Correction methods for measurement error in epidemiologic research

Nab, L.

8

Quantitative bias analysis for a misclassified confounder in marginal structural models

Observational data are increasingly used with the aim of estimating causal effects of treatments, through careful control for confounding. Marginal structural models estimated using inverse probability weighting (MSMs-IPW), like other methods to control for confounding, assume that confounding variables are measured without error. The average treatment effect estimator in a MSM-IPW may however be biased when a confounding variable is error-prone. Using the potential outcome framework, we derive expressions for the bias due to confounder misclassification in analyses that aim to estimate the average treatment effect using a MSM-IPW. We compare this bias with the bias due to confounder misclassification in analyses based on a conditional regression model. Focus is on a point-treatment study with a continuous outcome. Compared to bias in the average treatment effect estimator from a conditional model, the bias in MSM-IPW can be different in magnitude, but is equal in sign. Also, we use a simulation study to investigate the finite sample performance of MSM-IPW and conditional models when a confounding variable is misclassified. Simulation results indicate that confidence intervals of the treatment effect obtained from MSM-IPW are generally wider and coverage of the true treatment effect is higher compared to a conditional model, ranging from over-coverage if there is no confounder misclassification to under-coverage when there is confounder misclassification. We illustrate in a study of blood pressure lowering therapy, how the bias expressions can be used to inform a quantitative bias analysis to study the impact of confounder misclassification, supported by an online tool.

This chapter is based on: L. Nab, R.H.H. Groenwold, M. van Smeden and R.H. Keogh, Quantitative bias analysis for a misclassified confounder: A comparison between marginal structural models and conditional models for point treatments, Epidemiology, 31 (6) (2020) 796–805. doi:10.1097/EDE.0000000000001239

8.1. Introduction

The aim of many observational epidemiologic studies is to estimate a causal relation between an exposure and an outcome, through careful control for confounding. In the case of a point-treatment, that is estimating the effect of a treatment at a single time point on a subsequent outcome, many methods exist that aim to estimate average treatment effects. These include traditional conditional regression analysis as well as marginal structural models estimated using inverse probability weighting (MSMs-IPW) [1, 2]. Unlike conditional regression, MSMs extend to estimation of joint treatment effects over multiple time points in longitudinal settings with time-dependent confounding [1, 3].

To obtain valid inference, MSMs-IPW, like other methods to control for confounding, assume that confounding variables are measured without error, an assumption hardly ever warranted in observational epidemiologic research [4–7]. A type of measurement error is classification error, which occurs when categorical variables are misclassified. For instance, smoking status (smoker vs non-smoker) is prone to classification error, but has been used as a confounding variable in studies investigating dialysis on mortality [8] and iron supplement use during pregnancy on anemia at delivery [9]. Another example of the use of a potentially misclassified confounding variable is alcohol use during pregnancy (yes vs no) in studies investigating associations between exposure to triptans during fetal life and risk of externalizing and internalizing behaviors in children [10]. In all aforementioned examples, MSMs were used to estimate the exposure–outcome relation, but the assumption of error-free confounding variables is possibly violated and may lead to bias in the treatment effect estimator.

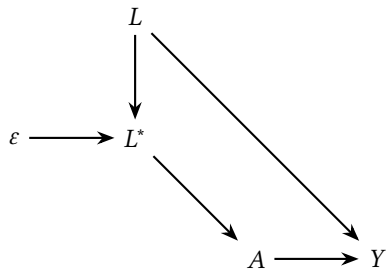
There is a substantial literature on bias due to measurement error in confounding variables in conditional regression analyses [11–15], but the impact of measurement error in confounding variables in causal inference methods, such as MSMs-IPW, has not received much attention. One exception is a study by Regier et al. that showed by means of a simulation study that measurement error in continuous confounding variables can introduce bias in the ATE in a point-treatment study [16]. McCaffrey et al. proposed a weighting method to restore the treatment effect estimator when covariates are measured with error [17].

We provide a discussion of measurement error in a confounding variable. In addition, we derive expressions that quantify the bias in the average treatment effect if a dichotomous confounding variable is misclassified, focusing on a point-treatment study with a continuous outcome. These expressions allow us 1) to quantify the bias due to classification error in a confounding variable in MSMs-IPW, and to compare this with the bias from a conditional regression analysis and 2) to inform quantitative bias analyses [18–20]. We use simulation results to study the finite sample performance of a MSM-IPW compared to that of conditional regression models if classification error in a confounding variable is present. We illustrate our quantitative bias analysis in a study of the effect of blood pressure lowering drugs on blood pressure.

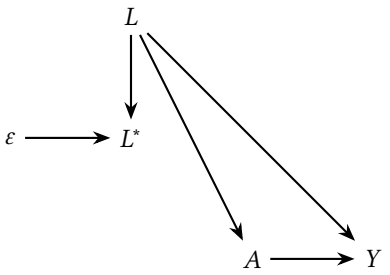
8.2. Settings and impact of measurement error, notation and assumptions

Let A denote the treatment indicator and Y the outcome. Let there be a variable L that confounds the association between treatment and outcome and suppose that, instead of confounding variable L, the error-prone confounding variable L^* is observed. We consider two settings in which measurement error in confounding variables may occur and discuss the impact of measurement error in both settings.

Settings and impact of measurement error. The directed acyclic graph (DAG) in Figure 8.1a illustrates **setting 1**. In this setting, treatment initiation is based on the error-prone confounding variable. Consider for example a study investigating the relation between the use of antidepressant drugs (A) and the risk of a hip fracture (Y) [21]. Benzodiazepine use may be a confounding variable, but is prone to classification error since only prescription information may be available and over-the-counter use is often unknown. The clinician initiating the antidepressant drugs might not know their patient's over-the-counter use and initiates treatment based on the observed error-prone benzodiazepine use (L^*) instead of actual use (L), as depicted in Figure 8.1a. Here, conditioning on the error-prone L^* will block the backdoor path from treatment A to outcome Y. Thus, it is sufficient to control for the error-prone confounding variable to estimate the causal effect of treatment on outcome. This means that measurement error in a confounding variable will not always lead to bias.



(a) **Setting 1**: treatment A is initiated based on the error-prone confounding variable L^*



(b) **Setting 2**: treatment A is initiated based on confounding variable L

Figure 8.1: Two settings of measurement error ε in variable L that confounds the association between treatment A and outcome Y illustrated in directed acyclic graphs

The DAG in Figure 8.1b illustrates **setting 2**, in which treatment initiation is based on L, but only a proxy of L is observed (L^*). An example here might be a study investigating the effect of influenza vaccination (A) on mortality (Y) in the elderly population [22]. Frailty (L) possibly confounds the association between influenza vaccination and mortality. Frailty is observed by a clinician, but only a proxy of frailty (L^*) may be available in electronic health records, as depicted in Figure 8.1b. Here, conditioning on L^* will not fully adjust for confounding by L, because conditioning on L^* does not block the backdoor path from A to Y via L.

