \mathbb{E}\left\{Y\left(e\right) - Y\left(e^*\right)|C\right\} = \sum_c \left[\mathbb{E}\left\{Y|e, c\right\} - \mathbb{E}\left\{Y|e^*, c\right\}\right] \Pr(C = c) \tag{6}$$

where E stands for expectation. Pearl[2] showed that under the NPSEM for the causal graph of Figure 1, the average natural direct effect conditional on C, is nonparametrically identified by:

$$\begin{aligned} NDE\left(e, e^*, c\right) &\equiv \mathbb{E}\left\{Y\left(e, M\left(e^*\right)|c\right) - Y\left(e^*, M\left(e^*\right)|c\right)\right\} \\ &= \sum_m \left[\mathbb{E}\left\{Y|e, m, c\right\} - \mathbb{E}\left\{Y|e^*, m, c\right\}\right] \Pr(M = m|E = e^*, C = c) \\ &= \sum_m \left[\mathbb{E}\left\{Y|e, m, c_1\right\} - \mathbb{E}\left\{Y|e^*, m, c_1\right\}\right] \Pr(M = m|E = e^*, C = c) \end{aligned} \tag{7}$$

Under the NPSEM defined above, we then have that:

$$NDE\left(e, e^*, c\right) = \sum_m \left[\mathbb{E}\left\{Y|e, m, c_1\right\} - \mathbb{E}\left\{Y|e^*, m, c_1\right\}\right] \Pr(M = m|E = e^*, C = c)$$

Therefore the average natural indirect effect is obtained under the NPSEM by $NIE\left(e, e^*, c\right) = TE\left(e, e^*, c\right) - NDE\left(e, e^*, c\right)$. A variety of statistical methods for estimating $NDE\left(e, e^*, c\right)$ and $NIE\left(e, e^*, c\right)$ have been proposed in recent literature.[5,6,9–14] Tchetgen Tchetgen and Shpitser compare several of these methods and develop a semiparametric approach with attractive robustness

and efficiency properties.[14] When, as we assume next, Y and M are both continuous, a prominent regression strategy for estimating $NDE\left(e,0,c\right)$ when one takes $e^{*}=0$, entails fitting linear models for Y and M, respectively, say:

$$Y=\mathbb{E}\left(Y|e,m,c_{1}\right)+\varepsilon_{Y}=\alpha_{0}+\alpha_{1}e+\alpha_{2}m+\alpha_{3}me+\alpha_{4}^{T}c_{1}+\varepsilon_{Y} \tag{8}$$

$$M=\mathbb{E}\left(M|e,c\right)+\varepsilon_{M}=\eta_{0}+\eta_{1}e+\eta_{2}^{T}c+\varepsilon_{M} \tag{9}$$

where we have allowed for an interaction between e and m in the outcome regression. In the absence of measurement error, ordinary least squares is then typically used to obtain unbiased estimates of the regression coefficients in the above display, and the natural direct effect is obtained by evaluating equation (7) using the above linear models, which yields the simple expression[10]

$$NDE\left(e,e^{*},c\right)=\left(\beta_{0}+\beta_{1}^{T}c\right)e, \tag{10}$$

where:

$$\beta_{0}=\alpha_{1}+\alpha_{3}\times\eta_{0} \tag{11}$$

$$\beta_{1}=\alpha_{3}\times\eta_{2} \tag{12}$$

Note that if either of models (8) or (9) is incorrect, equations (11) and (12) may not hold exactly, even though the linear form of equation (10) may actually be correctly specified. For instance, this would be the case if the outcome regression were missing an existing interaction between e and a component of c_1. Assuming no modeling error, suppose that M is subject to classical measurement error, and thus M_{ϵ} is observed instead of M, so that one observes data on (C,E,M_{ϵ},Y) as depicted

in Figure 2, where M is not directly observed:

$$M_\epsilon = M + \epsilon,$$

$$\epsilon \text{ independent mean zero error}$$

Then, it is well known that the standard OLS estimator of the coefficients of the mediator regression (9) remains unbiased, but some efficiency loss might be incurred due to additional variation in the regression outcome. In contrast, the OLS estimator of the coefficients of the regression model (8) will generally be biased, and inference about $NDE\,(e,0,c)$ may be incorrect. In the following section, a 3SLS approach is given, that unlike standard OLS, is robust to classical measurement error of the mediator, and that is guaranteed, under certain conditions, to give valid inferences about $NDE\,(e,0,c)$ regardless of such measurement error.

3SLS estimation of $NDE(1,0,c)$

To introduce the proposed strategy, suppose for a moment that both regression models (8) and (9) are correct, and assume for simplicity that the interaction parameter $\alpha_3 = 0$. Then, one can verify that models (8) and (9) induce the following linear model for the conditional mean of Y given (C,E), upon averaging over M :

$$\begin{aligned}
&\mathbb{E}\,(Y|e,c) \\
&= \mathbb{E}\left\{\mathbb{E}\,(Y|e,M,c)\,|e,c\right\} \\
&= \alpha_0 + \alpha_1 e + \alpha_2 \mathbb{E}\,(M|e,c) + \alpha_4^T c_1 \\
&= \alpha_0 + \alpha_1 e + \alpha_2 \mathbb{E}\,(M_\epsilon|e,c) + \alpha_4^T c_1
\end{aligned}$$

where the first equality is by the law of iterated expectations, and the last equality is by the

assumption of classical measurement error. Therefore, one can conclude that the outcome regression parameter α remains identified in the presence of measurement error, by regressing Y on $\left(1, E, \mathbb{E}\left(M_\epsilon|E, C\right), C_1^T\right)^T$ provided that $\mathbb{E}\left(M|e, c\right)$=$\mathbb{E}\left(M_\epsilon|e, c\right)$ can be consistently estimated, and as long as the predicted mean $\mathbb{E}\left(M|e, c\right)$ is not perfectly collinear with the vector $(1, e, c_1^T)^T$. The assumption that model (9) is correctly specified ensures the first condition, while the additional variation due to C_2 ensures that $\mathbb{E}\left(M|e, c\right)$ cannot be expressed exactly as a linear combination of the components of the vector $(1, e, c_1^T)^T$. Further insight about the second condition is gained upon noting that the condition essentially states the C_2 must be a valid instrumental variable for the $M-Y$ conditional relation give (E, C_1). In fact, recall that a key assumption was made at the outset, that the vector of covariates C partitions into a subset of variables C_1 that are known common causes of E, M and Y, and a subset of confounder C_2 of the E-M relation; and is otherwise unrelated to Y given (C_1, E, M). The assumption that C_2 is directly related to M, together with the assumption that upon conditioning on $(E, C_1), C_2$ is only related to Y through M, formally makes C_2 an instrumental variable for the effects of M on Y. Note however, that C_2 is somewhat more general than the typical instrumental variable, since it may also confound the $E-M$ relation, and may also be related to variables in C_1. Further note that since M is not directly observed, the second assumption needed to make C_2 a valid instrumental variable, mainly that it is independent of Y given (C_1, E, M), is not empirically verifiable using the observed data (C, M_ϵ, E, Y). This is because, although C_2 may be independent of Y given (C_1, E, M), it is not necessarily independent of Y given (C_1, M_ϵ, E). Thus, some care is required in forming the partition between C_1 and C_2 that must necessarily be made on the basis of expert knowledge about the nature and structure of the confounding operating. If the assumption is correct that C_2 is a valid instrumental variable for the $M-Y$ relation, then upon obtaining consistent estimates of (α, η), in principle, one could subsequently use equations (11) and (12) to obtain estimates of $NDE\left(1, 0, c\right)$ and $NIE\left(1, 0, c\right)$.

