Correction methods for measurement error in epidemiologic research

Nab, L.

S2

Supplementary material Chapter 2

These are the supplementary materials accompanying Chapter 2. The supplementary materials are structured as follows. In section S2.1 we discuss two more example trials for illustration of measurement error in an endpoint. In section S2.2 we explain why and under which assumptions ignoring measurement error will lead to incorrect inference. Section S2.3 provides an explanation of corrected effect estimators (and why these are consistent) and explains the methods used for confidence interval estimation. In section S2.4 a proof is given that measurement error depending on prognostic factors does not introduce bias in the treatment effect estimator. In section S2.5 an approximation for the bias and variance of the corrected estimator is derived.

S2.1. Illustrative examples

We introduce here two additional example trials from literature, hypothesize that these trial could also have used endpoints measured with error to illustrate how the use of an endpoint that is contaminated with error would affect trial inference. We assume that the original endpoints used in our example trials are measurement error free.

S2.1.1. Example trial 2: energy expenditure

Poehlman and colleagues [1] studied the effects of endurance and resistance training on total daily energy expenditure in a randomised trial of young sedentary women. Participants were randomized to one of three six-month during exercise programmes: endurance training, resistance training or the control arm. Some controversy regarding the effect of exercise training on total energy expenditure (TEE) existed at the time of the start of the trial, partly because of the difficulty to assess daily energy expenditure [1]. Starting 72 hours after completion of the training program, TEE of the participants was measured by doubly labelled water during a ten day period, which is considered the gold standard in measuring energy expenditure in humans [2]. In short, the study found no evidence for an effect of resistance and endurance training (compared to placebo) on total energy expenditure. Post-trial, measured TEE was higher in the control arm than in the two intervention arms. Table 1 shows the decrease in TEE of the women exposed to the

existence training programme versus the placebo arm.

S2.1.2. Example trial 3: rheumatoid arthritis disease activity

The U-Act-Early trial tested the efficacy of a new treatment strategy for rheumatoid arthritis (RA) in patients with newly diagnosed RA [3] in a three-arm trial: tocilizumab plus methotrexate versus tocilizumab only versus methotrexate only, all as initial treatment. For endpoint assessment, this trial used a validated RA disease activity measure (the Disease Activity Score 28, DAS28) [4]) which is commonly used and recommended to measure endpoints in RA clinical trials [5, 6]. In short, the trial showed that immediate initiation of tocilizumab with or without methotrexate is more effective than methotrexate alone to achieve sustained remission in newly diagnosed RA patients. The difference in mean DAS28 score in the tocilizumab plus methotrexate versus methotrexate only group after 24 weeks is shown in Table S2.1. The sample size of the former groups reported in Table S2.1 is based on measurements available at 24 weeks of follow up.

A common alternative approach to measure energy expenditure (example trial 2) is by a accelerometer, that measures body movement via motion sensors to assess energy expenditure (e.g. [2]). As compared to double labelled water (example trial 2), the accelerometer is cheaper, but less accurate [2]. Lastly, instead of endpoint assessment by DAS28 (example trial 3), where assessment is done by trained medical staff [4], trials could alternatively use the patient-based RA disease activity score (PDAS), where endpoint assessment is done by the patient [7].

For the example trial in the paper and each of the aforementioned example trials here, in Table S2.1 we show to what extent the Type-II of a test for treatment effect changes when a hypothetical lower standard of endpoint measurement would have been used introducing classical measurement error. The table clearly shows the anticipated increase in Type-II error with increasing error at the same sample size.

S2.2. Measurement error structures

Consider a two-arm randomized controlled trial that compares the effects of two treatments ($X \in \{0, 1\}$), where 0 may represent a placebo treatment or an active comparator. Let Y denote the true (or preferred) trial endpoint and Y^* an error prone operationalisation of Y. We will assume that both Y and Y^* are measured on a continuous scale. Throughout, we assume that Y^* is measured for all $i = 1, \ldots, N$ randomly allocated patients in the trial. We assume that the effect of allocated treatment ($X \in \{0, 1\}$) on preferred endpoint Y is defined by the linear model

$$Y = \alpha_Y + \beta_Y X + \varepsilon, \tag{S2.1}$$

where β_Y defines the treatment effect on the endpoint, and ε has expected mean 0 and variance σ^2. Throughout, we assume that X is fixed. Further, we assume that model S2.1 is inestimable from the observed data because the endpoint Y^* instead of Y was measured. We will assume that the relation between Y and Y^* is given by a linear model,

$$Y^* = \theta_0 + \theta_1 Y + e, \tag{S2.2}$$

where e is a random variable whose distribution is independent of ε, Y and X. The parameters θ_0 and θ_1 define the relation between Y and Y^*, where it is assumed that θ_1

Table S2.1: Impact of classical measurement error on Type-II error in the three example trials. Effect estimates, standard errors and sample sizes are based on results in the papers by Makridis et al. [8] (trial 1), Poehlman et al. [1] (trial 2) and Bijlsma et al. [3] (trial 3)

Example	Effect Estimate	Standard Error	Sample Size	ρ[a]	Type-II Error[b]
Trial 1	6.9	1.27	393	0	-
		2.43	108	0	20%
		2.71	108	1/5	29%
		2.45	132	1/5	20%
Trial 2	−246.0	369.00	35	0	-
		88.70	600	0	20%
		109.00	600	1/3	38%
		88.70	900	1/3	20%
Trial 3	−1.4	0.08	198	0	-
		0.41	8	0	18%
		0.50	8	3/7	41%
		0.44	12	3/7	18%

[a] Proportion of observed variance in endpoints due to measurement error.
[b] Type-II error calculations are based on results provided in section 3.1.

does not equal 0. We assume that both parameters θ_0 and θ_1 are estimable only in the external calibration sample comprising individuals not included in the trial ($j = 1, \ldots, K$).

Simple OLS regression estimators for β_Y, α_Y and σ^2 (the variance of the errors ε) in (S2.1) are,

$$\hat{\beta}_{Y^*} = \frac{\sum_i (X_i - \bar{X})(Y_i^* - \bar{Y}^*)}{\sum_i (X_i - \bar{X})^2}, \tag{S2.3}$$

$$\hat{\alpha}_{Y^*} = \bar{Y}^* - \hat{\beta}_{Y^*}\bar{X}, \tag{S2.4}$$

$$\omega_i = Y_i^* - \hat{\alpha}_{Y^*} - \hat{\beta}_{Y^*}X_i, \tag{S2.5}$$

$$s^2 = \frac{1}{N-2}\sum_i \omega_i^2, \tag{S2.6}$$

respectively. In a two-arm trial, the interest is in making inferences about β_Y, which cannot be directly estimated because in the trial the endpoint of interest Y was replaced by Y^*. In the following we will show: a) that $\hat{\beta}_{Y^*}$ may be a poor estimator for β_Y (section 3.1-3.4), and b) how adjustments to $\hat{\beta}_{Y^*}$ using information from the calibration model described by (S2.2) can improve inference about the treatment effect (section 4). As a starting point, in the following section relevant and known properties are defined for the special case that $Y^* = Y$, which is then followed by the properties under different measurement error structures for Y^* in subsequent sections.

S2.2.1. No measurement error

Consider the hypothetical case that Y^* is a perfect proxy for Y, i.e. $Y^* = Y$. By using that $Y = \alpha_Y + \beta_Y X + \varepsilon$, as defined in (S2.1), it follows that:

$$Y^* = \alpha_Y + \beta_Y X + \varepsilon.$$

From standard regression theory (e.g. [9]), we know that if the errors ε satisfy the regular Gauss-Markov assumptions [9] and their variance is defined by σ^2, the OLS estimators $\hat{\beta}_Y^*$, $\hat{\alpha}_Y^*$, and s^2 (defined by S2.3, S2.4, and S2.6, respectively) are Best Linear Unbiased Estimators (BLUE) for β_Y, α_Y, and σ^2, respectively.

Moreover, if the ε are independently and identically (iid) normally distributed, the OLS estimators $\hat{\beta}_{Y^*}$ and $\hat{\alpha}_{Y^*}$ (defined in S2.3 and S2.4, respectively) are the Maximum Likelihood Estimators (MLE) of β_Y and α_Y, respectively. Note that the errors ε satisfy the Gauss-Markov assumptions if we assume that they are iid normally distributed with mean 0 and constant variance σ^2.

Hypotheses for the treatment effect β_Y, can be defined by:

$$\begin{aligned} H_0 &: \quad \beta_Y = \beta_0, \\ H_A &: \quad \beta_Y \neq \beta_0. \end{aligned}$$

Under normality of the error terms ε, the OLS estimator $\hat{\beta}_Y^*$ defined in (S2.3) is the MLE for β_Y and s^2 is an unbiased estimator for σ^2, the following is known for the Wald test:

$$T = \frac{\hat{\beta}_{Y^*} - \beta_0}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_{Y^*})}} \sim t_{N-2}, \tag{S2.7}$$

where,

$$\widehat{\mathrm{Var}}(\hat{\beta}_{Y^*}) = \frac{s^2}{\sum_i (X_i - \bar{X})^2}. \tag{S2.8}$$

Assuming no measurement error in Y and X, under H_0, T follows a Student's t distribution with $N-2$ degrees of freedom [9]. Under H_A, T follows a Student's t distribution with $N-2$ degrees of freedom and non-centrality parameter $(\beta_Y - \beta_0)/\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_{Y^*})}$.

S2.2.2. Classical measurement error

There is classical measurement error in Y^* if Y^* is an unbiased proxy for Y [10]:

$$Y^* = Y + e, \tag{S2.9}$$

where $\mathrm{E}[e] = 0$ and $\mathrm{Var}(e) = \tau^2$ and e mutually independent of Y, X, ε (in (S2.1)). By using that $Y = \alpha_Y + \beta_Y X + \varepsilon$ from (S2.1), it follows that:

$$Y^* = \alpha_Y + \beta_Y X + \varepsilon + e.$$

Given the aforementioned assumptions, the sum of e and ε, $\delta_1 = e + \varepsilon$, has variance $\mathrm{Var}(\delta_1) = \sigma^2 + \tau^2$. It follows that if the errors δ_1 satisfy the Gauss-Markov assumptions, $\hat{\beta}_{Y^*}$ in (S2.3)

remains a BLUE estimator for β_Y. Also, $\hat{\alpha}_{Y^*}$ in (S2.4) and s^2 in (S2.6) remain BLUE estimators for α_Y and the variance of δ_1, respectively.