Notation and assumptions. We will now continue investigating the impact of classification error in setting 2, by focusing on the setting where L is a dichotomous confounding variable and Y a continuous outcome. We use the potential outcomes framework [23, 24]. Let $Y^{a=0}$ denote the outcome that an individual would have had if treatment A was set to $a = 0$, and let $Y^{a=1}$ denote the outcome if treatment A was set to $a = 1$. We assume that L^* is non-differentially misclassified with respect to the outcome ($L^* \perp\!\!\!\perp Y|L$) and to the treatment ($L^* \perp\!\!\!\perp A|L$). Let p_1 denote the sensitivity of L^* and $1 - p_0$ the specificity of L^* (i.e., $P(L^*|L = l) = p_l$). We also denote the probability of treatment given the level of L by $P(A = 1|L = l) = \pi_l$ and the prevalence of L by $P(L = 1) = \lambda$. Here, we assume that $0 < \lambda < 1$ since we are not interested in populations where L is present or absent in everyone. Finally, we assume no measurement error in exposure and outcome.

We also assume that the following causal assumptions are satisfied to recover the causal effect of treatment on the outcome. Under the consistency assumption, we require that we observe $Y = Y^{a=0}$ if the individual is not exposed, or $Y = Y^{a=1}$ if the individual is exposed [25]. Further, we assume that the potential outcome Y^a for an individual does not depend on treatments received by other individuals and that there are not multiple versions of treatment, also referred to as Stable-Unit-Treatment-Value-Assumption [26]. Additionally, we assume conditional exchangeability, i.e., given any level of L, if the untreated group had in fact received treatment, then their expected outcome would have been the same as that in the treated, and vice versa [25]. In notation, $A \perp\!\!\!\perp Y^a|L$, for $a = 0, 1$. Finally, we assume $\pi_L > 0$ for $L = 0, 1$ (positivity) [27].

For causal contrasts, we compare expected potential outcomes (i.e., counterfactual outcomes) under the two different treatments. The average causal effect of the treatment on the outcome is $\beta = \mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$. Under the above defined assumptions, the conditional effect of treatment A on outcome Y can be defined through the following linear model:

$$\mathrm{E}[Y^a|L] = \mathrm{E}[Y|A = a, L] = \alpha + \beta a + \gamma L. \tag{8.1}$$

Estimates for β in the above model can be obtained by fitting a conditional regression model. Alternatively, the effect of treatment A on outcome Y may be estimated by fitting a MSM:

$$\mathrm{E}[Y^a] = \alpha_{\mathrm{msm}} + \beta a, \qquad \text{where} \quad \alpha_{\mathrm{msm}} = \alpha + \gamma \mathrm{E}[L]. \tag{8.2}$$

Estimates for β in the above model can be obtained by IPW estimation: by fitting a linear regression model for Y on A where the contribution of each individual is weighted by 1 over the probability of that individual's observed treatment given L [28], estimating the marginal treatment effect. Since our focus is on linear models and we make the simplifying assumption that the effect of A on Y does not vary between strata of L, the conditional and marginal treatment effects, denoted by β in model (8.1) and (8.2), respectively, are equal. This is not generally true for non-linear models due to non-collapsibility [28]. We assume that the effect of A on Y does not vary between strata of L, to derive bias expressions that are easier to use in practice and require fewer parameters [29].

8.3. Quantification of bias due to classification error in a confounding variable

Our aim is to study the effect of using the misclassified confounding variable L^* in place of the confounding variable L in the conditional regression model or in the model for the

weights used to fit the MSM on the average treatment effect estimator in the setting where L, not L^*, influences treatment initiation (setting 2 above).

Conditional model. By the law of total expectation, the expected value of the outcome Y given treatment A and L^* is (see S8.1 section Conditional model for further detail),

$$\begin{aligned}\mathsf{E}[Y|A=a,L^*] = \mathsf{E}_{L|A=a,L^*}\big[\mathsf{E}[Y|A=a,L^*,L]\big] &= \{\alpha+\gamma\phi_{00}+\delta u_0\}\\ &+ \{\beta+\gamma(\phi_{10}-\phi_{00})+\delta u_A\}a\\ &+ \{\gamma(\phi_{01}-\phi_{00})+\delta u_{L^*}\}L^*,\end{aligned}$$

where $\phi_{al^*} = P(L=1|A=a,L^*=l^*)$, $\delta = \mathsf{E}[Y|A=1,L^*=1] = \gamma(\phi_{11}-\phi_{10}-\phi_{01}+\phi_{00})$ and u_0, u_A, u_{L^*} represent the coefficients of the linear model $\mathsf{E}[AL^*|A,L^*] = u_0 + u_A A + u_{L^*}L^*$, modelling the mean of A times L^* (i.e., AL^*) given A and L^* (see next paragraph for an explanation of why these appear). The coefficient for treatment A in the above model is $\beta+\gamma(\phi_{10}-\phi_{00})+\delta u_A$, and is therefore biased for the parameter of interest (i.e., β). By rewriting u_A in terms of λ, π_0, π_1, p_0 and p_1 (see S8.1 section Conditional model), we find that the bias due to classification error in L^* in the average treatment effect in a conditional regression model is,

$$\begin{aligned}\text{Bias}_{\text{cm}}(\beta) = \gamma(\phi_{10}-\phi_{00})&\left(1-\ell\times\left\{\frac{\pi_1^*(1-\pi_1^*)}{\pi_1^*(1-\pi_1^*)\ell+\pi_0^*(1-\pi_0^*)(1-\ell)}\right\}\right)\\ +\gamma(\phi_{11}-\phi_{01})&\left(\ell\times\left\{\frac{\pi_1^*(1-\pi_1^*)}{\pi_1^*(1-\pi_1^*)\ell+\pi_0^*(1-\pi_0^*)(1-\ell)}\right\}\right),\end{aligned} \tag{8.3}$$

where $\pi_{l^*}^* = P(A=1|L^*=l^*)$, $\ell = P(L^*=1)$ (see S8.1 section Conditional model for a derivation).

We focused on a model for Y conditional on A and L^* which includes only main effects of A and L^*, as this is typically done in practice when replacing L with L^*. In fact, it can be shown that when the model for Y given A and L includes only main effects of A and L, the implied correctly specified model for Y given A and L^* also includes an interaction between A and L^*, explaining the appearance of u_0, u_A and u_L in the above since the interaction is not modeled. See S8.1 section Conditional model for the bias in case an interaction is modelled.