The above strategy suggests the following 3SLS approach:

Stage 1:Using data $(E_i, C_i, M_{\epsilon,i})$, $i = 1, ..., n$, compute the OLS estimate of model (9) and compute the corresponding predicted values

$$\widehat{M_i} = \widehat{\eta}_0 + \widehat{\eta}_1 E_i + \widehat{\eta}_2^T C_i$$

$i = 1, ..., n$.

Stage 2: Compute the OLS $\widehat{\alpha} = (\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3, \widehat{\alpha}_4)$ of Y_i regressed on $\left(E_i, \widehat{M_i}, C_{1,i}\right)$, under the working model

$$\alpha_0 + \alpha_1 E_i + \alpha_2 \widehat{M_i} + \alpha_3 \widehat{M_i} E_i + \alpha_4^T C_{1,i}$$

and compute the following predicted value

$$\widehat{Y_i} = (\widehat{\alpha}_0 + \widehat{\alpha}_1) + (\widehat{\alpha}_2 + \widehat{\alpha}_3) M_{\epsilon,i} + \widehat{\alpha}_4^T C_{1,i}$$

For each unexposed person with $E_i = 0$, define the residual $\Delta Y_i = \widehat{Y}_i - Y_i$.

Stage 3: Compute the OLS $\widehat{\beta} = \left(\widehat{\beta}_0, \widehat{\beta}_1^T\right)$of ΔY_i regressed on $(1, C_i)$ using unexposed individuals only, under the working model

$$\mathbb{E}\left(\Delta Y_i | C_i, E_i = 0\right) = \beta_0 + \beta_1^T C_i,$$

then,

$$\widehat{NDE}\left(1, 0, c\right) = \widehat{\beta}_0 + \widehat{\beta}_1^T c$$

is the 3SLS estimator of $NDE\left(1, 0, c\right)$.

The following result states that, despite observing an error prone mediator, the 3SLS estimator

$\widehat{NDE}(1,0,c)$ is nonetheless a consistent estimator of $NDE(1,0,c)$.

Result 1: Suppose that the NPSEM $(1)-(4)$ *holds, and suppose that the outcome regression model* (8), *the mediator regression* (9) *and the natural direct effect model* (10) *all hold, then we have that the 3SLS estimator* $\widehat{NDE}(1,0,c)$ *is consistent for* $NDE(1,0,c)$.

According to Result 1, if C_2 is a valid instrumental variable of the $E-M$ relation, then 3SLS can be used to recover a consistent estimate of $NDE(1,0,c)$. However, the approach relies on correct specification of both models (8) and (9). In the following section, we describe a stratified 3SLS approach that relaxes modeling requirements, and only requires a correct regression for the outcome for exposed individuals.

Stratified 3SLS estimation of $NDE(1,0,c)$

Consider the following stratified 3SLS approach:

Stratified Stage 1:Using data $(C_i, M_{\epsilon,i})$, for exposed individuals only, i.e. with $E_i = 1$, $i = 1, ..., n$ compute the OLS estimate of the working model stratified on $E = 1$

$$\mathbb{E}(M_\epsilon | E = 1, c) = \eta_0^* + \eta_2^{*T} c,$$

and compute their corresponding predicted value

$$\widehat{M_i^*} = \widehat{\eta}_0^* + \widehat{\eta}_2^{*T} C_i.$$

Stratified Stage 2:Using data for exposed individuals only, compute the OLS estimate $\widehat{\alpha}^* = \left(\widehat{\alpha}_0^*, \widehat{\alpha}_2^*, \widehat{\alpha}_3^*, \widehat{\alpha}_4^{*T}\right)^T$ of Y_i regressed on $\left(\widehat{M_i^*}, C_{1,i}\right)$, under the working model

$$\alpha_0^* + \alpha_2^* \widehat{M_i^*} + \alpha_4^{*T} C_{1,i}$$

and compute the corresponding predicted values for unexposed individuals

$$\widehat{Y}_i^* = \widehat{\alpha}_0^* + \widehat{\alpha}_2^* M_{\epsilon,i} + \widehat{\alpha}_4^{*T} C_{1,i}.$$

For each unexposed person define the residual $\Delta Y_i^* = \widehat{Y}_i^* - Y_i$.

Stratified Stage 3: Compute the OLS $\widehat{\beta}^* = \left(\widehat{\beta}_0^*, \widehat{\beta}_1^{*T}\right)$ of ΔY_i^* regressed on $(1, C_i)$ using data on unexposed individuals only, under the model

$$\mathbb{E}\left(\Delta Y_i^* | C_i, E_i = 0\right) = \beta_0^* + \beta_1^{*T} C_i$$

$$\widehat{NDE}^*\left(1, 0, c\right) = \widehat{\beta}_0^* + \widehat{\beta}_1^{*T} c$$

is the stratified 3SLS estimator of $NDE\left(1, 0, c\right)$.

The appendix gives a proof of the following result:

Result 2: Suppose the NPSEM $(1) - (4)$ *holds, and assume the following outcome regression model restricted to exposed individuals:*

$$\mathbb{E}\left(Y_i | M, E_i = 1, C_{1,i}\right) = \alpha_0^* + \alpha_2^* M_i + \alpha_4^{*T} C_{1,i} \tag{13}$$

and the natural direct effect model (10) *both hold; then, we have that the stratified 3SLS estimator* $\widehat{NDE}^*\left(1, 0, c\right)$ *is consistent for* $NDE\left(1, 0, c\right)$.

Result 2 improves on Result 1 and, in addition to a correct natural direct effect model (10) only requires a correct working model for the outcome of exposed individuals. According to the result, stratified 3SLS is guaranteed to be asymptotically unbiased under many more data generating mechanisms than 3SLS. To see why, note that if model (8) were incorrect because it failed to

incorporate an interaction between exposure and at least one component of C_1; then, Model (13) would still be correctly specified and therefore stratified 3SLS would produce correct inferences about the natural direct effect, while both the standard estimator of direct effect, and 3SLS would not.

