Methodological obstacles in causal inference: confounding, missing data, and measurement error

Penning de Vries, B.B.L.

8

A weighting method for simultaneous adjustment for confounding and joint exposure-outcome misclassifications

Bas B. L. Penning de Vries
Maarten van Smeden
Rolf H. H. Groenwold

Statistical Methods in Medical Research 2021; 30(2): 473–487

Abstract

Joint misclassification of exposure and outcome variables can lead to considerable bias in epidemiological studies of causal exposure-outcome effects. In this paper, we present a new maximum likelihood based estimator for marginal causal effects that simultaneously adjusts for confounding and several forms of joint misclassification of the exposure and outcome variables. The proposed method relies on validation data for the construction of weights that account for both sources of bias. The weighting estimator, which is an extension of the outcome misclassification weighting estimator proposed by Gravel and Platt (Statistics in Medicine, 2018), is applied to reinfarction data. Simulation studies were carried out to study its finite sample properties and compare it with methods that do not account for confounding or misclassification. The new estimator showed favourable large sample properties in the simulations. Further research is needed to study the sensitivity of the proposed method and that of alternatives to violations of their assumptions. The implementation of the estimator is facilitated by a new R function (`ipwm`) in an existing R package (`mecor`).

8.1 Introduction

In epidemiological research on causal associations between a particular exposure and a certain outcome, erroneous information on either or both of these variables poses a serious methodological obstacle in making valid inferences. In particular, joint misclassification of exposure and outcome can lead to considerable bias of standard causal effect estimators, with direction and magnitude depending on various factors, including the misclassification mechanism and the direction and magnitude of the true effect (Kristensen, 1992; Brenner et al., 1993; Vogel et al., 2005; Jurek et al., 2008; VanderWeele and Hernán, 2012; Brooks et al., 2018).

Exposure and outcome misclassification is typically categorised according to two separate properties: whether or not the misclassification is differential and whether or not it is dependent relative to some covariate vector L containing patient characteristics (Kristensen, 1992; VanderWeele and Hernán, 2012). Joint misclassification of exposure and outcome is said to be *nondifferential* if (1) the sensitivity and specificity of exposure classification are constant across all categories of the (true) outcome given L and (2) the sensitivity and specificity of outcome classification are constant across all categories of the (true) exposure given L; otherwise it is *differential*. Misclassification is said to be *independent* if the joint probability of any exposure and outcome classification given any true exposure and outcome categories and L can be factored into the product of the corresponding probabilities for exposure and outcome separately; otherwise, it is *dependent*. In Dawid's notation (1979), that is, if true exposure level A and true outcome Y are (potentially mis)classified as B and Z, respectively, misclassification is nondifferential if and only if $B \perp\!\!\!\perp Y|A, L$ and $Z \perp\!\!\!\perp A|Y, L$ and independent if and only if $Z \perp\!\!\!\perp B|Y, A, L$.

Epidemiological research hampered by joint misclassification of some type is likely voluminous (Brooks et al., 2018). Examples of studies affected by exposure and outcome misclassification can be found, for example, in the literature on the causal effects of drug use, which is largely based on routinely collected data, where exposures are typically operationalised on the basis of prescription records and where outcomes are often self-reported (Marcum et al., 2013; Culver et al., 2012; Leong et al., 2013; Ni et al., 2017). In applied epidemiological research, misclassification or some of its potential consequences are often ignored (Jurek et al., 2006; Brakenhoff et al., 2018). The assertion often made in the discussion of study results that observed measures of association are biased toward the null under nondifferentiality, for example, is not generally true unless additional conditions are presupposed (Brenner et al., 1993; Brooks et al., 2018).

Methods to adjust for misclassification rely on additional information that

can be used to estimate or correct for bias. One potential source of information is validation data obtained through supposedly infallible measurement. Recently, Gravel and Platt (2018) proposed an inverse probability weighting (IPW) method to simultaneously address confounding and outcome misclassification by means of internal validation data. Other methods likewise suppose that either the exposure or the outcome is subject to misclassification (Babanezhad et al., 2010; Braun et al., 2017; Gravel and Platt, 2018; Shu and Yi, 2019). In what follows, we propose an extension of Gravel and Platt's method to allow for confounding adjustment and joint exposure and outcome misclassification. This flexible estimator allows for the misclassifications to be dependent, differential or both. In Section 8.2, inverse probability weights for confounding and joint misclassification are introduced through a hypothetical study based on the illustrative example of Gravel and Platt. Section 8.3 details methods for estimation of the various components of the proposed weights based on validation data. In Section 8.4, we describe a series of Monte Carlo simulations that were used to study properties of the proposed method in finite samples. We conclude with a summary and discussion of our findings in context of the existing literature.

8.2 Data distribution for illustration and development of weighting method

We first consider the data and setting described by Gravel and Platt and suppose that Table 8.1 represents a simple random (i.i.d.) sample from (or that its cell counts are proportional to the respective densities in) the population of interest. This illustration is based on a cohort study on the association between post-myocardial infarction statin use (A) and the 1-year risk of reinfarction (Y). In what follows, we will refer to this example as the 'reinfarction example'.

Throughout we take the counterfactual framework for causal inference, formal accounts of which are given for example by Neyman, Rubin, Holland, and Pearl (Neyman et al., 1935; Rubin, 1974; Holland, 1986, 1988; Pearl, 2009). The interest, we suppose, lies in estimating $g(\mathbb{E}[Y(0)], \mathbb{E}[Y(1)])$ for some function g, where $Y(0)$ and $Y(1)$ denote the counterfactual outcomes for hypothetical interventions setting A to 0 and 1, respectively. Common choices of g define $g(p_0, p_1) = p_1 - p_0$ (risk difference), $g(p_0, p_1) = p_1/p_0$ (risk ratio) or $g(p_0, p_1) = [p_1/(1-p_1)]/[p_0/(1-p_0)]$ (odds ratio). For our numerical example and simulation studies, we concentrate on the causal marginal odds ratio (OR) in particular, with

$$\text{OR} = g(\mathbb{E}[Y(0)], \mathbb{E}[Y(1)]) = \frac{\mathbb{E}[Y(1)]/(1 - \mathbb{E}[Y(1)])}{\mathbb{E}[Y(0)]/(1 - \mathbb{E}[Y(0)])}, \tag{8.1}$$

but the results naturally extend to other effect measures.

8.2.1 No misclassification

Under conditional exchangeability given L (i.e., $(Y(0), Y(1)) \perp\!\!\!\perp A|L$), consistency ($Y(a) = Y$ if $A = a$) and positivity ($\Pr(A = a|L = l) > 0$ for $a = 0, 1$ and all l in the support of L), the mean counterfactuals $E[Y(0)]$ and $E[Y(1)]$ can be expressed in terms of 'observables' (meaning, here, variables that would be observed in the absence of measurement error) as follows:

$$\mathbb{E}[Y(0)] = \mathbb{E}[WY|A = 0] \text{ and } \mathbb{E}[Y(1)] = \mathbb{E}[WY|A = 1],$$

where W denotes the inverse probability of the allocated exposure level A given L multiplied by the prevalence of the allocated exposure level A (i.e., $W = \Pr(A)/\Pr(A|L)$; Supplementary Appendix S8.1). We therefore have

$$g(\mathbb{E}[Y(0)], \mathbb{E}[Y(1)]) = g(\mathbb{E}[WY|A = 0], \mathbb{E}[WY|A = 1]). \tag{8.2}$$

Replacing components of the right-hand side of (8.2) with sample analogues, we obtain the following estimator for the setting where L is binary:

$$\begin{aligned}\widehat{\text{OR}} &:= g(\widehat{\mathbb{E}}[\widehat{W}Y|A = 0], \widehat{\mathbb{E}}[\widehat{W}Y|A = 1]) \\ &= \frac{\widehat{\mathbb{E}}[\widehat{W}Y|A = 1]/(1 - \widehat{\mathbb{E}}[\widehat{W}Y|A = 1])}{\widehat{\mathbb{E}}[\widehat{W}Y|A = 0]/(1 - \widehat{\mathbb{E}}[\widehat{W}Y|A = 0])} \\ &= \frac{(\widehat{W}_{10}n_{110} + \widehat{W}_{11}n_{111})/(n_{110} + n_{111} + n_{010} + n_{011} - \widehat{W}_{10}n_{110} - \widehat{W}_{11}n_{111})}{(\widehat{W}_{00}n_{100} + \widehat{W}_{01}n_{101})/(n_{100} + n_{101} + n_{000} + n_{001} - \widehat{W}_{00}n_{100} - \widehat{W}_{01}n_{101})},\end{aligned} \tag{8.3}$$

where n_{yal} denotes the number of subjects with $Y = y$, $A = a$, $L = l$ and where $\widehat{W}_{al}$ is the product of the proportion of subjects in the sample with $A = a$ and

Table 8.1: Cross-classification of the reinfarction data for 33,007 individuals as given by Gravel and Platt (2018).

	$L = 0$		$L = 1$	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0$	11602	13116	1302	5363
$Y = 1$	890	589	49	96

the inverse of the proportion of subjects with $A = a$ among those with $L = l$. For the data in Table 8.1, we obtain $\widehat{\text{OR}} \approx 0.573$. The corresponding crude odds ratio (i.e., with $\widehat{W} = 1$) is 0.509.

8.2.2 Joint misclassification

Suppose that rather than observing Y and A we observe Z and B, the misclassified versions of Y and A, respectively. The relation between Z and B on the one hand and Y, A and L on the other can be expressed as follows:

$$\begin{aligned}&\Pr(Z = z, B = b|Y = y, A = a, L = l)\\ &\quad = (\pi_{byal})^z(1 - \pi_{byal})^{1-z}(\lambda_{yal})^b(1 - \lambda_{yal})^{1-b}\end{aligned}$$

for $z, b \in \{0, 1\}$ and all possible realisations y, a, l of Y, A, L, and where $\pi_{byal} = \Pr(Z = 1|B = b, Y = y, A = a, L = l)$ and $\lambda_{yal} = \Pr(B = 1|Y = y, A = a, L = l)$.

To simulate (dependent differential) misclassification in the reinfarction dataset, we use the true positive and false positive rates given in Table 8.2. The expected cell counts for these rates are given in Table 8.3.

We redefine the weights in (8.2) as a function of B and L (as per Supplementary Appendix S8.1) such that

$$W = \frac{p(B)\varepsilon_{BL}}{\sum_y \sum_a \pi_{ByaL}(\lambda_{yaL})^B(1 - \lambda_{yaL})^{1-B}(\varepsilon_{aL})^y(1 - \varepsilon_{aL})^{1-y}(\delta_L)^a(1 - \delta_L)^{1-a}}, \tag{8.4}$$

where $p(B)$ is the prevalence of level B of the potentially misclassified version of the exposure variable and where $\varepsilon_{al} = \Pr(Y = 1|A = a, L = l)$ and $\delta_l =$

Table 8.2: True and false positive rates for reinfarction example. For $b, y, a, l \in \{0, 1\}$, $\lambda_{yal} = \Pr(B = 1|Y = y, A = a, L = l)$ and $\pi_{byal} = \Pr(Z = 1|B = b, Y = y, A = a, L = l)$.

$\pi_{0000} = 0.050$	$\pi_{0001} = 0.020$	$\lambda_{000} = 0.010$
$\pi_{1000} = 0.060$	$\pi_{1001} = 0.108$	$\lambda_{100} = 0.181$
$\pi_{0100} = 0.930$	$\pi_{0101} = 0.806$	$\lambda_{010} = 0.880$
$\pi_{1100} = 0.938$	$\pi_{1101} = 0.692$	$\lambda_{110} = 0.910$
$\pi_{0010} = 0.030$	$\pi_{0011} = 0.109$	$\lambda_{001} = 0.100$
$\pi_{1010} = 0.060$	$\pi_{1011} = 0.050$	$\lambda_{101} = 0.265$
$\pi_{0110} = 0.906$	$\pi_{0111} = 0.765$	$\lambda_{011} = 0.930$
$\pi_{1110} = 0.950$	$\pi_{1111} = 0.861$	$\lambda_{111} = 0.823$

$\Pr(A = 1|L = l)$ for all possible realisations a and l of A and L, respectively. In Supplementary Appendix S8.1, it is shown that

$$\mathbb{E}[Y(0)] = \mathbb{E}[WZ|B = 0] \quad \text{and} \quad \mathbb{E}[Y(1)] = \mathbb{E}[WZ|B = 1], \tag{8.5}$$

which suggests the plug-in estimator

$$\begin{aligned} \widehat{\text{OR}} &:= g(\widehat{\mathbb{E}}[\widehat{W}Z|B = 0], \widehat{\mathbb{E}}[\widehat{W}Z|B = 1]) \\ &= \frac{\widehat{\mathbb{E}}[\widehat{W}Z|B = 1]/(1 - \widehat{\mathbb{E}}[\widehat{W}Z|B = 1])}{\widehat{\mathbb{E}}[\widehat{W}Z|B = 0]/(1 - \widehat{\mathbb{E}}[\widehat{W}Z|B = 0])}, \end{aligned} \tag{8.6}$$

where $\widehat{\mathbb{E}}$ denotes the sample mean operator and $\widehat{W}$ the sample analogue (i.e., consistent estimator) of W in (8.4). For other effect measures (i.e., other choices of g), the same plug-in strategy can be implemented.