MSM-IPW. A MSM-IPW proceeds by fitting a linear regression for outcome Y on treatment A where the contribution of each individual is weighted by 1 over the probability of that individual's observed treatment given misclassified L^* [28]. An estimator for the average treatment effect β is,

$$\hat{\beta} = \frac{\sum_{i=1}^n \frac{1}{P(A_i|L_i^*)}(Y_i-\overline{Y}_w)(A_i-\overline{A}_w)}{\sum_{i=1}^n \frac{1}{P(A_i|L_i^*)}(A_i-\overline{A}_w)^2} \quad \text{where,} \quad \overline{Y}_w = \frac{\sum_{i=1}^n Y_i/P(A_i|L_i^*)}{\sum_{i=1}^n 1/P(A_i|L_i^*)}$$

$$\text{and,} \quad \overline{A}_w = \frac{\sum_{i=1}^n A_i/P(A_i|L_i^*)}{\sum_{i=1}^n 1/P(A_i|L_i^*)}.$$

It can be shown that $\mathsf{E}[\hat{\beta}] = \beta+\gamma(\phi_{10}-\phi_{00})(1-\ell)+\gamma(\phi_{11}-\phi_{01})\ell$. Consequently, the bias in the average treatment effect β in a MSM-IPW is,

$$\text{Bias}_{\text{msm}}(\beta) = \gamma(\phi_{10}-\phi_{00})(1-\ell)+\gamma(\phi_{11}-\phi_{01})\ell. \tag{8.4}$$

We refer to S8.1 section Marginal structural model estimated using inverse probability weighting for a derivation of the above formula.

8.3.1. Exploration of bias

To study the bias due to misclassification from the conditional model and MSM-IPW, we explore bias expressions (8.3) and (8.4).

Null-bias. To confirm the derived bias expressions, we consider three trivial conditions where bias in the average treatment effect is expected to be null, in line with general understanding of causal inference [30]. (1) If there is no classification error in L^*, i.e., specificity is 1 ($p_0 = 0$) and sensitivity is 1 ($p_1 = 1$), it follows that L corresponds to L^*, irrespective of treatment level (i.e., $\phi_{10} = 0$, $\phi_{00} = 0$, $\phi_{11} = 1$ and $\phi_{01} = 1$). (2) If the true relation between L and Y is null (i.e., γ is zero, thus there is no arrow from L to Y in Figure (8.1b)). (3) If L does not affect the probability of receiving treatment (i.e., $\pi_0 = \pi_1$, thus there is no arrow from L to A in Figure (8.1b)), the probability that L is 1 depends on the value of L^* but no longer on A (i.e., $\phi_{00} = \phi_{10}$ and $\phi_{01} = \phi_{11}$). Bias is null under these conditions for both models (MSM-IPW and conditional model). Since the bias expressions are strictly monotonic, the bias in a MSM-IPW cannot be negative if the bias in the conditional model is positive and vice versa (i.e., the bias will be in the same direction for both models).

Equal biases. The bias in the average treatment effect from the conditional regression analysis is equal to that from the MSM-IPW if bias in both models is null (see above). We also see that bias expressions (8.3) and (8.4) show that bias for the two methods is equal if the term between curly brackets in equation (8.3) is equal to 1, which is the case if: (i) $\ell = 1$; (ii) $\pi_0^* = \pi_1^*$; (iii) $\pi_0^* = 1 - \pi_1^*$. If conditions i and/or ii are met, there is no bias in a MSM-IPW nor in a conditional model. Under condition iii, bias is generally non-null (except if for example $\gamma = 0$, see null-bias).

Sign and magnitude of bias. Figures 8.2-8.4 illustrate the contributions to bias in the average treatment effect estimator due to misclassification components (sensitivity and specificity) and due to confounding components (prevalence of confounding variable, strength of association between confounding variable and treatment and outcome) in a conditional model and a MSM-IPW, obtained by using the bias expressions.

Figure 8.2 shows that: (1) the bias is positive if both the association between L and treatment and, L and outcome are positive (i.e., $\pi_1 > \pi_0$ and $\gamma = 2$, respectively), and (2) the bias is greater if the difference between π_1 and π_0 is greater (i.e., if the strength of the association between L and treatment is greater). In contrast, the bias is negative if $\pi_1 < \pi_0$, while γ is positive. In case $\gamma = -2$, Figure 8.2 is mirrored in $y = 0$ and consequently, bias is negative if $\pi_1 > \pi_0$ and positive if $\pi_1 < \pi_0$. An increment in γ will result in greater bias and steeper curves in Figure 8.2. Figure 8.3 shows that the magnitude of the bias depends on the prevalence of L. Further, it shows that bias is greater if the strength of association between L and treatment is greater. Figure 8.4 shows that, generally, the bias is greater if L^* has lower specificity and sensitivity. Moreover, for a fixed sensitivity, bias is minimal if specificity equals 1 and is maximal if 1 minus specificity equals sensitivity; by fixing specificity, bias is minimal if sensitivity equals 1 and is maximal if sensitivity equals 1 minus specificity. Figure 8.4 shows that the bias is greater if the strength of the association between L and treatment is greater. An increment in γ will result in greater bias and steeper curves in Figure 8.4. An online application can be used to obtain bias plots for other combinations of the parameters available at: https://lindanab.shinyapps.io/SensitivityAnalysis.

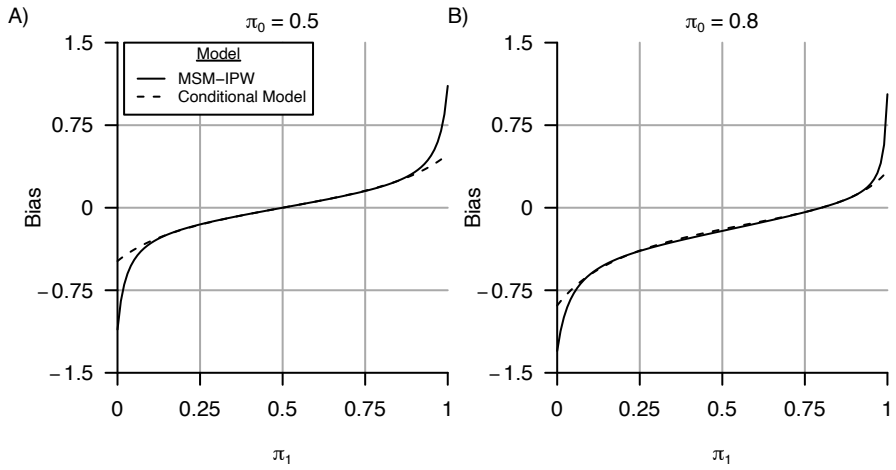


Figure 8.2: Visualisation of the direction and magnitude of the bias in the average treatment effect in relation to the prevalence of treatment among individuals with the confounding variable present. In this visualisation, the confounding variable L is misclassified with a sensitivity of 0.9 and specificity of 0.95. Consequently, the average treatment effect estimated in a MSM-IPW or conditional regression model is biased, independent of true average treatment effect. The prevalence of L is 50% (i.e., $P(L = 1) = 0.5$). The direction and magnitude of the bias depend on: (1) the strength and direction of the association between L and treatment (denoted by $\pi_1 = P(\text{treatment} = 1|L = 1)$ and $\pi_0 = P(\text{treatment} = 1|L = 0)$, here set at $\pi_0 = 0.5$ in the left-hand-side plot and $\pi_0 = 0.8$ in the right-hand-side plot); and (2) the strength and direction of the association between L and the outcome (denoted by γ in the text and here set at $\gamma = 2$). Larger values of γ will result in steeper curves; $\gamma = -2$ will mirror the graph in $y = 0$.