Statistical inference about $NDE\,(1,0,c)$ requires a consistent estimator of the variance-covariance matrix of $\widehat{\beta}$ or of $\widehat{\beta}^*$, which we give in the Appendix, denoted $\widehat{\Sigma}$ and $\widehat{\Sigma}^*$, respectively. Then, a Wald-type 95% confidence intervals, say for instance $\widehat{\beta}_0^*$ is given by $\widehat{\beta}_0^* \pm 1.96\widehat{\sigma}_{11}^*$, where $\widehat{\sigma}_{jk}^{*2}$ is the element of $\widehat{\Sigma}^*$ in row j and column k. Alternatively, one could use the nonparametric bootstrap to estimate the variance-covariance matrices of $\widehat{\beta}$ and $\widehat{\beta}^*$, respectively.

Estimation of $NIE(1,0,c)$

Because the potential mediator M is not directly needed to estimate total effects, inference about the latter quantity is completely unaffected by measurement error of the mediator and can be obtained by using standard regression techniques using only data on (Y,E,C). For example, consider the linear model for the mean of Y given (E,C) :

$$\mathbb{E}\,(Y|e,c) = \gamma_0 + \gamma_1 e + \gamma_2^T ec + \gamma_3^T c$$

then, under the previous assumption that C includes all confounders of the effects of E, it is straightforward to verify that this corresponds to the following linear model for the conditional total effect of E on Y given C :

$$TE\,(1,0,c) = \gamma_1 + \gamma_2^T c$$

and therefore, standard OLS can be used to obtain an unbiased estimate of $\gamma = \left(\gamma_0, \gamma_1, \gamma_2^T, \gamma_3^T\right)^T$, which in turn gives an unbiased estimate of the total effect $TE\,(1,0,c)$. Finally, the natural indirect

effect $NIE\,(1,0,c)$ is obtained by computing $NIE\,(1,0,c) = TE\,(1,0,c) - NDE\,(1,0,c)$ using estimates of (γ, β).

A simulation study

In this section, we report a simulation study that illustrates the finite sample performance of estimators introduced in previous sections. We generated 1000 samples of size $n = 500, 1000, 2000$ from the following model:

$$\text{(Model.C)}$$

$$C_1 \sim Normal(3, 2);$$

$$[C_2|C_1] \sim \ Normal(0.5 + 1.3 \times C_1, 3);$$

$$\text{(Model.E)}$$

$$[E|C_1, C_2] \sim Bernoulli\left([1 + \exp\{-\left(0.01 \times C_1 - 0.03 \times C_2\right)\}]^{-1}\right);$$

$$\text{(Model.M)}$$

$$[M|E, C_1, C_2] \sim Normal(1 + 10 \times E + C_1 + C_2 + \eta_4 \times C_1 \times E, 4);$$

$$\eta_4 = 0, 5;$$

$$\text{(Model.Y)}$$

$$[Y|M, A, X_1, X_2, X_3] \sim Normal(0.2 + 3 \times E + 0.4 \times M + \alpha_3 \times E \times M + C_1 + \alpha_5 \times E \times C_1, 3);$$

$$(\alpha_3, \alpha_5) = (0, 0)\,; (3, 0)\,; (3, -4)\,;$$

$$\text{(Measurement Error Model)}$$

$$M_\epsilon = M + Normal(0, 4 \times k/(1-k));$$

$$k = 0, 0.05\,, 0.15, 0.3, 0.5.$$

By evaluating equation (7) under the model in the above display, we obtain the following expression for the conditional natural direct effect:

$$NDE\,(1,0,c) = \beta_0 + \beta_1 \times C_1 + \beta_2 \times C_2$$

where $\beta_0 = 6$; $\beta_1 = 3 + \alpha_5$; $\beta_2 = 3$

The simulation study evaluated and compared the performance of the prominent estimator of direct effect, which uses equations (11) and (12) , with that of 3SLS and Stratified 3SLS. To assess the impact of measurement error, the three estimators were evaluated with no measurement error, i.e. $k = 0$, and were compared as the degree of measurement error was increased, i.e. $k = 0, 0.05\,, 0.15, 0.3, 0.5$, and $\alpha_5 = \eta_4 = 0$ such that all models were correctly specified. To compare the various methods in terms of robustness to partial modeling error, data was generated with $\eta_4 = 5$ and $\alpha_5 = -4$, such that the working models used by the standard approach as well as 3SLS were mis-specified by virtue of omitting certain nonzero interactions, and we assessed whether the proposed stratified 3SLS approach was likewise affected by such modeling error.

Insert Tables 1-5.

Tables 1-5 summarize the simulation results regarding inferences about $\beta = (\beta_0, \beta_1, \beta_2)$. For the most part, the results largely agree with the theory developed in the previous sections. Mainly, all three estimators performed well at both moderate and large sample size in the absence of both measurement and modeling error, see rows for $k = 0$ in Tables 1-4. Furthermore, when all models were correct and measurement error was absent, the proposed estimators were somewhat less efficient than the standard estimator. This is not surprising since the latter essentially amounts in this particular setting to the maximum likelihood estimator. However, introducing a moderate amount of measurement error can have severe implications for the standard approach, compromising both bias and coverage of the estimator and corresponding confidence intervals. Such effects are particularly notable when there is an $E - M$ interaction in the outcome regression, as illustrated in

Tables 3 and 4. In contrast, the proposed 3SLS estimators seemed quite robust to measurement error, with bias considerably smaller than the standard approach and coverage preserved at the nominal level of 95%. The disastrous performance of the standard approach became worse with increasing measurement error, irrespective of sample size.

Modeling error due to omission of an $E - C_1$ interaction in the outcome regression was likewise observed to be detrimental to performance of the standard approach, even when the mediator was observed without error, i.e. $k = 0$, see rows of Table 5 with $\alpha_5 = 5$. As dictated by theory, the proposed 3SLS approach was likewise affected by such modeling error as illustrated in Table 5, however, stratified 3SLS was found not to be affected by this form of modeling error. Modeling error due to omission of an $E - C_1$ interaction in the mediator regression likewise gave biased results using the standard approach and 3SLS as can be seen in the rows of Table 5 where $\eta_4 \neq 0$, however, as predicted by our theoretical results, stratified 3SLS was not affected by this form of mis-specification of the mediator regression. More extensive simulation results are provided in the Online Appendix under a variety of additional settings, further confirming that both at moderate and large sample sizes, the estimators essentially behaved as theorized by Results 1 and 2.