In the absence of exposure misclassification, (8.4) reduces to

$$W = \left(\frac{(\delta_L)^A (1 - \delta_L)^{1-A}}{p(A)} \left[\pi_{A0AL} \frac{1 - \varepsilon_{AL}}{\varepsilon_{AL}} + \pi_{A1AL} \right] \right)^{-1}. \tag{8.7}$$

The first term within the round brackets corrects for confounding and represents the propensity of the received exposure level A divided by prevalence of exposure level A. The term within square brackets is a factor that corrects for misclassification in the outcome variable. This correction factor is similar to

Table 8.3: Expected cell counts (rounded to integers) for reinfarction example after misclassification was introduced. Because of rounding, the sum of all cell entries is 33,006 rather than 33,007, the size of the reinfarction dataset.

	$L = 0$		$L = 1$	
	$A = 0$	$A = 1$	$A = 0$	$A = 1$
$Y = 0,\ A = 0,\ L = 0$	10912	109	574	7
$Y = 1,\ A = 0,\ L = 0$	51	10	678	151
$Y = 0,\ A = 1,\ L = 0$	1527	10850	47	693
$Y = 1,\ A = 1,\ L = 0$	5	27	48	509
$Y = 0,\ A = 0,\ L = 1$	1148	116	23	14
$Y = 1,\ A = 0,\ L = 1$	7	4	29	9
$Y = 0,\ A = 1,\ L = 1$	334	4738	41	249
$Y = 1,\ A = 1,\ L = 1$	4	11	13	68

that proposed by Gravel and Platt (2018). The only difference is that where in (8.7) it does not depend on the fallible measurement Z of Y, Gravel and Platt define different weights for subjects with $Z = 0$. Note, however, that the choice of weights for subjects with $Z = 0$ does not affect the population quantity in (8.5) or the estimator defined by (8.6), because the weights only appear in products with Z, which equal zero if $Z = 0$.

As for the reinfarction example, the odds ratio estimate for the exposure-outcome effect based on inverse probability weighting that assumes absence of exposure or outcome misclassification is 1.120, while the corresponding misclassification naive crude odds ratio is 1.031. Estimation of the population weights W from observables using validation data is discussed in the next section. As shown below, weighting using the proposed weights that account for confounding and outcome and exposure misclassification results in an odds ratio of $\mathrm{OR} = \widehat{\mathrm{OR}} \approx 0.573$. Inference based on (8.7) rather than (8.4), i.e., using Gravel and Platt's method and ignoring misclassification in the exposure but correcting for outcome misclassification, yields an odds ratio estimate of 0.934.

8.2.3 Parameterisation based on positive and negative predictive values

In the foregoing discussion, the proposed weights were expressed in terms of sensitivity and specificity parameters. The sensitivity and specificity of Z with respect to Y, given (B, A, L), are π_{B1AL} and $1 - \pi_{B0AL}$, respectively. Similarly, λ_{Y1L} and $1-\lambda_{Y0L}$ reflect the sensitivity and specificity, respectively, with respect to A, conditional on Y and L.

As discussed below, it may be more convenient to choose a parameterisation that is based on (positive and negative) predictive values. Define $\delta_l^* = \Pr(B = 1|L = l)$, $\varepsilon_{bl}^* = \Pr(Z = 1|B = b, L = l)$, $\lambda_{zbl}^* = \Pr(A = 1|Z = z, B = b, L = l)$ and $\pi_{azbl}^* = \Pr(Y = 1|A = a, Z = z, B = b, L = l)$. The weights in (8.4) can be rewritten as

$$W = \frac{\sum_y \sum_a \pi^*_{ByaL}(\lambda^*_{yaL})^B(1-\lambda^*_{yaL})^{1-B}(\varepsilon^*_{aL})^y(1-\varepsilon^*_{aL})^{1-y}(\delta^*_L)^a(1-\delta^*_L)^{1-a}}{\sum_y \sum_a (\lambda^*_{yaL})^B(1-\lambda^*_{yaL})^{1-B}(\varepsilon^*_{aL})^y(1-\varepsilon^*_{aL})^{1-y}(\delta^*_L)^a(1-\delta^*_L)^{1-a}} \times \frac{p(B)}{\varepsilon^*_{BL}(\delta^*_L)^B(1-\delta^*_L)^{1-B}}. \tag{8.8}$$

In the absence of exposure misclassification, these weights simplify to

$$W = \frac{p(A)}{(\delta_L)^A(1-\delta_L)^{1-A}} \frac{\varepsilon_{AL}}{\varepsilon^*_{AL}}.$$

8.3 Estimation of weights based on validation data

Estimation of the proposed weights can be done using a number of approaches and we will here consider a maximum likelihood approach that assumes the availability of internal validation data, i.e., that some study participants have their observed exposure or outcome measured by an 'infallible' or 'gold standard' (100% accurate) classifier, and that all participants have the misclassified exposure and outcome variables measured.

8.3.1 Validation subset inclusion mechanism

Let R_Y be the indicator variable that takes the value of 1 if the outcome is observed (i.e., measured by an infallible classifier) and 0 otherwise. Similarly, define R_A to be the indicator variable that takes the value of 1 if the exposure variable is observed and 0 otherwise. R_Y and R_A reflect which subjects have validation data available on Y and A, respectively. The subset of subjects with validation data on Y need not fully overlap with the subset with validation data on A.

The validation subsets can be approached from the missing data framework of Rubin (1976) Provided that Z, B, L are free of missing values, Rubin's missing at random (MAR) condition is met if the vector (R_Y, R_A) is conditionally independent of (Y, A) given (Z, B, L).

8.3.2 Full likelihood approach based on parameterisation in terms of sensitivities and specificities

Simultaneous estimation of the whole vector of δ, ε, λ and π parameters can be done via maximum likelihood estimation as follows. Assuming i.i.d. observations $(Z_i, B_i, Y_i, A_i, L_i)$ and ignorable missingness in the sense of Rubin (1976) (MAR and distinctness), for valid likelihood-based inference it is appropriate to maximise the following log-likelihood over the parameter space of θ, the vector of δ, ε, λ and π parameters:

$$\begin{aligned}\ell(\theta) = &\sum_{i:R_{Yi}=R_{Ai}=1} \log f(\theta; Z_i, B_i, Y_i, A_i, L_i) \\ &+ \sum_{i:R_{Yi}=1\wedge R_{Ai}=0} \log \sum_{A_i} f(\theta; Z_i, B_i, Y_i, A_i, L_i) \\ &+ \sum_{i:R_{Yi}=0\wedge R_{Ai}=1} \log \sum_{Y_i} f(\theta; Z_i, B_i, Y_i, A_i, L_i)\end{aligned}$$

$$+ \sum_{i:R_{Yi}=R_{Ai}=0} \log \sum_{Y_i} \sum_{A_i} f(\theta; Z_i, B_i, Y_i, A_i, L_i),$$

where

$$\begin{aligned} & f(\theta; Z_i, B_i, Y_i, A_i, L_i) \\ & \quad = (\pi_{B_i Y_i A_i L_i})^{Z_i} (1 - \pi_{B_i Y_i A_i L_i})^{1-Z_i} (\lambda_{Y_i A_i L_i})^{B_i} (1 - \lambda_{Y_i A_i L_i})^{1-B_i} \\ & \qquad \times (\varepsilon_{A_i L_i})^{Y_i} (1 - \varepsilon_{A_i L_i})^{1-Y_i} (\delta_{L_i})^{A_i} (1 - \delta_{L_i})^{1-A_i}. \end{aligned}$$

Evaluating this log-likelihood involves marginalising over unobserved quantities in the last three terms of $\ell(\theta)$. The log-likelihood equations may become considerably more tractable if we choose a parameterisation of the likelihood that is based on predictive values rather than sensitivities and specificities.

8.3.3 Full likelihood approach based on parameterisation in terms of predictive values

Inference may alternatively be based on a log-likelihood that is parameterised in terms of the vector θ^* of the δ^*, ε^*, λ^* and π^* parameters, i.e.,

$$\begin{aligned} \ell^*(\theta^*) = & \sum_{i:R_{Yi}=R_{Ai}=1} \log h(\theta^*; Z_i, B_i, Y_i, A_i, L_i) \\ & + \sum_{i:R_{Yi}=1 \wedge R_{Ai}=0} \log \sum_{A_i} h(\theta^*; Z_i, B_i, Y_i, A_i, L_i) \\ & + \sum_{i:R_{Yi}=0 \wedge R_{Ai}=1} \log \sum_{Y_i} h(\theta^*; Z_i, B_i, Y_i, A_i, L_i) \\ & + \sum_{i:R_{Yi}=R_{Ai}=0} \log \sum_{Y_i} \sum_{A_i} h(\theta^*; Z_i, B_i, Y_i, A_i, L_i), \end{aligned}$$

where

$$\begin{aligned} & h(\theta^*; Z_i, B_i, Y_i, A_i, L_i) \\ & \quad = (\pi^*_{A_i Z_i B_i L_i})^{Y_i} (1 - \pi^*_{A_i Z_i B_i L_i})^{1-Y_i} (\lambda^*_{Z_i B_i L_i})^{A_i} (1 - \lambda^*_{Z_i B_i L_i})^{1-A_i} \\ & \qquad \times (\varepsilon^*_{B_i L_i})^{Z_i} (1 - \varepsilon^*_{B_i L_i})^{1-Z_i} (\delta^*_{L_i})^{B_i} (1 - \delta^*_{L_i})^{1-B_i}. \end{aligned}$$

If validation data is available on Y if and only if it is available on A, the complete data log-likelihood ignoring the missing data mechanism can be conveniently expressed as follows:

$$\ell^*(\theta^*) = \ell^*_1(\theta^*) + \ell^*_2(\theta^*) + \ell^*_3(\theta^*) + \ell^*_4(\theta^*), \tag{8.9}$$

with θ^* denoting the vector of δ^*, ε^*, λ^* and π^* parameters and where

$$\begin{aligned}
\ell_1^*(\theta^*) &= \sum_{i:R_{Yi}=R_{Ai}=1} Y_i \log(\pi^*_{A_iZ_iB_iL_i}) + (1-Y_i)\log(1-\pi^*_{A_iZ_iB_iL_i}),\\
\ell_2^*(\theta^*) &= \sum_{i:R_{Yi}=R_{Ai}=1} A_i \log(\lambda^*_{Z_iB_iL_i}) + (1-A_i)\log(1-\lambda^*_{Z_iB_iL_i}),\\
\ell_3^*(\theta^*) &= \sum_i Z_i \log(\varepsilon^*_{B_iL_i}) + (1-Z_i)\log(1-\varepsilon^*_{B_iL_i}),\\
\ell_4^*(\theta^*) &= \sum_i B_i \log(\delta^*_{L_i}) + (1-B_i)\log(1-\delta^*_{L_i}).
\end{aligned}$$

Now, assuming distinct parameter spaces for the vectors of π^*, λ^*, ε^*, and δ^* parameters, the parameter values that maximise $\ell^*(\theta^*)$ can be found by separately maximising $\ell_1^*(\theta^*)$ and $\ell_2^*(\theta^*)$ in the validation subset with respect to the π^* and λ^* parameters, respectively, and $\ell_3^*(\theta^*)$ and $\ell_4^*(\theta^*)$ in the entire dataset with respect to ε^* and δ^*. Following Gravel and Platt (2018) and Tang et al. (2013), the sum of the first and last two terms are therefore suitably labelled the internal validation and main study log-likelihood, respectively. With this parameterisation, finding the maximum likelihood estimates is readily achieved by taking advantage of standard statistical software.