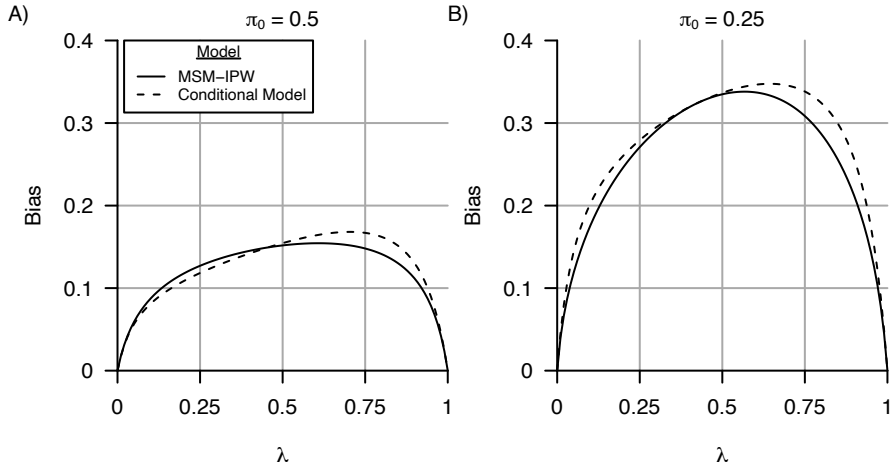


Figure 8.3: Visualisation of the magnitude of the bias in the average treatment effect in relation to the prevalence of a confounding variable. In this visualisation, the confounding variable L is misclassified with a sensitivity of 0.9 and specificity of 0.95. Consequently, the average treatment effect estimated in a MSM-IPW or conditional regression model is biased, independent of true average treatment effect. The confounding variable is positively associated with treatment (i.e., here $\pi_1 > \pi_0$, where $\pi_1 = P(\text{treatment} = 1|L = 1)$ and $\pi_0 = P(\text{treatment} = 1|L = 0)$), and outcome (denoted by γ in the text and here set at $\gamma = 2$). The magnitude of the bias depends on the prevalence of the confounding variable (i.e., $P(L = 1)$). Larger values of γ will result in steeper curves.

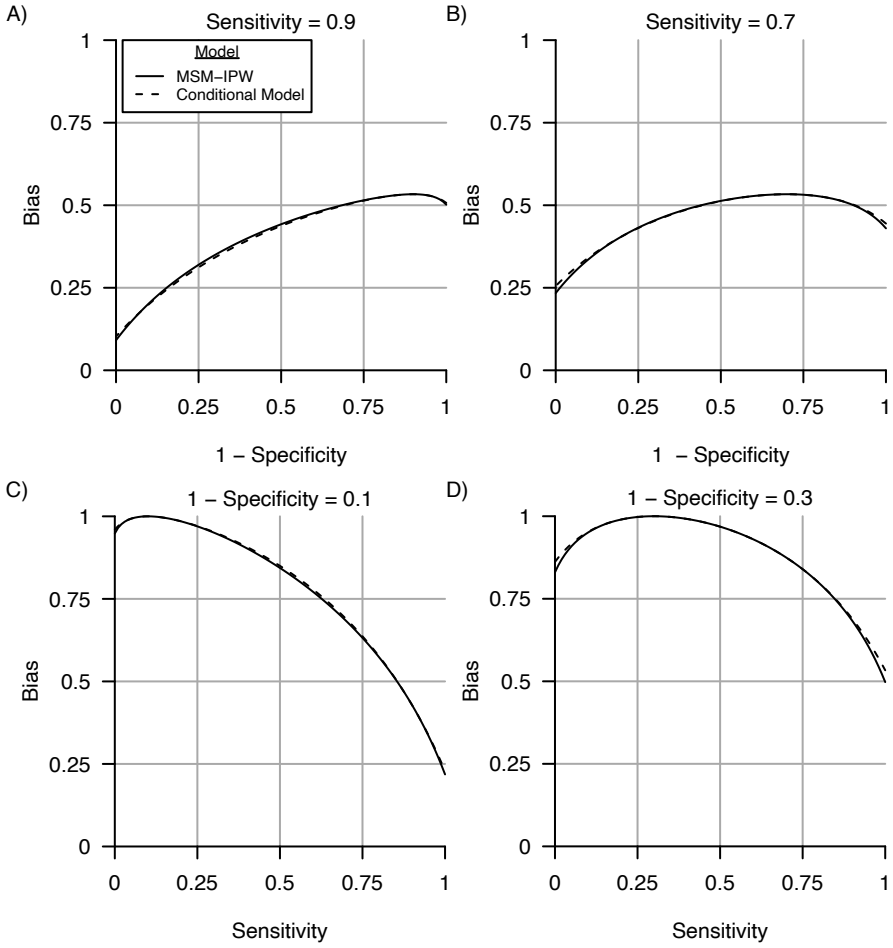


Figure 8.4: Visualisation of the magnitude of the bias in the average treatment effect in relation to specificity and sensitivity of a misclassified confounding variable. In this visualisation, the prevalence of the confounding variable L is 50% (i.e., $P(L = 1) = 0.5$), the association between L and treatment (denoted by $\pi_1 = P(\text{treatment} = 1|L = 1)$ and $\pi_0 = P(\text{treatment} = 1|L = 0)$) and outcome is positive (denoted by γ in the text and here set at $\gamma = 2$). Given these values, if L is misclassified, the average treatment effect estimated in a MSM-IPW or conditional regression model is biased, independent of true average treatment effect. The magnitude of the bias depends on the specificity and sensitivity of L and is maximal if sensitivity equals 1 minus specificity. The strength of the association between L and treatment is greater in the right-hand-side plot ($\pi_0 = 0.25, \pi_1 = 0.75$) compared to the left-hand-side plot ($\pi_0 = 0.5, \pi_1 = 0.75$) and consequently, bias is greater. Larger values of γ will result in steeper curves.

8.3.2. Simulation study

We conducted a simulation study to study the finite sample properties of MSMs-IPW and conditional models if there is classification error in the confounding variable. Five-thousand data sets were generated with sample sizes of 1,000 and 100, using the following data generating mechanisms:

$$\begin{aligned} L &\sim \text{Bern}(\lambda), \quad A|L \sim \text{Bern}\left(\pi_0^{(1-L)}\pi_1^{L}\right), \\ L^*|L &\sim \text{Bern}\left(p_0^{(1-L)}p_1^{L}\right) \quad \text{and,} \quad Y|A,L \sim \text{N}(1+\beta A+\gamma L, 1). \end{aligned}$$

We studied five different scenarios, of which the parameters values can be found in Table 8.1. In all scenarios, the average treatment effect β (estimand) is 1 and the association between the confounding variable L and outcome Y is 2 (i.e., $\gamma = 2$). In scenario 0, we assume no classification error. In scenarios 1-4, we assume that L^* has a specificity of 0.95 (i.e., $p_0 = 0.05$) and a sensitivity of 0.90 (i.e., $p_1 = 0.9$). In scenario 1, bias in the average treatment effect β is expected to be negative since the probability of receiving treatment given that L is not present is greater than receiving treatment given that L is present, and the association between L and Y is positive (i.e., $\pi_0 > \pi_1$ and $\gamma = 2$). In contrast, in scenario 2 and 3, bias in the average treatment effect is expected to be positive, since $\pi_0 < \pi_1$ and $\gamma = 2$. Further, after investigation of Figure 8.3, we expect that bias in the average treatment effect estimated in a conditional model is greater than that in a MSM-IPW in scenario 2 and 3. Finally, in scenario 4, we expect that bias in the average treatment effect from the conditional model is equal to that in a MSM-IPW.