Further, if δ_1 is iid normally distributed with mean 0 and variance $\sigma^2 + \tau^2$, then $\hat{\alpha}_{Y^*}$ is the MLE for α_Y and $\hat{\beta}_{Y^*}$ is the MLE for β_Y. Obviously, given that $\sigma^2 > 0$ and $\tau^2 > 0$, the variance of the OLS regression estimator $\hat{\beta}_{Y^*}$ is larger if there is classical measurement error in the outcome compared to the case when there is no measurement error. Under the null hypothesis, the Wald test-statistic T defined in (S2.7) still follows a Student's t distribution with $N - 2$ degrees of freedom. However, under the alternative hypothesis, the non-centrality parameter of T, $(\beta_Y - \beta_0)/\sqrt{\widehat{\text{Var}}(\hat{\beta}_{Y^*})}$, will be smaller in the presence of classical measurement error.

To summarize, in the presence of only classical measurement error, Type-II error for detecting any given treatment effect increases, Type-I error is unaffected and the treatment effect estimator is unbiased MLE under standard regularity conditions.

Heteroscedastic classical measurement error

In the preceding we assumed that the Gauss-Markov assumptions were met. But notably, in the case that the variance of the errors e in (S2.9) varies per treatment arm, the errors are no longer homoscedastic (as needed to satisfy the Gauss-Markov assumptions) but heteroscedastic. In the case of this type of heteroscedastic classical measurement error, it can be shown that the variance of β_{Y^*} will be underestimated by the default estimator of the variance of $\hat{\beta}_{Y^*}$ defined by (S2.8), affecting both Type-I and Type-II error.

S2.2.3. Systematic measurement error

There is systematic measurement error in Y^*, if Y^* systematically depends on Y. Assuming this dependence is linear, the relation between Y^* and Y can be defined as:

$$Y^* = \theta_0 + \theta_1 Y + e, \tag{S2.10}$$

where $\text{E}[e] = 0$ and $\text{Var}(e) = \tau^2$. Throughout, we assume systematic measurement error if $\theta_0 \neq 0$ or $\theta_1 \neq 1$ (and of course, $\theta_1 \neq 0$ in all cases). We assume mutual independence between e and Y, X, ε (in S2.1). Naturally, if $\theta_0 = 0$ and $\theta_1 = 1$ the measurement error is of the classical form.

By using that $Y = \alpha_Y + \beta_Y X + \varepsilon$ from (S2.1), it follows that:

$$Y^* = \theta_0 + \theta_1\alpha_Y + \theta_1\beta_Y X + \theta_1\varepsilon + e.$$

Given the aforementioned assumptions, $\delta_2 = \theta_1\varepsilon + e$ with expected variance $\theta_1^2\sigma^2 + \tau^2$. It follows that under the Gauss-Markov assumptions, $\hat{\beta}_{Y^*}$ defined in (S2.3) is BLUE for $\theta_1\beta_Y$, and $\hat{\alpha}_{Y^*}$ defined in (S2.4) is BLUE for $\theta_0 + \alpha_Y$ and s^2 defined in (S2.6) is BLUE for the variance of δ_2 (i.e. $\theta_1^2\tau^2 + \sigma^2$). Conversely, $\hat{\beta}_{Y^*}$ is no longer BLUE for β_Y. Note that in this case s^2 is BLUE for $\theta_1^2\sigma^2 + \tau^2$, that is, depending on θ_1, smaller or larger than σ^2 (the variance of the error terms if there is no measurement error).

If we further assume that δ_2 is iid normally distributed, we can conclude that $\hat{\alpha}_{Y^*}$ is the MLE for $\theta_0 + \alpha_Y$ and $\hat{\beta}_{Y^*}$ is the MLE for $\theta_1\beta_Y$. Conversely, $\hat{\beta}_{Y^*}$ is no longer the MLE for β_Y, if there is systematic measurement error in Y^*. In the absence of a treatment effect, as $\theta_1\beta_Y = 0$ if $\beta_Y = 0$, T defined in (S2.7) still follows a Student's t distribution with $N - 2$

degrees of freedom. In the presence of any given treatment effect, T follows a non-central Student's t distribution with $N - 2$ degrees of freedom and non-centrality parameter $(\theta_1\beta_Y - \beta_0)/\sqrt{\widehat{\text{Var}}(\hat{\beta}_{Y^*})}$. Depending on the value of θ_1, the non-centrality parameter will be smaller or larger than the non-centrality parameter in the absence of measurement error (see section 3.2).

In summary, if there is systematic measurement error in the endpoints, the Type-I error is unaffected under standard regularity conditions and hence testing whether there is no effect is still valid under the null hypothesis [11]). Type-II, however, is affected (it may increase or decrease) and the treatment effect estimator is a biased MLE.

S2.2.4. Differential measurement error

There is differential measurement error in Y^* when measurement error varies with X. Assuming a linear model for this variation, formally:

$$Y^* = \theta_{00} + (\theta_{01} - \theta_{00})X + \theta_{10}Y + (\theta_{11} - \theta_{10})XY + e_X, \tag{S2.11}$$

where $\text{E}[e_X] = 0$ and $\text{Var}(e_X) = \tau_X^2$ and e_X independent of the endpoint of interest Y, and ε in (S2.1). From the equations it becomes clear that systematic error (equation (S2.10)) can be seen as a special case of differential error, where $\theta_{00} = \theta_{01}$ and $\theta_{10} = \theta_{11}$.

By using that $Y = \alpha_Y + \beta_Y X + \varepsilon$ from (S2.1), it follows from equation (S2.11) that,

$$Y^* = \theta_{00} + \theta_{10}\alpha_Y + \left[\theta_{01} - \theta_{00} + (\theta_{11} - \theta_{10})\alpha_Y + \theta_{11}\beta_Y\right]X + \left[\theta_{10} + (\theta_{11} - \theta_{10})X\right]\varepsilon + e_X.$$

Let $\delta_{3X} = \left[\theta_{10} + (\theta_{11} - \theta_{10})X\right]\varepsilon + e_X$, with expected variance $\left[\theta_{10}^2 + (\theta_{11}^2 - \theta_{10}^2)X\right]\sigma^2 + \tau_X^2$. Since the the error term δ_{3X} is no longer homoscedastic, the OLS estimators defined in (S2.3) and (S2.4) are no longer BLUE. However, the OLS estimator $\hat{\beta}_{Y^*}$ in (S2.3) is consistent (although not efficient) for $\theta_{01} - \theta_{00} + (\theta_{11} - \theta_{10})\alpha_Y + \theta_{11}\beta_Y$. The OLS estimator $\hat{\alpha}_{Y^*}$ defined in (S2.4) is consistent (although not efficient) for $\theta_{00} + \theta_{10}\alpha_Y$. Nevertheless, the estimator for the variance of $\hat{\beta}_{Y^*}$ defined in (S2.8) is no longer valid.

By using the residuals ω_i defined in (S2.6), a heteroscedastic consistent estimator for the variance of $\hat{\beta}_{Y^*}$ is:

$$\widehat{\text{Var}}(\hat{\beta}_{Y^*}) = \frac{\sum_i \left[(X_i - \bar{X})^2\omega_i^2\right]}{\left[\sum_i (X_i - \bar{X})^2\right]^2},$$

which is known as the White estimator [12]. From standard regression theory, it is known that using the above defined estimator, T defined in (S2.7) is still valid. Yet, under differential measurement error no longer $\left[\theta_{01} - \theta_{00} + (\theta_{11} - \theta_{10})\alpha_Y + \theta_{11}\beta_Y\right] = 0$ if $\beta_Y = 0$. Thus, under the null hypothesis, T defined in (S2.7) follows a Student's t distribution with $N - 2$ degrees of freedom and non-centrality parameter $(\left[\theta_{01} - \theta_{00} + \theta_{11}\alpha_Y - \theta_{10}\alpha_Y + \theta_{11}\beta_0\right] - \beta_0)/\sqrt{\widehat{\text{Var}}(\hat{\beta}_{Y^*})}$. Consequently, Type-I error changes if there is differential measurement error in Y^* and test about contrast under the null hypothesis are invalid [11]. Moreover, under the alternative hypothesis, T follows a non-central Student's t distribution with $N - 2$ degrees of freedom and non-centrality parameter $(\left[\theta_{01} - \theta_{00} + (\theta_{11} - \theta_{10})\alpha_Y + \theta_{11}\beta_Y\right] - \beta_0)/\sqrt{\widehat{\text{Var}}(\hat{\beta}_{Y^*})}$. Depending on the values of the θ's and α_Y, the non-centrality parameters will be smaller or larger than 0 and the non-centrality parameter if there is no measurement error, respectively (see section 3.2). Hence, Type-I

error and Type-II error could increase or decrease if there is differential measurement error in Y^*.

To summarize, Type-I error is not expected nominal (α) if there is differential measurement error in Y^* (see also [11]). Also, similar to systematic error in Y^*, Type-II error is affected (may increase or decrease) and the treatment effect estimator is biased.

S2.3. Correction methods for measurement error in a continuous trial endpoint

To accommodate measurement error correction, we assume that Y and Y^* are both measured for a smaller set of different individuals not included in the trial ($j = 1, \dots, K, K < N$), hereinafter referred to as the external calibration sample. In all but one case, it is assumed that only Y^* and Y are measured in the external calibration sample. In the case that the error in Y^* is different for the two treatment groups, it is assumed that the external calibration sample is in the form of a small pilot study where both treatments are allocated (i.e., Y^* and Y are both measured after assignment of X).