8.3.4 Equivalence of likelihood approaches based on different parameterisations

Without restrictions imposed on

$$\begin{aligned}
\theta_l &:= (\pi_{000l}, \pi_{100l}, \pi_{010l}, \pi_{110l}, \pi_{001l}, \pi_{101l}, \pi_{011l}, \pi_{111l}, \lambda_{00l}, \lambda_{10l}, \lambda_{01l}, \lambda_{11l},\\
&\qquad \varepsilon_{0l}, \varepsilon_{1l}, \delta_l) \text{ and}\\
\theta_l^* &:= (\pi^*_{000l}, \pi^*_{100l}, \pi^*_{010l}, \pi^*_{110l}, \pi^*_{001l}, \pi^*_{101l}, \pi^*_{011l}, \pi^*_{111l}, \lambda^*_{00l}, \lambda^*_{10l}, \lambda^*_{01l}, \lambda^*_{11l},\\
&\qquad \varepsilon^*_{0l}, \varepsilon^*_{1l}, \delta^*_l),
\end{aligned}$$

other than that $\theta_l, \theta_l^* \in (0,1)^{15}$, it can be shown that the maximum likelihood estimator based on the internal validation design is invariant to its parameterisation (sensitivities/specificities versus positive and negative predictive values). This is because there exists a function mapping every $\theta_l \in (0,1)^{15}$ to a unique $\theta_l^* \in (0,1)^{15}$ and vice versa. Maximising $\ell(\theta)$ with respect to θ is then equivalent to maximising $\ell(\sigma(\theta^*))$ $(= \ell^*(\theta^*))$ with respect to θ^* for some bijection σ such that $\theta = \sigma(\theta^*)$; that is,

$$\arg\max_{\theta} \ell(\theta) = \sigma\left(\arg\max_{\theta^*} \ell(\sigma(\theta^*))\right).$$

If more restrictions are imposed on θ or θ^*, e.g., if we assume non-saturated logistic models for the components of θ and θ^*, this equivalence no longer holds and the resulting weight estimates may differ depending on the parameterisation.

8.3.5 Application

For the re-infarction data example, we assume validation data are available according to a MAR mechanism characterised by

$$\begin{aligned}\Pr(R_Y = 1 | R_A = s, Z = z, B = b, Y = y, A = a, L = l) &= s,\\ \Pr(R_A = 1 | Z = z, B = b, Y = y, A = a, L = l) &= 0.25 + 0.10b.\end{aligned}$$

This mechanism assigns validation data to an individual on either both Y and A (30% of all individuals) or neither depending on their realisation of B, the misclassified version of the exposure variable A (Supplementary Table S8.1). Supplementary Tables S8.2 and S8.3 give the likelihood contributions for the parameterisation based on predictive values and the closed form maximum likelihood expressions, respectively. Maximum likelihood estimates can also be found by fitting to the data the saturated logistic regression models of B and Z on L and (B, L), respectively, and to the validation subset the fully saturated logistic regression models of A and Y on (Z, B, L) and (A, Z, B, L), respectively. Estimated weights are then obtained by plugging in the maximum likelihood estimates into (8.8). As in the complete data setting where we assumed the weights to be known, evaluating (8.6) then yields an odds ratio of $\widehat{\text{OR}} = \text{OR} \approx 0.573$.

8.4 Simulations

We performed a series of Monte Carlo simulation experiments to illustrate the implementation of the proposed method, to study its finite sample properties and to compare the method to estimators that ignore the presence of confounding or joint exposure and outcome misclassification. All simulations were conducted using R-3.5.0 ((R Core Team, 2018)) on x86_64-pc-linux-gnu platforms of the high performance computer cluster of Leiden University Medical Center.

8.4.1 Methods

For all 54 simulation experiments, we generated $n_{\text{sim}} = 1000$ samples of size n according to the data generating mechanisms depicted in the directed acyclic graphs of Figure 8.1. This multi-step data generating process included generating

values on measurement error-free variables, introducing misclassification and allocating individuals validation data. We applied various estimators to each of the simulation samples to yield, for each scenario, an empirical distribution of each point estimator and corresponding precision estimators. These distributions were then summarised into various performance metrics. These metrics include the empirical bias of the estimator on the log-scale (i.e., the mean estimated log-OR minus the target log-OR across the n_{sim} samples), the empirical standard error (SE) of the estimator on the log-scale (i.e., the square root of the mean squared deviation of the estimated log-OR from the mean log-OR), the empirical mean squared error (MSE) (i.e., the sum of the squared SE and the squared bias), the square root of the mean estimated variance (SSE, sample standard error) and the empirical coverage probability (CP) (i.e., the fraction of simulation runs per scenario where the 95% confidence interval (95%CI) contained the target quantity).

1 Distribution of measurement error-free variables

Following Gravel and Platt (2018), we consider a setting based on that of "Scenario A" in the work of Setoguchi et al. (2008) with slight modifications to the propensity score and outcome models. We consider a fully observed covariate vector $L = (L_0, ..., L_{10})$ whose distribution coincides with that of $h(V)$, where $V = (V_1, ..., V_{10})$ has the multivariate normal distribution with zero means, unit variances and correlations equal to zero except for the correlations between W_1 and V_5, V_2 and V_6, V_3 and V_8, and V_4 and V_9, which were set to 0.2, 0.9, 0.2, and 0.9, respectively. Function h was defined such that

$$h(V) = (I(V_1 > 0), V_2, I(V_3 > 0), V_4, I(V_5 > 0), I(V_6 > 0), V_7, I(V_8 > 0), \\ I(V_9 > 0), V_{10}).$$

Thus, sampling from the distribution of L is equivalent to sampling from the multivariate normal distribution with the given parameter values and dichotomising the 1st, 3rd, 5th, 6th, 8th and 9th elements.

Next, let U_1 and U_2 be binary variables distributed according to the following logistic models:

$$\text{logit}\,\Pr(U_1 = 1|L) = \eta_0, \tag{8.10}$$

$$\text{logit}\,\Pr(U_2 = 1|L, U_1) = \mu_0. \tag{8.11}$$

The distribution of the binary exposure variable A was defined according to the model

$$\text{logit}\,\Pr(A = 1|L, U_1, U_2) = \alpha_0 + \textstyle\sum_{j=1}^{10} \alpha_j L_j + \alpha_{11} U_1. \tag{8.12}$$

Letting U_3 be a scalar random variable that is independent of $(A, L_1, ..., L_{10}, U_1, U_2)$ and uniformly distributed over the interval $[0, 1]$, we defined the counterfactual outcome $Y(a)$, under the intervention setting A to a, as

$$Y(a) = I\Big(U_3 < \text{expit}\Big\{\beta_0 + \gamma a + \textstyle\sum_{j=1}^{10} \beta_j L_j + \beta_{11} U_2\Big\}\Big). \tag{8.13}$$

With $Y := Y(A)$, the above implies consistency, conditional exchangeability given L and structural positivity.

2 Misclassification mechanism

For scenarios with joint misclassification, we defined $B = U_1$ and $Z = U_2$, so that the predictive values take a standard logistic form:

$$\text{logit}\Pr(Y = 1|A, B, L, Z) = \beta_0 + \gamma A + \textstyle\sum_{j=1}^{10} \beta_j L_j + \beta_{11} Z \tag{8.14}$$

$$\text{logit}\Pr(A = 1|B, L, Z) = \alpha_0 + \textstyle\sum_{j=1}^{10} \alpha_j L_j + \alpha_{11} B. \tag{8.15}$$

For scenarios without exposure misclassification, we set $\alpha_{11} = 0$ and defined $B = A$ and $Z = U_2$, so that

$$\text{logit}\Pr(Y = 1|A, B, L, Z) = \beta_0 + \gamma A + \textstyle\sum_{j=1}^{10} \beta_j L_j + \beta_{11} Z, \tag{8.16}$$

$$\text{logit}\Pr(B = 1|L, Z) = \alpha_0 + \textstyle\sum_{j=1}^{10} \alpha_j L_j. \tag{8.17}$$

For simplicity, we removed any marginal dependence of Z on the covariates L and U_1 as well as any marginal dependence of U_1 on L (cf. equations (8.10) and (8.11)). Although models (8.10) through (8.15) take a standard logistic form, they do not imply that the corresponding sensitivities and specificities can be written in the same form. We chose the predictive values rather than the sensitivities and specificities to take a standard logistic form so as to ensure correct model specification in the estimation of the weights in the simulation experiments, in which a likelihood approach based on predictive values was adopted (cf. (8.9)).

3 Missing data mechanism

For these simulations, we stipulated L, B and Z to be observed for all subjects. We consider scenarios where the dataset can be partitioned into a subset with validation data on all misclassified variables (denoted $R = 1$) and a dataset with validation data on neither ($R = 0$). That is, we simulated data such that subjects

have validation data on both A and Y or neither on A nor on Y. Values for the response indicator R were generated according to the following (MAR) model:

$$\begin{aligned}\operatorname{logit}\Pr(R=1|Z,B,Y,A,L) &= \operatorname{logit}\Pr(R=1|Z,B,L)\\ &= \xi_0+\xi_1 Z+\xi_2 B+\xi_3 ZB.\end{aligned}$$

4 Scenarios

We initially fixed most parameters of models (8.12) and (8.13) at the respective values of "Scenario A" of Setoguchi et al. (2008): $\alpha_1 = 0.8$, $\alpha_2 = -0.25$, $\alpha_3 = 0.6$, $\alpha_4 = -0.4$, $\alpha_5 = -0.8$, $\alpha_6 = -0.5$, $\alpha_7 = 0.7$, $\alpha_8 = 0$, $\alpha_9 = 0$, $\alpha_{10} = 0$, $\beta_0 = -3.85$, $\beta_1 = 0.3$, $\beta_2 = -0.36$, $\beta_3 = -0.73$, $\beta_4 = -0.2$, $\beta_5 = 0$, $\beta_6 = 0$, $\beta_7 = 0$, $\beta_8 = 0.71$, $\beta_9 = -0.19$ and $\beta_{10} = 0.26$. Parameters η_0 and α_0 were fixed at zero and ξ_1, ξ_2 and ξ_3 at 2, 1 and -1, respectively. The remaining parameters and β_0 were allowed to vary across scenarios as per Table 8.4.

Scenarios differ by sample size n, the presence of outcome misclassification, potentially misclassified outcome prevalence (via μ_0), the associations between the exposure and outcome on the one hand and the respective misclassified versions on the other (via α_{11} and β_{11}), outcome model intercept β_0, the conditional log-OR γ, or the size of the validation subset (via ξ_0). Based on an iterative Monte Carlo integration approach (Austin and Stafford, 2008), we specified γ so as to keep the target marginal log odds ratio at -0.4.

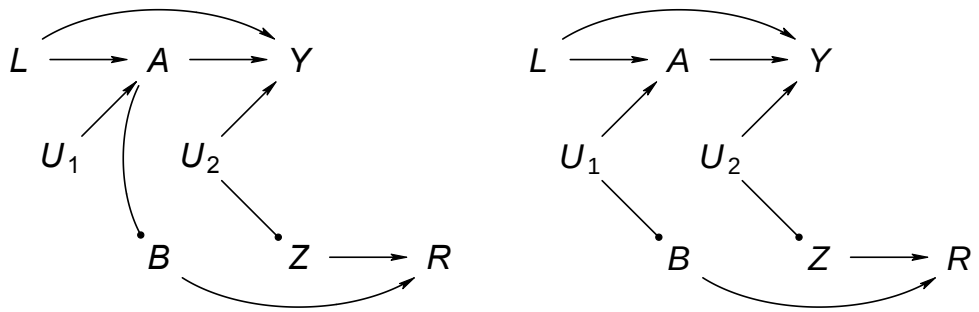

Figure 8.1: Data structure for scenarios with misclassification on the outcome only (left) or on both the exposure and outcome (right). Bullet arrowheads represent deterministic relationships.

5 *Estimators*

We considered five estimators of the OR for the marginal exposure-outcome effect: a crude estimator (labeled Crude) that ignores both confounding and misclassication of any variable, a misclassification naive estimator (labeled PS) that addresses confounding through IPW, complete cases analysis (CCA) in which IPW is applied only to the subset of subjects with validation data, the Gravel and Platt estimator (GP) that ignores exposure misclassification, and the method proposed in this article (labeled IPWM). Both GP and IPWM are implemented using the R function `mecor::ipwm` (Nab, 2019; Nab et al., 2018), which in the simulation settings considered uses iteratively reweighted least squares via the `stats::glm` function for maximum likelihood estimation. GP coincides with the approach of Gravel and Platt where it concerns point estimation, but they

Table 8.4: Simulation parameter values used in the Monte Carlo studies. Scenarios indicated with 'a' have $n = 10000$, those with 'b' have $n = 5000$ and those with 'c' have $n = 1000$.