Model estimation and performance measures. We obtained the average treatment effect β (estimand) by fitting a conditional model using conditional regression and by fitting a MSM-IPW, both using the misclassified L^* instead of L from the data generating mechanism. For the MSM-IPW analysis we used the R package `ipw` [31] [32]. Performance of both models was evaluated in terms of the bias, the mean squared error of the estimated treatment effect (MSE), the percentages of 95% confidence intervals that contain the true value of the estimand (coverage), the empirical standard deviation of the estimated treatment effects (empSE) and mean model based standard error of the estimated treatment effect. We estimated robust model based standard errors of the average treatment effect in a MSM-IPW using the R package `survey` [33]. We calculated Monte Carlo standard errors for all performance measures [34], using the R package `rsimsum` [35]. Additionally, we calculated the theoretical bias of the average treatment effect in both methods based on the bias expressions (8.3) and (8.4).

Table 8.1: Values of the parameters in the five different simulation scenarios

Scenario Number	**Parameters**						
	p_0	p_1	λ	π_0	π_1	β	γ
0	0	1	0.50	0.50	0.75	1	2
1	0.05	0.90	0.50	0.90	0.45	1	2
2	0.05	0.90	0.80	0.25	0.75	1	2
3	0.05	0.90	0.80	0.50	0.75	1	2
4	0.05	0.90	0.45	0.50	0.75	1	2

Table 8.2: Results of simulation study studying the finite-sample properties of a marginal structural models estimated using inverse probability weighting (MSM-IPW) and a conditional model (CM) if there is classification error in the confounding variable. Bias formula indicates the bias based on bias expressions (8.3) and (8.4) in the text, MSE mean squared error, EmpSE empirical SE and ModelSE model based SE.

Method	Sample Size	Scenario[a]	Bias Formula	Bias	MSE	Coverage	EmpSE	ModelSE
MSM-IPW	1,000	0	0.00	0.00 (0.001)	0.00 (0.000)	0.99 (0.001)	0.07 (0.001)	0.10 (0.000)
		1	-0.42	-0.42 (0.001)	0.18 (0.001)	0.03 (0.002)	0.10 (0.001)	0.11 (0.000)
		2	0.14	0.14 (0.001)	0.03 (0.000)	0.67 (0.007)	0.08 (0.001)	0.09 (0.000)
		3	0.29	0.29 (0.001)	0.09 (0.001)	0.08 (0.004)	0.08 (0.001)	0.09 (0.000)
		4	0.15	0.15 (0.001)	0.03 (0.000)	0.68 (0.007)	0.08 (0.001)	0.10 (0.000)
	100	**0**	0.00	0.00 (0.003)	0.05 (0.001)	0.99 (0.001)	0.22 (0.002)	0.31 (0.000)
		1	-0.42	-0.42 (0.005)	0.29 (0.005)	0.78 (0.006)	0.34 (0.003)	0.37 (0.001)
		2	0.14	0.14 (0.004)	0.08 (0.002)	0.94 (0.003)	0.25 (0.003)	0.29 (0.000)
		3	0.29	0.29 (0.004)	0.15 (0.002)	0.84 (0.005)	0.26 (0.003)	0.28 (0.000)
		4	0.15	0.15 (0.004)	0.08 (0.002)	0.95 (0.003)	0.25 (0.002)	0.31 (0.000)
CM	1,000	0	0.00	0.00 (0.001)	0.00 (0.000)	0.95 (0.003)	0.07 (0.001)	0.07 (0.000)
		1	-0.34	-0.34 (0.001)	0.12 (0.001)	0.02 (0.002)	0.09 (0.001)	0.08 (0.000)
		2	0.16	0.16 (0.001)	0.03 (0.000)	0.46 (0.007)	0.08 (0.001)	0.08 (0.000)
		3	0.32	0.32 (0.001)	0.11 (0.001)	0.02 (0.002)	0.08 (0.001)	0.08 (0.000)
		4	0.15	0.15 (0.001)	0.03 (0.000)	0.49 (0.007)	0.08 (0.001)	0.07 (0.000)
	100	0	0.00	0.00 (0.003)	0.05 (0.001)	0.95 (0.003)	0.22 (0.002)	0.22 (0.000)
		1	-0.34	-0.33 (0.004)	0.19 (0.003)	0.73 (0.006)	0.29 (0.003)	0.27 (0.000)
		2	0.16	0.16 (0.004)	0.09 (0.002)	0.90 (0.004)	0.25 (0.003)	0.25 (0.000)
		3	0.32	0.32 (0.004)	0.17 (0.003)	0.74 (0.006)	0.26 (0.003)	0.25 (0.000)
		4	0.15	0.15 (0.003)	0.08 (0.002)	0.90 (0.004)	0.24 (0.002)	0.24 (0.000)

[a] In all scenarios, the average treatment effect (estimand) is 1 ($\beta = 1$) and the effect of the confounding variable on the outcome is 2 ($\gamma = 2$). Five-thousand data sets were generated. Monte Carlo standard errors are shown between brackets. In scenario 0, there is no classification error (specificity and sensitivity of the misclassified confounding variable are 1, i.e., $p_0 = 0$ and $p_1 = 1$). In scenarios 1-4, the specificity of the misclassified confounding variable is 0.95 (i.e., $p_0 = 0.05$) and the sensitivity is 0.9 (i.e., $p_1 = 0.9$). The prevalence of the confounding variable (λ), and the probability of receiving treatment if the confounding is not present or present (π_0 and π_1, respectively) are set as follows in the scenarios: scenario 0: $\lambda = 0.5$, $\pi_0 = 0.5$, $\pi_1 = 0.75$; scenario 1: $\lambda = 0.5$, $\pi_0 = 0.9$, $\pi_1 = 0.45$; scenario 2: $\lambda = 0.8$, $\pi_0 = 0.25$, $\pi_1 = 0.75$; scenario 3: $\lambda = 0.8$, $\pi_0 = 0.5$, $\pi_1 = 0.75$; scenario 4: $\lambda = 0.45$, $\pi_0 = 0.5$, $\pi_1 = 0.75$.