S2.3.1. Systematic measurement error

Using an external calibration set and assuming that the errors e in (S2.10) are iid normal, the MLE of the measurement error parameters in (S2.10) are:

$$\begin{aligned} \hat{\theta}_1 &= \frac{\sum_j (Y_j^{(c)} - \bar{Y}^{(c)})(Y_j^{*(c)} - \bar{Y}^{*(c)})}{\sum (Y_j^{(c)} - \bar{Y}^{(c)})^2}, \\ \hat{\theta}_0 &= \bar{Y}^{*(c)} - \hat{\theta}_1 \bar{Y}^{(c)}, \\ t^2 &= \frac{1}{K-2} \sum_j (Y_j^{*(c)} - \hat{\theta}_0 - \hat{\theta}_1 Y_j^{(c)})^2. \end{aligned} \tag{S2.12}$$

The superscript (c) is used to indicate that the measurement is obtained in the calibration set. From section 3.4, under systematic measurement error and assuming that ε in (S2.1) and e in (S2.10) iid normal and independent, the estimator $\hat{\beta}_{Y^*}$ defined in (S2.3) is the MLE of $\theta_1 \beta_Y$ and, the estimator $\hat{\alpha}_{Y^*}$ defined in (S2.4) is the MLE of $\theta_0 + \theta_1 \alpha_Y$. Natural sample estimators for α_Y and β_Y are then

$$\hat{\alpha}_Y = (\hat{\alpha}_{Y^*} - \hat{\theta}_0)/\hat{\theta}_1 \quad \text{and} \quad \hat{\beta}_Y = \hat{\beta}_{Y^*}/\hat{\theta}_1, \tag{S2.13}$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are the estimated error parameters from the calibration data set. From equation (S2.13), it becomes apparent that $\hat{\theta}_1$ needs to be assumed bounded away from zero for finite estimates of $\hat{\alpha}_Y$ and $\hat{\beta}_Y$ [13].

The first moment of estimators $\hat{\alpha}_Y$ and $\hat{\beta}_Y$ can be approximated by using multivariate Taylor expansions and assuming that ($\hat{\alpha}_{Y^*}$, $\hat{\beta}_{Y^*}$, $\hat{\theta}_0$, $\hat{\theta}_1$) are normally distributed [13],

$$\mathrm{E}[\hat{\alpha}_Y] \approx \alpha_Y + \frac{[\alpha_Y - \bar{y}^*]\tau^2}{\theta_1^2 S_{yy}^{(c)}} \quad \text{and} \quad \mathrm{E}[\hat{\beta}_Y] \approx \beta_Y + \frac{\beta_Y \tau^2}{\theta_1^2 S_{yy}^{(c)}},$$

where $S_{yy}^{(c)} = \sum(Y_j^{(c)} - \bar{Y}^{(c)})^2$, the total sum of squares of $Y^{(c)}$. In conclusion, the estimators $\hat{\alpha}_Y$ and $\hat{\beta}_Y$ are consistent. Formal derivations for the presented formulas are provided in section S2.5.

In the following we will focus on specifying confidence limits for the treatment effect estimator $\hat{\beta}_Y$ defined in (S2.13). We make use of the fact that this estimator is a ratio, which motivates the use of the Delta method, Fieller method and Zero-variance method [14]. We also present a non-parametric bootstrap method for specifying confidence limits [15].

Delta method

Assuming that $\hat{\beta}_{Y^*}$ and $\hat{\theta}_1$ are both normally distributed and applying the Delta method, the second moment of $\hat{\beta}_Y$ can be approximated [11]. Formal derivations of the presented formulas are provided in section S2.5. The Delta method variance of $\hat{\beta}_Y$ is given by:

$$\mathrm{Var}(\hat{\beta}_Y) \approx \frac{1}{\theta_1^2}\Big[\frac{\theta_1^2\sigma^2 + \tau^2}{S_{xx}} + \frac{\beta_Y^2\tau^2}{S_{yy}^{(c)}}\Big],$$

where $S_{xx} = \sum_i (X_i - \bar{X})^2$, the total sum of squares of X. An approximation of the above defined variance, denoted by $\widehat{\mathrm{Var}}(\hat{\beta}_Y)$, is provided by approximating θ_1, $\theta_1^2\sigma^2 + \tau^2$, τ^2 and β_Y respectively by $\hat{\theta}_1$, s^2, t^2 and $\hat{\beta}_Y$ [11].

An approximate confidence interval for the estimator $\hat{\beta}_Y$ is then given by

$$\hat{\beta}_Y \pm t_{(\alpha/2,n-2)}\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_Y)}. \tag{S2.14}$$

Fieller method

A second method to construct confidence intervals for the estimator $\hat{\beta}_Y$ in (S2.13), described by Buonaccorsi, is the Fieller method [11, 16]. In the case that $\hat{\theta}_1$ is significantly different from zero at a significance level of α (that is, $\hat{\theta}_1/\sqrt{t^2/S_{yy}^{(c)}} > t_{N-2}$), the $(1-\alpha)$ confidence intervals of $\hat{\beta}_Y$ are defined by the Fieller method by:

$$l_{upper,lower} = \frac{\hat{\beta}_{Y^*}\hat{\theta}_1 \pm \sqrt{\hat{\beta}_{Y^*}^2\hat{\theta}_1^2 - (\frac{t^2}{S_{yy}^{(c)}}t_q^2 - \hat{\theta}_1^2)(\frac{s^2}{S_{xx}}t_q^2 - \hat{\beta}_{Y^*}^2)}}{\frac{\tau^2}{S_{yy}^{(c)}}t_q^2 + \hat{\theta}_1^2}. \tag{S2.15}$$

A formal derivation can be found in section S2.5.

Zero-variance method

The zero-variance method adjusts the observed endpoints Y_i^* by

$$\hat{Y}_i = (Y_i^* - \hat{\theta}_0)/\hat{\theta}_1,$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are derived from (S2.10). The adjusted endpoints are regressed on the treatment variable X, which yields,

$$\hat{\beta}_{\hat{Y}} = \frac{\sum_i (X_i - \bar{X})(\hat{Y}_i - \bar{\hat{Y}})}{\sum_i (X_i - \bar{X})^2} = \frac{\sum_i (X_i - \bar{X})(Y_i^* - \bar{Y}^*)/\hat{\theta}_1}{\sum_i (X_i - \bar{X})^2} = \hat{\beta}_{Y^*}/\hat{\theta}_1,$$

$$\hat{\alpha}_{\hat{Y}} = \bar{\hat{Y}} - \hat{\beta}_{\hat{Y}}\bar{X} = \frac{\bar{Y^*} - \hat{\beta}_{Y^*}\bar{X} - \hat{\theta}_0}{\hat{\theta}_1} = (\hat{\alpha}_{Y^*} - \hat{\theta}_0)/\hat{\theta}_1,$$
$$s^2_{\hat{Y}} = \frac{1}{N-2}\sum_i (\hat{Y}_i - \hat{\alpha}_{\hat{Y}} - \hat{\beta}_{\hat{Y}}X_i)^2 = \frac{1}{\hat{\theta}_1^2}s^2,$$

with $\hat{\beta}_{Y^*}$, $\hat{\alpha}_{Y^*}$ and s^2 as in equations (S2.3, S2.4 and S2.6), respectively. Thus, $\hat{\beta}_{\hat{Y}}$ equals $\hat{\beta}_Y$ and $\hat{\alpha}_{\hat{Y}}$ equals $\hat{\alpha}_Y$ defined in (S2.13).

When the value of $\hat{\theta}_1$ (i.e. θ_1) is known, the variance of the estimator $\hat{\beta}_{\hat{Y}}$ is equal to:

$$\text{Var}(\hat{\beta}_{\hat{Y}}) = \text{Var}(\hat{\beta}_{Y^*})/\theta_1^2 = \frac{\sigma^2 + \tau^2/\theta_1^2}{\sum_i (X_i - \bar{X})^2}.$$

Using the standard OLS regression framework the variance of $\hat{\beta}_{\hat{Y}}$ can be estimated by:

$$\widehat{\text{Var}}(\hat{\beta}_{\hat{Y}}) = \frac{s^2_{\hat{Y}}}{\sum_i (X_i - \bar{X})^2} = \frac{s^2/\hat{\theta}_1^2}{\sum_i (X_i - \bar{X})^2}. \tag{S2.16}$$

By replacing $\hat{\theta}_1$ by θ_1 in the above, the quantity in (S2.16) is in expectation equal to $\text{Var}(\hat{\beta}_{\hat{Y}})$ (defined above). The quantity in (S2.16) is used in the zero-variance method to construct confidence intervals for $\hat{\beta}_{\hat{Y}}$, by replacing $\widehat{\text{Var}}(\hat{\beta}_{\hat{Y}})$ for $\widehat{\text{Var}}(\hat{\beta}_Y)$ in equation S2.14. In conclusion, this zero-variance approach will provide confidence intervals for the treatment effect estimator while assuming there is no variance in $\hat{\theta}_1$ (giving it its name zero-variance method). Although the zero-variance approach wins in terms of simplicity, it may underestimate the variability of the ratio since the variance in $\hat{\theta}_1$ is assumed zero.

Bootstrap

An alternative for defining confidence intervals for the corrected treatment effect estimator $\hat{\beta}_Y$ is by using a non-parametric bootstrap [15]. We propose the following stepwise procedure:

1. Draw a random sample with replacement of size K of the calibration sample $(Y^{*(c)}, Y^{(c)})$ to estimate $\hat{\theta}_{1_B}$ defined in (S2.12).
2. Draw a random sample with replacement of size N of the trial data (Y^*, X) to calculate the corrected treatment effect estimate by $\hat{\beta}_{Y_B} = \beta_{Y_B^*}/\hat{\theta}_{1_B}$. Where $\beta_{Y_B^*}$ is defined in (S2.3).
3. Repeat step 1-2 B times, with B large (e.g. 999 times).
4. Approximate confidence intervals are given by the $(\alpha/2, 1-\alpha/2)$ percentile of the distribution of $\hat{\beta}_{Y_B}$.

