Biostatistics (2014), **15**, 3, *pp.* 498–512
doi:10.1093/biostatistics/kxu007
Advance Access publication on March 26, 2014

The estimation of direct and indirect causal effects in the presence of misclassified binary mediator

LINDA VALERI*, TYLER J. VANDERWEELE

Department of Biostatistics and Epidemiology, Harvard School of Public Health, 655 Huntington avenue, Boston, MA 02115, USA

lvaleri@hsph.harvard.edu

Summary

Mediation analysis serves to quantify the effect of an exposure on an outcome mediated by a certain intermediate and to quantify the extent to which the effect is direct. When the mediator is misclassified, the validity of mediation analysis can be severely undermined. The contribution of the present work is to study the effects of non-differential misclassification of a binary mediator in the estimation of direct and indirect causal effects when the outcome is either continuous or binary and exposure–mediator interaction can be present, and to allow the correction of misclassification. A hybrid of likelihood-based and predictive value weighting method for misclassification correction coupled with sensitivity analysis is proposed and a second approach using the expectation–maximization algorithm is developed. The correction strategy requires knowledge of a plausible range of sensitivity and specificity parameters. The approaches are applied to a perinatal epidemiological study of the determinants of pre-term birth.

Keywords: EM algorithm; Iteratively re-weighted least squares; Mediation analysis; Misclassification; Predictive value weighting; Pre-eclampsia; Pre-term birth; Sensitivity analysis.

1. Introduction

Causal mediation analysis investigates the role of intermediate variables (mediators) in explaining the mechanisms through which an exposure variable exerts a causal effect on an outcome variable. A mediational model hypothesizes that the exposure variable causes the mediator variable, which in turn causes the outcome variable (MacKinnon, 2008). The use of mediation analysis in biomedical and social sciences is widespread and has been strongly influenced by the seminal paper of Baron and Kenny (1986). More recently, new advances in mediation analysis have been made by applying the counterfactual framework in this field (Robins and Greenland, 1992; Pearl, 2001; VanderWeele and Vansteelandt, 2009, 2010; Imai *and others*, 2010). The use of the counterfactual framework has allowed definitions of direct and indirect effects and decomposition of a total effect into direct and indirect effects even in models with interactions and non-linearities (VanderWeele and Vansteelandt, 2009, 2010).

The problem of measurement error in mediation analysis has been explored for the simple and linear mediation model by Hoyle and Kenny (1999). le Cessie *and others* (2012) studied the bias of direct effect

*To whom correspondence should be addressed.

© The Author 2014. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

when a continuous mediator is measured with error and the outcome is modeled using linear or logistic regression. VanderWeele *and others* (2012) and Valeri *and others* (2014) studied the impact of measurement error on a continuous mediator when direct and indirect causal effects are estimated using generalized linear models in the presence of exposure–mediator interaction. The direction of bias due to the misclassification of a binary mediator has been considered by Ogburn and VanderWeele (2012) in a non-parametric setting.

In the context of a regression-based approach to mediation analysis, the investigator needs to estimate the parameters from the *outcome* and *mediator regressions*. Then, direct and indirect causal effects are recovered as functions of those regression parameters, provided the models have been correctly specified and the no confounding assumptions described below hold (VanderWeele and Vansteelandt, 2009, 2010). When a binary mediator is misclassified, it is crucial to investigate how misclassification affects the estimation of outcome and mediator regressions' parameters. We use results derived about the consequences of outcome or covariate misclassification on parameter estimators in regression models (Neuhaus, 1999; Gustafson, 2004; Carroll *and others*, 2006).

The present work makes two contributions. First, we study the implications of non-differential misclassification of the mediator variable (i.e. misclassification mechanism is independent of the outcome, the exposure, and the covariates) on the validity of mediation analysis. Assuming a continuous outcome modeled using linear regression and that exposure and mediator may interact in their effect on the outcome, we derive the asymptotic bias of direct and indirect causal effect estimators in closed form. We demonstrate that even if the error is assumed to be non-differential, regression coefficient estimators obtained in mediation analysis ignoring misclassification can sometimes be severely biased and therefore induce bias in the estimation of direct and indirect causal effects. Secondly, we propose correction strategies for non-differential misclassification of a binary mediator that yield consistent or approximately consistent estimators of the direct and indirect effects when the outcome is continuous or binary, allowing for interactions.

The paper is organized as follows. Section 2 defines direct and indirect causal effects and discusses some results from mediation analysis. Section 3 studies the asymptotic bias in direct and indirect causal effects when the binary mediator is misclassified. In Section 4, we describe the approaches for misclassification correction and we evaluate their performance in estimating direct and indirect causal effects via a simulation study. In Section 5, we apply the proposed methods to a perinatal epidemiological study, followed by discussion in Section 6.

2. Mediation analysis within the counterfactual framework in the absence of misclassification

Let A be an exposure or treatment, Y be an outcome, M be a mediator, and $\boldsymbol{C}$ be a k-dimensional vector of covariates. The causal diagram in Figure 1 captures how the role of a mediator variable can be conceptualized. In the figure, the exposure can have an effect on the outcome either by exerting a causal effect on the mediator which in turn is causally related to the outcome or by affecting the level of the outcome independently of its impact on the intermediate variable.

Let Y_a and M_a denote the values of the outcome and mediator that would have been observed had the exposure A been set to level a. Let Y_{am} denote the value of the outcome that would have been observed had the exposure, A, and mediator, M, been set to levels a and m, respectively. Given these counterfactual variables, the following causal effects can be defined. The natural direct effect (NDE), defined by $E[Y_{aM_{\tilde{a}}} - Y_{\tilde{a}M_{\tilde{a}}} \mid \boldsymbol{C}]$, measures how much the mean of the outcome would change if the exposure were set at level a versus level $\tilde{a}$ but the mediator were kept at the level it would have taken under $\tilde{a}$. The natural indirect effect (NIE), defined by $E[Y_{aM_a} - Y_{aM_{\tilde{a}}} \mid \boldsymbol{C}]$, measures how much the mean of the outcome would

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

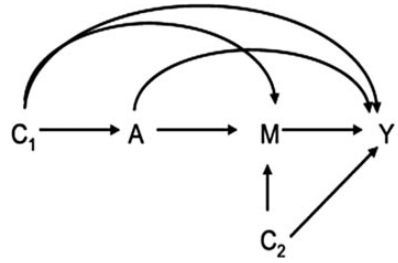

Fig. 1. Mediation directed acyclic graph.

change if the exposure were controlled at level a, but the mediator were changed from the level it would take under $\tilde{a}$ to the level it would take under a (Robins and Greenland, 1992; Pearl, 2001).