Scenarios	Exposure misclassification	μ_0	α_{11}	β_0	β_{11}	γ	ξ_0
1a,1b,1c	Absent	−2	0	−3.85	2	−0.431	−1.5
2a,2b,2c	Absent	−3	0	−3.85	2	−0.417	−1.5
3a,3b,3c	Absent	−2	0	−3.85	4	−0.624	−1.5
4a,4b,4c	Absent	−2	0	−3.85	2	−0.431	−2.5
5a,5b,5c	Present	−2	2	−3.85	2	−0.431	−1.5
6a,6b,6c	Present	−3	2	−3.85	2	−0.417	−1.5
7a,7b,7c	Present	−2	4	−3.85	2	−0.431	−1.5
8a,8b,8c	Present	−2	2	−3.85	4	−0.624	−1.5
9a,9b,9c	Present	−2	2	−3.85	2	−0.431	−2.5
10a,10b,10c	Absent	−2	0	−2	2	−0.470	−1.5
11a,11b,11c	Absent	−3	0	−2	2	−0.445	−1.5
12a,12b,12c	Absent	−2	0	−2	4	−0.641	−1.5
13a,13b,13c	Absent	−2	0	−2	2	−0.470	−2.5
14a,14b,14c	Present	−2	2	−2	2	−0.470	−1.5
15a,15b,15c	Present	−3	2	−2	2	−0.445	−1.5
16a,16b,16c	Present	−2	4	−2	2	−0.470	−1.5
17a,17b,17c	Present	−2	2	−2	4	−0.641	−1.5
18a,18b,18c	Present	−2	2	−2	2	−0.470	−2.5

differ in the construction of confidence intervals. Unlike Gravel and Platt (2018), we used a non-parametric rather than a semi-parametric bootstrap procedure for estimating standard errors and constructing confidence intervals. Semi-parametrically generating response indicators would preferably require modelling of (or making additional assumptions about) the missing data mechanism. In particular, to obtain a bootstrap dataset, we defined the record of a unit as their observed data and response indicators, imposed a uniform distribution across all records in the original dataset, and drew independently as many records from this distribution as the total number of records in the original dataset. For all methods and each original dataset, we drew 1000 bootstrap datasets for variance estimation and the construction of percentile confidence intervals.

All estimators are based on a function of the estimated outcome probability P_1 in the exposed group and the estimated outcome probability P_0 in the unexposed group. However, since P_1 and P_0 may take a value of 0 or 1, the crude odds ratio $[P_1/(1-P_1)]/[P_0/(1-P_0)]$ need not exist. In contrast to what is often (implicitly) done in simulation studies—i.e., studying the properties of the estimators after conditioning on datasets where $[P_1/(1-P_1)]/[P_0/(1-P_0)]$ is defined—we first define $P_1^* = (P_1 s+1)/(s+2)$ and $P_0^* = (P_0 s+1)/(s+2)$ for a large positive number s (here set to 10^6) and then regard $[P_1^*/(1-P_1^*)]/[P_0^*/(1-P_0^*)]$ as the estimator of the OR for the exposure-outcome association. This ensures the estimator is always defined and effectively shrinks the outcome probabilities towards 0.5 and the OR towards 1 (Supplementary Appendix S8.2).

For PS and CCA, we used a logistic regression of B and A, respectively, on covariates L_1 through L_{10} as main effects to estimate the propensity scores. Taking the crude OR for the association between B and Z (PS) or A and Y (CCA) over the data weighted by the reciprocal of the propensity of the received exposure level provided an estimate of the target OR. R code for the methods GP and IPWM is given in Supplementary Appendix S8.3.

8.4.2 Results

The treatment assignment mechanism detailed above resulted in average exposure rates ranging from 17% to 51%, whereas average outcome rates ranged from 3% to 22%. Across all simulation studies, the average outcome rate ranged from 6% to 18%. Across all simulation studies with exposure misclassification, exposure and joint misclassification rates ranged from 16% to 33% and from 2% to 6%, respectively. Approximately 16% to 32% of subjects were allocated validation data.

The results on the performance of the various methods in simulations studies 1-9 are provided in Table 8.5 (see Supplementary Table S8.4 for the results on all scenarios).

As expected, Crude, PS and CCA clearly showed bias with respect to the target log OR of -0.4. The bias associated with restricting the analysis to records with validation data is likely brought on to a large extent by collider stratification, with R acting as the collider here (cf. Figure 8.1). Both Crude and PS indicated a null effect, as one would anticipate in view of the marginal and L-conditional independence of B and Z implied by the simulation set-up. The empirical coverage probabilities were, although low for both estimators, similar to substantially larger for PS as compared with Crude. Paralleling this is that Crude, whose (implicit) propensity score model is inherently at least as parsimonious, yielded similar to smaller empirical and sample standard errors as compared with PS. With the average fraction of subjects with validation data being as low as 16% (in scenarios with low ξ_0) to 32%, it is not unsurprising that Crude was subject to the largest degree of variability.

The results for the IPWM approach are generally favourable for large samples and in line with its theoretical (large sample) properties. For scenarios with smaller samples (scenarios 1c, 2c and 4c, 6c and 9c in particular), however, we observed considerable bias (see Supplementary Table S8.4). Comparing CCA with IPWM, we note a strong linear association between the methods in terms of the absolute within-method differences in estimated bias between scenarios of size 10000 (scenarios labeled 'a') and the respective scenarios of size 1000 (scenarios labeled 'c') (Pearson correlation 0.997). Note that the results for GP and IPWM are identical for scenarios labeled 1-4 and 10-13 since the methods are equivalent in terms of point estimation in the absence of exposure misclassification. In all other scenarios, i.e., scenarios for which GP was not developed, GP performed substantially worse than IPWM. The non-zero, albeit relatively small, systematic deviations of the IPWM point estimates from the target -0.4, notably the estimated bias of -0.097 (scenario 2b), may be attributable in part to the outcome being rare (with prevalence ranging from 3% to 8% across scenarios labeled 1-9). This is indicated by the superior performance of IPWM in scenarios where the outcome is more prevalent (cf. scenarios labeled 1-9b versus 10-18b, which have prevalence up to 22%). A similar observation was made by Gravel and Platt (2018).

The standard errors for GP and IPWM were noticeably higher than those of Crude and PS, which is unsurprising in view of the discrepancies in the number of estimated parameters. As expected, increasing the sample size, the true outcome rate (via β_0) or both led to a decrease in the variability of IPWM (cf. Table 8.4

and Supplementary Table S8.4). However, despite the large discrepancies between SSE and SE for some scenarios, the empirical coverage probabilities of IPWM were close to the nominal level of 0.95, except for scenarios 1c, 2c and 4c, where we observed considerable bias.

8.5 Discussion

The analysis of epidemiologic data is often complicated by the presence of confounding and misclassifications in exposure and outcome variables. In this paper we propose a new estimator for estimating a marginal odds-ratio in the presence of confouding and joint misclassification of the exposure and outcome variables. In simulation studies, this weighting estimator showed promising finite sample performance, reducing bias and mean squared error as compared with simpler methods.

The proposed IPWM estimator is an extension of the inverse probability weighting estimator recently proposed by Gravel and Platt (GP) which only addresses the misclassification in the outcome (Gravel and Platt, 2018). IPWM and GP are (mathematically) equivalent when the exposure is (assumed to be) measured without misclassification error.

Like the Gravel and Platt approach, IPWM relies on estimates of sensitivity and specificity or positive and negative predictive values for the misclassified variables. In this paper, we used an internal validation approach where a portion of subjects would receive error-free ('gold standard') measurements on either or both the outcome and exposure. However, we anticipate that in some settings the likelihood may not be fully identifiable from the data at hand. In these settings, it may be possible to incorporate external rather than internal information on the misclassification rates, possibly through a Bayesian approach using prior assumptions about misclassification probabilities. When validation data is external, however, it may be necessary to assume misclassification to be independent of covariates L, because external studies seldom consider the same covariates as the main study (Lyles et al., 2011). External validation approaches also require the assumption that the misclassification parameters targeted in the validation sample are transportable to the main study.

In the absence of internal and external validation data, it is possible to conduct a sensitivity analysis within the weighting framework. Formula (8.8) for the weights can readily be used in a sensitivity analysis in which the terms describing the distribution of true exposure and outcome variables in relation to the observed data (positive and negative predictive values) serve as sensitivity parameters of

Table 8.5: Results for simulation studies 1-9b on the performance of different causal estimators in various scenarios of confounding and misclassification in exposure and outcome. Abbreviations: PS, propensity score method ignoring misclassification; CCA, complete case analysis; GP, Gravel and Platt estimator ignoring exposure misclassification, consistent with the methodology of Gravel and Platt (2018) for point (but not for variance) estimation; IPWM, inverse probability weighting method for confounding and joint exposure and outcome misclassification; BSE, estimated standard error for the bias due to Monte Carlo error; SE, empirical standard error; SSE, sample standard error; CP, empirical coverage probability. In all scenarios, the true marginal log OR (estimand) was −0.4.

	Crude					
Scenario	Bias	BSE	MSE	SE	SSE	CP
1b	0.394	0.004	0.169	0.119	0.118	0.122
2b	0.382	0.006	0.179	0.183	0.184	0.492
3b	0.394	0.004	0.169	0.117	0.118	0.116
4b	0.401	0.004	0.174	0.117	0.118	0.102
5b	0.401	0.003	0.169	0.090	0.088	0.007
6b	0.407	0.004	0.183	0.132	0.134	0.133
7b	0.396	0.003	0.164	0.086	0.088	0.009
8b	0.395	0.003	0.164	0.086	0.088	0.005
9b	0.398	0.003	0.166	0.089	0.088	0.005
	PS					
Scenario	Bias	BSE	MSE	SE	SSE	CP
1b	0.392	0.005	0.182	0.168	0.169	0.382
2b	0.379	0.008	0.213	0.264	0.258	0.738
3b	0.389	0.006	0.182	0.175	0.169	0.402
4b	0.389	0.006	0.182	0.176	0.168	0.392
5b	0.402	0.003	0.170	0.090	0.088	0.010
6b	0.407	0.004	0.183	0.131	0.135	0.136
7b	0.396	0.003	0.164	0.086	0.088	0.009
8b	0.395	0.003	0.164	0.086	0.088	0.004
9b	0.398	0.003	0.166	0.089	0.088	0.005

the sensitivity analysis. The models for the predictive values can take complex forms, however, thus complicating the analysis and presentation of results.

If internal validation is available, the subjects with validation data need not form a completely random subset. The proposed method, IPWM, was developed under the assumption that validation data allocation occurs in an 'ignorable' fashion (Rubin, 1976). In practice, it may be that the researchers have limited control over the validation data allocation mechanism. For instance, it is conceivable that individuals with specific indications (e.g., with a realisation of L, B or Z) are practically ineligible to be assigned a double measurement of the exposure (A and B) and outcome (Y and Z). Further, the estimator also allows for validation subjects to receive either the double exposure or double

Table 8.5 continued.

	CCA					
Scenario	Bias	BSE	MSE	SE	SSE	CP
1b	−0.078	0.015	0.226	0.469	0.491	0.899
2b	−0.117	0.019	0.375	0.601	0.900	0.887
3b	−0.020	0.010	0.091	0.301	0.300	0.919
4b	−0.093	0.020	0.407	0.631	1.158	0.899
5b	−0.145	0.009	0.103	0.286	0.286	0.903
6b	−0.109	0.011	0.131	0.345	0.362	0.930
7b	−0.213	0.007	0.101	0.237	0.250	0.865
8b	−0.209	0.006	0.079	0.187	0.186	0.775
9b	−0.175	0.012	0.184	0.392	0.411	0.902
	GP					
Scenario	Bias	BSE	MSE	SE	SSE	CP
1b	−0.036	0.011	0.130	0.359	0.428	0.958
2b	−0.097	0.016	0.265	0.505	0.861	0.938
3b	−0.019	0.007	0.055	0.233	0.240	0.939
4b	−0.045	0.016	0.253	0.501	1.087	0.944
5b	0.269	0.008	0.132	0.244	0.244	0.799
6b	0.280	0.010	0.177	0.314	0.339	0.862
7b	0.134	0.008	0.076	0.241	0.252	0.926
8b	0.259	0.004	0.087	0.140	0.144	0.570
9b	0.263	0.010	0.174	0.325	0.339	0.883

Table 8.5 continued.