Differential additive measurement error

Previous sections have assumed that measurement error of the mediator is nondifferential and thus, independent of the outcome measure. In this section, it is shown that this assumption can be relaxed, and that progress can still be made if we allow for a certain form of differential additive measurement error. Specifically, suppose that the correlation, denoted ρ, between the continuous outcome residual error ε_Y and the measurement error ϵ of M is not zero, such that the measurement error of the mediator is correlated with the outcome Y, but the measurement error ϵ is otherwise independent of (E, C, M). This situation is depicted in the causal diagram of Figure 2, where the

double-headed edge represents the correlation between the measurement error and the continuous outcome, and ensures that the measurement error is d-separated from (E, C, M), thus graphically encoding their independence. In the appendix, we show that Results 1 and 2 continue to hold even if $\rho \neq 0$, and therefore allowing for the magnitude of measurement error operating to depend on the outcome. Intuitively, the results continue to hold principally because, as confirmed by inspecting the causal diagram of Figure 2, the association between ε_Y and the measurement error ϵ does not invalidate C_2 as an instrumental variable for the effects of M on Y. This is in fact the case since C_2 continues to only affect Y through its effects on M, even after the double-headed edge has been added to the diagram of Figure 2 to allow for differential measurement error.

Interestingly, the measurement error depicted in Figure 2 could arise if the intermediate were not measured exactly at the moment that is relevant for occurrence of the event, but were measured at a later stage. For example, it is sometimes only possible in epidemiologic studies, to measure the mediator concurrently with the outcome from cross-sectional data. If such cross-sectional data on the outcome and intermediate variables were available, but the exposure data were collected at an earlier time such that the temporal ordering between E and (M, M_ϵ, Y) were ensured, then one would expect the mismeasured mediator to generally be correlated with the outcome for reasons not directly related to the latent intermediate M, exposure, and pre-exposure confounders C_1. In principle, the differential additive measurement error model depicted in Figure 2 could be used to model such measurement error due to cross-sectional measurements of the mediating and outcome variables, in which case, the methodology developed in this paper could be used to recover valid mediation inferences that account for the mediator measured concurrently with the outcome. The diagram of Figure 2 is closely related to the "intraindividual variation over time models" of le Cessie et al[16], however, whereas le Cessie et al.[16] assumed that measurement error of the intermediate is independent of the outcome, such an assumption is not made here.

Discussion

In this paper, the authors have proposed two-stage and three-stage regression techniques for estimating natural direct and indirect effects in the presence of measurement error of the mediator. The proposed methods which are shown to apply quite generally for a continuous outcome, can also be used for a binary outcome under a log-linear model for the risk of the outcome; see the Online Appendix. The basic assumption made by the methods, is that the structure of confounding is such that a subset of the confounders of the exposure-mediator relation, is known to only affect the outcome through its effects on the mediator. However, one should also note that when available, a more conventional instrumental variable for the effects of the mediator on the outcome can also be used in this capacity. For instance, if available, replicate measurements of the mediator can be used in 2SLS and 3SLS, even though, such replicates do not usually confound the exposure-mediator relation.

The paper has mainly considered the context of a prospective study, but the approach can be adapted to accommodate other study designs often encountered in epidemiologic practice. For instance, the methods described above could be used in a case-control study by taking one of two strategies. If the outcome were rare, one could simply restrict 2SLS or 3SLS to controls and discard data on cases; but more generally, if sampling fractions were known, one could simply re-weight cases and controls by the inverse of their associated probability of selection into the sample in conjunction with 2SLS or 3SLS estimation.

Finally, an important direction future work could consider, is whether the methods can be developed and extended to the context of a survival outcome.[13]

Appendix

Proof of Result 1: Assuming that the first stage regression of M is correctly specified, we have that $\widehat{M_i}$ is consistent for $\mathbb{E}\left(M|E_i, C_i\right)$, and therefore, as long as $\mathbb{E}\left(WW^T\right)$ is not singular, where $W^T = (1, E, \mathbb{E}\left(M|E, C\right), \mathbb{E}\left(M|E, C\right)E, C_1^T)$, the second stage regression estimate $\widehat{\alpha}$ converges in probability to the solution of the population normal equation: $\mathbb{E}\left\{W\left(Y - \alpha^{\dagger T}W\right)\right\} = 0$ which $\alpha^\dagger = \alpha$ solves, since $\alpha^T W = \mathbb{E}\left\{\mathbb{E}\left(Y|E, M, C\right)|C, E\right\}$. The assumption that C_2 is a valid instrumental variable guarantees that $\mathbb{E}\left(WW^T\right)$ is not singular. As a result, $\widehat{Y_i}$ is consistent for $\alpha_0 + \alpha_1 E_i + \alpha_2 M_{\epsilon,i} + \alpha_3 M_{\epsilon,i} E_i + \alpha_4^T C_{1,i}$.

The Stage 3 regression estimate $\widehat{\beta}$ converges in probability to the solution of the population normal equation

$$
\begin{aligned}
0 &= \mathbb{E}\left\{\left(1, C^T\right)^T (1-E)\left(\left(\alpha_0 + \alpha_1 \ + \alpha_2 M_\epsilon + \alpha_3 M_\epsilon + \alpha_4^T C_1\right) - Y - \beta_0^\dagger - \beta_1^{\dagger T} C\right)\right\} \\
&= \mathbb{E}\left\{\left(1, C^T\right)^T (1-E)\right. \\
&\left.\times\left(\left(\alpha_0 + \alpha_1 + (\alpha_2 + \alpha_3)\,\mathbb{E}\left(M_\epsilon|M, C, E=0\right) + \alpha_4^T C_1\right) - \mathbb{E}\left(Y|E=0, M, C\right) - \beta_0^\dagger - \beta_1^{\dagger T} C\right)\right\} \\
&= \mathbb{E}\left\{\left(1, C^T\right)^T (1-E)\left(\left(\alpha_0 + \alpha_1 + (\alpha_2 + \alpha_3) M + \alpha_4^T C_1\right) - \mathbb{E}\left(Y|E=0, M, C\right) - \beta_0^\dagger - \beta_1^{\dagger T} C\right)\right\} \\
&= \mathbb{E}\left\{\left(1, C^T\right)^T (1-E)\left(\mathbb{E}\left(Y|E=1, M, C\right) - \mathbb{E}\left(Y|E=0, M, C\right) - \beta_0^\dagger - \beta_1^{\dagger T} C\right)\right\} \\
&= \mathbb{E}\left\{\left(1, C^T\right)^T (1-E)\left(\mathbb{E}\left\{\mathbb{E}\left(Y|E=1, M, C\right) - \mathbb{E}\left(Y|E=0, M, C\right)|E=0, C\right\} - \beta_0^\dagger - \beta_1^{\dagger T} C\right)\right\}
\end{aligned}
$$

which $\beta_0^\dagger - \beta_1^{\dagger T} C = \beta_0 - \beta_1^T C$ solves since, under the assumptions stated in the result,

$$
\mathbb{E}\left\{\mathbb{E}\left(Y|E=1, M, C\right) - \mathbb{E}\left(Y|E=0, M, C\right)|E=0, C\right\} = NDE(1, 0, C).
$$