Let A and $\boldsymbol{C}$ be either continuous or categorical. In the context of a regression approach to mediation analysis, for binary mediator and continuous outcome, consider the following models:

$$\text{logit}\{P(M=1 \mid A=a, \boldsymbol{C}=\boldsymbol{c})\} = \beta_0 + \beta_1 a + \boldsymbol{\beta_2'c}, \tag{2.1}$$

$$E(Y \mid A=a, M=m, \boldsymbol{C}=\boldsymbol{c}) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}. \tag{2.2}$$

Let $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ denote the vector of mediator and outcome regression parameters. Under models (2.1) and (2.2) and under the confounding control assumptions described below $\text{NDE}(\boldsymbol{\theta}, \boldsymbol{\beta})$ and $\text{NIE}(\boldsymbol{\theta}, \boldsymbol{\beta})$ for a change in exposure from level $\tilde{a}$ to level a can be estimated has (Valeri and VanderWeele, 2013):

$$\text{NDE}(\boldsymbol{\theta}, \boldsymbol{\beta}) = E(Y_{aM_{\tilde{a}}} - Y_{\tilde{a}M_{\tilde{a}}} \mid C=c) = \{\theta_1(a-\tilde{a})\} + \{\theta_3(a-\tilde{a})\}\frac{\exp(\beta_0 + \beta_1\tilde{a} + \boldsymbol{\beta_2'c})}{1+\exp(\beta_0 + \beta_1\tilde{a} + \boldsymbol{\beta_2'c})},$$

$$\begin{aligned}\text{NIE}(\boldsymbol{\theta}, \boldsymbol{\beta}) &= E(Y_{aM_a} - Y_{aM_{\tilde{a}}} \mid C=c) \\ &= (\theta_2 + \theta_3 a)\left\{\frac{\exp(\beta_0 + \beta_1 a + \boldsymbol{\beta_2'c})}{1+\exp(\beta_0 + \beta_1 a + \boldsymbol{\beta_2'c})} - \frac{\exp(\beta_0 + \beta_1\tilde{a} + \boldsymbol{\beta_2'c})}{1+\exp(\beta_0 + \beta_1\tilde{a} + \boldsymbol{\beta_2'c})}\right\}.\end{aligned}$$

When the outcome is binary modeled with a logit link, (2.2) can be replaced by

$$\text{logit}\{P(Y=1 \mid A=a, M=m, \boldsymbol{C}=\boldsymbol{c})\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}. \tag{2.3}$$

If the outcome is binary and rare, then from models (2.1) and (2.3) NDE and NIE for a change in exposure from level $\tilde{a}$ to level a are given in terms of odds ratios by Valeri and VanderWeele (2013):

$$\text{OR}^{\text{NDE}}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{\exp(\theta_1 a)\{1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1\tilde{a} + \boldsymbol{\beta_2'c})\}}{\exp(\theta_1\tilde{a})\{1+\exp(\theta_2 + \theta_3\tilde{a} + \beta_0 + \beta_1\tilde{a} + \boldsymbol{\beta_2'c})\}},$$

$$\text{OR}^{\text{NIE}}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{\{1+\exp(\beta_0 + \beta_1\tilde{a} + \boldsymbol{\beta_2'c})\}\{1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \boldsymbol{\beta_2'c})\}}{\{1+\exp(\beta_0 + \beta_1 a + \boldsymbol{\beta_2'c})\}\{1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1\tilde{a} + \boldsymbol{\beta_2'c})\}}.$$

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

The expressions above in terms of regression coefficients will be equal to the counterfactual direct and indirect effects, and therefore have a causal interpretation, provided that the parametric models are correctly specified and that conditional on covariates $\boldsymbol{C}$ there is no unmeasured confounding of (i) the exposure–outcome relationship, (ii) the mediator–outcome relationship, (iii) the exposure–mediator relationship, and (iv) that there are no mediator–outcome confounders affected by the exposure. In the counterfactual notation, this is: (i) $Y_{am} \amalg A \mid \boldsymbol{C}$, (ii) $Y_{am} \amalg M \mid A, \boldsymbol{C}$, (iii) $M_a \amalg A \mid \boldsymbol{C}$, and (iv) $Y_{am} \amalg M_{\tilde{a}} \mid \boldsymbol{C}$. (See Pearl, 2001 and Robins and Richardson, 2010 for discussion of these assumptions.)

3. Results on direct and indirect effects naive estimators when the mediator is misclassified

3.1 *Mediator and outcome regressions when mediator is misclassified*

Using the notation in Section 2, assume that both A and $\boldsymbol{C}$, as well as the outcome Y, are correctly measured. Let M be the binary mediator at its true level and $M^* = M + U$ be the misclassified version of M. In the following, we assume that the misclassification error, U, is independent of the outcome, the exposure, and the covariates so that $P(M^* \mid M, Y, A, \boldsymbol{C}) = P(M^* \mid M)$ (i.e. non-differential). The misclassification error U takes values $(-1, 0, 1)$. Under the assumption of non-differential misclassification, the moments of U are characterized by sensitivity ($\mathrm{SN} = P(M^* = 1 \mid M = 1)$), specificity ($\mathrm{SP} = P(M^* = 0 \mid M = 0)$), and the prevalence of the mediator, $p = P(M = 1)$ (Aigner, 1973). When the true intermediate M is replaced by the observed intermediate M^* in models (2.1) and (2.2), *observed* outcome and mediator regressions are given by

$$\mathrm{logit}\{P(M^* = 1 \mid A = a, \boldsymbol{C} = \boldsymbol{c})\} = \beta_0^* + \beta_1^* a + \boldsymbol{\beta_2^{*'}} \boldsymbol{c}, \tag{3.1}$$

$$E(Y \mid A = a, M^* = m^*, \boldsymbol{C} = \boldsymbol{c}) = \theta_0^* + \theta_1^* a + \theta_2^* m^* + \theta_3^* a m^* + \boldsymbol{\theta_4^{*'}} \boldsymbol{c}. \tag{3.2}$$

Misclassification typically causes parameter estimates of the mediator and outcome regression to be asymptotically biased (Gustafson, 2004; Carroll *and others*, 2006). In Appendix A of supplementary material available at *Biostatistics* online, we derive the asymptotic limit for the naive estimators of the mediator regression coefficients assuming a logistic model and for the naive estimators of the outcome regression coefficients assuming a linear model allowing for mediator–exposure interaction. The asymptotic bias of the naive direct and indirect causal effects estimators is given below.