	IPWM					
Scenario	Bias	BSE	MSE	SE	SSE	CP
1b	−0.036	0.011	0.130	0.359	0.428	0.958
2b	−0.097	0.016	0.265	0.505	0.861	0.938
3b	−0.019	0.007	0.055	0.233	0.240	0.939
4b	−0.045	0.016	0.253	0.501	1.087	0.944
5b	−0.017	0.009	0.082	0.286	0.284	0.942
6b	−0.014	0.011	0.129	0.359	0.386	0.958
7b	0.004	0.008	0.059	0.243	0.261	0.969
8b	−0.004	0.006	0.032	0.180	0.181	0.958
9b	−0.025	0.012	0.141	0.374	0.415	0.956

outcome measurement. We simulated data such that subjects have validation data on both the exposure and outcome variables or on neither. Although this may greatly simplify analysis and enhance efficiency, in practice it is not necessary to assume that this condition holds. An interesting scenario is where subjects have validation data on at most one variable, i.e., on the exposure variable or the outcome variable but not both. In this case, valid estimation would require additional modelling assumptions; for example, the error-free outcome variable cannot then be regressed on the error-free exposure variable.

To accommodate settings where validation data allocation is not completely at random, we deviated from the semi-parametric bootstrap procedure for variance estimation proposed by Gravel and Platt. Instead, the non-parametric procedure we used requires less assumptions regarding the validation subset sampling procedure. The non-parametric procedure showed good performance in our simulations.

Whilst we have discussed under what conditions the proposed method consistently estimates or at least identifies the target quantity, the assumptions may be untenable in particular settings. Particularly, an infallible measurement tool for the exposure and outcome that can be performed on a subset of the data need not always exist. The robustness to deviations of infallibility is an interesting and important direction for further research. This is especially relevant where there exists considerable uncertainty about the tenability of the assumptions that is difficult to incorporate in the analysis. An obvious and flexible alternative to IPWM is to multiply impute missing values including absent measurement

error-free variables before implementing IPW (MI+IPW). Although MI+IPW and IPWM may be comparable in terms of their assumptions, it is yet unclear how they behave under assumption violations such as misspecification of the outcome model.

An advantageous property of MI+IPW is that it can easily accommodate missing covariate values. Other alternatives that can accommodate missing covariates were recently developed by Shu and Yi (2018). Their proposed weighting estimators simultaneously addresses confounding, misclassification of the outcome (but not of the exposure) and measurement error on the covariates under a classical additive measurement error model. The methods can be implemented using validation data or repeated measurements and use a simple misclassification model (in which the outcome surrogate is independent of exposure or covariates given the target outcome) that is suitable for performing sensitivity analyses.

Another interesting area for further research is where the researchers do have control over who is referred for further testing by the assumed infallible measurement tool(s). An obvious choice is to adopt a completely at random strategy (simple random sampling). However, other referral (sampling) strategies exist and it is not clear what strategy leads to the most favourable estimator properties for the given setting.

In summary, we have developed an extension to an existing method, to allow for valid estimation of a marginal causal OR in the presence of confounding and a commonly ignored and misunderstood source of bias—joint exposure and outcome misclassification. The R function `mecor::ipwm` has been made available to facilitate implementation (Nab, 2019; Nab et al., 2018).

References

Austin, P. C. and J. Stafford (2008): "The performance of two data-generation processes for data with specified marginal treatment odds ratios," *Communications in Statistics - Simulation and Computation*, 37, 1039–1051.

Babanezhad, M., S. Vansteelandt, and E. Goetghebeur (2010): "Comparison of causal effect estimators under exposure misclassification," *Journal of Statistical Planning and Inference*, 140, 1306–1319.

Brakenhoff, T. B., M. Mitroiu, R. H. Keogh, K. G. Moons, R. H. Groenwold, and M. van Smeden (2018): "Measurement error is often neglected in medical literature: a systematic review," *Journal of Clinical Epidemiology*, 98, 89–97.

Braun, D., M. Gorfine, G. Parmigiani, N. D. Arvold, F. Dominici, and C. Zigler (2017): "Propensity scores with misclassified treatment assignment: a likelihood-based adjustment," *Biostatistics*, 18, 695–710.

Brenner, H., D. A. Savitz, and O. Gefeller (1993): "The effects of joint misclassification of exposure and disease on epidemiologic measures of association exposure and disease on epidemiologic measures of association," *Journal of Clinical Epidemiology*, 46, 1195–1202.

Brooks, D. R., K. D. Getz, A. T. Brennan, A. Z. Pollack, and M. P. Fox (2018): "The impact of joint misclassification of exposures and outcomes on the results of epidemiologic research," *Current Epidemiology Reports*, 5, 166–174.

Culver, A. L., I. S. Ockene, R. Balasubramanian, B. C. Olendzki, D. M. Sepavich, J. Wactawski-Wende, J. E. Manson, Y. Qiao, S. Liu, P. A. Merriam, et al. (2012): "Statin use and risk of diabetes mellitus in postmenopausal women in the women's health initiative," *Archives of internal medicine*, 172, 144–152.

Dawid, A. (1979): "Conditional independence in statistical theory," *Journal of the Royal Statistical Society, Series B (Methodological)*, 1–31.

Gravel, C. A. and R. W. Platt (2018): "Weighted estimation for confounded binary outcomes subject to misclassification," *Statistics in Medicine*, 37, 425–436.

Holland, P. (1986): "Statistics in causal inference," *Journal of the American Statistical Association*, 81, 945–960.

Holland, P. (1988): "Causal inference, path analysis, and recursive structural equations models," *Sociological Methodology*, 18, 449–484.

Jurek, A. M., S. Greenland, and G. Maldonado (2008): "Brief report: how far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null?" *International Journal of Epidemiology*, 37, 382–385.

Jurek, A. M., G. Maldonado, S. Greenland, and T. R. Church (2006): "Exposure-measurement error is frequently ignored when interpreting epidemiologic study results," *European journal of epidemiology*, 21, 871–876.

Kristensen, P. (1992): "Bias from nondifferential but dependent misclassification of exposure and outcome," *Epidemiology*, 210–215.

Leong, A., K. Dasgupta, S. Bernatsky, D. Lacaille, A. Avina-Zubieta, and E. Rahme (2013): "Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records," *PloS one*, 8, e75256.

Lyles, R. H., L. Tang, H. M. Superak, C. C. King, D. D. Celentano, Y. Lo, and J. D. Sobel (2011): "Validation data-based adjustments for outcome misclassification in logistic regression: an illustration," *Epidemiology*, 22, 589.

Marcum, Z. A., M. A. Sevick, and S. M. Handler (2013): "Medication nonadherence: a diagnosable and treatable medical condition," *Jama*, 309, 2105–2106.

Nab, L. (2019): *mecor: Measurement Error Corrections*, r package version 0.1.0. Available from: https://github.com/LindaNab/mecor.git.

Nab, L., R. H. Groenwold, P. M. Welsing, and M. van Smeden (2018): "Measurement error in continuous endpoints in randomised trials: problems and solutions," *arXiv preprint arXiv:1809.07068*.

Neyman, J., K. Iwaszkiewicz, and St. Kolodziejczyk (1935): "Statistical problems in agricultural experimentation," *Supplement to the Journal of the Royal Statistical Society*, 2, 107–180.

Ni, J., A. Leong, K. Dasgupta, and E. Rahme (2017): "Correcting hazard ratio estimates for outcome misclassification using multiple imputation with internal validation data," *Pharmacoepidemiology and drug safety*, 26, 925–934.

Pearl, J. (2009): *Causality: Models, Reasoning and Inference*, New York: Cambridge University Press.

R Core Team (2018): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL `https://www.R-project.org/`.

Rubin, D. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66, 688–701.

Rubin, D. (1976): "Inference and missing data," *Biometrika*, 63, 581–592.

Setoguchi, S., S. Schneeweiss, M. B. MA, R. Glynn, and E. Cook (2008): "Evaluating uses of data mining techniques in propensity score estimation: a simulation study," *Pharmacoepidemiology and Drug Safety*, 17, 546–555.

Shu, D. and G. Y. Yi (2018): "Weighted causal inference methods with mismeasured covariates and misclassified outcomes," *Statistics in Medicine*.

Shu, D. and G. Y. Yi (2019): "Causal inference with measurement error in outcomes: Bias analysis and estimation methods," *Statistical Methods in Medical Research*, 28, 2049–2068.

Tang, L., R. H. Lyles, Y. Ye, Y. Lo, and C. C. King (2013): "Extended matrix and inverse matrix methods utilizing internal validation data when both disease and exposure status are misclassified," *Epidemiologic Methods*, 2, 49–66.

VanderWeele, T. J. and M. A. Hernán (2012): "Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs," *American journal of epidemiology*, 175, 1303–1310.

Vogel, C., H. Brenner, A. Pfahlberg, and O. Gefeller (2005): "The effects of joint misclassification of exposure and disease on the attributable risk," *Statistics in Medicine*, 24, 1881–1896.

Supplementary Material

S8.1

Suppose A, B, Y and Z are random variables that take values in $\{0, 1\}$.

Theorem 8.1. *For any a, l, let*

$$\varphi(a,l) = \frac{\varphi^*(a,l)}{\mathbb{E}[\varphi^*(A,L)|A=a]} \quad \textit{and} \quad \varphi^*(a,l) = \frac{1}{\Pr(A=a|L=l)}.$$

If $Y(A) = Y$ (consistency), $(Y(0), Y(1)) \perp\!\!\!\perp A|L = l$ (conditional exchangeability), $\Pr(A = a) > 0$ and $\Pr(A = a|L = l) > 0$ (positivity) for all a and every l in the support of L, then

$$\mathbb{E}[Y(a)] = \mathbb{E}[\varphi(A,L)I(Y=1)|A=a].$$

Proof. We begin by considering $\mathbb{E}[\varphi^*(A,L)|A=a]$. By the law of the unconscious statistician and Bayes' theorem, we have

$$\begin{aligned}
\mathbb{E}[\varphi^*(A,L)|A=a] &= \sum_l \frac{\Pr(L=l|A=a)}{\Pr(A=a|L=l)} \\
&= \sum_l \frac{\Pr(A=a|L=l)\Pr(L=l)}{\Pr(A=a)\Pr(A=a|L=l)} \\
&= \frac{1}{\Pr(A=a)} \sum_l \Pr(L=l) \\
&= \frac{1}{\Pr(A=a)}.
\end{aligned}$$

Hence, for all a, y, we have

$$\begin{aligned}
&\sum_l \varphi(a,l)\Pr(Y=y, L=l|A=a) \qquad (8.18)\\
&\quad= \sum_l \frac{\Pr(Y=y, L=l|A=a)\Pr(A=a)}{\Pr(A=a|L=l)} \\
&\quad= \sum_l \frac{\Pr(Y=y|A=a, L=l)\Pr(A=a|L=l)\Pr(L=l)}{\Pr(A=a|L=l)} \\
&\quad= \sum_l \Pr(Y=y|A=a, L=l)\Pr(L=l)
\end{aligned}$$

$$= \sum_l \Pr(Y(a) = y | A = a, L = l) \Pr(L = l) \tag{8.19}$$
$$= \sum_l \Pr(Y(a) = y | L = l) \Pr(L = l) \tag{8.20}$$
$$= \Pr(Y(a) = y),$$

where (8.19) and (8.20) hold under consistency and conditional exchangeability given L, respectively. Positivity ensures the weights are defined/exist. Hence, $\mathbb{E}[\varphi(A, L)I(Y = 1)|A = a] = \sum_l \varphi(a, l) \Pr(Y = 1, L = l|A = a) = \mathbb{E}[Y(a)]$, as desired. □

Corollary 8.1. *For any y, a, l, let*

$$\varphi(a, l) = \frac{\varphi^*(a, l)}{\mathbb{E}[\varphi^*(A, L)|A = a]}, \quad \varphi^*(a, l) = \frac{1}{\Pr(A = a|L = l)}, \quad \textit{and}$$
$$\phi(a, l) = \frac{\Pr(Y = 1, L = l|A = a)}{\Pr(Z = 1, L = l|B = a)}.$$

If $Y(A) = Y$, $(Y(0), Y(1)) \perp\!\!\!\perp A|L$ and positivity holds, then

$$\begin{aligned}\mathbb{E}[Y(a)] &= \sum_l \varphi(a, l) \Pr(Y = 1, L = l|A = a) \\ &= \sum_l \varphi(a, l)\phi(a, l) \Pr(Z = 1, L = l|B = a) \\ &= \mathbb{E}[\varphi(B, L)\phi(B, L)I(Z = 1)|B = a].\end{aligned}$$

S8.2

Theorem 8.2. *Fix some $s > 0$ and let $P^* = (Ps + 1)/(s + 2)$ for all $P \in [0, 1]$. If $(P_0, P_1) \in (0, 1) \times (0, 1)$, then*

$$1 < \frac{P_1^*/(1 - P_1^*)}{P_0^*/(1 - P_0^*)} < \frac{P_1/(1 - P_1)}{P_0/(1 - P_0)} \quad \textit{if } P_1 > P_0,$$
$$1 = \frac{P_1^*/(1 - P_1^*)}{P_0^*/(1 - P_0^*)} = \frac{P_1/(1 - P_1)}{P_0/(1 - P_0)} \quad \textit{if } P_1 = P_0, \textit{ and}$$
$$1 > \frac{P_1^*/(1 - P_1^*)}{P_0^*/(1 - P_0^*)} > \frac{P_1/(1 - P_1)}{P_0/(1 - P_0)} \quad \textit{if } P_1 < P_0$$