Results. Table 8.2 shows the results of the simulation study. Bias found in the simulation study corresponds to the theoretical bias derived from the bias expressions. The empirical standard deviation of the average treatment effect estimates (empSE) from the MSM-IPW is equal to or greater than that from the conditional model. Yet, in the scenarios where bias in the average treatment effect in the MSM-IPW was smaller than bias in the conditional model (scenarios 2 and 3), empSE of both methods was equal, and hence, MSE is smaller for one method if also bias is smaller. Furthermore, the (robust) model based standard errors of the average treatment effect in a MSM-IPW are conservative and greater than the empirical standard errors, since the uncertainty in estimating the treatment weights is not taken into account. Allowing for the estimation of the weights will shrink the standard errors [2, 28]. We chose not to use a less conservative standard error estimation for MSM-IPW, such as bootstrapping, since our goal was to frame this simulation as investigating the properties of the commonly used MSM-IPW estimation procedure. Consequently, confidence intervals of the treatment effect obtained in a MSM-IPW are generally wider and coverage of the true treatment effect is higher compared to a conditional model, ranging from over coverage if there is no classification error to smaller under coverage when there is classification error.

8.4. Illustration: quantitative bias analysis

Quantitative bias analysis provides a tool to incorporate uncertainty in study results due to systematic errors [18, 20]. Using an example study of blood pressure lowering therapy, we illustrate how the bias expressions (8.3) and (8.4) can be used to perform a quantitative bias analysis for misclassification of a confounding variable.

Application. For our illustration we use data of the National Health And Nutritional Examination Survey (NHANES) [36, 37], more details can be found in the supplementary material section S8.2. Specifically, we study the effect of diuretic use ($A = 1$) in comparison to beta blocker use ($A = 0$) on systolic blood pressure (Y) using two approaches: by inverse weighting with the propensity for diuretic or beta blocker use given self-reported categorical body mass index (BMI) (L^*), and using a conditional linear regression with adjustment for self-reported categorical BMI. For this illustration, we categorize self-reported BMI into two distinct categories: underweight/normal weight (BMI < 25 ($L^* = 0$)) and overweight/obese (BMI ≥ 25 ($L^* = 1$)). However, we stress that one should preferably not categorise BMI in most practical applications [38]. Moreover, we assume that dichotomizing self-reported BMI does not introduce differential misclassification [7].

We assume that blood pressure lowering therapy is initiated based on the true BMI (L) instead of the observed self-reported BMI (setting 2, Figure 8.1b). Further, we consider BMI the only confounding variable, and treatment and outcome to be measured without error, which is a simplification of reality. Additionally, we assume that the classification error in self-reported BMI category is non-differential for the subject's treatment or blood pressure (given true BMI category). Expert knowledge is needed to inform this assumption. To quantify how large the bias in the average treatment effect is expected to be due to classification error in self-reported BMI category, we perform a quantitative bias analysis using the bias expressions (8.3) and (8.4).

Average treatment effect. Table 8.3 shows the average treatment effect of diuretics use in comparison to beta blocker use on mean systolic blood pressure. In a MSM-IPW,

Table 8.3: Average treatment effect of diuretics use compared to beta blocker use on mean systolic blood pressure in NHANES [36, 37]. CI indicates confidence interval.

Model	Effect Size (95% CI)
Unadjusted	−4.03(−6.30; −1.76)
Marginal Structural Model[a]	−3.52(−5.74; −1.21)
Conditional Model[b]	−3.48(−5.76; −1.27)

[a] Estimated in a marginal structural model, by inverse weighting with the propensity for diuretic or beta blocker use given self-reported categorised body mass index (BMI).
[b] Estimated in a conditional regression model with adjustment for self-reported categorical BMI.

we estimated an average treatment effect (95 % CI) of −3.52 (−1.21; −5.74). In a conditional regression model, we estimated an average treatment effect (95 % CI) of −3.48 (−1.27; −5.76).

Quantitative bias analysis. To inform the quantitative bias analysis, we need to make assumptions on the sensitivity and specificity of the self-reported BMI as well as that classification errors are non-differential with respect to blood pressure and treatment. For the purpose of this illustration, we speculate ranges for the sensitivity and specificity of self-reported BMI category of 0.90 to 0.98. In practice, these parameters should be informed by reports in the literature and/or a researcher's expert experience. Researchers may also decide to investigate how extreme the misclassification (measured using sensitivity and specificity) would need to be to change the conclusions of their study. We refer to the Shiny application (introduced in the subsequent section) for other choices for the sensitivity and specificity of self-reported BMI category.

By uniformly sampling from the range of plausible values of p_0 and p_1 and using the bias expressions (8.3) and (8.4), a distribution of possible biases is obtained (see supplementary material section S8.2 for further details). The solid line in Figure 8.5 shows the distribution of bias in a MSM-IPW. Mean bias is -0.31 and median bias is -0.30 (interquartile range -0.40 to -0.20). We also considered sampling p_0 and p_1 from a trapezoidal (with modes at one third and two thirds between the minimum and maximum) or a symmetrical triangular distribution. Sampling from these distributions results in mean bias approximately equal to when uniform sampling is applied, but with less spread (panels B and C in Figure 8.5). This result suggests that the results in Table 8.3 are not affected much by the classification error in self-reported BMI category. In the NHANES, anthropometric measures were also taken by trained technicians. See S8.2 for the average treatment effect when BMI measures taken by trained technicians were used instead of self-reported BMI measures.

8.5. Shiny application: an online tool

We developed an online tool for creating bias plots (Figure 8.2-8.4) and performing quantitative bias analyses (illustrated in the previous section), available at `https://lindanab.shinyapps.io/SensitivityAnalysis`. The bias plots can be used to predict the implications of classification error in a confounding variable in specific study settings by varying: the strength of association between the confounding variable and treatment and between the confounding variable and outcome; prevalence of the confounding variable; specificity and sensitivity of the misclassified confounding variable. The quantitative bias analysis can be used for studying the impact of classification error

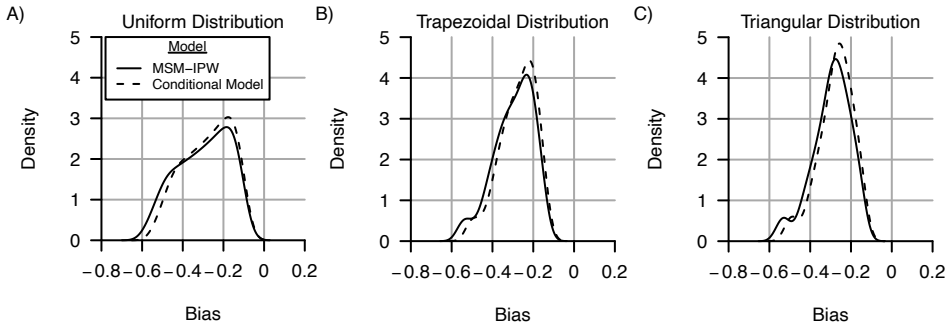


Figure 8.5: Density of predicted bias due to classification error in self-reported BMI category in NHANES [37]. Bias in the average treatment effect of diuretics use compared to beta blocker use on mean systolic blood pressure by inverse weighting with the propensity for diuretic or beta blocker use given self-reported categorical BMI (MSM-IPW), and using a conditional linear regression with adjustment for self-reported categorical BMI. The specificity and sensitivity of self-reported BMI category range from 0.90 to 0.98 and are sampled from a uniform distribution, trapezoidal (with modes on one-third and two-third), and symmetrical triangular distribution.

in a confounding variable at the analysis stage of a study, and to investigate how sensitive conclusions are to the assumption of no classification error. These bias plots can also be used to inform decisions about measurement methods or choice of variables to be extracted in the planning stage of studies.