S2.3.2. Differential measurement error

For corrections for endpoints that suffer from differential measurement error we will here assume the existence of a pilot trial, which serves as an external calibration set, where both

treatments are allocated at random that serves as an external calibration set to estimate the measurement error model in (S2.11). For notational convenience we rewrite the linear model in equation (S2.11) in matrix form as:

$$Y^* = X\theta + e, \tag{S2.17}$$

where $\mathrm{E}(e) = 0$ and $\mathrm{E}(ee') = \Sigma$, a positive definite matrix, with τ_X^2 on its diagonal. Further, $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\theta_{00}, \theta_{01} - \theta_{00}, \theta_{10}, \theta_{11} - \theta_{10})$. In the external calibration set, the measurement error parameters $\hat{\theta}$ can be estimated by,

$$\hat{\theta} = (X^{(c)\prime}X^{(c)})^{-1}X^{(c)\prime}Y^{(c)}, \tag{S2.18}$$

with variance,

$$\mathrm{Var}(\hat{\theta}) = (X^{(c)\prime}X^{(c)})^{-1}X^{(c)\prime}\Sigma X^{(c)}(X^{(c)\prime}X^{(c)})^{-1}.$$

See [12] for a discussion on different estimators for the above defined variance. From section 2.5 it follows that natural estimators for α_Y and β_Y are,

$$\hat{\alpha}_Y = (\hat{\alpha}_{Y^*} - \hat{\theta}_{00})/\hat{\theta}_{10} \quad \text{and} \quad \hat{\beta}_Y = (\hat{\beta}_{Y^*} + \hat{\alpha}_{Y^*} - \hat{\theta}_{01})/\hat{\theta}_{11} - \hat{\alpha}_Y, \tag{S2.19}$$

where $\hat{\theta}_{00}$, $\hat{\theta}_{10}$, $\hat{\theta}_{01}$ and $\hat{\theta}_{11}$ are estimated from the external calibration set. Here it is assumed that both $\hat{\theta}_{10}$ and $\hat{\theta}_{11}$ are bounded away from zero (for reasons similar to those mentioned in section 3.1).

By multivariate Taylor expansions, the first moments of the estimators $\hat{\alpha}_Y$ and $\hat{\beta}_Y$ defined in (S2.19) can be approximated [11], in the same way as the estimators for systematic measurement error (section 4.1),

$$\begin{aligned}
\mathrm{E}\big[\hat{\alpha}_Y\big] &\approx \alpha_Y + \frac{1}{\theta_{10}^2}\Big[\alpha_Y \mathrm{Var}\big(\hat{\theta}_{10}\big) + \mathrm{Cov}\big(\hat{\theta}_{00}, \hat{\theta}_{10}\big)\Big], \\
\mathrm{E}\big[\hat{\beta}_Y\big] &\approx \beta_Y + \frac{1}{\theta_{11}^2}\Big[(\beta_Y + \alpha_Y)\mathrm{Var}\big(\hat{\theta}_{11}\big) + \mathrm{Cov}\big(\hat{\theta}_{01}, \hat{\theta}_{11}\big)\Big] \\
&\quad - \frac{1}{\theta_{10}^2}\Big[\alpha_Y \mathrm{Var}\big(\hat{\theta}_{10}\big) + \mathrm{Cov}\big(\hat{\theta}_{00}, \hat{\theta}_{10}\big)\Big].
\end{aligned}$$

From this, it is apparent that the estimators $\hat{\alpha}_Y$ and $\hat{\beta}_Y$ defined in (S2.19) are consistent (details are found in section S2.5). In the subsequent sections we review the Delta method, zero-variance and propose a bootstrap for specifying confidence limits for the estimator of the treatment effect under differential measurement error of the endpoints.

Delta method

The variance of the estimator $\hat{\beta}_Y$ defined in (S2.19) can be approximated by the Delta method [11]:

$$\begin{aligned}
\mathrm{Var}\big(\hat{\beta}_Y\big) \approx \frac{1}{\theta_{11}^2}\Big[&\big(\beta_Y + \alpha_Y\big)^2 \mathrm{Var}\big(\hat{\theta}_{11}\big) + \mathrm{Var}\big(\hat{\beta}_{Y^*}\big) + \mathrm{Var}\big(\hat{\alpha}_{Y^*}\big) + \\
&2\mathrm{Cov}\big(\hat{\alpha}_{Y^*}, \hat{\beta}_{Y^*}\big) + \mathrm{Var}\big(\hat{\theta}_{01}\big) + 2\big(\beta_Y + \alpha_Y\big)\mathrm{Cov}\big(\hat{\theta}_{11}, \hat{\theta}_{01}\big)\Big] +
\end{aligned}$$

$$\mathrm{Var}\big(\hat{\alpha}_Y\big),$$

where $\mathrm{Var}\big(\hat{\alpha}_Y\big)$ is approximated by:

$$\mathrm{Var}\Big(\frac{\hat{\alpha}_{Y^*} - \hat{\theta}_{00}}{\hat{\theta}_{10}}\Big) \approx \frac{1}{\theta_{10}^2}\Big[\mathrm{Var}\big(\hat{\alpha}_{Y^*}\big) + \alpha_Y^2\mathrm{Var}\big(\hat{\theta}_{10}\big) + \mathrm{Var}\big(\hat{\theta}_{00}\big) + 2\alpha_Y\mathrm{Cov}\big(\hat{\theta}_{00}, \hat{\theta}_{10}\big)\Big].$$

An approximate confidence interval for the estimator $\hat{\beta}_Y$ in (S2.19) is:

$$\hat{\beta}_Y \pm t_{(\alpha/2,n-2)}\sqrt{\mathrm{Var}\big(\hat{\beta}_Y\big)}. \tag{S2.20}$$

An approximation of θ_{11}, θ_{10}, $\theta_{11}^2\sigma^2 + \tau_1^2$, $\theta_{10}^2\sigma^2 + \tau_0^2$, τ_1^2, τ_0^2, β_Y and α_Y in the above is provided by: $\hat{\theta}_{11}$, $\hat{\theta}_{10}$, s_1^2, s_0^2, t_1^2, t_0^2, $\hat{\beta}_Y$ and $\hat{\alpha}_Y$ [11].

Zero-variance method

The zero-variance method adjusts the observed endpoints Y_i^* by

$$\hat{Y}_{ix} = (Y_{ix}^* - \hat{\theta}_{0x})/\hat{\theta}_{1x},$$

for $x \in \{0, 1\}$ and $\hat{\theta}_{0x}$) and $\hat{\theta}_{1x}$ derived from (S2.18). In the zero-variance method the above defined adjusted values are regressed on the treatment variable X, yielding in estimators $\hat{\alpha}_{\hat{Y}}$ and $\hat{\beta}_{\hat{Y}}$, which are, respectively, equal to the estimators $\hat{\alpha}_Y$ and $\hat{\beta}_Y$ defined in (S2.19). The variance of these estimators can be approximated with a heteroscedastic consistent covariance estimator (see [12] for an overview). Confidence intervals for $\hat{\beta}_{\hat{Y}}$ are subsequently constructed by using formula S2.20. Similar to what is described in section 4.1.3 discussing the zero-variance method for systematic measurement error, this way of constructing confidence intervals neglects the variance of the θ's from the calibration data set, and will thus often yield in confidence intervals that are too narrow.

Bootstrap

We here alternatively propose a non-parametric bootstrap procedure to specify confidence limits. This entails the following steps:

1. Draw a random sample with replacement of size K of the calibration sample and estimate $\hat{\theta}$ as defined in (S2.18).

2. Draw a random sample (with replacement) of size N of the study population and calculate the effect estimate by $\hat{\alpha}_{Y_B} = (\alpha_{Y_B^*} - \hat{\theta}_{00_B})/\hat{\theta}_{10_B}$ and $\hat{\beta}_{Y_B} = (\beta_{Y_B^*} + \alpha_{Y_B^*} - \hat{\theta}_{01_B})/\hat{\theta}_{11_B} - \hat{\alpha}_{Y_B}$. Where $\beta_{Y_B^*}$ and $\alpha_{Y_B^*}$ are defined in (S2.3) and (S2.4), respectively.

3. Repeat step 1-2 B times, with B large (e.g. 999 times).

4. Approximate confidence intervals are given by the $(\alpha/2, 1 - \alpha/2)$ percentile of the distribution of $\hat{\beta}_{Y_B}$.

S2.4. Measurement error depending on prognostic factors

Assume that, $\mathrm{E}[Y|X,S] = \alpha + \beta X + \gamma S$, $\mathrm{E}[Y^*|Y,S] = Y + \zeta S$, $Y^* \perp\!\!\!\perp X|Y$ (non-differential measurement error) and $S \perp\!\!\!\perp X$ (randomization is well-performed).

Suppose that we want to estimate the effect of Y on X (i.e., β), but instead of Y we have only measured the with measurement error contaminated Y^*. If one is aware that there is a prognostic factor that confounds the relation between Y^* and Y (and this factor is measured), one could decide to regress Y^* on X and S. The regression of Y^* on X and S equals,

$$\begin{aligned}
\mathrm{E}[Y^*|X,S] &= \mathrm{E}_{Y|X,S}\{\mathrm{E}_{Y^*|X,S,Y}[Y^*|X,S,Y]|X,S\} \\
&= \mathrm{E}_{Y|X,S}\{\mathrm{E}_{Y^*|S,Y}[Y^*|S,Y]|X,S\} \\
&= \mathrm{E}_{Y|X,S}\{Y + \boxtimes S|X,S\} \\
&= \alpha + \beta X + (\gamma + \zeta)S.
\end{aligned}$$

Thus, using the with measurement error contaminated endpoint Y^* instead of the preferred endpoint Y will provide an unbiased estimation of β.

However, if one is not aware of the prognostic factor, one might naively regress Y^* on X, which equals:

$$\begin{aligned}
\mathrm{E}[Y^*|X] &= \mathrm{E}_{S|X}\{\mathrm{E}_{Y|X,S}\{\mathrm{E}_{Y^*|X,S,Y}[Y^*|X,S,Y]|X,S\}|X\} \\
&= \mathrm{E}_{S|X}\{\alpha + \beta X + (\gamma + \zeta)S|X\} \\
&= \alpha + \beta X + (\gamma + \zeta)\mathrm{E}[S].
\end{aligned}$$

In conclusion, with ignoring the prognostic factor and using the with measurement error contaminated endpoint Y^* instead of the preferred endpoint Y, the regression of Y^* on X still results in an unbiased estimation of β.