Proof. Suppose $(P_0, P_1) \in (0, 1) \times (0, 1)$. If and only if

$$\frac{P_1^*/(1 - P_1^*)}{P_0^*/(1 - P_0^*)} < \frac{P_1/(1 - P_1)}{P_0/(1 - P_0)}, \tag{8.21}$$

then

$$\frac{P_1 s+1}{s+1-P_1 s}\frac{s+1-P_0 s}{P_0 s+1} < \frac{P_1}{1-P_1}\frac{1-P_0}{P_0},$$
$$\frac{P_1 s+1}{s+1-P_1 s}\frac{1-P_1}{P_1} < \frac{P_0 s+1}{s+1-P_0 s}\frac{1-P_0}{P_0}.$$

Now, since

$$\frac{\partial}{\partial P}\left\{\frac{Ps+1}{s+1-Ps}\frac{1-P}{P}\right\} = \frac{(-2P^2+2P-1)S-1}{P^2(1-(P-1)S)^2} < 0$$

over the interval $(0,1)$ for P, it follows that inequality (8.21) holds if $P_1 > P_0$. Also, if $P_1 > P_0$, then, since $\partial/(\partial P)\{(Ps+1)/(s+1-Ps)\} > 0$ if $P \in (0,1)$, we have

$$1 < \frac{P_1^*/(1-P_1^*)}{P_0^*/(1-P_0^*)}.$$

Similar arguments establish the assertion for the case where $P_1 < P_0$. It is easily verified that if $P_1 = P_0$, then

$$\frac{P_1^*/(1-P_1^*)}{P_0^*/(1-P_0^*)} = \frac{P_1 s+1}{s+1-P_1 s}\frac{s+1-P_0 s}{P_0 s+1} = 1$$
$$= \frac{P_1}{1-P_1}\frac{1-P_0}{P_0} = \frac{P_1/(1-P_1)}{P_0/(1-P_0)},$$

as desired. □

S8.3

GP and IPWM were applied to every dataset `data` in R using the function `mecor::ipwm` and the following code:

```
# GP:
formulasGP <- list(
   Y~Z+B+L1+L2+L3+L4+L5+L6+L7+L8+L9+L10,
   B~Z+L1+L2+L3+L4+L5+L6+L7+L8+L9+L10,
   Z~L1+L2+L3+L4+L5+L6+L7+L8+L9+L10
)
mecor::ipwm(
   formulas=formulasGP, data=data, outcome_true=‘‘Y’’,
   outcome_mis=‘‘Z’’, exposure_true=‘‘B’’, exposure_mis=NULL, sp=1e6
```

```
)

# IPWM:
formulasIPWM <- list(
   Y~A+Z+B+L1+L2+L3+L4+L5+L6+L7+L8+L9+L10,
   A~Z+B+L1+L2+L3+L4+L5+L6+L7+L8+L9+L10,
   Z~B+L1+L2+L3+L4+L5+L6+L7+L8+L9+L10,
   B~L1+L2+L3+L4+L5+L6+L7+L8+L9+L10
)
mecor::ipwm(
   formulas=formulasIPWM, data=data, outcome_true=``Y'',
   outcome_mis=``Z'', exposure_true=``A'', exposure_mis=``B'', sp=1e6
)
```

S8.4 Supplementary Tables

Table S8.1: Expected cell counts (rounded to integers) for illustrative study setting after misclassification and formation of validation subsets

					$L=0$		$L=1$	
R_Y	R_A	Y	A	L	$A=0$	$A=1$	$A=0$	$A=1$
0	0			0	$m_1 = 9371$	$m_2 = 7147$	$m_3 = 1011$	$m_4 = 884$
0	0			1	$m_5 = 1120$	$m_6 = 3165$	$m_7 = 80$	$m_8 = 221$
0	1				$m_9 = 0$	$m_{10} = 0$	$m_{11} = 0$	$m_{12} = 0$
1	0				$m_{13} = 0$	$m_{14} = 0$	$m_{15} = 0$	$m_{16} = 0$
1	1	0	0	0	$m_{17} = 2728$	$m_{18} = 38$	$m_{19} = 144$	$m_{20} = 2$
1	1	1	0	0	$m_{21} = 13$	$m_{22} = 3$	$m_{23} = 169$	$m_{24} = 53$
1	1	0	1	0	$m_{25} = 382$	$m_{26} = 3797$	$m_{27} = 12$	$m_{28} = 242$
1	1	1	1	0	$m_{29} = 1$	$m_{30} = 9$	$m_{31} = 12$	$m_{32} = 178$
1	1	0	0	1	$m_{33} = 287$	$m_{34} = 41$	$m_{35} = 6$	$m_{36} = 5$
1	1	1	0	1	$m_{37} = 2$	$m_{38} = 1$	$m_{39} = 7$	$m_{40} = 3$
1	1	0	1	1	$m_{41} = 84$	$m_{42} = 1658$	$m_{43} = 10$	$m_{44} = 87$
1	1	1	1	1	$m_{45} = 1$	$m_{46} = 4$	$m_{47} = 3$	$m_{48} = 24$

Table S8.2: Log-likelihood contributions for all possible types of observations under internal validation sampling.

Type	R_Y	R_A	Z	B	Y	A	L	Count	Log-likelihood contribution
1	0	0	0	0			0	m_1	$\log(1-\varepsilon^*_{00}) + \log(1-\delta^*_0)$
2	0	0	1	0			0	m_2	$\log(\varepsilon^*_{00}) + \log(1-\delta^*_0)$
3	0	0	0	1			0	m_3	$\log(1-\varepsilon^*_{10}) + \log(\delta^*_0)$
4	0	0	1	1			0	m_4	$\log(\varepsilon^*_{10}) + \log(\delta^*_0)$
5	0	0	0	0			1	m_5	$\log(1-\varepsilon^*_{01}) + \log(1-\delta^*_1)$
6	0	0	1	0			1	m_6	$\log(\varepsilon^*_{01}) + \log(1-\delta^*_1)$
7	0	0	0	1			1	m_7	$\log(1-\varepsilon^*_{11}) + \log(\delta^*_1)$
8	0	0	1	1			1	m_8	$\log(\varepsilon^*_{11}) + \log(\delta^*_1)$
9	0	1						$m_9 + ... + m_{12} = 0$	0
10	1	0						$m_{13} + ... + m_{16} = 0$	0
11	1	1	0	0	0	0	0	m_{17}	$\log(1-\pi^*_{0000}) + \log(1-\lambda^*_{000}) + \log(1-\varepsilon^*_{00}) + \log(1-\delta^*_0)$
12	1	1	1	0	0	0	0	m_{18}	$\log(1-\pi^*_{0100}) + \log(1-\lambda^*_{100}) + \log(\varepsilon^*_{00}) + \log(1-\delta^*_0)$
13	1	1	0	1	0	0	0	m_{19}	$\log(1-\pi^*_{0010}) + \log(1-\lambda^*_{010}) + \log(1-\varepsilon^*_{10}) + \log(\delta^*_0)$
14	1	1	1	1	0	0	0	m_{20}	$\log(1-\pi^*_{0110}) + \log(1-\lambda^*_{110}) + \log(\varepsilon^*_{10}) + \log(\delta^*_0)$

Table S8.2 continued.

Type	R_Y	R_A	Z	B	Y	A	L	Count	Log-likelihood contribution
15	1	1	0	0	1	0	0	m_{21}	$\log(\pi^*_{0000}) + \log(1-\lambda^*_{000}) + \log(1-\varepsilon^*_{00}) + \log(1-\delta^*_0)$
16	1	1	1	0	1	0	0	m_{22}	$\log(\pi^*_{0100}) + \log(1-\lambda^*_{100}) + \log(\varepsilon^*_{00}) + \log(1-\delta^*_0)$
17	1	1	0	1	1	0	0	m_{23}	$\log(\pi^*_{0010}) + \log(1-\lambda^*_{010}) + \log(1-\varepsilon^*_{10}) + \log(\delta^*_0)$
18	1	1	1	1	1	0	0	m_{24}	$\log(\pi^*_{0110}) + \log(1-\lambda^*_{110}) + \log(\varepsilon^*_{10}) + \log(\delta^*_0)$
19	1	1	0	0	0	1	0	m_{25}	$\log(1-\pi^*_{1000}) + \log(\lambda^*_{000}) + \log(1-\varepsilon^*_{00}) + \log(1-\delta^*_0)$
20	1	1	1	0	0	1	0	m_{26}	$\log(1-\pi^*_{1100}) + \log(\lambda^*_{100}) + \log(\varepsilon^*_{00}) + \log(1-\delta^*_0)$
21	1	1	0	1	0	1	0	m_{27}	$\log(1-\pi^*_{1010}) + \log(\lambda^*_{010}) + \log(1-\varepsilon^*_{10}) + \log(\delta^*_0)$
22	1	1	1	1	0	1	0	m_{28}	$\log(1-\pi^*_{1110}) + \log(\lambda^*_{110}) + \log(\varepsilon^*_{10}) + \log(\delta^*_0)$
23	1	1	0	0	1	1	0	m_{29}	$\log(\pi^*_{1000}) + \log(\lambda^*_{000}) + \log(1-\varepsilon^*_{00}) + \log(1-\delta^*_0)$
24	1	1	1	0	1	1	0	m_{30}	$\log(\pi^*_{1100}) + \log(\lambda^*_{100}) + \log(\varepsilon^*_{00}) + \log(1-\delta^*_0)$
25	1	1	0	1	1	1	0	m_{31}	$\log(\pi^*_{1010}) + \log(\lambda^*_{010}) + \log(1-\varepsilon^*_{10}) + \log(\delta^*_0)$
26	1	1	1	1	1	1	0	m_{32}	$\log(\pi^*_{1110}) + \log(\lambda^*_{110}) + \log(\varepsilon^*_{10}) + \log(\delta^*_0)$

Table S8.2 continued.

Type	R_Y	R_A	Z	B	Y	A	L	Count	Log-likelihood contribution
27	1	1	0	0	0	0	1	m_{33}	$\log(1-\pi^*_{0001})+\log(1-\lambda^*_{001})+\log(1-\varepsilon^*_{01})+\log(1-\delta^*_1)$
28	1	1	1	0	0	0	1	m_{34}	$\log(1-\pi^*_{0101})+\log(1-\lambda^*_{101})+\log(\varepsilon^*_{01})+\log(1-\delta^*_1)$
29	1	1	0	1	0	0	1	m_{35}	$\log(1-\pi^*_{0011})+\log(1-\lambda^*_{011})+\log(1-\varepsilon^*_{11})+\log(\delta^*_1)$
30	1	1	1	1	0	0	1	m_{36}	$\log(1-\pi^*_{0111})+\log(1-\lambda^*_{111})+\log(\varepsilon^*_{11})+\log(\delta^*_1)$
31	1	1	0	0	1	0	1	m_{37}	$\log(\pi^*_{0001})+\log(1-\lambda^*_{001})+\log(1-\varepsilon^*_{01})+\log(1-\delta^*_1)$
32	1	1	1	0	1	0	1	m_{38}	$\log(\pi^*_{0101})+\log(1-\lambda^*_{101})+\log(\varepsilon^*_{01})+\log(1-\delta^*_1)$
33	1	1	0	1	1	0	1	m_{39}	$\log(\pi^*_{0011})+\log(1-\lambda^*_{011})+\log(1-\varepsilon^*_{11})+\log(\delta^*_1)$
34	1	1	1	1	1	0	1	m_{40}	$\log(\pi^*_{0111})+\log(1-\lambda^*_{111})+\log(\varepsilon^*_{11})+\log(\delta^*_1)$
35	1	1	0	0	0	1	1	m_{41}	$\log(1-\pi^*_{1001})+\log(\lambda^*_{001})+\log(1-\varepsilon^*_{01})+\log(1-\delta^*_1)$
36	1	1	1	0	0	1	1	m_{42}	$\log(1-\pi^*_{1101})+\log(\lambda^*_{101})+\log(\varepsilon^*_{01})+\log(1-\delta^*_1)$
37	1	1	0	1	0	1	1	m_{43}	$\log(1-\pi^*_{1011})+\log(\lambda^*_{011})+\log(1-\varepsilon^*_{11})+\log(\delta^*_1)$
38	1	1	1	1	0	1	1	m_{44}	$\log(1-\pi^*_{1111})+\log(\lambda^*_{111})+\log(\varepsilon^*_{11})+\log(\delta^*_1)$

Table S8.2 continued.