3.2 *Asymptotic bias of the direct and indirect causal effects*

Let the vector $\boldsymbol{\beta^*}$ and $\boldsymbol{\theta^*}$ denote the limit of the vector of the naive mediator and outcome regression parameter estimators $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{\theta}}^*$. Let $\widehat{\mathrm{NIE}}^* = \mathrm{NDE}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\beta}}^*)$ and $\widehat{\mathrm{NIE}}^* = \mathrm{NIE}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\beta}}^*)$ denote the naive estimators for the NDE, and the NIE, respectively, obtained by substituting regressions (3.1) and (3.2) for (2.1) and (2.2). Let $X^* = (A, M^*, C)$ denote the matrix of observed centered covariates and let $\Delta^* = E[X^{*T} X^*]$ denote the variance–covariance matrix of the observed covariates. Let δ_{A,M^*} denote the first off-diagonal elements of Δ^{*-1} and δ_{M^*,M^*} denote the second diagonal element of Δ^{*-1}. Let $p = P(M = 1)$, $q = 1 - p$, $p^* = P(M^* = 1)$, $q^* = 1 - p^*$, and the covariance of M^* and U can be computed as $\mathrm{Cov}(M^*, U) = \{(1 - SN)p/q^* + (1 - SP)q/p^*\}p^*q^*$ (Aigner, 1973). The asymptotic bias of the direct and indirect effects naive estimators when the mean of the outcome follows a linear model in the absence of exposure–mediator interaction (i.e. $\theta_3 = 0$) can be derived as (for

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

proof and expressions allowing for interactions see Appendix A of supplementary material available at *Biostatistics* online):

$$\begin{aligned}
\text{ABIAS}(\widehat{\text{NDE}}^{*}) &= -\theta_2 \delta_{A,M^*} \text{Cov}(M^*, U)(a - \tilde{a}),\\
\text{ABIAS}(\widehat{\text{NIE}}^{*}) &\approx \theta_2 (1 - \delta_{M^*,M^*} \text{Cov}(M^*, U))\\
&\quad \times \left\{ \frac{\exp(\beta_0^* + \beta_1^* a + \boldsymbol{\beta_2^{*\prime}} \boldsymbol{c})}{1 + \exp(\beta_0^* + \beta_1^* a + \boldsymbol{\beta_2^{*\prime}} \boldsymbol{c})} - \frac{\exp(\beta_0^* + \beta_1^* \tilde{a} + \boldsymbol{\beta_2^{*\prime}} \boldsymbol{c})}{1 + \exp(\beta_0^* + \beta_1^* \tilde{a} + \boldsymbol{\beta_2^{*\prime}} \boldsymbol{c})} \right\}\\
&\quad - \theta_2 \left\{ \frac{\exp(\beta_0 + \beta_1 a + \boldsymbol{\beta}_2' \boldsymbol{c})}{1 + \exp(\beta_0 + \beta_1 a + \boldsymbol{\beta}_2' \boldsymbol{c})} - \frac{\exp(\beta_0 + \beta_1 \tilde{a} + \boldsymbol{\beta}_2' \boldsymbol{c})}{1 + \exp(\beta_0 + \beta_1 \tilde{a} + \boldsymbol{\beta}_2' \boldsymbol{c})} \right\}.
\end{aligned}$$

Misclassification of a binary mediator exerts its impact on the estimation of direct and indirect effects by inducing bias in both the mediator and the outcome regression parameter estimation. This contrasts with the effect of measurement error on a continuous mediator which typically induces bias on the outcome regression parameter naive estimators while leaving unbiased the naive estimators of the mediator linear regression parameters (Valeri *and others*, 2014). In the absence of exposure–mediator interaction, the asymptotic bias of direct and indirect effects have a particular direction (proofs of the statements below and additional results are given in Appendix B of supplementary material available at *Biostatistics* online). The indirect effect naive estimator will be biased toward the null under the assumption that the effect of the exposure on the mediator, β_1, and the effect of the mediator on the outcome, θ_2, have the same sign. The direct effect will be biased away from the null under the same assumption. The bias of direct and indirect effects will be larger as SN and SP decrease, as the effect of the mediator on the outcome increases (θ_2) as well as the larger is the correlation between the exposure and the mediator. In the presence of exposure–mediator interaction, it has been shown that the direction of the bias remains the same at least when non-parametric methods are employed (Ogburn and VanderWeele, 2012). In a non-parametric setting, VanderWeele *and others* (2012) show that although measurement error in the mediator induces biased direct and indirect effects, the sum or the product, depending on the scale, of naive direct and indirect effects non-parametric estimators that ignore misclassification will still yield an unbiased total effect. However, these results do not necessarily hold in a parametric regression-approach to mediation analysis. This is because in a parametric setting misclassification of a binary mediator induces misspecification in the mediator and outcome regressions, creating a source of bias beyond a mere measurement error. In Appendix B4 of supplementary material available at *Biostatistics* online, we illustrate how departures to this result can occur using parametric models via a numerical study. The results of the numerical study matched our theoretical findings.

4. CORRECTION STRATEGY FOR DIRECT AND INDIRECT EFFECTS ESTIMATORS

In this section, we propose to carry out mediation analysis employing correction approaches coupled with sensitivity analysis to account for misclassification of a binary mediator. Two methods for misclassification correction are proposed. The first method, called here iteratively re-weighted least squares/predictive value weighting (IRLS/PVW), is a two-stage approach that recovers estimates for outcome and mediator regression parameters by maximizing two weighted score equations separately. For each equation maximization is done with respect to all parameters simultaneously; weights are estimated only once. The second approach is developed in the spirit of the expectation–maximization (EM) algorithm. The EM algorithm approach is an iterative algorithm that maximizes the incomplete data weighted likelihood for the mediator and the outcome jointly with weights re-estimated at each iteration. Estimates of direct and indirect causal effects

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

are recovered by plugging the corrected estimates of outcome and mediator regression parameters in the causal effects formulae given in Section 2. Standard errors for the corrected estimates can be obtained using jackknife or bootstrap procedure or analytical standard errors for the corrected causal effects estimators can be obtained employing the multivariate delta method as illustrated in Appendix C of supplementary material available at *Biostatistics* online. We describe the approaches below and technical details are given in Appendix C and Appendix D2 of supplementary material available at *Biostatistics* online. We have also developed a direct maximum likelihood (dMLE) method, details of which are given in Appendix D1 of supplementary material available at *Biostatistics* online. The dMLE approach did not perform well in some settings. Therefore, we focus our attention and evaluation on the IRLS/PVW method and EM algorithm (see Appendix D3 of supplementary material available at *Biostatistics* online for evaluation of dMLE method). In Section 6, we discuss the relative advantages and disadvantages of the IRLS/PVW and EM algorithm approaches and guide the investigator in the choice of the appropriate method in different settings.

4.1 *IRLS estimators for mediator regression*

Asymptotically unbiased estimators for the vector of true mediator regression parameters $\boldsymbol{\beta}$, described in Section 2, can be found by adjusting the link function and programming IRLS as proposed by Neuhaus (1999). This is a widely popular approach for misclassified outcome regression correction and is known to perform well in most settings (Carroll *and others*, 2006; Lyles and Lin, 2010). Provided that the misclassification probabilities are known or set as sensitivity analysis parameters, the IRLS method entails constructing the likelihood for the probability that the true mediator is equal to one in terms of misclassification probabilities and the conditional probability of the mediator given the observed covariates using a logistic model and maximizing it with respect to the regression's parameters.