S2.5. Approximation of bias and variance in corrected estimator

S2.5.1. Systematic measurement error

Obvious estimators for α_Y and β_Y are:

$$\hat{\alpha}_Y = (\hat{\alpha}_{Y^*} - \hat{\theta}_0)/\hat{\theta}_1 \quad \text{and} \quad \hat{\beta}_Y = \hat{\beta}_{Y^*}/\hat{\theta}_1.$$

These estimators can be approximated with a second order Taylor expansion by:

$$\begin{aligned}
\frac{\hat{\beta}_{Y^*}}{\hat{\theta}_1} &\approx \frac{\beta_{Y^*}}{\theta_1} - \frac{\beta_{Y^*}}{\theta_1^2}(\hat{\theta}_1 - \theta_1) + \frac{1}{\theta_1}(\hat{\beta}_{Y^*} - \beta_{Y^*}) \\
&\quad + \frac{1}{2!}\Big[\frac{2\beta_{Y^*}}{\theta_1^3}(\hat{\theta}_1 - \theta_1)^2 - \frac{2}{\theta_1^2}(\hat{\theta}_1 - \theta_1)(\hat{\beta}_{Y^*} - \beta_{Y^*})\Big], \\
\frac{\hat{\alpha}_{Y^*}}{\hat{\theta}_1} &\approx \frac{\alpha_{Y^*}}{\theta_1} - \frac{\alpha_{Y^*}}{\theta_1^2}(\hat{\theta}_1 - \theta_1) + \frac{1}{\theta_1}(\hat{\alpha}_{Y^*} - \alpha_{Y^*}) \\
&\quad + \frac{1}{2!}\Big[\frac{2\alpha_{Y^*}}{\theta_1^3}(\hat{\theta}_1 - \theta_1)^2 - \frac{2}{\theta_1^2}(\hat{\theta}_1 - \theta_1)(\hat{\alpha}_{Y^*} - \alpha_{Y^*})\Big],
\end{aligned}$$

$$\frac{\hat{\theta}_0}{\hat{\theta}_1} \approx \frac{\theta_0}{\theta_1} - \frac{\theta_0}{\theta_1^2}(\hat{\theta}_1 - \theta_1) + \frac{1}{\theta_1}(\hat{\theta}_0 - \theta_0) + \frac{1}{2!}\Big[\frac{2\theta_0}{\theta_1^3}(\hat{\theta}_1 - \theta_1)^2 - \frac{2}{\theta_1^2}(\hat{\theta}_1 - \theta_1)(\hat{\theta}_0 - \theta_0)\Big].$$

Simplifying these terms and substraction of the latter two, will lead to the following approximations for $\hat{\alpha}_Y$ and $\hat{\beta}_Y$:

$$\frac{\hat{\beta}_{Y^*}}{\hat{\theta}_1} \approx \frac{\beta_{Y^*}}{\theta_1} + \frac{1}{\theta_1}\Big[-\frac{\beta_{Y^*}}{\theta_1}(\hat{\theta}_1 - \theta_1) + (\hat{\beta}_{Y^*} - \beta_{Y^*})\Big] + \frac{1}{\theta_1^2}\Big[\frac{\beta_{Y^*}}{\theta_1}(\hat{\theta}_1 - \theta_1)^2 - (\hat{\beta}_{Y^*} - \beta_{Y^*})(\hat{\theta}_1 - \theta_1)\Big],$$

$$\frac{\hat{\alpha}_{Y^*} - \hat{\theta}_0}{\hat{\theta}_1} \approx \frac{\alpha_{Y^*} - \theta_0}{\theta_1} + \frac{1}{\theta_1}\Big[-\frac{\alpha_{Y^*} - \theta_0}{\theta_1}(\hat{\theta}_1 - \theta_1) + (\hat{\alpha}_{Y^*} - \alpha_{Y^*}) - (\hat{\theta}_0 - \theta_0)\Big] + \frac{1}{\theta_1^2}\Big[\frac{\alpha_{Y^*} - \theta_0}{\theta_1}(\hat{\theta}_1 - \theta_1)^2 - (\hat{\alpha}_{Y^*} - \alpha_{Y^*})(\hat{\theta}_1 - \theta_1) + (\hat{\theta}_0 - \theta_0)(\hat{\theta}_1 - \theta_1)\Big].$$

Since $\mathrm{E}[\hat{\theta}_1 - \theta_1] = 0$, $\mathrm{E}[\hat{\theta}_0 - \theta_0] = 0$, $\mathrm{E}[\hat{\alpha}_{Y^*} - \alpha_{Y^*}] = 0$ and $\mathrm{E}[\hat{\beta}_{Y^*} - \beta_{Y^*}] = 0$ an approximation of the expected value of the estimator $\hat{\alpha}_Y$ is given by:

$$\begin{aligned}\mathrm{E}\Big[\frac{\hat{\alpha}_{Y^*} - \hat{\theta}_0}{\hat{\theta}_1}\Big] &\approx \frac{\alpha_{Y^*} - \theta_0}{\theta_1} + \frac{1}{\theta_1^2}\Big[\frac{\alpha_{Y^*} - \theta_0}{\theta_1}\mathrm{E}\big[(\hat{\theta}_1 - \theta_1)^2\big] \\ &\quad - \mathrm{E}\big[(\hat{\alpha}_{Y^*} - \alpha_{Y^*})(\hat{\theta}_1 - \theta_1)\big] + \mathrm{E}\big[(\hat{\theta}_0 - \theta_0)(\hat{\theta}_1 - \theta_1)\big]\Big] = \\ &= \frac{\alpha_{Y^*} - \theta_0}{\theta_1} + \frac{1}{\theta_1^2}\Big[\frac{\alpha_{Y^*} - \theta_0}{\theta_1}\mathrm{Var}\big(\hat{\theta}_1\big) - \mathrm{Cov}\big(\hat{\alpha}_{Y^*}, \hat{\theta}_1\big) + \mathrm{Cov}\big(\hat{\theta}_0, \hat{\theta}_1\big)\Big] = \\ &= \alpha_Y + \frac{1}{\theta_1^2}\Big[\frac{\tau^2[\alpha_Y - \bar{Y}^{(c)}]}{\sum(Y_j^{(c)} - \bar{Y}^{(c)})^2}\Big].\end{aligned}$$

Congruently, an approximation of the expected value of the estimator $\hat{\beta}_Y$ is given by:

$$\begin{aligned}\mathrm{E}\Big[\frac{\hat{\beta}_{Y^*}}{\hat{\theta}_1}\Big] &\approx \frac{\beta_{Y^*}}{\theta_1} + \frac{1}{\theta_1^2}\Big[\frac{\beta_{Y^*}}{\theta_1}\mathrm{E}\big[(\hat{\theta}_1 - \theta_1)^2\big] - \mathrm{E}\big[(\hat{\beta}_{Y^*} - \beta_{Y^*})(\hat{\theta}_1 - \theta_1)\big]\Big] = \\ &= \frac{\beta_{Y^*}}{\theta_1} + \frac{1}{\theta_1^2}\Big[\frac{\beta_{Y^*}}{\theta_1}\mathrm{Var}\big(\hat{\theta}_1\big)\Big] = \\ &= \beta_Y + \frac{1}{\theta_1^2}\Big[\frac{\tau^2\beta_Y}{\sum(Y_j^{(c)} - \bar{Y}^{(c)})^2}\Big].\end{aligned}$$

Only using the first order Taylor expansion of the estimators, approximations of the variance of $\hat{\alpha}_Y$ and $\hat{\beta}_Y$ are respectively:

$$\mathrm{Var}\Big(\frac{\hat{\alpha}_{Y^*} - \hat{\theta}_0}{\hat{\theta}_1}\Big) \approx \frac{1}{\theta_1^2}\Big[\alpha_Y^2\mathrm{Var}\big(\hat{\theta}_1\big) + \mathrm{Var}\big(\hat{\alpha}_{Y^*} - \hat{\theta}_0\big) - 2\alpha_Y\mathrm{Cov}\big(\hat{\theta}_1, \hat{\alpha}_{Y^*} - \hat{\theta}_0\big)\Big] =$$

S2

$$
\begin{aligned}
&= \frac{1}{\theta_1^2}\Big[\alpha_Y^2 \mathrm{Var}\big(\hat{\theta}_1\big) + \mathrm{Var}\big(\hat{\alpha}_{Y^*}\big) + \mathrm{Var}\big(\hat{\theta}_0\big) - 2\mathrm{Cov}\big(\hat{\alpha}_{Y^*}, \hat{\theta}_0\big) \\
&- 2\alpha_Y \mathrm{Cov}\big(\hat{\theta}_1, \hat{\alpha}_{Y^*}\big) + 2\alpha_Y \mathrm{Cov}\big(\hat{\theta}_1, \hat{\theta}_0\big)\Big] = \\
&= \frac{1}{\theta_1^2}\Big[\frac{(\theta_1^2\sigma^2 + \tau^2)\sum X_i^2}{N\sum (X_i - \bar{X})^2} + \alpha_Y^2 \frac{\tau^2}{\sum (Y_j^{(c)} - \bar{Y}^{(c)})^2} \\
&+ \frac{\tau^2 \sum (Y_j^{(c)})^2}{K\sum (Y_j^{(c)} - \bar{Y}^{(c)})^2} \\
&+ 2\alpha_Y \frac{-\tau^2 \bar{Y}^{(c)}}{\sum (Y_j^{(c)} - \bar{Y}^{(c)})^2}\Big] = \\
&= \frac{1}{\theta_1^2}\Big[\frac{(\theta_1^2\sigma^2 + \tau^2)\sum X_i^2}{N\sum (X_i - \bar{x})^2} + \alpha_Y^2 \frac{\tau^2}{\sum (y_j^{(c)} - \bar{y}^{(c)})^2} \\
&+ \frac{\tau^2(\sum (y_j^{(c)} - \bar{y}^{(c)})^2 + K(\bar{y}^{(c)})^2)}{K\sum (y_j^{(c)} - \bar{y}^{(c)})^2} \\
&- 2\alpha_Y \frac{\tau^2 \bar{y}^{(c)}}{\sum (y_j^{(c)} - \bar{y}^{(c)})^2}\Big] = \\
&= \frac{1}{\theta_1^2}\Big[\frac{(\theta_1^2\sigma^2 + \tau^2)\sum x_i^2}{N\sum (x_i - \bar{x})^2} + \tau^2\Big(\frac{1}{K} + \frac{(\bar{y}^{(c)} - \alpha_Y)^2}{\sum (y_j^{(c)} - \bar{y}^{(c)})^2}\Big)\Big],
\end{aligned}
$$

$$
\mathrm{Var}\Big(\frac{\hat{\beta}_{Y^*}}{\hat{\theta}_1}\Big) \approx \frac{1}{\theta_1^2}\Big[\frac{\theta_1^2\sigma^2 + \tau^2}{\sum (x_i - \bar{x})^2} + \frac{\beta_Y^2 \tau^2}{\sum (y_j^{(c)} - \bar{y}^{(c)})^2}\Big].
$$