Proof of Result 2: The second Stage regression solves the normal equation:

$$\sum_i E_i \left(Y_i - \left(\widehat{\alpha}_0^* + \widehat{\alpha}_2^* \widehat{M}_i^* + \widehat{\alpha}_4^{*T} C_{1,i} \right) \right) \left(1, \widehat{M}_i^*, C_{1,i}^T \right)^T = 0 \tag{14}$$

where $\widehat{M}_i^*$ satisfies the first Stage normal equation:

$$\sum_i E_i \left(M_{\epsilon,i} - \widehat{M}_i^* \right) \left(1, C_i^T \right)^T = 0 \tag{15}$$

Therefore, equation (14) can be written

$$\begin{aligned} 0 &= \sum_i E_i \left(Y_i - \left(\widehat{\alpha}_0^* + \widehat{\alpha}_2^* \widehat{M}_i^* + \widehat{\alpha}_4^{*T} C_{1,i} \right) \right) \left(1, \widehat{M}_i^*, C_{1,i}^T \right)^T \\ &= \sum_i E_i \left(Y_i - \left(\widehat{\alpha}_0^* + \widehat{\alpha}_2^* M_{\epsilon,i} + \widehat{\alpha}_4^{*T} C_{1,i} \right) \right) \left(1, \widehat{M}_i^*, C_{1,i}^T \right)^T \\ &+ \underbrace{\widehat{\alpha}_2^* \sum_i E_i \left(M_{\epsilon,i} - \widehat{M}_i^* \right) \left(1, \widehat{M}_i^*, C_{1,i}^T \right)^T}_{=0} \end{aligned}$$

by equation (15) and the fact that $\left(1, \widehat{M}_i^*, C_{1,i}^T \right)$ can be expressed as a linear transformation of $\left(1, C_i^T \right)$. Let $\eta_0^* + \eta_2^{*T} c$ denote the limiting value of $\widehat{\eta}_0^* + \widehat{\eta}_2^{*T} c$ which is not necessarily equal to $\mathbb{E}\left(M | E = 1, c \right)$, and let $W_1 = \left(1, \eta_0^* + \eta_2^{*T} C, C_{1,i}^T \right)^T$ and $W_2 = \left(1, M_{\epsilon,i}, C_{1,i}^T \right)^T$, then, as long as $\mathbb{E}\left(E W_1 W_2^T \right)$ is not singular, $\widehat{\alpha}^*$ converges in probability to the solution α^* of the population equation:

$$\mathbb{E}\left(E_i \left(Y_i - \left(\alpha_0^* + \alpha_2^* M_{\epsilon,i} + \alpha_4^{*T} C_{1,i} \right) \right) \left(1, \eta_0^* + \eta_2^{*T} C_i, C_{1,i}^T \right)^T \right) = 0$$

which is solved by $\alpha^* = \alpha$. The assumption that $\mathbb{E}\left(E W_1 W_2^T \right)$ is not singular is ensured by the assumption that C_2 is a valid instrumental variable. Finally, $\widehat{\beta}^* = \left(\widehat{\beta}_0^*, \widehat{\beta}_1^{*T} \right)$ is consistent for β by the same argument used in the last part of the proof of Result 1.

Variance Estimators: We give analytic expressions of consistent variance-covariance estimators:

$\underline{\text{Var}(\widehat{\beta})}$: Consider the first stage OLS $\widehat{\eta}$, its large sample expansion gives

$$\widehat{\eta} - \eta \approx n^{-1} \sum_i U_{1,i}$$

$$\text{where } U_{1,i} = \mathbb{E}\left(X_{1,i} X_{1,i}^T\right)^{-1} X_{1,i} \varepsilon_{M,i}^*$$

$$X_{1,i} = (1, E_i, C_i)^T$$

$$\varepsilon_{M,i}^* = M_{\epsilon,i} - \left(\eta_0 + \eta_1 E_i + \eta_2^T C_i\right)$$

A standard Taylor approximation gives the large sample expansion of the second stage OLS $\widehat{\alpha}$

$$\widehat{\alpha} - \alpha \approx n^{-1} \sum_i U_{2,i}$$

$$\text{where } U_{2,i} = \mathbb{E}\left(X_{2,i} X_{2,i}^T\right)^{-1} \left\{X_{2,i} \varepsilon_{Y,i}^* - \mathbb{E}\left(\left(\alpha_2 + \alpha_3 E_i\right) X_{2,i} X_{1,i}^T\right) U_{1,i}\right\}$$

$$X_{2,i} = \left(1, E_i, \mathbb{E}\left(M_i | E_i, C_i\right), \mathbb{E}\left(M_i | E_i, C_i\right)_i E_i, C_{1,i}^T\right)^T$$

$$\varepsilon_{Y,i}^* = Y_i - X_{2,i}^T \alpha$$

Finally, the Stage 3 regression gives $\widehat{\beta}$ with large sample expansion:

$$\widehat{\beta} - \beta \approx n^{-1} \sum_i U_{3,i}$$

$$\begin{aligned}
\text{where } U_{3,i} &= \mathbb{E}\left((1-E_i)X_{3,i}X_{3,i}^T\right)^{-1} \\
&\quad \times \left\{(1-E_i)X_{3,i}\varepsilon_{NDE,i} + \mathbb{E}\left[(1-E_i)\,X_{3,i}X_{4,i}^T\right] \times U_{2,i}\right\} \\
X_{3,i} &= \left(1, C_i^T\right)^T \\
X_{4,i} &= \left(1, 1, M_{\epsilon,i}, M_{\epsilon,i}, C_{1,i}^T\right)^T \\
\varepsilon_{NDE,i} &= X_{4,i}^T\alpha - Y_i - \beta_0 - \beta_1^T C_i
\end{aligned}$$

The variance-covariance matrix of $\widehat{\beta}$ is therefore approximately given by

$$\begin{aligned}
& n^{-1}\mathbb{E}\left(U_{3,i}U_{3,i}^T\right) \\
& = n^{-1}\mathbb{E}\left((1-E_i)X_{3,i}X_{3,i}^T\right)^{-1} \\
& \times \mathbb{E}\left[\left\{(1-E_i)X_{3,i}\varepsilon_{NDE,i} + \mathbb{E}\left[(1-E_i)\,X_{3,i}X_{4,i}^T\right] \times U_{2,i}\right\}\right. \\
& \left. \times \left\{(1-E_i)X_{3,i}^T\varepsilon_{NDE,i} + U_{2,i}^T \times \mathbb{E}\left[(1-E_i)\,X_{3,i}X_{4,i}^T\right]\right\}\right] \\
& \mathbb{E}\left((1-E_i)X_{3,i}X_{3,i}^T\right)^{-1}
\end{aligned}$$