It can be shown that the likelihood for the true mediator parameters given the observed misclassified mediator, SN, SP, and the observed covariates A and $\boldsymbol{C}$ is given by

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ (1 - \mathrm{SP}) \frac{1}{1 + \exp(\beta_0 + \beta_1 A_i + \boldsymbol{\beta_2'} \boldsymbol{C}_i)} + \mathrm{SN} \frac{\exp(\beta_0 + \beta_1 A_i + \boldsymbol{\beta_2'} \boldsymbol{C}_i)}{1 + \exp(\beta_0 + \beta_1 A_i + \boldsymbol{\beta_2'} \boldsymbol{C}_i)} \right\}^{M_i^*}$$
$$\times \left\{ \mathrm{SP} \frac{1}{1 + \exp(\beta_0 + \beta_1 \boldsymbol{A}_i + \boldsymbol{\beta_2'} \boldsymbol{C}_i)} + (1 - \mathrm{SN}) \frac{\exp(\beta_0 + \beta_1 A_i + \boldsymbol{\beta_2'} \boldsymbol{C}_i)}{1 + \exp(\beta_0 + \beta_1 A_i + \boldsymbol{\beta_2'} \boldsymbol{C}_i)} \right\}^{(1 - M_i^*)}.$$

By setting a plausible range of values for the sensitivity analysis parameters SP and SN and maximizing this likelihood, consistent estimators for the true parameters $\boldsymbol{\beta}$ can be recovered when the true SP and SN lie in the specified range. Standard errors for the parameters can be obtained via close numerical approximations to the observed information matrix or via the bootstrapping procedure.

4.2 *PVW estimators for outcome regression*

PVW proposed by Lyles and Lin (2010) is a method for misclassification correction that shares similarities with imputation methods for missing data. We consider this methodology because it is intuitive, easy to program, and not too computationally intensive. The PVW approach maximizes a weighted log-likelihood for the true outcome regression parameters that arises exploiting the relationship between the conditional density of the outcome given the *observed* covariates and the conditional density of the outcome given the *true* covariates and positive and negative predicted values (PPV and NPV):

$$\ell(\boldsymbol{\theta}; Y, A, M^*, \boldsymbol{C}) = \sum_{i=1}^{n} \sum_{m=0}^{1} P(M_i = m \mid Y_i, A_i, M_i^*, \boldsymbol{C}_i) \ell(\boldsymbol{\theta}; Y_i, A_i, M_i = m, \boldsymbol{C}_i),$$

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

where $P(M_i = m \mid Y_i, A_i, M_i^*, \boldsymbol{C}_i)$ depends on $\text{PPV} = P(M = 1 \mid Y = y, A = a, M^* = 1, \boldsymbol{C} = \boldsymbol{c})$, and $\text{NPV} = P(M = 0 \mid Y = y, A = a, M^* = 0, \boldsymbol{C} = \boldsymbol{c})$; and $\ell(\boldsymbol{\theta}; Y_i, A_i, M_i = m, \boldsymbol{C}_i)$ is the likelihood for the true $\boldsymbol{\theta}$, which arises from the outcome regression models described in Section 2. The above function is maximized with respect to the true $\boldsymbol{\theta}$ provided PPV and NPV are correctly estimated. PPV and NPV are found as solutions to a system of two linear equations (Lyles and Lin, 2010):

$$\begin{pmatrix} \text{PPV} \\ \text{NPV} \end{pmatrix} = \begin{pmatrix} (\text{SN} - 1)p^*_{yac}\{\text{SN}[p^*_{yac} - 1]\}^{-1} & 1 \\ 1 & (\text{SP} - 1)[1 - p^*_{yac}][\text{SP}p^*_{yac}]^{-1} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{4.1}$$

where $p^*_{yac} = P(M^* = 1 \mid Y, A, C)$. Consistent estimates of $\widehat{\text{PPV}}$ and $\widehat{\text{NPV}}$ can be obtained by setting SN and SP as sensitivity analysis parameters and estimating consistently $\widehat{p}^*_{yac}$ from the observed data. Lyles and Lin (2010) propose to estimate $p^*_{yac} = P(M^* \mid Y, A, \boldsymbol{C})$ via a logistic regression:

$$\text{logit}\{P(M^* = 1 \mid Y = y, A = a, \boldsymbol{C} = \boldsymbol{c})\} = \xi_0 + \xi_1 y + \xi_2 a + \boldsymbol{\xi}_3' \boldsymbol{c}. \tag{4.2}$$

Given the data generating process for M, M^*, and Y, it can be shown that (4.2) is misspecified for the probability of $M^* \mid Y, A, \boldsymbol{C}$, which will not follow a logistic model. We refer to this approach as IRLS/PVW. In order to overcome the issue of model misspecification, we considered two modified PVW estimators. In the first modified approach, we used splines to model more flexibly the relationship between M^* and $(Y, A, \boldsymbol{C})$. We implemented this approach by modeling each term in (4.2) and all possible interactions with splines

$$\begin{aligned} \text{logit}\{P(M^* = 1 \mid Y = y, A = a, \boldsymbol{C} = \boldsymbol{c})\} = {} & s(y, k) + s(a, k) + s(\boldsymbol{c}, k) + s(y * a, k) \\ & + s(y * \boldsymbol{c}, k) + s(a * \boldsymbol{c}, k), \end{aligned}$$

where $s(\cdot)$ denotes a smoothing function and k indicates the number of knots (the more knots, the more flexible is the smoothing function). We refer to this approach as IRLS/**s**PVW. In the second approach (IRLS/**t**PVW), we recover the outcome regression parameters by maximizing the likelihood arising from the conditional distribution $P(M^* \mid Y, A, \boldsymbol{C})$. For a data generating process typical in mediation analysis context, this conditional probability is given by (see Appendix C of supplementary material available at *Biostatistics* online for proof):

$$\begin{aligned} P(M^* = 1 \mid Y, A, C; \boldsymbol{\theta}, \boldsymbol{\beta}) = {} & \{\text{SN} * f(Y \mid A, M = 1, C; \boldsymbol{\theta}) P(M = 1 \mid A, C; \boldsymbol{\beta}) \\ & + (1 - \text{SP}) * f(Y \mid A, M = 0, C; \boldsymbol{\theta}) P(M = 0 \mid A, C; \boldsymbol{\beta})\} \\ & \times \left\{ \sum f(Y \mid A, M = m, C; \boldsymbol{\theta}) P(M = m \mid A, C; \boldsymbol{\beta}) \right\}^{-1} \end{aligned}$$

where $f(\cdot)$ indicates a probability density function. Given this probability, the likelihood function for the binary observed mediator conditional on the outcome, the exposure, and the covariates, $M^* \mid Y, A, \boldsymbol{C}$, can be constructed and upon numerical maximization, the parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$ can be estimated. In the simulation section, we evaluate the performance of the three approaches. To overcome the model incompatibility, we also developed an EM algorithm approach described below.

Note that (4.1) implies two restrictions on the values that SN and SP could take, for which $P(M^* \mid Y, A, \boldsymbol{C}) < \text{SN}$ and $P(M^* \mid Y, A, \boldsymbol{C}) > 1 - \text{SP}$ for all $(Y, A, \boldsymbol{C})$ combinations (Lyles and Lin, 2010). Failing to consider these restrictions may result in numerical problems and produce PPV and NPV that are

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

outside the (0, 1) range. Standard errors for all parameters in the model can be recovered using jackknife or bootstrap procedures (Lyles and Lin, 2010) or can be analytically derived as shown in Appendix C of supplementary material available at *Biostatistics* online. The approach can be extended to allow for the presence of exposure–mediator interaction and can be applied both to continuous and discrete outcomes.