Fieller method

Assume that $\hat{\beta}_{Y^*}$ and $\hat{\theta}_1$ are normally distributed (note that this assumption is satisfied with large study samples (N) and large calibration samples (K)). The sum of two normally distributed variables is normally distributed, hence, $\hat{\beta}_{Y^*} - \beta_Y \hat{\theta}_1$ is normally distributed. Furthermore, we have,

$$
\mathrm{Var}(\hat{\beta}_{Y^*} - \beta_Y \hat{\theta}_1) = \mathrm{Var}(\hat{\beta}_{Y^*}) + \beta_Y^2 \mathrm{Var}(\hat{\theta}_1).
$$

Where,

$$
\begin{aligned}
\mathrm{Var}(\hat{\beta}_{Y^*}) &= \frac{\theta_1^2\sigma^2 + \tau^2}{\sum (x_i - \bar{x})^2} \\
\mathrm{Var}(\hat{\theta}_1) &= \frac{\tau^2}{\sum (y_j^{(c)} - \bar{y}^{(c)})^2}
\end{aligned}
$$

If we now divide the term $\hat{\beta}_{Y^*} - \beta_Y \hat{\theta}_1$ by its standard deviation, we get:

$$T_0 = \frac{\hat{\beta}_{Y^*} - \beta_Y \hat{\theta}_1}{\sqrt{\frac{\theta_1^2\sigma^2+\tau^2}{\sum(x_i-\bar{x})^2} + \frac{\tau^2}{\sum(y_j^{(c)}-\bar{y}^{(c)})^2}\beta_Y^2}} \tag{S2.21}$$

We are interested to find the set of β_Y values for which the corresponding T_0 values lie within the $(1-\alpha)$ quantiles of the t-distribution with $N-2$ degrees of freedom (this only holds approximately, see for details [14]). Let us denote these values by t_q, from (S2.21) we have,

$$(\frac{\tau^2}{\sum(y_j^{(c)} - \bar{y}^{(c)})^2} t_q^2 - \hat{\theta}_1^2)\beta_Y^2 + 2\hat{\beta}_{Y^*}\hat{\theta}_1\beta_Y + (\frac{\theta_1^2\sigma^2 + \tau^2}{\sum(x_i - \bar{x})^2} t_q^2 - \hat{\beta}_{Y^*}^2) = 0.$$

In the case that $\hat{\theta}_1$ is significantly different from zero at a significance level of α (that is, $\hat{\theta}_1 / \sqrt{\frac{\tau^2}{\sum(y_j^{(c)}-\bar{y}^{(c)})^2}} > t_q$), solving this for β_Y results in the following $(1-\alpha)$ confidence intervals:

$$\beta_Y = \frac{-\hat{\beta}_{Y^*}\hat{\theta}_1 \pm \sqrt{\hat{\beta}_{Y^*}^2\hat{\theta}_1^2 - (\frac{\tau^2}{\sum(y_j^{(c)}-\bar{y}^{(c)})^2} t_q^2 - \hat{\theta}_1^2)(\frac{\theta_1^2\sigma^2+\tau^2}{\sum(x_i-\bar{x})^2} t_q^2 - \hat{\beta}_{Y^*}^2)}}{\frac{\tau^2}{\sum(y_j^{(c)}-\bar{y}^{(c)})^2} t_q^2 - \hat{\theta}_1^2}.$$

In the other case, the confidence intervals are unbounded, see for more details [14].

S2.5.2. Differential measurement error

Obvious estimators for α_Y and β_Y are:

$$\hat{\alpha}_Y = (\hat{\alpha}_{Y^*} - \hat{\theta}_{00})/\hat{\theta}_{10} \quad \text{and} \quad \hat{\beta}_Y = (\hat{\beta}_{Y^*} + \hat{\alpha}_{Y^*} - \hat{\theta}_{01})/\hat{\theta}_{11} - \hat{\alpha}_Y.$$

These estimators can be approximated with a second order Taylor expansion by:

$$\begin{aligned}
\frac{\hat{\alpha}_{Y^*} - \hat{\theta}_{00}}{\hat{\theta}_{10}} &\approx \frac{\alpha_{Y^*} - \theta_{00}}{\theta_{10}} + \frac{1}{\theta_{10}}\Big[- \frac{\alpha_{Y^*} - \theta_{00}}{\theta_{10}}(\hat{\theta}_{10} - \theta_{10}) + (\hat{\alpha}_{Y^*} - \alpha_{Y^*}) - (\hat{\theta}_{00} - \theta_{00})\Big] \\
&\quad + \frac{1}{\theta_{11}^2}\Big[\frac{\alpha_{Y^*} - \theta_{00}}{\theta_{10}}(\hat{\theta}_{10} - \theta_{10})^2 - (\hat{\alpha}_{Y^*} - \alpha_{Y^*})(\hat{\theta}_{10} - \theta_{10}) \\
&\quad + (\hat{\theta}_{00} - \theta_{00})(\hat{\theta}_{10} - \theta_{10})\Big], \\
\frac{\hat{\beta}_{Y^*} - \hat{\theta}_{01}}{\hat{\theta}_{11}} &\approx \frac{\beta_{Y^*} - \theta_{01}}{\theta_{11}} + \frac{1}{\theta_{11}}\Big[- \frac{\beta_{Y^*} - \theta_{01}}{\theta_{11}}(\hat{\theta}_{11} - \theta_{11}) + (\hat{\beta}_{Y^*} - \beta_{Y^*}) - (\hat{\theta}_{01} - \theta_{01})\Big] \\
&\quad + \frac{1}{\theta_{11}^2}\Big[\frac{\beta_{Y^*} - \theta_{01}}{\theta_{11}}(\hat{\theta}_{11} - \theta_{11})^2 - (\hat{\beta}_{Y^*} - \beta_{Y^*})(\hat{\theta}_{11} - \theta_{11}) \\
&\quad + (\hat{\theta}_{01} - \theta_{01})(\hat{\theta}_{11} - \theta_{11})\Big], \\
\frac{\hat{\alpha}_{Y^*}}{\hat{\theta}_{11}} &\approx \frac{\alpha_{Y^*}}{\theta_{11}} + \frac{1}{\theta_{11}}\Big[- \frac{\alpha_{Y^*}}{\theta_{11}}(\hat{\theta}_{11} - \theta_{11}) + (\hat{\alpha}_{Y^*} - \alpha_{Y^*}))\Big]
\end{aligned}$$

$$+ \frac{1}{\theta_{11}^2}\Big[\frac{\alpha_{Y^*}}{\theta_{11}}(\hat{\theta}_{11} - \theta_{11})^2 - (\hat{\alpha}_{Y^*} - \alpha_{Y^*})(\hat{\theta}_{11} - \theta_{11})\Big].$$

S2

Congruent to the results for the estimators under systematic measurement error, we can conclude:

$$\mathrm{E}\Big[\frac{\hat{\alpha}_{Y^*} - \hat{\theta}_{00}}{\hat{\theta}_{10}}\Big] \approx \alpha_Y + \frac{1}{\theta_{10}^2}\Big[\alpha_Y \mathrm{Var}(\hat{\theta}_{10}) + \mathrm{Cov}(\hat{\theta}_{00}, \hat{\theta}_{10})\Big].$$

Congruently, an approximation of the expected value of the estimator $\hat{\beta}_Y$ is given by:

$$\begin{aligned}\mathrm{E}\Big[\frac{\hat{\beta}_{Y^*} + \hat{\alpha}_{Y^*} - \hat{\theta}_{01}}{\hat{\theta}_{11}} - \hat{\alpha}_Y\Big] \approx \beta_Y &+ \frac{1}{\theta_{11}^2}\Big[(\beta_Y + \alpha_Y)\mathrm{Var}(\hat{\theta}_{11}) + \mathrm{Cov}(\hat{\theta}_{01}, \hat{\theta}_{11})\Big] \\ &- \frac{1}{\theta_{10}^2}\Big[\alpha_Y \mathrm{Var}(\hat{\theta}_{10}) + \mathrm{Cov}(\hat{\theta}_{00}, \hat{\theta}_{10})\Big].\end{aligned}$$

And the variance of the estimators is approximated by:

$$\begin{aligned}\mathrm{Var}\Big(\frac{\hat{\alpha}_{Y^*} - \hat{\theta}_{00}}{\hat{\theta}_{10}}\Big) \approx\ & \frac{1}{\theta_{10}^2}\Big[\mathrm{Var}(\hat{\alpha}_{Y^*}) \\ &+ \alpha_Y^2 \mathrm{Var}(\hat{\theta}_{10}) + \mathrm{Var}(\hat{\theta}_{00}) + 2\alpha_Y \mathrm{Cov}(\hat{\theta}_{00}, \hat{\theta}_{10})\Big], \\ \mathrm{Var}\Big(\frac{\hat{\beta}_{Y^*} + \hat{\alpha}_{Y^*} - \hat{\theta}_{01}}{\hat{\theta}_{11}} - \hat{\alpha}_Y\Big) \approx\ & \frac{1}{\theta_{11}^2}\Big[(\beta_Y + \alpha_Y)^2 \mathrm{Var}(\hat{\theta}_{11}) + \mathrm{Var}(\hat{\beta}_{Y^*}) + \mathrm{Var}(\hat{\alpha}_{Y^*}) \\ &+ 2\mathrm{Cov}(\hat{\alpha}_{Y^*}, \hat{\beta}_{Y^*}) + \mathrm{Var}(\hat{\theta}_{01}) \\ &+ 2(\beta_Y + \alpha_Y)\mathrm{Cov}(\hat{\theta}_{11}, \hat{\theta}_{01})\Big] \\ &+ \mathrm{Var}(\hat{\alpha}_Y).\end{aligned}$$

Note that in the case of differential measurement error, we assume that $\mathrm{Cov}(\hat{\theta}_{11}, \hat{\theta}_{00}) = 0$, $\mathrm{Cov}(\hat{\theta}_{11}, \hat{\theta}_{10}) = 0$, $\mathrm{Cov}(\hat{\theta}_{01}, \hat{\theta}_{00}) = 0$ and $\mathrm{Cov}(\hat{\theta}_{01}, \hat{\theta}_{10}) = 0$.

References

[1] E. T. Poehlman, Effects of endurance and resistance training on total daily energy expenditure in young women: A controlled randomized trial, Journal of Clinical Endocrinology & Metabolism 87 (3) (2002) 1004–1009. doi:10.1210/jc.87.3.1004.