Type	R_Y	R_A	Z	B	Y	A	L	Count	Log-likelihood contribution
39	1	1	0	0	1	1	1	m_{45}	$\log(\pi^*_{1001}) + \log(\lambda^*_{001}) + \log(1-\varepsilon^*_{01}) + \log(1-\delta^*_1)$
40	1	1	1	0	1	1	1	m_{46}	$\log(\pi^*_{1101}) + \log(\lambda^*_{101}) + \log(\varepsilon^*_{01}) + \log(1-\delta^*_1)$
41	1	1	0	1	1	1	1	m_{47}	$\log(\pi^*_{1011}) + \log(\lambda^*_{011}) + \log(1-\varepsilon^*_{11}) + \log(\delta^*_1)$
42	1	1	1	1	1	1	1	m_{48}	$\log(\pi^*_{1111}) + \log(\lambda^*_{111}) + \log(\varepsilon^*_{11}) + \log(\delta^*_1)$

Table S8.3: Closed form expressions for the maximum likelihood estimators (MLE) of the parameters of the likelihood parameterised in terms of predictive values for the hypothetical study setting.

Parameter	MLE
δ_0^*	$\hat{\delta}_0^* = (m_3+m_4+m_{19}+m_{20}+m_{23}+m_{24}+m_{27}+m_{28}+m_{31}+m_{32})/(\{\sum_{j=1}^{4} m_j\}+\{\sum_{j=17}^{32} m_j\})$
δ_1^*	$\hat{\delta}_1^* = (m_7+m_8+m_{35}+m_{36}+m_{39}+m_{40}+m_{43}+m_{44}+m_{47}+m_{48})/(\{\sum_{j=5}^{8} m_j\}+\{\sum_{j=33}^{48} m_j\})$
ε_{00}^*	$\hat{\varepsilon}_{00}^* = (m_2+m_{18}+m_{22}+m_{26}+m_{30})/(m_1+m_2+m_{17}+m_{18}+m_{21}+m_{22}+m_{25}+m_{26}+m_{29}+m_{30})$
ε_{10}^*	$\hat{\varepsilon}_{10}^* = (m_4+m_{20}+m_{24}+m_{28}+m_{32})/(m_3+m_4+m_{19}+m_{20}+m_{23}+m_{24}+m_{27}+m_{28}+m_{31}+m_{32})$
ε_{01}^*	$\hat{\varepsilon}_{01}^* = (m_6+m_{34}+m_{38}+m_{42}+m_{46})/(m_5+m_6+m_{33}+m_{34}+m_{37}+m_{38}+m_{41}+m_{42}+m_{45}+m_{46})$
ε_{11}^*	$\hat{\varepsilon}_{11}^* = (m_8+m_{36}+m_{40}+m_{44}+m_{48})/(m_7+m_8+m_{35}+m_{36}+m_{39}+m_{40}+m_{43}+m_{44}+m_{47}+m_{48})$
λ_{000}^*	$\hat{\lambda}_{000}^* = (m_{25}+m_{29})/(m_{17}+m_{21}+m_{25}+m_{29})$
λ_{100}^*	$\hat{\lambda}_{100}^* = (m_{26}+m_{30})/(m_{18}+m_{22}+m_{26}+m_{30})$
λ_{010}^*	$\hat{\lambda}_{010}^* = (m_{27}+m_{31})/(m_{19}+m_{23}+m_{27}+m_{31})$
λ_{110}^*	$\hat{\lambda}_{110}^* = (m_{28}+m_{32})/(m_{20}+m_{24}+m_{28}+m_{32})$
λ_{001}^*	$\hat{\lambda}_{001}^* = (m_{41}+m_{45})/(m_{33}+m_{37}+m_{41}+m_{45})$
λ_{101}^*	$\hat{\lambda}_{101}^* = (m_{42}+m_{46})/(m_{34}+m_{38}+m_{42}+m_{46})$
λ_{011}^*	$\hat{\lambda}_{011}^* = (m_{43}+m_{47})/(m_{35}+m_{39}+m_{43}+m_{47})$
λ_{111}^*	$\hat{\lambda}_{111}^* = (m_{44}+m_{48})/(m_{36}+m_{40}+m_{44}+m_{48})$

Table S8.3 continued.

Parameter	MLE
π^*_{0000}	$\hat{\pi}^*_{0000} = m_{21}/(m_{17} + m_{21})$
π^*_{1000}	$\hat{\pi}^*_{1000} = m_{29}/(m_{25} + m_{29})$
π^*_{0100}	$\hat{\pi}^*_{0100} = m_{22}/(m_{18} + m_{22})$
π^*_{1100}	$\hat{\pi}^*_{1100} = m_{30}/(m_{26} + m_{30})$
π^*_{0010}	$\hat{\pi}^*_{0010} = m_{23}/(m_{19} + m_{23})$
π^*_{1010}	$\hat{\pi}^*_{1010} = m_{31}/(m_{27} + m_{31})$
π^*_{0110}	$\hat{\pi}^*_{0110} = m_{24}/(m_{20} + m_{24})$
π^*_{1110}	$\hat{\pi}^*_{1110} = m_{32}/(m_{28} + m_{32})$
π^*_{0001}	$\hat{\pi}^*_{0001} = m_{37}/(m_{33} + m_{37})$
π^*_{1001}	$\hat{\pi}^*_{1001} = m_{45}/(m_{41} + m_{45})$
π^*_{0101}	$\hat{\pi}^*_{0101} = m_{38}/(m_{34} + m_{38})$
π^*_{1101}	$\hat{\pi}^*_{1101} = m_{46}/(m_{42} + m_{46})$
π^*_{0011}	$\hat{\pi}^*_{0011} = m_{39}/(m_{35} + m_{39})$
π^*_{1011}	$\hat{\pi}^*_{1011} = m_{47}/(m_{43} + m_{47})$
π^*_{0111}	$\hat{\pi}^*_{0111} = m_{40}/(m_{36} + m_{40})$
π^*_{1111}	$\hat{\pi}^*_{1111} = m_{48}/(m_{44} + m_{48})$

Table S8.4: Results for simulation studies 1a-18a,1b-18b,1c-18c on the performance of different causal estimators in various scenarios of confounding and misclassification in exposure and outcome. Abbreviations: PS, propensity score method ignoring misclassification; CCA, complete case analysis; GP, Gravel and Platt estimator ignoring exposure misclassification; IPWM, inverse probability weighting method for confounding and joint exposure and outcome misclassification; BSE, estimated standard error for the bias due to Monte Carlo error; SE, empirical standard error; SSE, sample standard error; CP, empirical coverage probability. In all scenarios, the true marginal log OR (estimand) was −0.4.

	Crude						PS					
Scenario	Bias	BSE	MSE	SE	SSE	CP	Bias	BSE	MSE	SE	SSE	CP
1a	0.401	0.003	0.167	0.081	0.083	0.004	0.399	0.004	0.173	0.117	0.120	0.080
2a	0.392	0.004	0.170	0.127	0.127	0.189	0.391	0.006	0.184	0.177	0.181	0.436
3a	0.400	0.003	0.167	0.083	0.083	0.005	0.391	0.004	0.167	0.119	0.119	0.104
4a	0.394	0.003	0.162	0.081	0.083	0.007	0.392	0.004	0.169	0.122	0.120	0.106
5a	0.398	0.002	0.162	0.061	0.062	0.000	0.398	0.002	0.162	0.061	0.062	0.000
6a	0.404	0.003	0.172	0.094	0.094	0.010	0.404	0.003	0.172	0.094	0.095	0.011
7a	0.399	0.002	0.163	0.062	0.062	0.000	0.399	0.002	0.163	0.062	0.062	0.000
8a	0.401	0.002	0.165	0.064	0.062	0.000	0.401	0.002	0.165	0.064	0.062	0.000
9a	0.400	0.002	0.164	0.064	0.062	0.000	0.400	0.002	0.164	0.064	0.062	0.000
10a	0.396	0.003	0.164	0.085	0.083	0.004	0.395	0.004	0.171	0.123	0.119	0.101
11a	0.396	0.004	0.173	0.128	0.127	0.176	0.388	0.006	0.185	0.187	0.182	0.455
12a	0.398	0.003	0.165	0.081	0.083	0.007	0.398	0.004	0.173	0.120	0.120	0.096
13a	0.399	0.003	0.166	0.083	0.083	0.004	0.395	0.004	0.171	0.120	0.119	0.102
14a	0.404	0.002	0.167	0.061	0.062	0.000	0.404	0.002	0.167	0.061	0.062	0.000
15a	0.398	0.003	0.167	0.092	0.094	0.011	0.398	0.003	0.167	0.092	0.095	0.012
16a	0.404	0.002	0.167	0.063	0.062	0.000	0.404	0.002	0.167	0.063	0.062	0.000
17a	0.399	0.002	0.163	0.061	0.062	0.000	0.399	0.002	0.163	0.061	0.062	0.000
18a	0.401	0.002	0.164	0.059	0.062	0.000	0.401	0.002	0.164	0.059	0.062	0.000

Table S8.4 continued.

	Crude						PS					
Scenario	Bias	BSE	MSE	SE	SSE	CP	Bias	BSE	MSE	SE	SSE	CP
1b	0.394	0.004	0.169	0.119	0.118	0.122	0.392	0.005	0.182	0.168	0.169	0.382
2b	0.382	0.006	0.179	0.183	0.184	0.492	0.379	0.008	0.213	0.264	0.258	0.738
3b	0.394	0.004	0.169	0.117	0.118	0.116	0.389	0.006	0.182	0.175	0.169	0.402
4b	0.401	0.004	0.174	0.117	0.118	0.102	0.389	0.006	0.182	0.176	0.168	0.392
5b	0.401	0.003	0.169	0.090	0.088	0.007	0.402	0.003	0.170	0.090	0.088	0.010
6b	0.407	0.004	0.183	0.132	0.134	0.133	0.407	0.004	0.183	0.131	0.135	0.136
7b	0.396	0.003	0.164	0.086	0.088	0.009	0.396	0.003	0.164	0.086	0.088	0.009
8b	0.395	0.003	0.164	0.086	0.088	0.005	0.395	0.003	0.164	0.086	0.088	0.004
9b	0.398	0.003	0.166	0.089	0.088	0.005	0.398	0.003	0.166	0.089	0.088	0.005
10b	0.397	0.004	0.171	0.117	0.118	0.100	0.396	0.005	0.185	0.167	0.170	0.387
11b	0.391	0.006	0.185	0.179	0.183	0.466	0.362	0.008	0.199	0.261	0.253	0.732
12b	0.401	0.004	0.174	0.118	0.118	0.109	0.391	0.005	0.182	0.173	0.169	0.394
13b	0.404	0.004	0.176	0.111	0.117	0.080	0.396	0.005	0.185	0.169	0.167	0.367
14b	0.400	0.003	0.168	0.087	0.088	0.008	0.400	0.003	0.168	0.087	0.088	0.006
15b	0.397	0.004	0.176	0.135	0.134	0.161	0.397	0.004	0.176	0.135	0.135	0.161
16b	0.401	0.003	0.168	0.087	0.088	0.006	0.400	0.003	0.168	0.087	0.088	0.006
17b	0.403	0.003	0.170	0.087	0.088	0.003	0.403	0.003	0.170	0.087	0.088	0.004
18b	0.400	0.003	0.168	0.087	0.088	0.004	0.400	0.003	0.168	0.088	0.088	0.003

Table S8.4 continued.