4.3 *EM algorithm approach to misclassification correction*

The EM algorithm introduced in 1977 by Dempster, Laird, and Rubin has become one of the methods of choice for maximum likelihood estimation when outcome or covariates are missing. Misclassification of a binary mediator in mediation analysis can be seen as a type of missing data problem. We can reformulate the problem starting from the joint distribution of the true outcome and the mediator, $f_{Y,M|A,C}(y, m \mid a, c)$, which can be factorized as

$$f_{Y,M|A,C}(y, m \mid a, c) = f_{Y|A,M,C}(y \mid a, m, c) f_{M|A,C}(m \mid a, c).$$

The complete data log-likelihood is obtained as

$$\begin{aligned}\ell_{y,m|a,c}(\boldsymbol{z}, \boldsymbol{\beta} \mid a, m, c, y) &= \sum_{i=1}^{n} \ell_{y,m|a,c}(\boldsymbol{z}, \boldsymbol{\beta}; a_i, m_i, c_i, y_i) \\ &= \sum_{i=1}^{n} \ell_{y|a,m,c}(\boldsymbol{z}; a_i, m_i, c_i, y_i) + \sum_{i=1}^{n} \ell_{m|a,c}(\boldsymbol{\beta}; a_i, m_i, c_i),\end{aligned}$$

where $\boldsymbol{z} = (\theta_1, \ldots, \theta_4', \sigma)$ if the outcome is normally distributed and $\boldsymbol{z} = (\theta_1, \ldots, \theta_4')$ if the outcome is binary modeled using logistic regression and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2')$. We observe that this likelihood factorizes in two components. The EM algorithm is a general purpose iterative algorithm for maximizing incomplete data likelihoods. The EM algorithm proceeds in two steps, the E-step and the M-step. At the E-step, one calculates the expectation of $\ell_{y,m|a,c}(\boldsymbol{z}, \boldsymbol{\beta})$ conditioning on the current parameter vector $(\boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)})$, say, and on the observed data, in this case (y_i, a_i, c_i, m_i^*), $i = 1, \ldots, n$. In the case of misclassification, we assume that the mediator is missing for each individual. Denoting this expectation by $Q(\boldsymbol{z}, \boldsymbol{\beta} \mid \boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)})$, we have that the E-step can be calculated using (Ibrahim, 1990):

$$\begin{aligned}Q(\boldsymbol{z}, \boldsymbol{\beta} \mid \boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)}) &= \sum_{i=1}^{n} Q_i(\boldsymbol{z}, \boldsymbol{\beta} \mid \boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)}) = \sum_{i=1}^{n} E[\ell_{y,m|a,c}(\boldsymbol{z}, \boldsymbol{\beta}; a_i, m_i, c_i, y_i) \mid a_i, m_i^*, c_i, y_i, \boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)}] \\ &= \sum_{i=1}^{n} \sum_{m_i=0}^{1} w_i(\boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)}) \ell_{y,m|a,c}(\boldsymbol{z}, \boldsymbol{\beta}; a_i, m_i, c_i, y_i).\end{aligned}$$

In the expectation step, we obtain a weighted log-likelihood. Observe the connection between the PVW approach for covariate misclassification described above and the EM approach to maximum likelihood estimation. The latter can be seen as an iterative version of the former. An important difference between the hybrid approach and the EM algorithm approach is that the latter allows to incorporate the outcome information in the estimation of the mediator regression parameters, improving the efficiency of estimation. Note that both the expected log-likelihoods corresponding to the mediator and the outcome distribution are weighted by the same function w_i which depends on the outcome. The weights can be calculated by

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

the use of Bayes' rule and assuming non-differential misclassification:

$$w_i(\boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)}) = P(m_i \mid a_i, m_i^*, c_i, y_i, \boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)}) = \frac{P(y_i \mid a_i, m_i, c_i, \boldsymbol{z}^{(t)}) P(m_i^* \mid m_i) P(m_i \mid a_i, c_i, \boldsymbol{\beta}^{(t)})}{\sum_{m_i=0}^{1} P(y_i \mid a_i, m_i, c_i, \boldsymbol{z}^{(t)}) P(m_i^* \mid m_i) P(m_i \mid a_i, c_i, \boldsymbol{\beta}^{(t)})}.$$

For the M-step, we maximize $\sum_{i=1}^{n} Q_i(\boldsymbol{z} \mid \boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)})$ as a function of $\boldsymbol{z}^{(t)}$ and $\boldsymbol{\beta}^{(t)}$ by finding the solution of the complete-data likelihood of the expanded dataset in which, for every individual m_i can take value either 0 or 1 with probability depending on the weights $w_i(\boldsymbol{z}^{(t)}, \boldsymbol{\beta}^{(t)})$. For the first iteration, the weights can be calculated from a naive analysis that ignores misclassification. By repeating until convergence, we obtain our parameter estimates of interest. Standard errors for these parameters can be obtained through bootstrapping procedures.

4.4 *Simulations*

We conducted simulation studies to evaluate the estimates produced by the iteratively re-weighted least squares/predictive value weighting method (IRLS/PVW, IRLS/**s**PVW, IRLS/**t**PVW) and the EM algorithm approach and to compare those with naive estimators. We implemented the iteratively re-weighted approach with splines (IRLS/**s**PVW) allowing for $k = 4$ knots for the smoothing terms and analyses were performed using the package *gam* built in the R software. Here, we report the results of one such study. The simulation implies a scenario in which the indirect effect of A on Y through M is small relative to the direct effect of A on Y, while the exposure–mediator interaction, if present, is particularly strong. See Appendix D3 of supplementary material available at *Biostatistics* online for other scenarios. Table 1 displays the simulation results. In all the simulation scenarios, IRLS/**s**PVW, IRLS/**t**PVW, and EM algorithm behaved similarly and outperformed IRLS/PVW in terms of bias. For continuous outcome, the IRLS/**s**PVW, IRLS/**t**PVW, and EM algorithm methods eliminated the bias and appeared consistent. In the binary outcome scenario, we observe that the proposed approaches did not completely eliminate the bias even for a sample size of $n = 10\,000$. However, as the sample size increases the estimators are shown to converge to the true parameter values. For a sample size of $n = 50\,000$, the IRLS/**s**PVW, IRLS/**t**PVW, and EM algorithm yielded unbiased direct and indirect effect estimators (Table 2). The EM algorithm estimators outperformed the hybrid approach in terms of variance across all scenarios and is therefore to be preferred. We note, however, that in the simulation study the IRLS/PVW approaches were found be to more computationally efficient than the EM algorithm. While the latter converged in about 10 min, the formers only required about 10 s.