[2] G. Plasqui, K. R. Westerterp, Physical activity assessment with accelerometers: An evaluation against doubly labeled water, Obesity 15 (10) (2007) 2371–2379. doi:10.1038/oby.2007.281.

[3] J. W. J. Bijlsma, P. M. J. Welsing, T. G. Woodworth, L. M. Middelink, A. Pethö-Schramm, C. Bernasconi, M. E. A. Borm, C. H. Wortel, E. J. ter Borg, Z. N. Jahangier, W. H. van der Laan, G. A. W. Bruyn, P. Baudoin, S. Wijngaarden, P. A. J. M. Vos, R. Bos, M. J. F. Starmans, E. N. Griep, J. R. M. Griep-Wentink, C. F. Allaart, A. H. M. Heurkens, X. M. Teitsma, J. Tekstra, A. C. A. Marijnissen, F. P. J. Lafeber, J. W. G. Jacobs, Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial, The Lancet 388 (10042) (2016) 343–355. doi:10.1016/S0140-6736(16)30363-4.

[4] M. L. L. Prevoo, M. A. Van'T Hof, H. H. Kuper, M. A. Van Leeuwen, L. B. A. Van De Putte, P. L. C. M. Van Riel, Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis, Arthritis & Rheumatism 38 (1) (1995) 44–48. doi:10.1002/art.1780380107.

[5] T. Pincus, Y. Yazici, T. Sokka, Complexities in assessment of rheumatoid arthritis: Absence of a single gold standard measure, Rheumatic Disease Clinics of North America 35 (4) (2009) 687–697. doi:10.1016/j.rdc.2009.10.002.

[6] J. Anderson, L. Caplan, J. Yazdany, M. L. Robbins, T. Neogi, K. Michaud, K. G. Saag, J. R. O'dell, S. Kazi, Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice, Arthritis Care & Research 64 (5) (2012) 640–647. arXiv:NIHMS150003, doi:10.1002/acr.21649.

[7] E. H. Choy, B. Khoshaba, D. Cooper, A. MacGregor, D. L. Scott, Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice, Arthritis & Rheumatism 59 (2) (2008) 192–199. doi:10.1002/art.23342.

[8] M. Makrides, C. Crowther, R. Gibson, R. Gibson, C. Skeaff, Efficacy and tolerability of low-dose iron supplements during pregnancy: A randomized controlled trial, American Journal of Clinical Nutrition 78 (1) (2003) 145–153. doi:10.1093/ajcn/78.1.145.

[9] R. Davidson, J. MacKinnon, Econometric Theory and Methods, Oxford University Press, New York, NY, 2004.

[10] 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.

[11] J. Buonaccorsi, Measurement errors, linear calibration and inferences for means, Computational Statistics & Data Analysis 11 (3) (1991) 239–257. doi:10.1016/0167-9473(91)90083-E.

[12] J. S. Long, L. H. Ervin, Using heteroscedasticity consistent standard errors in the linear regression model, The American Statistician 54 (3) (2000) 217–224. doi:10.2307/2685594.

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

[14] V. H. Franz, Ratios: A short guide to confidence limits and proper use (2007). arXiv:arXiv:0710.2024v1.

[15] B. Efron, Bootstrap methods: Another look at the jackknife, Annals of Statistics 7 (1) (1979) 1–26. doi:10.1214/aos/1176344552.

[16] E. C. Fieller, The biological standardization of insulin, Supplement to the Journal of the Royal Statistical Society 7 (1) (1940) 1–64. doi:10.2307/2983630.

S3

Supplementary material Chapter 3

These are the supplementary materials accompanying Chapter 3. The supplementary materials are structured as follows. In section S3.1, the variance of the regression calibration estimator is derived. In section S3.2, the variance of the maximum likelihood estimator for replicates study is derived.

S3.1. Variance estimation: standard regression calibration

Covariate measurement error. The variance–covariance matrix for the standard regression estimator $\hat{\beta}_{\text{RC}}$ can be approximated by using the multivariate delta method as described by [1], given by

$$\hat{\Sigma}_{\beta_{\text{RC}}}(j_1, j_2) = (\hat{A}' \hat{\Sigma}_{\beta^*} \hat{A})_{j_1, j_2} + \hat{\beta}^* \hat{\Sigma}_{A, j_1, j_2} \hat{\beta}^{*\prime}, \qquad j_1, j_2 = 1, \dots, (k+2), \tag{S3.1}$$

where $\hat{A}$ is the inverse of the calibration model matrix $\hat{\Lambda}$. Further, $\hat{\Sigma}_{\beta^*}$ is the variance–covariance matrix obtained from the naive regression defined in equation (3.2) in the main chapter and $\hat{\Sigma}_{A, j_1, j_2}$ is the $(k+2) \times (k+2)$ matrix relating the j_1th and j_2th column of A (we refer to Appendix of [1] for a derivation). Additionally, the so-called zero-variance variance–covariance matrix for $\hat{\beta}$ can be estimated by $\hat{A}' \Sigma_{\beta^*} \hat{A}$ (i.e., by omitting the variance in the calibration model matrix).

A $100(1 - \alpha)$ percent confidence interval for the jth element of $\hat{\beta}_{\text{RC}}$ is then

$$\hat{\beta}_{\text{RC}\,j} \pm \sqrt{\text{Var}(\hat{\beta}_{\text{RC}\,j})}, \tag{S3.2}$$

where $\text{Var}(\hat{\beta}_{\text{RC}\,j})$ is the jth element on the diagonal of $\hat{\Sigma}_{\beta_{\text{RC}}}$. The variance–covariance matrix $\hat{\Sigma}_{\beta_{\text{RC}}}$ can be obtained by either using the delta variance–covariance matrix or zero-variance variance–covariance matrix. In general, the zero-variance variance–covariance matrix will underestimate the true variance–covariance matrix and thus lead to too narrow confidence intervals.

Other methods to construct confidence intervals include stratified bootstrap [2] and the Fieller method [3–6]. In case of covariate measurement error, the Fieller method can

only be applied to construct a $100(1-\alpha)$ percent confidence interval for the first element of $\hat{\beta}_{\text{RC}}$, i.e., $\hat{\phi}_{\text{RC}}$. From [6] we obtain:

$$\{f_1 \pm \sqrt{f_1^2 - f_0 f_2}/f_2\}, \tag{S3.3}$$

where $f_0 = z^2_{\alpha/2}\text{Var}(\hat{\phi}^*) - \hat{\phi}^*$, $f_1 = z^2_{\alpha/2}\text{Cov}(\hat{\phi}^*, \hat{\lambda}_1) - \hat{\phi}^*\hat{\lambda}_1$, $f_2 = z^2_{\alpha/2}\text{Var}(\hat{\lambda}_1) - \hat{\lambda}_1^2$. Where it is assumed that $\text{Cov}(\hat{\phi}^*, \hat{\lambda}_1)$ is null. If the $(1-\alpha)\times 100\%$ confidence interval of $\hat{\lambda}_1$ includes 0, the Fieller method does not lead to bounded confidence intervals. Bootstrap confidence intervals are obtained by sampling the people in the validation set separately from the people not included in the validation set [2] and taking the $(100-\alpha)$ percentiles of the obtained distribution.

Outcome measurement error. The variance–covariance matrix for the standard regression estimator $(\hat{\beta_{\text{RC}}}, 1)$ can be approximated by applying the multivariate delta method similar to the variance obtained for the corrected estimator for covariate maesurement error,

$$\hat{\Sigma}_{(\beta_{\text{RC}},1)}(j_1, j_2) = (B'\hat{\Sigma}_{(\beta^*,1)}B)_{j_1,j_2} + (\hat{\beta}^*, 1)\hat{\Sigma}_{B,j_1,j_2}(\hat{\beta}^*, 1)', \qquad j_1, j_2 = 1, \ldots, (k+3),$$

where $\hat{B}$ is the inverse of the measurement error model matrix $\hat{\Theta}$. $\hat{\Sigma}_{(\beta^*,1)}$ is a $(k+3)\times(k+3)$ matrix where the upper $(k+2)\times(k+2)$ comprises the variance–covariance matrix obtained from the uncorrected regression defined by model (3.6) and the last row and column contain zeros. Further, $\hat{\Sigma}_{B,j_1,j_2}$ is the $(k+3)\times(k+3)$ matrix relating the j_1th and j_2th column of B (similar to [1]). The so-called zero-variance variance–covariance matrix for $\hat{\beta}$ can be estimated by $B'\hat{\Sigma}_{(\beta^*,1)}B$.

A $100(1-\alpha)$ percent confidence interval can be obtained from equation (S3.2). Further, $100(1-\alpha)$ percent confidence intervals for $\hat{\phi}$ and $\hat{\gamma}$ can be approximated by the Fieller method as defined in model S3.3, where $f_0 = \hat{\phi}^* - z^2_{\alpha/2}\text{Var}(\hat{\phi}^*)$, $f_1 = \hat{\phi}^*/\hat{\theta}_1 - z^2_{\alpha/2}\text{Cov}(\hat{\phi}^*, 1/\hat{\theta}_1)$, $f_2 = 1/\hat{\lambda}_1^2 - z^2_{\alpha/2}\text{Var}(1/\hat{\lambda}_1)$ and idem for $\hat{\gamma}$. Additionally, bootstrap can be used to construct confidence intervals for $\hat{\beta_{\text{RC}}}$. Bootstrap confidence intervals are obtained by sampling the individuals in the internal adjustment set separately from the individuals not included in the internal adjustment set and taking the $(100-\alpha)$ percentiles of the obtained distribution.

Differential outcome measurement error in univariable analyses. The varia-nce–covariance matrix for the standard regression estimator $(\hat{\beta_{\text{RC}}}, 1)$ can be estimated similar to non-differential outcome measurement error defined above (by using the measurement error matrices for differential outcome measurement error). Confidence intervals can then be obtained from equation (S3.2). Bootstrap confidence intervals are obtained by sampling the individuals in the internal adjustment set separately from the individuals not included in the internal adjustment set and taking the $(100-\alpha)$ percentiles of the obtained distribution.