An estimator of this matrix is obtained upon substituting all unknown parameters by the corresponding estimates, and by replacing population expectations with empirical expectations. For instance, $\mathbb{E}\left((1-E_i)X_{3,i}X_{3,i}^T\right)$ is consistenty estimated by $n^{-1}\sum_i \left((1-E_i)X_{3,i}X_{3,i}^T\right)$ ...etc

Var($\widehat{\beta^*}$): Consider the first stage stratified OLS $\widehat{\eta}^*$, its large sample expansion gives

$$\widehat{\eta}^* - \eta^* \approx n^{-1} \sum_i U^*_{1,i}$$

$$\text{where } U^*_{1,i} = \mathbb{E}\left(X^*_{1,i} X^{*T}_{1,i}\right)^{-1} X^*_{1,i} \varepsilon^*_{M,i}$$

$$X^*_{1,i} = (E_i, E_i C_i)^T$$

$$\varepsilon^*_{M,i} = M_{\epsilon,i} - \eta^{*T} X^*_{1,i}$$

A standard Taylor approximation gives the large sample expansion of the second stage OLS $\widehat{\alpha}$

$$\widehat{\alpha}^* - \alpha^* \approx n^{-1} \sum_i U_{4,i}$$

$$\text{where } U_{4,i} = \mathbb{E}\left(X^*_{2,i} X^{*T}_{2,i}\right)^{-1} \left\{ X^*_{2,i} \varepsilon^\dagger_{Y,i} - \mathbb{E}\left(\alpha^*_2 X^*_{2,i} X^{*T}_{1,i}\right) U^*_{1,i} \right\}$$

$$X^*_{2,i} = \left(E_i, E_i \mathbb{E}\left(M_i | E_i, C_i\right), E_i C^T_{1,i}\right)^T$$

$$\varepsilon^\dagger_{Y,i} = Y_i - X^{*T}_{2,i} \alpha^*$$

$$\widehat{\beta} - \beta \approx n^{-1} \sum_i U_{5,i}$$

$$\text{where } U_{5,i} = \mathbb{E}\left((1 - E_i) X_{3,i} X^T_{3,i}\right)^{-1}$$

$$\times \left\{ (1 - E_i) X_{3,i} \varepsilon_{NDE,i} + \mathbb{E}\left[(1 - E_i) X_{3,i} X^T_{7,i}\right] \times U_{4,i} \right\}$$

$$\varepsilon_{NDE,i} = X^T_{7,i} \alpha^* - Y_i - \beta_0 - \beta^T_1 C_i$$

$$X_{7,i} = (1, M_{\epsilon,i}, C^T_{1,i})^T$$

The variance-covariance matrix of $\widehat{\beta}$ is therefore approximately given by

$$
\begin{aligned}
& n^{-1}\mathbb{E}\left(U_5 U_5^T\right) \\
& = \mathbb{E}\left((1-E_i)X_{3,i}X_{3,i}^T\right)^{-1} \\
& \left\{\mathbb{E}\left[(1-E_i)X_{3,i}\varepsilon_{NDE,i}^2 X_{3,i}^T\right] + \mathbb{E}\left[(1-E_i)\,X_{3,i}X_{6,i}^T\right]\mathbb{E}\left[U_{4,i}U_{4,i}^T\right]\mathbb{E}\left[(1-E_i)\,X_{3,i}^T X_{6,i}\right]\right\} \\
& \mathbb{E}\left((1-E_i)X_{3,i}X_{3,i}^T\right)^{-1}
\end{aligned}
$$

An estimator of this matrix is obtained upon substituting all unknown parameters by the corresponding estimates, and by replacing population expectations with empirical expectations.

References

[1] Robins, J. M. and S. Greenland (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology 3, 143–155.

[2] Pearl, J. (2001). Direct and indirect effects. In Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01), San Francisco, CA, pp. 411–42. Morgan Kaufmann.

[3] Robins, J. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. Green, N. Hjort, and S. Richardson (Eds.), Highly Structured Stochastic Systems, pp. 70–81. Oxford, UK: Oxford University Press.

[4] Avin, C., I. Shpitser, and J. Pearl (2005). Identifiability of path-specific effects. In IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005, pp. 357–363.

[5] Petersen M, Sinisi, S; van der Laan, M.(2006) Estimation of Direct Causal Effects.Epidemiology. Volume 17, 3, 276-284.

[6] van der Laan, M, Petersen, M. (2005) Direct Effect Models. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 187. http://www.bepress.com/ucbbiostat/paper187

[7] Goetgeluk, S., Vansteelandt, S. and Goetghebeur, E. (2008). Estimation of controlled direct effects. Journal of the Royal Statistical Soceity, Series B, 70:1049-1066.

[8] Vansteelandt, S. (2009). Estimating direct effects in cohort and case-control studies. Epidemiology, 20:851-860.

[9] VanderWeele, T.J. (2009). Marginal structural models for the estimation of direct and indirect effects. Epidemiology, 20:18-26.

[10] VanderWeele, T.J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. Statistics and its Interface, 2, 457-468.

[11] VanderWeele, T.J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome - with discussion. American Journal of Epidemiology, 172, 1339-1348.

[12] Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. Statistical Science 25, 51–71.

[13] Tchetgen Tchetgen, E J (2011) On Causal Mediation Analysis with a Survival Outcome.International Journal of Biostatistics: Vol. 7: Iss. 1, Article 33.

[14] Tchetgen Tchetgen EJ and Shpitser I (2012) Semiparametric Theory for Causal Mediation Analysis: efficiency bounds, multiple robustness, and sensitivity analysis. Annals of Statistics. In press.

[15] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu C. (2006) Measurement Error in Nonlinear Models. 2nd ed. Boca Ratin, FL: Chapman and Hall.

[16] le Cessie S, Debeij J, Rosendaal FR,„Cannegieter SC, Vandenbroucke JP. Quantification of Bias in Direct Effects Estimates Due to Different Types of Measurement Error in the Mediator. (2012).Epidemiology;23: 551–560.

[17] Rubin DB. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal Educational Psychology, 66:688-701.

[18] Rubin DB. (1978). Bayesian inference for causal effects: The role of randomization. Annals of Statistics, 6:34-58.

[19] Pearl, J. *Causality: Models, Reasoning, and Inference*, 2nd ed. New York: Cambridge University Press; 2009.

[20] Robins JM. (1997). Causal Inference from Complex Longitudinal Data. Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120), M. Berkane, Editor. NY: Springer Verlag, pp. 69-117.