5. EXAMPLE

We applied the proposed correction methodologies for misclassification of a binary mediator to a perinatal epidemiological study of the causal mechanisms leading to pre-term birth using National Center for Health Statistics (NCHS) birth certificate data. Pre-term birth is strongly associated with perinatal mortality and long-term morbidity (McCormick, 1985). The risk of pre-term delivery has been found higher for mothers above age 35 (Jacobsson *and others*, 2004). A potential intermediate of the age–pre-term birth causal relationship is pre-eclampsia. Several studies confirmed pre-eclampsia as a risk factor for medically induced pre-term birth (Goldenberg *and others*, 2008). Maternal age, in turn, has been found to be a risk factor for pre-eclampsia (Lamminpaa *and others*, 2012). Pre-eclampsia is a multisystem hypertensive disorder of pregnancy that affects approximately 3% to 5% of all pregnancies worldwide (World Health Organization,

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

Table 1. *Bias and variance of naive*, *EM algorithm*, *and IRLS/PVW*, *IRLS/**s**PVW*, *IRLS/**t**PVW estimators of NDE*, *NIE*, *and TE when sample size is* $n = 10\,000$, *number of runs is* $r = 1000$, *marginal probability of the true mediator is* 50%

		Naive	EM	IRLS/PVW	IRLS/**t**PVW	IRLS/**s**PVW
No Int and $Y \sim N$						
NDE = 1, NIE = 0.05						
SP = SN = 0.9	NDE	0.02 (0.00)	0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)
	NIE	−0.02 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	TE	0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)
SP = SN = 0.85	NDE	0.03 (0.00)	0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)
	NIE	−0.03 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	TE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)
SP = SN = 0.8	NDE	0.03 (0.00)	0.00 (0.00)	0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)
	NIE	−0.03 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	TE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	−0.00 (0.00)
No Int and $Y \sim$ Ber						
NDE = 2.7, NIE = 1.07						
SP = SN = 0.9	NDE	−0.03 (0.02)	0.03 (0.02)	−0.02 (0.02)	0.03 (0.09)	−0.02 (0.02)
	NIE	−0.03 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)
	TE	−0.10 (0.02)	0.03 (0.02)	−0.02 (0.02)	0.04 (0.10)	−0.02 (0.02)
SP = SN = 0.85	NDE	−0.04 (0.02)	0.03 (0.03)	−0.03 (0.02)	0.05 (0.14)	−0.01 (0.02)
	NIE	−0.04 (0.00)	0.00 (0.00)	−0.01 (0.00)	−0.00 (0.00)	−0.01 (0.00)
	TE	−0.14 (0.02)	0.03 (0.03)	−0.06 (0.02)	0.06 (0.16)	−0.04 (0.02)
SP = SN = 0.8	NDE	−0.06 (0.02)	0.03 (0.03)	−0.03 (0.02)	0.05 (0.18)	−0.03 (0.02)
	NIE	−0.04 (0.00)	−0.00 (0.00)	−0.01(0.00)	−0.00 (0.00)	−0.01 (0.00)
	TE	−0.18 (0.02)	0.03 (0.03)	−0.09 (0.02)	0.07 (0.20)	−0.04 (0.02)
Int and $Y \sim N$						
NDE = 1.5, NIE = 0.1						
SP = SN = 0.9	NDE	0.04 (0.00)	0.00 (0.00)	0.01(0.00)	0.00 (0.00)	−0.00 (0.00)
	NIE	−0.04 (0.00)	−0.00 (0.00)	−0.01 (0.00)	0.00 (0.00)	0.00 (0.00)
	TE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	−0.00 (0.00)
SP = SN = 0.85	NDE	0.05 (0.00)	0.00 (0.00)	0.01 (0.00)	0.00 (0.00)	−0.00 (0.00)
	NIE	−0.05 (0.00)	−0.00 (0.00)	−0.01 (0.00)	−0.00 (0.00)	0.00 (0.00)
	TE	0.00 (0.00)	−0.00 (0.00)	0.01 (0.00)	0.00 (0.00)	−0.00 (0.00)
SP = SN = 0.8	NDE	0.07 (0.00)	0.00 (0.00)	0.02 (0.00)	0.00 (0.00)	−0.00 (0.00)
	NIE	−0.07 (0.00)	−0.00 (0.00)	−0.01 (0.00)	−0.00 (0.00)	0.00 (0.00)
	TE	0.00 (0.00)	−0.00 (0.00)	0.01 (0.00)	0.00 (0.00)	−0.00 (0.00)
Int and $Y \sim$ Ber						
NDE = 3, NIE = 1.07						
SP = SN = 0.9	NDE	−0.11 (0.04)	0.02 (0.11)	−0.08 (0.05)	0.04 (0.18)	0.07 (0.10)
	NIE	−0.03 (0.00)	−0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	TE	−0.19 (0.05)	0.03 (0.14)	−0.09 (0.04)	0.05 (0.20)	0.08 (0.12)
SP = SN = 0.85	NDE	−0.14 (0.04)	0.01 (0.15)	−0.12 (0.06)	0.09 (0.20)	0.06 (0.15)
	NIE	−0.04 (0.00)	−0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	TE	−0.26 (0.04)	0.02 (0.17)	−0.12 (0.05)	0.10 (0.25)	0.05 (0.17)

(continued)

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

Table 1. *Continued.*

		Naive	EM	IRLS/PVW	IRLS/**t**PVW	IRLS/**s**PVW
SP = SN = 0.8	NDE	−0.18 (0.04)	0.05 (0.21)	−0.14 (0.04)	0.12 (0.35)	0.07 (0.20)
	NIE	−0.05 (0.00)	−0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	TE	−0.32 (0.04)	0.05 (0.24)	−0.12 (0.05)	0.11 (0.52)	0.05 (0.24)

Table 2. *Bias and variance of EM algorithm, and IRLS/PVW, IRLS/**s**PVW, IRLS/**t**PVW estimators of NDE and NIE when sample size is* $n = 50\,000$, *marginal probability of the true mediator is* 50%, *and the outcome is binary in the absence of exposure–mediator interaction* ($NDE = 2.7$, $NIE = 1.07$) *or in presence of exposure–mediator interaction* ($NDE = 3$, $NIE = 1.07$)

		EM	IRLS/PVW	IRLS/**s**PVW	IRLS/**t**PVW
No Int and $Y \sim$ Ber					
SP = SN = 0.9	NDE	0.01 (0.00)	−0.02 (0.00)	−0.00 (0.00)	0.01 (0.02)
	NIE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
SP = SN = 0.85	NDE	0.02 (0.00)	−0.03 (0.00)	0.01 (0.00)	0.02 (0.03)
	NIE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
SP = SN = 0.8	NDE	0.03 (0.00)	−0.04 (0.00)	−0.00 (0.01)	0.03 (0.03)
	NIE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Int and $Y \sim$ Ber					
SP = SN = 0.9	NDE	0.01 (0.02)	−0.15 (0.00)	−0.00 (0.01)	0.00 (0.04)
	NIE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
SP = SN = 0.85	NDE	0.01 (0.02)	−0.13 (0.00)	0.00 (0.03)	0.02 (0.06)
	NIE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
SP = SN = 0.8	NDE	0.03 (0.04)	−0.15 (0.00)	−0.00 (0.03)	0.01 (0.07)
	NIE	0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

2005). Due to the non-specificity of signs and symptoms, diagnosis of pre-eclampsia is typically subject to misclassification (Turner, 2010).