8.6. Discussion

Inverse probability weighting and conditional models are both important and frequently used tools to adjust for confounding variables in observational studies. In this article, we derived expressions for the bias in the average treatment effect in a MSM-IPW and a conditional model. These expressions can inform quantitative bias analyses for bias due to a misclassified confounding variable.

Quantitative bias analysis of misclassified confounding variables is one example of quantitative bias analyses for observational epidemiologic studies. Several approaches exist to assess sensitivity of causal conclusions to unmeasured confounding [29, 39, 40]. These aim to quantify the impact of violations of the assumption of no unmeasured confounding, while our approach aims to quantify the impact of violations of the assumption that all confounding variables are measured without error.

Several methods have been proposed to adjust for measurement error in covariates in MSMs-IPW. Pearl developed a general framework for causal inference in the presence of error-prone covariates, which yields weighted estimators in the case of a dichotomous confounding variable measured with error [41]. The framework relies on a joint distribution of the outcome and the confounding variable. Conversely, the weighting method proposed by McCaffrey et al. does not require a model for the outcome [17]. Additionally, regression calibration [42], simulation-extrapolation [43, 44] and multiple imputation [45] have been proposed for correcting for measurement error in covariates of MSMs. These methods assume that the measurement error model is known, which may often be unrealistic. In this context it is also important to mention previous studies of the impact of measurement error in the exposure or the endpoint in MSMs, which has been studied by Babanezhad et

al. [46] and Shu et al. [47], respectively.

If treatment is allocated based on an error-prone confounding variable, the treatment effect will not be biased (see DAG in Figure 8.1a). However, investigators should be careful in concluding that covariate measurement error will not affect their analysis. Suppose that there is an unmeasured variable U that acts as a confounding variable between the error-prone covariate L^* and treatment A. Conditioning on L^* will then open a path between A and L via U and thus confound the relation between A and Y.

This article considered classification error in a dichotomous confounding variable in a point-treatment study with a continuous outcome. The same principles apply to measurement error in a categorical or continuous confounding variable or when multiple confounding variables are considered, although more elaborate assumptions should then be made [48]. Moreover, we assumed that the relation between exposure and outcome does not vary between strata of the confounding variable, i.e. that there is no treatment effect modification. Future research could extend our bias expressions by relaxing this simplifying assumption, therefore extending our results to more general settings.

MSMs-IPW are increasingly applied to longitudinal data to estimate the joint effects of treatment at multiple time points on a subsequent outcome, including time-dependent outcomes, addressing the problem of time-dependent confounding [1, 3]. There has been little work to understand or correct for the impact of misclassified or mismeasured confounding variables in this more complex setting. Our results extend directly to the time-dependent setting when the aim is to estimate the effect of a current treatment on a time-dependent outcome measured at the next time point [49]. An area for future work is to extend our results to the setting in which the aim is to estimate the joint effects of treatment at multiple time points. and to the time-dependent setting with time varying treatments and confounding variables. An additional factor to consider in the time-varying setting is the impact of stabilized vs unstabilized weights on the bias if both numerator and denominator of the stabilized weights involve conditioning on an error-prone covariate.

The bias expressions derived in this paper can be used to assess bias due to classification error in a dichotomous confounding variable. If classification error in confounding variables is suspected, a quantitative bias analysis provides an opportunity to quantitatively inform readers on the possible impact of such errors on causal conclusions.

References

[1] M. A. Hernán, B. A. Brumback, J. M. Robins, Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures, Statistics in Medicine 21 (12) (2002) 1689–1709. doi:10.1002/sim.1144.

[2] J. Robins, M. Hernán, B. Brumback, Marginal structural models and causal inference in Epidemiology, Epidemiology 11 (5) (2000) 550–560.

[3] R. Daniel, S. Cousens, B. De Stavola, M. G. Kenward, J. A. C. Sterne, Methods for dealing with time-dependent confounding, Statistics in Medicine 32 (9) (2013) 1584–1618. doi:10.1002/sim.5686.

[4] D. B. Rubin, For objective causal inference, design trumps analysis, The Annals of Applied Statistics 2 (3) (2008) 808–840. doi:10.1214/08-AOAS187.

[5] P. M. Steiner, T. D. Cook, W. R. Shadish, On the importance of reliable covariate measurement in selection bias adjustments using propensity scores, Journal of Educational and Behavioral Statistics 36 (2) (2011) 213–236. doi:10.3102/1076998610375835.

[6] K. B. Michels, A renaissance for measurement error, International Journal of Epidemiology 30 (3) (2001) 421–422. doi:10.1093/ije/30.3.421.

[7] M. van Smeden, T. L. Lash, R. H. H. Groenwold, Reflection on modern methods: Five myths about measurement error in epidemiological research, International Journal of Epidemiology 49 (1) (2020) 338–347. doi:10.1093/ije/dyz251.

[8] J. Kasza, K. R. Polkinghorne, M. R. Marshall, S. P. McDonald, R. Wolfe, Clustering and residual confounding in the application of marginal structural models: Dialysis modality, vascular access, and mortality, American Journal of Epidemiology 182 (6) (2015) 535–543. doi:10.1093/aje/kwv090.

[9] L. M. Bodnar, Marginal structural models for analyzing causal effects of time-dependent treatments: An application in perinatal epidemiology, American Journal of Epidemiology 159 (10) (2004) 926–934. doi:10.1093/aje/kwh131.

[10] M. E. Wood, K. Lapane, J. A. Frazier, E. Ystrom, E. O. Mick, H. Nordeng, Prenatal triptan exposure and internalising and externalising behaviour problems in 3-year-old children: Results from the Norwegian mother and child cohort study, Paediatric and Perinatal Epidemiology 30 (2) (2016) 190–200. doi:10.1111/ppe.12253.

[11] B. G. Armstrong, Effect of measurement error on epidemiological studies of environmental and occupational exposures, Occupational and Environmental Medicine 55 (10) (1998) 651–656. doi:10.1136/oem.55.10.651.

[12] J. Buonaccorsi, Measurement error: Models, methods, and applications, Chapman & Hall/CRC, Boca Raton, FL, 2010.

[13] R. Carroll, D. Ruppert, L. Stefanski, C. Crainiceanu, Measurement error in nonlinear models: A modern perspective, 2nd Edition, Chapman & Hall/CRC, Boca Raton, FL, 2006.

[14] P. Gustafson, Measurement error and misclassification in statistics and epidemiology: Impacts and Bayesian adjustments, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[15] W. Fuller, Measurement error models, John Wiley & Sons, New York, NY, 1987.

[16] M. D. Regier, E. E. M. Moodie, R. W. Platt, The effect of error-in-confounders on the estimation of the causal parameter when using marginal structural models and inverse probability-of-treatment weights: A simulation study, The International Journal of Biostatistics 10 (1) (2014) 1–15. doi:10.1515/ijb-2012-0039.