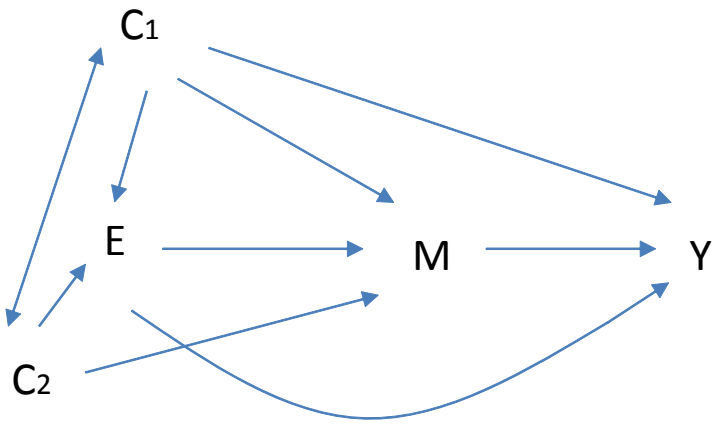

Figure 1. No measurement error in the mediator

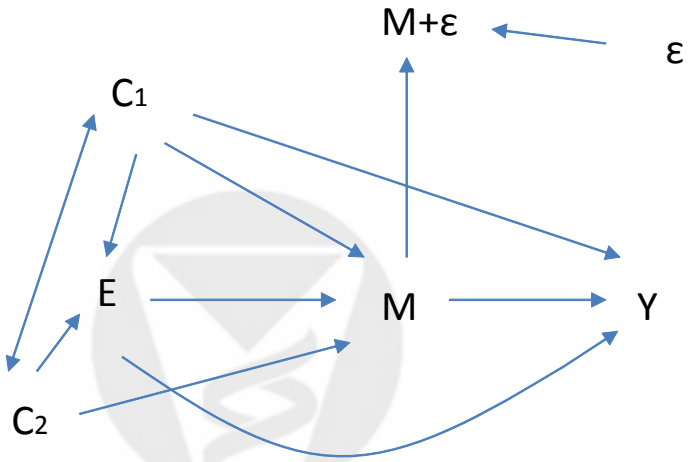

Figure 2. Nondifferential additive measurement error of the mediator M

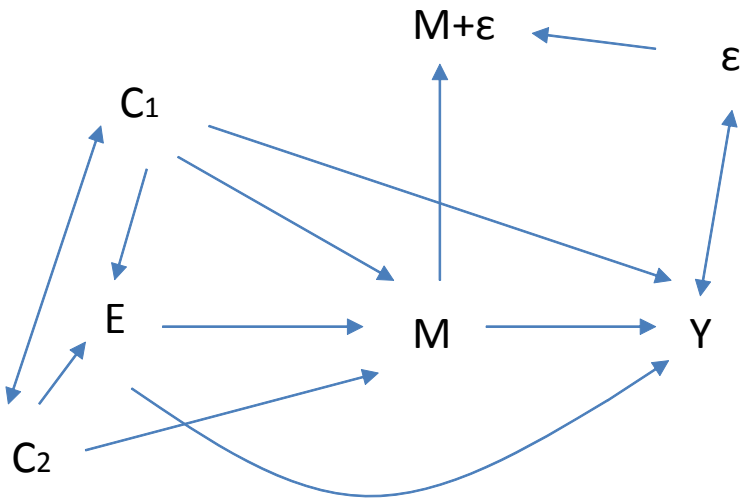

Figure 3. Differential measurement error for the mediator M

Table 1. Simulation results for $\beta_0 = 3$ and $(\alpha_3, \alpha_5, \eta_4) = (0, 0, 0)$ which corresponds to $\beta_1 = \beta_2 = 0$

n=500										
		Standard	Estimator		3SLS	Estimator		Stratified	3SLS	
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
$k = 0$	bias	0.007	-0.001	-0.001	0.013	-0.003	-0.003	0.022	-0.003	-0.003
	variance	0.644	0.0414	0.040	0.863	0.095	0.062	0.852	0.178	0.088
	Coverage[#]	0.956	0.951	0.946	0.936	0.947	0.946	0.946	0.952	0.956
$k = 0.05$	bias	0.014	-0.001	-0.001	0.013	-0.003	-0.003	0.022	-0.002	-0.004
	variance	0.643	0.041	0.040	0.863	0.095	0.062	0.852	0.174	0.088
	Coverage	0.955	0.950	0.945	0.937	0.947	0.947	0.948	0.952	0.955
$k = 0.15$	bias	0.083	-0.001	-0.001	0.013	-0.003	-0.003	0.023	-0.002	-0.004
	variance	0.641	0.041	0.040	0.866	0.095	0.062	0.854	0.175	0.088
	Coverage	0.955	0.948	0.946	0.938	0.948	0.949	0.950	0.953	0.959
$k = 0.3$	bias	0.417	-0.001	-0.001	0.011	-0.00264	-0.003	0.021	-0.002	0.004
	variance	0.632	0.040	0.039	0.876	0.098	0.062	0.869	0.179	0.090
	Coverage	0.912	0.948	0.944	0.938	0.949	0.957	0.949	0.955	0.953
$k = 0.5$	bias	1.526	-0.001	-0.001	5.8×10^{-4}	-0.001	-0.004	0.003	-0.004	-0.002
	variance	0.593	0.038	0.037	0.922	0.108	0.065	0.957	0.201	0.101
	Coverage	0.314	0.956	0.950	0.940	0.948	0.957	0.950	0.956	0.948

[#]Coverage of 95% Wald-type Confidence intervals using asymptotic variance estimator derived in the appendix

Table 2. Simulation results for $\beta_0 = 3$ and $(\alpha_3, \alpha_5, \eta_4) = (0, 0, 0)$ which corresponds to $\beta_1 = \beta_2 = 0$

n=2000										
		Standard	Estimator		3SLS	Estimator		Stratified	3SLS	
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
$k = 0$	bias	-0.005	1.9×10^{-4}	2.5×10^{-4}	7.6×10^{-4}	-2.4×10^{-4}	4.6×10^{-5}	-0.001	-3.2×10^{-4}	3.0×10^{-4}
	variance	0.308	0.020	0.020	0.411	0.046	0.031	0.417	0.085	0.045
	Coverage[#]	0.959	0.944	0.942	0.944	0.961	0.946	0.944	0.958	0.950
$k = 0.05$	bias	0.001	1.9×10^{-4}	2.5×10^{-4}	4.9×10^{-4}	-2.8×10^{-4}	7.4×10^{-5}	-0.002	-4.0×10^{-4}	3.7×10^{-4}
	variance	0.308	0.020	0.020	0.411	0.046	0.031	0.418	0.085	0.045
	Coverage	0.959	0.943	0.942	0.946	0.961	0.945	0.944	0.958	0.948
$k = 0.15$	bias	0.070	1.9×10^{-4}	2.5×10^{-4}	-2.0×10^{-4}	-3.7×10^{-4}	1.4×10^{-4}	-0.003	5.4×10^{-4}	-6.3×10^{-4}
	variance	0.307	0.020	0.020	0.412	0.046	0.031	0.421	0.045	0.086
	Coverage	0.956	0.94	0.940	0.946	0.961	0.943	0.946	0.946	0.96
$k = 0.3$	bias	0.402	1.8×10^{-4}	2.5×10^{-4}	-2.0×10^{-4}	-5.6×10^{-4}	2.9×10^{-4}	-0.008	9.7×10^{-4}	-0.001
	variance	0.306	0.019	0.018	0.417	0.047	0.032	0.433	0.047	0.088
	Coverage	0.782	0.942	0.94	0.950	0.961	0.942	0.943	0.946	0.959
$k = 0.5$	bias	1.514	4.4×10^{-5}	1.4×10^{-4}	-0.007	-9.9×10^{-4}	6.0×10^{-4}	-0.021	0.002	-0.002
	variance	0.295	0.018	0.018	0.440	0.051	0.034	0.485	0.051	0.099
	Coverage	0.001	0.95	0.948	0.949	0.962	0.948	0.942	0.957	0.959