We, therefore, carried out mediation analysis to quantify the effects through and independent of pre-eclampsia. We used data from NCHS for 2003 ($N = 3\,918\,542$) derived from all birth certificates in the USA. For the purposes of this illustration, we used a subsample of 200 000 individuals due to computational time constraints. Pre-term birth was categorized according to the gestational age variable in the NCHS data derived from the last menstrual period. Pre-eclampsia was diagnosed according to blood pressure and protein in the urine and was potentially subject to misclassification. We tried to adjust for factors that may confound age–pre-term birth relationship, age–pre-eclampsia relationship, and pre-eclampsia–pre-term birth relationship including mother's ethnicity (categorized as White Caucasian, Black non-Hispanic, Hispanic, Asian, and Native American), marital status, as well as smoking status, drinking status, and whether the mother went to college. We also assumed the absence of pre-eclampsia and pre-term birth relationship confounders that are affected by age. We computed total effect and conditional direct and indirect effects of a white married woman, who smokes but does not drink, and has not attended college. Evaluating the effects at the mean level of the covariates yielded approximately the same results and we here report results of conditional effects only. Running a logistic regression of the outcome on the exposure and the covariates, we found that maternal age exerts a positive, significant total effect on pre-term birth ($\widehat{\text{OR}}^{\text{TE}} = 1.34$). Computing naive direct and indirect causal effects for such an individual

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

Table 3. *Results of the example. Naive and misclassification-corrected odds ratio conditional direct and indirect effect of young (age <35) versus old (age >= 35) mothers on pre-term birth, potentially mediated by pre-eclampsia status allowing for exposure–mediator interaction (SP = 0.99, SN = (0.8, 0.9, 0.95, 0.99), percentile bootstrap CI, r = 1000 bootstrap replications).*

	$\widehat{\text{NDE}}$	$\widehat{\text{NIE}}$	$\widehat{\text{PM}}$
Naive			
	1.336	1.001	2.6%
	(1.29,1.38)	(0.99,1.00)	(0.60,4.50)
IRLS/PVW			
SP = SN = 0.99	1.327	1.013	5.1%
	(1.27, 1.37)	(1.00, 1.04)	(0.41, 14.1)
SP = 0.99 and SN = 0.95	1.327	1.016	6.0%
	(1.28, 1.37)	(1.00, 1.03)	(−0.01 ,13.1)
SP = 0.99 and SN = 0.90	1.333	1.025	8.9%
	(1.28, 1.40)	(1.00, 1.05)	(0.00, 18.4)
SP = 0.99 and SN = 0.80	1.334	1.030	9.0%
	(1.28, 1.39)	(0.99, 1.11)	(−14.6, 32.3)
EM			
SP = SN = 0.99	1.335	1.007	2.6%
	(1.29, 1.41)	(1.00, 1.012)	(0.2, 4.8)
SP = 0.99 and SN = 0.95	1.333	1.007	3.0%
	(1.29, 1.41)	(1.00, 1.013)	(0.2, 5.0)
SP = 0.99 and SN = 0.90	1.333	1.008	3.0%
	(1.29, 1.41)	(1.00, 1.013)	(0.2, 5.2)
SP = 0.99 and SN = 0.80	1.330	1.010	3.5%
	(1.29, 1.41)	(1.01, 1.015)	(0.2, 6.0)

allowing for exposure–mediator interaction (Table 3), we found that maternal age exerts a positive, significant direct effect on pre-term birth through pathways independent of pre-eclampsia ($\widehat{\text{OR}}^{*\text{NDE}} = 1.33$, CI = (1.28, 1.38)). We observed a small, non-significant indirect causal effect of age on pre-term birth through pre-eclampsia ($\widehat{\text{OR}}^{*\text{NIE}} = 1.006$, CI = (0.999, 1.007)). A useful measure to quantify the proportion of the total causal effect of an exposure on an outcome that is explained by the hypothesized mechanism is the proportion mediated (PM). When direct and indirect effects are expressed in terms of odds ratios, $\text{PM} = \text{OR}^{\text{NDE}} \times (\text{OR}^{\text{NIE}} - 1)/(\text{OR}^{\text{NDE}} \times \text{OR}^{\text{NIE}} - 1)$ (VanderWeele and Vansteelandt, 2010). The naive analysis yielded $\widehat{\text{PM}}^{*} = 2.6\%$ (CI = (0.8%, 5.8%)), indicating that the effect of maternal age on pre-term birth is primarily through pathways independent of pre-eclampsia.

Aware that the naive analyses might be biased due to misclassification of the binary mediator, we implemented the IRLS/PVW, modeling the probability $P(M^* \mid Y, A, \boldsymbol{C})$ with a saturated logistic model (i.e. including all the possible interaction terms between Y, A, and $\boldsymbol{C}$ in the logistic model (4.2) described in Section 4.2) and the EM algorithm. We assumed non-differential misclassification. Pre-eclampsia information was obtained from medical records, reducing the risk of differential recall bias. Plausible range of values of SN and SP parameters were determined to satisfy the constraints $P(M^* = 1 \mid Y, A, \boldsymbol{C}) < \text{SN}$ and $P(M^* = 1 \mid Y, A, \boldsymbol{C}) > 1 - \text{SP}$ for all $(Y, A, \boldsymbol{C})$ combinations (Lyles and Lin, 2010). The maximum value that the estimated conditional probability $\widehat{P}(M^* = 1 \mid Y, A, \boldsymbol{C})$ took in the sample was 0.03, and therefore,

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

the minimum value that SP can take to satisfy the constraint is $1 - 0.03 = 0.97$. We chose to report sensitivity analyses assuming SP $= 0.99$ and SN $= (0.8, 0.9, 0.95, 0.99)$. In Table 3, we display the results of the analysis, accounting for the presence of exposure–mediator interaction. The misclassification-corrected analyses accounting for interaction revealed that the indirect effect of maternal age on pre-term birth mediated by pre-eclampsia status was underestimated in the naive analyses and became significant when accounting for misclassification. However, the indirect effect remained small relative to the direct effect due to the low prevalence of the mediator. The PM varied from 2.6% up to 9.0% depending on the values of SN and SP considered. Therefore, for all the values considered in the sensitivity analysis, the association between maternal age and pre-term birth was primarily through pathways other than pre-eclampsia status. Further analyses of this example are given in supplementary material available at *Biostatistics* online.

6. DISCUSSION

In this paper, we have studied the problem of misclassification of a binary mediator in the context of causal mediation analysis, when the outcome is either binary or continuous, allowing for the presence of exposure–mediator interaction. We demonstrated that when non-differential misclassification of a binary mediator is ignored in the analyses, the parametric estimators of direct and indirect causal effects can be severely biased. The theoretical and numerical results illustrate that the misclassification bias can take unintuitive directions in the presence of non-linearities.