	Crude						PS					
Scenario	Bias	BSE	MSE	SE	SSE	CP	Bias	BSE	MSE	SE	SSE	CP
1c	0.394	0.009	0.232	0.277	0.275	0.698	0.366	0.013	0.292	0.398	0.391	0.871
2c	0.334	0.018	0.423	0.558	0.844	0.873	0.256	0.022	0.563	0.706	0.924	0.916
3c	0.383	0.009	0.222	0.274	0.276	0.739	0.371	0.013	0.297	0.399	0.393	0.875
4c	0.375	0.009	0.218	0.278	0.277	0.732	0.332	0.013	0.276	0.407	0.392	0.880
5c	0.405	0.006	0.204	0.200	0.199	0.470	0.405	0.006	0.205	0.201	0.199	0.474
6c	0.410	0.010	0.261	0.304	0.317	0.724	0.410	0.010	0.263	0.308	0.318	0.729
7c	0.406	0.006	0.203	0.196	0.199	0.469	0.406	0.006	0.204	0.198	0.200	0.469
8c	0.404	0.006	0.204	0.202	0.199	0.474	0.405	0.006	0.205	0.201	0.200	0.470
9c	0.406	0.006	0.202	0.192	0.198	0.468	0.404	0.006	0.201	0.193	0.199	0.470
10c	0.384	0.009	0.222	0.272	0.276	0.717	0.359	0.013	0.288	0.399	0.388	0.873
11c	0.358	0.014	0.324	0.443	0.825	0.864	0.296	0.020	0.471	0.619	0.902	0.923
12c	0.377	0.008	0.212	0.265	0.277	0.749	0.343	0.013	0.284	0.407	0.393	0.878
13c	0.377	0.008	0.210	0.259	0.276	0.741	0.341	0.013	0.274	0.397	0.390	0.888
14c	0.411	0.006	0.206	0.192	0.199	0.446	0.411	0.006	0.206	0.192	0.200	0.458
15c	0.393	0.009	0.241	0.294	0.315	0.764	0.393	0.009	0.242	0.296	0.316	0.770
16c	0.399	0.006	0.198	0.196	0.198	0.484	0.399	0.006	0.198	0.196	0.200	0.482
17c	0.395	0.006	0.193	0.191	0.199	0.471	0.394	0.006	0.191	0.190	0.199	0.474
18c	0.402	0.006	0.201	0.197	0.199	0.478	0.403	0.006	0.202	0.199	0.200	0.482

Table S8.4 continued.

	CCA						GP					
Scenario	Bias	BSE	MSE	SE	SSE	CP	Bias	BSE	MSE	SE	SSE	CP
1a	−0.011	0.010	0.091	0.302	0.315	0.932	−0.016	0.008	0.059	0.242	0.257	0.960
2a	−0.038	0.013	0.165	0.404	0.398	0.909	−0.024	0.010	0.108	0.328	0.353	0.956
3a	0.004	0.007	0.044	0.210	0.208	0.939	−0.013	0.005	0.028	0.167	0.165	0.943
4a	−0.050	0.014	0.189	0.432	0.441	0.905	−0.022	0.011	0.116	0.341	0.371	0.944
5a	−0.128	0.006	0.054	0.194	0.199	0.890	0.271	0.005	0.100	0.163	0.168	0.633
6a	−0.097	0.007	0.066	0.237	0.245	0.926	0.269	0.007	0.121	0.221	0.225	0.772
7a	−0.232	0.005	0.082	0.168	0.173	0.736	0.118	0.005	0.043	0.171	0.173	0.904
8a	−0.197	0.004	0.056	0.130	0.130	0.646	0.263	0.003	0.079	0.098	0.101	0.261
9a	−0.173	0.008	0.101	0.266	0.270	0.883	0.257	0.007	0.116	0.224	0.229	0.795
10a	0.017	0.005	0.029	0.169	0.170	0.953	0.003	0.005	0.022	0.147	0.152	0.946
11a	0.007	0.006	0.040	0.200	0.193	0.947	−0.014	0.006	0.039	0.196	0.203	0.952
12a	0.058	0.005	0.028	0.157	0.154	0.928	−0.003	0.004	0.019	0.138	0.136	0.943
13a	0.018	0.007	0.056	0.236	0.237	0.946	−0.003	0.006	0.037	0.192	0.194	0.940
14a	−0.092	0.003	0.018	0.099	0.105	0.864	0.265	0.003	0.079	0.091	0.095	0.191
15a	−0.051	0.004	0.016	0.115	0.119	0.933	0.264	0.004	0.084	0.121	0.124	0.421
16a	−0.166	0.003	0.036	0.094	0.092	0.559	0.138	0.003	0.028	0.096	0.096	0.710
17a	−0.116	0.003	0.023	0.095	0.094	0.762	0.264	0.003	0.076	0.080	0.082	0.110
18a	−0.115	0.005	0.035	0.149	0.144	0.859	0.266	0.004	0.088	0.131	0.128	0.455
1b	−0.078	0.015	0.226	0.469	0.491	0.899	−0.036	0.011	0.130	0.359	0.428	0.958
2b	−0.117	0.019	0.375	0.601	0.900	0.887	−0.097	0.016	0.265	0.505	0.861	0.938
3b	−0.020	0.010	0.091	0.301	0.300	0.919	−0.019	0.007	0.055	0.233	0.240	0.939
4b	−0.093	0.020	0.407	0.631	1.158	0.899	−0.045	0.016	0.253	0.501	1.087	0.944
5b	−0.145	0.009	0.103	0.286	0.286	0.903	0.269	0.008	0.132	0.244	0.244	0.799
6b	−0.109	0.011	0.131	0.345	0.362	0.930	0.280	0.010	0.177	0.314	0.339	0.862
7b	−0.213	0.007	0.101	0.237	0.250	0.865	0.134	0.008	0.076	0.241	0.252	0.926
8b	−0.209	0.006	0.079	0.187	0.186	0.775	0.259	0.004	0.087	0.140	0.144	0.570
9b	−0.175	0.012	0.184	0.392	0.411	0.902	0.263	0.010	0.174	0.325	0.339	0.883

Table S8.4 continued.

	CCA						GP					
Scenario	Bias	BSE	MSE	SE	SSE	CP	Bias	BSE	MSE	SE	SSE	CP
10b	0.011	0.007	0.056	0.237	0.244	0.957	−0.002	0.007	0.050	0.223	0.221	0.939
11b	0.001	0.009	0.083	0.288	0.276	0.918	−0.019	0.010	0.093	0.304	0.304	0.938
12b	0.058	0.007	0.050	0.216	0.223	0.953	−0.007	0.006	0.038	0.194	0.197	0.949
13b	−0.015	0.011	0.122	0.350	0.345	0.934	−0.023	0.009	0.077	0.277	0.287	0.950
14b	−0.092	0.005	0.030	0.146	0.148	0.889	0.263	0.004	0.088	0.136	0.136	0.505
15b	−0.060	0.005	0.033	0.170	0.170	0.929	0.263	0.006	0.101	0.177	0.183	0.712
16b	−0.171	0.004	0.047	0.132	0.131	0.741	0.139	0.004	0.038	0.136	0.138	0.820
17b	−0.121	0.004	0.032	0.134	0.135	0.842	0.263	0.004	0.082	0.115	0.118	0.388
18b	−0.113	0.007	0.055	0.206	0.207	0.904	0.264	0.006	0.102	0.178	0.185	0.702
1c	−1.163	0.098	10.972	3.101	3.092	0.792	−0.994	0.095	10.003	3.003	3.085	0.878
2c	−2.086	0.131	21.614	4.155	3.415	0.733	−1.979	0.130	20.875	4.118	3.689	0.835
3c	−0.184	0.029	0.880	0.920	1.477	0.887	−0.102	0.024	0.574	0.751	1.412	0.939
4c	−2.730	0.148	29.275	4.671	3.710	0.722	−2.295	0.142	25.436	4.491	3.974	0.916
5c	−0.254	0.029	0.904	0.916	1.764	0.891	0.288	0.018	0.409	0.571	1.106	0.951
6c	−0.548	0.067	4.832	2.129	2.684	0.893	0.402	0.046	2.276	1.454	2.782	0.969
7c	−0.223	0.020	0.467	0.646	1.303	0.912	0.109	0.020	0.424	0.642	1.322	0.952
8c	−0.236	0.014	0.258	0.450	0.508	0.891	0.279	0.011	0.197	0.345	0.369	0.883
9c	−0.662	0.079	6.624	2.487	3.186	0.896	0.314	0.053	2.915	1.678	2.454	0.972
10c	−0.105	0.019	0.376	0.604	0.740	0.897	−0.098	0.017	0.294	0.534	0.790	0.943
11c	−0.113	0.025	0.649	0.798	1.174	0.906	−0.183	0.031	1.006	0.986	1.758	0.931
12c	0.010	0.018	0.313	0.560	0.598	0.938	−0.053	0.021	0.455	0.673	0.703	0.937
13c	−0.170	0.036	1.330	1.140	1.967	0.919	−0.129	0.030	0.920	0.950	1.983	0.948
14c	−0.107	0.011	0.133	0.349	0.360	0.922	0.286	0.010	0.187	0.324	0.351	0.882
15c	−0.088	0.013	0.187	0.424	0.417	0.913	0.276	0.015	0.314	0.488	0.664	0.951
16c	−0.159	0.009	0.115	0.299	0.311	0.924	0.143	0.010	0.124	0.323	0.357	0.952
17c	−0.133	0.010	0.114	0.311	0.323	0.918	0.259	0.009	0.144	0.277	0.304	0.872
18c	−0.148	0.016	0.264	0.492	0.593	0.938	0.280	0.014	0.278	0.447	0.510	0.926

Table S8.4 continued.

	IPWM					
Scenario	Bias	BSE	MSE	SE	SSE	CP
1a	−0.016	0.008	0.059	0.242	0.257	0.960
2a	−0.024	0.010	0.108	0.328	0.353	0.956
3a	−0.013	0.005	0.028	0.167	0.165	0.943
4a	−0.022	0.011	0.116	0.341	0.371	0.944
5a	−0.004	0.006	0.035	0.186	0.194	0.954
6a	−0.010	0.008	0.060	0.244	0.252	0.952
7a	−0.013	0.005	0.030	0.172	0.179	0.951
8a	0.003	0.004	0.015	0.122	0.125	0.952
9a	−0.019	0.008	0.065	0.255	0.265	0.948
10a	0.003	0.005	0.022	0.147	0.152	0.946
11a	−0.014	0.006	0.039	0.196	0.203	0.952
12a	−0.003	0.004	0.019	0.138	0.136	0.943
13a	−0.003	0.006	0.037	0.192	0.194	0.940
14a	−0.005	0.003	0.011	0.104	0.106	0.962
15a	−0.001	0.004	0.017	0.129	0.134	0.963
16a	0.010	0.003	0.010	0.099	0.099	0.947
17a	0.001	0.003	0.009	0.096	0.095	0.943
18a	0.001	0.005	0.022	0.148	0.144	0.949
1b	−0.036	0.011	0.130	0.359	0.428	0.958
2b	−0.097	0.016	0.265	0.505	0.861	0.938
3b	−0.019	0.007	0.055	0.233	0.240	0.939
4b	−0.045	0.016	0.253	0.501	1.087	0.944
5b	−0.017	0.009	0.082	0.286	0.284	0.942
6b	−0.014	0.011	0.129	0.359	0.386	0.958
7b	0.004	0.008	0.059	0.243	0.261	0.969
8b	−0.004	0.006	0.032	0.180	0.181	0.958
9b	−0.025	0.012	0.141	0.374	0.415	0.956

Table S8.4 continued.

	IPWM					
Scenario	Bias	BSE	MSE	SE	SSE	CP
10b	−0.002	0.007	0.050	0.223	0.221	0.939
11b	−0.019	0.010	0.093	0.304	0.304	0.938
12b	−0.007	0.006	0.038	0.194	0.197	0.949
13b	−0.023	0.009	0.077	0.277	0.287	0.950
14b	−0.003	0.005	0.022	0.147	0.152	0.960
15b	−0.006	0.006	0.035	0.187	0.198	0.963
16b	0.010	0.004	0.020	0.142	0.143	0.956
17b	−0.003	0.004	0.017	0.131	0.136	0.956
18b	0.010	0.006	0.042	0.205	0.207	0.955
1c	−0.994	0.095	10.003	3.003	3.085	0.878
2c	−1.979	0.130	20.875	4.118	3.689	0.835
3c	−0.102	0.024	0.574	0.751	1.412	0.939
4c	−2.295	0.142	25.436	4.491	3.974	0.916
5c	−0.101	0.029	0.849	0.916	1.771	0.950
6c	−0.373	0.069	4.896	2.181	3.041	0.978
7c	−0.027	0.022	0.470	0.685	1.298	0.961
8c	−0.019	0.014	0.200	0.447	0.527	0.953
9c	−0.372	0.068	4.769	2.152	2.579	0.989
10c	−0.098	0.017	0.294	0.534	0.790	0.943
11c	−0.183	0.031	1.006	0.986	1.758	0.931
12c	−0.053	0.021	0.455	0.673	0.703	0.937
13c	−0.129	0.030	0.920	0.950	1.983	0.948
14c	−0.003	0.011	0.130	0.360	0.396	0.967
15c	−0.018	0.016	0.263	0.512	0.705	0.984
16c	0.014	0.011	0.114	0.338	0.371	0.966
17c	−0.005	0.010	0.101	0.318	0.349	0.952
18c	−0.002	0.016	0.249	0.499	0.613	0.962