[#]Coverage of 95% Wald-type Confidence intervals using asymptotic variance estimator derived in the appendix

Table 3. Simulation results for $\beta_0 = 6$ and $(\alpha_3, \alpha_5, \eta_4) = (3, 0, 0)$ which corresponds to $\beta_1 = \beta_2 = 3$

n=500										
		Standard	Estimator		3SLS	Estimator		Stratified	3SLS	
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
$k = 0$	bias	-0.025	0.006	-0.005	-0.049	0.008	-0.013	0.026	-0.003	-0.011
	variance	1.343	0.373	0.183	2.175	0.408	0.204	1.624	0.516	0.264
	Coverage[#]	0.946	0.940	0.95	0.937	0.936	0.946	0.950	0.944	0.945
$k = 0.05$	bias	0.044	0.002	-0.011	-0.049	0.009	-0.013	0.030	-0.002	-0.011
	variance	1.355	0.373	0.264	2.181	0.408	0.204	1.633	0.518	0.264
	Coverage	0.941	0.943	0.947	0.939	0.936	0.947	0.950	0.942	0.948
$k = 0.15$	bias	0.732	-0.027	0.012	-0.049	0.009	-0.014	0.036	-0.001	-0.012
	variance	1.426	0.376	0.272	2.215	0.415	0.206	1.740	0.537	0.272
	Coverage	0.895	0.941	0.947	0.942	0.943	0.947	0.952	0.942	0.947
$k = 0.3$	bias	4.125	-0.177	-0.182	-0.054	0.010	-0.015	0.022	-0.002	-0.012
	variance	1.680	0.383	0.193	2.377	0.448	0.221	2.242	0.627	0.315
	Coverage	0.261	0.908	0.831	0.939	0.943	0.949	0.956	0.947	0.949
$k = 0.5$	bias	16.689	-0.774	0.775	-0.091	0.012	-0.012	-0.127	0.020	7×10^{-4}
	variance	2.204	0.390	0.217	3.081	0.593	0.315	4.036	0.983	0.500
	Coverage	0	0.479	0.06	0.942	0.938	0.946	0.958	0.954	0.946

[#]Coverage of 95% Wald-type Confidence intervals using asymptotic variance estimator derived in the appendix

Table 4. Simulation results for $\beta_0 = 6$ and $(\alpha_3, \alpha_5, \eta_4) = (3, 0, 0)$ which corresponds to $\beta_1 = \beta_2 = 3$

n=1000										
		Standard	Estimator		3SLS	Estimator		Stratified	3SLS	
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
$k = 0$	bias	-0.010	0.001	4.6×10^{-4}	-0.020	0.008	-0.002	-0.018	0.009	-6.0×10^{-4}
	variance	0.925	0.248	0.128	1.503	0.275	0.143	1.148	0.365	0.187
	Coverage[#]	0.946	0.957	0.957	0.942	0.957	0.946	0.937	0.957	0.953
$k = 0.05$	bias	0.057	-3.7×10^{-4}	-0.031	-0.019	0.008	-0.002	-0.020	0.008	-2.0×10^{-4}
	variance	0.932	0.249	0.130	1.504	0.276	0.143	1.136	0.367	0.189
	Coverage	0.942	0.955	0.956	0.940	0.954	0.947	0.943	0.955	0.949
$k = 0.15$	bias	0.741	-0.028	-0.031	-0.017	0.009	-0.002	-0.029	0.007	0.001
	variance	0.980	0.251	0.130	1.522	0.281	0.146	1.224	0.383	0.197
	Coverage	0.876	0.957	0.952	0.943	0.950	0.949	0.945	0.946	0.947
$k = 0.3$	bias	4.128	-0.172	-0.180	-0.015	0.010	-0.003	-0.059	0.003	0.004
	variance	1.164	0.260	0.136	1.623	0.306	0.159	1.586	0.452	0.231
	Coverage	0.046	0.902	0.726	0.952	0.947	0.943	0.949	0.941	0.951
$k = 0.5$	bias	16.686	-0.762	-0.774	-0.023	0.012	-0.003	-0.193	-0.014	0.020
	variance	1.542	0.275	0.149	2.090	0.413	0.211	2.884	0.712	0.362
	Coverage	0	0.209	0.001	0.951	0.945	0.947	0.955	0.944	0.949

[#]Coverage of 95% Wald-type Confidence intervals using asymptotic variance estimator derived in the appendix

Table 5. Simulation results in the presence of modeling error with and without measurement error

n=1000										
		Standard	Estimator		3SLS	Estimator		Stratified	3SLS	
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
$k=0$	bias	-0.012	0.001	4.0×10^{-4}	-0.020	0.008	-0.002	-0.018	0.009	$-6.0\times10^{-}$
$(\alpha_3,\alpha_5,\eta_4)=(3,0,0)$	variance	0.925	0.248	0.128	1.503	0.275	0.143	1.148	0.365	0.187
$(\beta_0,\beta_1,\beta_2)=(6,3,3)$	Coverage#	0.946	0.957	0.957	0.942	0.957	0.946	0.937	0.957	0.953
$k=0$	bias	-15.508	9.295	-0.533	-36.520	6.431	0.608	-0.018	0.018	-0.0006
$(\alpha_3,\alpha_5,\eta_4)=(3,-4,5)$	variance	1.582	0.433	0.171	4.176	0.527	0.213	1.148	0.467	0.187
$(\beta_0,\beta_1,\beta_2)=(6,-1,3)$	Coverage	0	0	0.114	0	0	0.177	0.940	0.959	0.953
$k=0.3$	bias	-13.587	9.384	-0.509	-36.518	6.431	0.608	-0.059	-0.011	0.004
$(\alpha_3,\alpha_5,\eta_4)=(3,-4,5)$	variance	1.682	0.449	0.179	4.278	0.542	0.232	1.586	0.779	0.231
$(\beta_0,\beta_1,\beta_2)=(6,-1,3)$	Coverage	0	0	0.188	0	0	0.247	0.950	0.949	0.951

#Coverage of 95% Wald-type Confidence intervals using asymptotic variance estimator derived in the appendix