We considered a hybrid of likelihood-based and PVW method as a possible strategy of correction for misclassification as well as an EM algorithm approach. We implemented the PVW approach with three different estimators for $P(M^* \mid Y, A, \boldsymbol{C})$, namely logistic regression (IRLS/PVW), logistic regression with splines (IRLS/**s**PVW), and using the true model for the conditional distribution that we seek to estimate (IRLS/**t**PVW). We compared the performance of corrected estimators for direct and indirect effects in a simulation study. The hybrid of likelihood-based and PVW performed best when logistic regression with splines (IRLS/**s**PVW) or the true model (IRLS/**t**PVW) were employed. In the simulation study, the IRLS/**s**PVW and IRLS/**t**PVW estimators and the EM algorithm approach appeared approximately consistent in all cases. However, with only moderate sample sizes, we expect some finite sample bias occurring for all the approaches proposed when the outcome is binary. The proposed approaches, which maximize weighted score equations, share theoretical similarities. The EM algorithm, by modeling the joint distribution of the outcome and latent intermediate, arises naturally from a regression-based mediation analysis and is found to be more efficient. The IRLS/PVW approach, by separating the estimation of the outcome and mediator regression parameters, is slightly computationally faster. In the simulation study, the IRLS/PVW approach gave results in about 10 s, while the EM algorithm converged in approximately 10 min. In the example section, for a sample size of 200 000 individuals, IRLS/PVW completed one bootstrap iteration in about 4 min, while the EM algorithm converged in about 30 min. Moreover, in cases in which the assumption of non-differential misclassification is unlikely to hold the hybrid approach has been extended to relax this assumption (Lyles and Lin, 2010). However, a very important disadvantage of IRLS/PVW is that the model for the conditional probability for the observed mediator given the outcome, the exposure, and covariates could be incompatible with the models for the mediator and the outcome. We proposed two approaches that alleviate the problem: either deriving the true model or using spline models. The spline model is more intuitive and easy to apply, however, smoothing may not work well when the number of covariates is large relative to the sample size and therefore the IRLS/**t**PVW approach should be preferred in this case. Importantly, the EM algorithm does not suffer from the issue of model incompatibility and is therefore the preferred approach. When the sample size is very large and also large with respect to the number of covariates, the IRLS/**s**PVW using splines or IRLS/**t**PVW can be also considered, being slightly more computationally efficient approaches.

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

In many instances, auxiliary information on the mismeasured intermediate is not available in mediation studies and sensitivity analysis must be employed. We illustrated in a real data example the correction strategy coupled with sensitivity analysis for the unknown SN and SP for which no validation data or replicates for the mismeasured mediator are needed. Although the correction strategy using sensitivity analysis does not require validation data or replicates for the mismeasured mediator, estimators for correction could make use of this information, if available. Some possible extensions of our study should be mentioned. Derivation of closed form bias formulae of direct and indirect effects when the outcome is binary in the presence of exposure–mediator interaction is of interest. We made the strong assumption that the misclassification mechanism was non-differential with respect to the outcome, exposure, and covariates; moreover, we assumed that these variables were measured without error. It would be of interest to study the bias of naive direct and indirect causal effects when misclassification of a binary mediator is differential and relaxing the assumption that the other variables are measured without error. Finally, more work is needed to improve the misclassification-corrected estimators for direct and indirect effects with binary outcome when the sample size is small.

7. Software

Software in the form of R code is available on request from the corresponding author.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

We thank the Editor, the Associate Editor, and two Reviewers for helpful comments. *Conflict of Interest*: None declared.

Funding

The research was funded by National Institute of Health grant ES017876.

References

Aigner, D. J. (1973). Regression with a binary independent variable subject to errors of observation. *International Journal of Econometrics* **1**, 49–60.

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.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Non-linear Models.* Boca Raton, FL: Chapman & Hall/CRC.

Goldenberg R. L., Culhane, J. F., Iams, J. D. and Romero, R. (2008). Epidemiology and causes of preterm birth. *Lancet* **371**(9606), 75–84.

Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology*. Boca Raton, FL: Chapman & Hall/CRC.

Hoyle, R. H. and Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In: Hoyle, R H. (editor), *Statistical Strategies for Small Sample Research*. Thousand Oaks, CA: Sage.

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association*, **85**, 765–769.

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021

IMAI, K., KEELE, L. AND TINGLEY, D. (2010). A general approach to causal mediation analysis. *Psychological Methods* **15**, 309–334.

JACOBSSON B., LADFORS L. AND MILSON I. (2004). Advanced maternal age and adverse perinatal outcome. *Obstetrics & Gynecology* **104**, 727–33.

LAMMINPAA, R., VEHVILAINEN-JULKUNEN, K., GISSLER, M. AND HEINONEN, S. (2012). Preeclampsia complicated by advanced maternal age: a registry-based study on primiparous women in Finland 1997–2008. *BMC Pregnancy Childbirth* **12**(1), 47.

LE CESSIE, S., DEBEIJ, J., ROSENDAAL, F. R., CANNEGIETER, S. C. AND VANDENBROUCKE, J. P. (2012). Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology* **23**, 551–560.

LYLES, R. H. AND LIN, J. (2010). Sensitivity analysis for misclassification in logistic regression via likelihood methods and predictive value weighting. *Statistics in Medicine* **29**(22), 2297–309.

MACKINNON, D. P. (2008). *Introduction to Statistical Mediation Analysis*. New York: Erlbaum.

MCCORMICK, M. C. (1985). The contribution of low birth weight to infant mortality and childhood morbidity. *The New England Journal of Medicine* **312**, 82–90.

NEUHAUS, J. M. (1999). Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika* **86**(4), 843–855.

OGBURN, E. L. AND VANDERWEELE, T. J. (2012). Analytic results on the bias due to nondifferential misclassification of a binary mediator. *American Journal of Epidemiology* **176**, 555–561.

PEARL, J. (2001). Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, pp. 411–420.

ROBINS, J. M. AND GREENLAND S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–155.

ROBINS, J. M. AND RICHARDSON, T. S. (2010). Alternative graphical causal models and the identification of direct effects. In: Shrout, P. (editor), *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*. Oxford: Oxford University Press, to appear.

TURNER, J. A. (2010). Diagnosis and management of pre-eclampsia: an update. *International Journal of Women's Health* **2**(1), 327–337.

VALERI, L., LIN, X. AND VANDERWEELE, T. J. (2014). Mediation analysis in generalized linear models when the mediator is measured with error. *Technical Report*.

VALERI, L. AND VANDERWEELE, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, doi:10.1037/a0031034.

VANDERWEELE, T. J., VALERI, L. AND OGBURN, E. L. (2012). The role of measurement error and misclassification in mediation analysis: mediation and measurement error. *Epidemiology* **23**, 561–564.

VANDERWEELE, T. J. AND VANSTEELANDT S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface* **2**(4), 457–468.

VANDERWEELE, T. J. AND VANSTEELANDT S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology* **172**(12), 1339–1348.

WORLD HEALTH ORGANIZATION. (2005). World health report 2005: make every mother and child count. Geneva: WHO. p. 63.

[*Received March 14, 2013; revised February 14, 2014; accepted for publication February 16, 2014*]

Downloaded from https://academic.oup.com/biostatistics/article/15/3/498/224455 by guest on 26 October 2021