[17] D. F. McCaffrey, J. R. Lockwood, C. M. Setodji, Inverse probability weighting with error-prone covariates, Biometrika 100 (3) (2013) 671–680. doi:10.1093/biomet/ast022.

[18] S. Greenland, Basic methods for sensitivity analysis of biases, International Journal of Epidemiology 25 (6) (1996) 1107–1116. doi:10.1093/ije/25.6.1107-a.

[19] T. Lash, M. Fox, A. Fink, Applying quantitative bias analysis to epidemiologic data, Springer, New York, NY, 2009.

[20] T. L. Lash, M. P. Fox, R. F. MacLehose, G. Maldonado, L. C. McCandless, S. Greenland, Good practices for quantitative bias analysis, International Journal of Epidemiology 43 (6) (2014) 1969–1985. doi:10.1093/ije/dyu149.

[21] M. S. Ali, R. H. H. Groenwold, S. V. Belitser, P. C. Souverein, E. Martín, N. M. Gatto, C. Huerta, H. Gardarsdottir, K. C. B. Roes, A. W. Hoes, A. de Boer, O. H. Klungel, Methodological comparison of marginal structural model, time-varying Cox regression, and propensity score methods: The example of antidepressant use and the risk of hip fracture, Pharmacoepidemiology and Drug Safety 25 (Suppl. 1) (2016) 114–121. doi:10.1002/pds.3864.

[22] L. A. Jackson, M. L. Jackson, J. C. Nelson, K. M. Neuzil, N. S. Weiss, Evidence of bias in estimates of influenza vaccine effectiveness in seniors, International Journal of Epidemiology 35 (2) (2006) 337–344. doi:10.1093/ije/dyi274.

[23] D. B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies, Journal of Educational Psychology 66 (5) (1974) 688–701. doi:10.1037/h0037350.

[24] D. B. Rubin, Formal mode of statistical inference for causal effects, Journal of Statistical Planning and Inference 25 (3) (1990) 279–292. doi:10.1016/0378-3758(90)90077-8.

[25] M. Hernán, J. Robins, Causal inference: What if, Chapman & Hall/CRC, Boca Raton, 2020.

[26] D. B. Rubin, Randomization analysis of experimental data: The Fisher randomization test comment, Journal of the American Statistical Association 75 (371) (1980) 591. doi:10.2307/2287653.

[27] S. R. Cole, M. A. Hernan, Constructing inverse probability weights for marginal structural models, American Journal of Epidemiology 168 (6) (2008) 656–664. doi:10.1093/aje/kwn164.

[28] J. Robins, Marginal structural models versus structural nested models as tools for causal inference, in: E. Halloran, D. Berry (Eds.), Statistical models in epidemiology, the environment, and clinical trials, Springer-Verlag, New York, 2000, Ch. 2, pp. 95–133.

[29] T. VanderWeele, O. Arah, Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders, Epidemiology 22 (1) (2011) 42–52. doi:10.1097/EDE.0b013e3181f74493.

[30] P. M. Steiner, Y. Kim, The mechanics of omitted variable bias: Bias amplification and cancellation of offsetting biases, Journal of Causal Inference 4 (2) (2016). doi:10.1515/jci-2016-0009.

[31] R Core Team, R: A language and environment for statistical computing (2020). URL https://www.r-project.org/

[32] W. M. van der Wal, R. B. Geskus, Ipw : An R package for inverse probability weighting, Journal of Statistical Software 43 (13) (2011) 1–23. doi:10.18637/jss.v043.i13.

[33] T. Lumley, Analysis of complex survey samples, Journal of Statistical Software 9 (8) (2004) 1–19. doi:10.18637/jss.v009.i08.

[34] T. P. Morris, I. R. White, M. J. Crowther, Using simulation studies to evaluate statistical methods, Statistics in Medicine 38 (11) (2019) 2074–2102. doi:10.1002/sim.8086.

[35] A. Gasparini, Rsimsum: Summarise results from Monte Carlo simulation studies, Journal of Open Source Software 3 (26) (2018) 739. doi:10.21105/joss.00739.

[36] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS), National health and nutrition examination survey data (2011). URL https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2011

[37] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS), National health and nutrition examination survey data (2013). URL https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2013

[38] D. G. Altman, P. Royston, The cost of dichotomising continuous variables, BMJ 332 (7549) (2006) 1080. doi:10.1136/bmj.332.7549.1080.

[39] R. H. H. Groenwold, D. B. Nelson, K. L. Nichol, A. W. Hoes, E. Hak, Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research, International Journal of Epidemiology 39 (1) (2010) 107–117. doi:10.1093/ije/dyp332.

[40] P. Ding, T. J. VanderWeele, Sensitivity analysis without assumptions, Epidemiology 27 (3) (2016) 368–377. doi:10.1097/EDE.0000000000000457.

[41] J. Pearl, On measurement bias in causal inference, in: P. Grunwald P, Spirtes (Ed.), Proceedings of uncertainty in artificial intelligence, AUAI, AUAI, Corvallis, OR, 2010, pp. 425–432.

[42] S. R. Cole, L. P. Jacobson, P. C. Tien, L. Kingsley, J. S. Chmiel, K. Anastos, Using marginal structural measurement-error models to estimate the long-term effect of antiretroviral therapy on incident AIDS or death, American Journal of Epidemiology 171 (1) (2010) 113–122. doi:10.1093/aje/kwp329.

[43] R. Kyle, E. Moodie, M. Klein, M. Abrahamowicz, Correcting for measurement error in time-varying covariates in marginal structural models, American Journal of Epidemiology 184 (3) (2016) 249–258. doi:10.1093/aje/kww068.

[44] J. Lockwood, D. F. McCaffrey, Simulation-extrapolation for estimating means and causal effects with mismeasured covariates, Observational Studies 1 (2) (2015) 241–290. doi:10.1353/obs.2015.0007.

[45] Y. Webb-Vargas, K. Rudolph, D. Lenis, P. Murakami, E. Stuart, An imputation-based solution to using mismeasured covariates in propensity score analysis, Statistical Methods in Medical Research 26 (4) (2017) 1824–1837. doi:10.1177/0962280215588771.

[46] M. Babanezhad, S. Vansteelandt, E. Goetghebeur, Comparison of causal effect estimators under exposure misclassification, Journal of Statistical Planning and Inference 140 (5) (2010) 1306–1319. doi:10.1016/j.jspi.2009.11.015.

[47] D. Shu, G. Y. Yi, Causal inference with measurement error in outcomes: Bias analysis and estimation methods, Statistical Methods in Medical Research 28 (7) (2019) 2049–2068. doi:10.1177/0962280217743777.

[48] R. H. Keogh, I. R. White, A toolkit for measurement error correction, with a focus on nutritional epidemiology, Statistics in Medicine 33 (12) (2014) 2137–2155. doi:10.1002/sim.6095.

[49] R. H. Keogh, R. M. Daniel, T. J. VanderWeele, S. Vansteelandt, Analysis of longitudinal studies with repeated outcome measures: Adjusting for time-dependent confounding using conventional methods, American Journal of Epidemiology 187 (5) (2018) 1085–1092. doi:10.1093/aje/kwx311.