S3.2. Variance estimation: maximum likelihood for replicates studies

The variance–covariance matrix for the maximum likelihood estimator $\hat{\beta}_{\text{MLE}}$ can be approximated by the multivariate delta method [7]. Denote $\zeta^* = (\delta_0, \delta_Z, \sigma^2_{Y|Z}, \kappa_0, \kappa_Y, \kappa_Z, \sigma^2_{X|Y,Z})$, leaving the τ^2 from ζ in the main chapter (see section 3.3.3) out as this parameter is not needed for the estimation of $\beta = (\alpha, \phi, \gamma)$. A standard result from linear mixed models is that the estimators of fixed parameters are asymptotically uncorrelated with the estimators of the variance component parameters [7]. If one further assumes that the estimators from the linear model of Y given Z are uncorrelated with the parameters estimated in the linear mixed model, it follows for large samples that $\hat{\zeta}^*$ is multivariate normal with mean ζ and variance covariance matrix $\text{Var}(\hat{\zeta})$ equal to:

$$\begin{pmatrix}
\text{Var}(\hat{\delta}_0) & \text{Cov}(\hat{\delta}_0, \hat{\delta}_Z) & 0 & 0 & 0 & 0 & 0 \\
\text{Cov}(\hat{\delta}_Z, \hat{\delta}_0) & \text{Var}(\hat{\delta}_Z) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \text{Var}(\hat{\sigma}^2_{Y|Z}) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \text{Var}(\hat{\kappa}_0) & \text{Cov}(\hat{\kappa}_0, \hat{\kappa}_Y) & \text{Cov}(\hat{\kappa}_0, \hat{\kappa}_Z) & 0 \\
0 & 0 & 0 & \text{Cov}(\hat{\kappa}_Y, \hat{\kappa}_0) & \text{Var}(\hat{\kappa}_Y) & \text{Cov}(\hat{\kappa}_Y, \hat{\kappa}_Z) & 0 \\
0 & 0 & 0 & \text{Cov}(\hat{\kappa}_Z, \hat{\kappa}_0) & \text{Cov}(\hat{\kappa}_Z, \hat{\kappa}_Y) & \text{Var}(\hat{\kappa}_Z) & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \text{Var}(\hat{\sigma}^2_{X|Y,Z})
\end{pmatrix}$$

If $g : \mathbb{R}^{5+2k} \longrightarrow \mathbb{R}^{2+k}$ is the function that transforms ζ^* to $\beta_{\text{ML}} = (\alpha_{\text{ML}}, \phi_{\text{ML}}, \gamma_{\text{ML}})$, as defined in the main chapter, then by the multivariate delta method it follows that in large samples:

$$\hat{\beta}_{\text{ML}} \sim N(\beta_{\text{ML}}, Jg\text{Var}(\hat{\zeta})(Jg)'),$$

Where J is the Jacobian matrix of g:

$$Jg = \begin{pmatrix}
\frac{\partial \phi}{\partial \delta_0} & \frac{\partial \phi}{\partial \delta_Z} & \frac{\partial \phi}{\partial \sigma^2_{Y|Z}} & \cdots & \frac{\partial \phi}{\partial \sigma^2_{X|Y,Z}} \\
\frac{\partial \alpha}{\partial \delta_0} & \frac{\partial \alpha}{\partial \delta_Z} & \frac{\partial \alpha}{\partial \sigma^2_{Y|Z}} & \cdots & \frac{\partial \alpha}{\partial \sigma^2_{X|Y,Z}} \\
\frac{\partial \gamma}{\partial \delta_0} & \frac{\partial \gamma}{\partial \delta_Z} & \frac{\partial \gamma}{\partial \sigma^2_{Y|Z}} & \cdots & \frac{\partial \gamma}{\partial \sigma^2_{X|Y,Z}}
\end{pmatrix}.$$

Confidence intervals can then be obtained from equation (S3.2). Bootstrap confidence intervals are obtained by sampling the individuals in the internal adjustment set separately from the individuals not included in the internal adjustment set and taking the $(100 - \alpha)$ percentiles of the obtained distribution.

References

[1] B. Rosner, D. Spiegelman, W. C. Willett, Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error, American Journal of Epidemiology 132 (4) (1990) 734–745. doi:10.1093/oxfordjournals.aje.a115715.

[2] 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.

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

[4] V. H. Franz, Ratios: A short guide to confidence limits and proper use (2007). arXiv:arXiv:0710.2024v1.

[5] C. Frost, S. G. Thompson, Correcting for regression dilution bias: comparison of methods for a single predictor variable, Journal of the Royal Statistical Society: Series A (Statistics in Society) 163 (2) (2000) 173–189. doi:10.1111/1467-985X.00164.

[6] L. Nab, R. H. H. Groenwold, P. M. J. Welsing, M. Smeden, Measurement error in continuous endpoints in randomised trials: Problems and solutions, Statistics in Medicine 38 (27) (2019) 5182–5196. doi:10.1002/sim.8359.

[7] J. W. Bartlett, B. L. De Stavola, C. Frost, Linear mixed models for replication data to efficiently allow for covariate measurement error, Statistics in Medicine 28 (25) (2009) 3158–3178. doi:10.1002/sim.3713.

S5

Supplementary material Chapter 5

These are the supplementary materials accompanying Chapter 5. The supplementary materials are structured as follows. In section S5.1 we introduce notation, describe the implications of exposure measurement error and describe the different analyses used in the main chapter. In section S5.2 the parameters of the simulation study from the main chapter are presented. Section S5.3 contains the additional results from the simulation study in the main chapter that were left out for brevity there.

S5.1. Notation, impact of measurement error and different analysis strategies

Throughout the main paper, our interest is the causal effect of the exposure VAT on the outcome insulin resistance IR, adjusted for a predefined set of k confounders, jointly written as Z (e.g., age, sex and total body fat). We assume a linear model for the outcome without interaction between exposure and covariates:

$$\text{IR} = \text{intercept} + \beta\text{VAT} + \gamma^{'} Z + \varepsilon. \tag{S5.1}$$

Here, we assume that the residuals errors ε are independent of VAT and confounders Z, with mean 0 and variance σ^2. Additionally, γ is assumed a $k \times 1$ vector of regression coefficients. The parameter β in equation (S5.1) is the parameter of interest. We consider the setting that instead of the exposure of interest, VAT, WC is measured. The variable WC is the error-prone substitute measure for VAT, where we assume that $\text{WC} = \theta_1\text{VAT} + U$, where U is a random variable, with mean 0 and variance τ^2, and U is assumed independent of VAT. The factor θ is a scalar, used to scale VAT to the same scale as WC. We also assume *non-differential* measurement error, i.e., $\text{WC}|\text{VAT} \perp\!\!\!\perp Y$. This form of measurement error is referred to as random (or sometimes classical) measurement error if $\theta = 1$ and systematic (or sometimes linear) measurement error otherwise [1, 2]. Since the substitute measure is often measured on a different scale than the true measure, measurement error will often be of the systematic form. Using WC instead of VAT in the linear model yields:

$$\text{E}[\text{IR}|\text{WC}, Z] = \text{intercept}^{*} + \beta^{*}\text{WC} + \gamma^{*'} Z. \tag{S5.2}$$

Under this model, by the law of total expectation, we have $E[IR|WC, Z]$ = intercept + $\beta \times E[VAT|WC, Z] + \gamma^* Z$, which relies on the assumption that the measurement error is non-differential [3]. It follows that,

$$\beta^* = \alpha\beta \quad \text{with} \quad \alpha = \frac{\text{Var}(WC, VAT|Z)}{\text{Var}(WC|Z)}. \tag{S5.3}$$

In conclusion, the ordinary least squared estimator for β^* is biased for β by a factor α. This factor is sometimes referred to as the attenuation factor in case of random measurement error, because in that case $\text{Var}(WC, VAT|Z) < \text{Var}(WC|Z)$ and hence, $\alpha < 1$.

S5.1.1. The different analyses with internal validation samples

When a study contains an internal validation sample for which information is available on both WC and VAT, different analyses can be conducted. Five different estimators are explained below. The variance of these estimators can be obtained from standard output of statistical software when no further details on variance estimation are provided below. The internal validation sample restricted analysis relies on the assumption that the VAT measures in the main study are completely missing at random and the regression calibration methods rely on the assumption that measurement error in WC is non-differential.

Uncorrected analysis. The measurement error is ignored and the relation between VAT and IR is estimated using the error-prone substitute measure WC. Under the assumptions in section S5.1, as shown in equation (S5.3), this estimator is biased by a factor α.

Internal validation sample restricted analysis. The association between VAT and IR is determined using only the data from the internal validation sample (in which a direct measure of VAT is available). This approach will naturally yield unbiased estimates if measures of VAT are missing completely at random in the main study, but power of the study will substantially decrease as only a part of the data available in the main study is used.

Standard regression calibration. The basis of regression calibration is the replacement of WC by a corrected version of WC, based on the regression of VAT on WC and the confounders Z. In this way, the induced measurement error in the uncorrected analysis is corrected by regressing the outcome IR on the confounders Z and $E[VAT|WC, Z]$ instead of WC (i.e., by using the predicted values from regressing VAT on WC and Z, instead of WC). This method is identical to dividing the least squares estimator β^* in equation (S5.2) by the correction factor α defined in equation (S5.3) [2]. The variance of this estimator can be estimated by applying the Delta method described by Rosner et al. [4].

Efficient regression calibration. This analysis pools the estimator of the internal validation sample restricted analysis with the regression calibration estimator, by using weights equal to the inverse of the variance of the two estimates, and was described by Spiegelman et al. [5]. This approach is called efficient regression calibration since it makes use of the fact that in the individuals included in the internal validation sample, VAT is actually known and does not neglect this information. The variance of this estimator can be estimated by taking the inverse of: the sum of the inverse of the variance of the internal validation sample restricted estimator and the inverse of the variance of the regression calibration estimator, as described by Spiegelman et al. [5].

Validation regression calibration. This analysis uses the predicted values from regressing VAT on WC and Z for individuals outside the internal validation sample and VAT otherwise. We call this approach validation regression calibration approach since this is the standard regression calibration approach in internal validation studies [1]. Validation regression calibration treats the predicted values as if they were known and therefore neglects their uncertainty.

S5.2. Simulation study parameters

In the simulation study presented in the main chapter, the measurement error variance τ and the parameter λ in the gamma distribution of the residual errors of VAT were varied according to the R-squared of the measurement error model and skewness of the residuals errors, respectively. The corresponding values for τ and λ in the data generating mechanism found in the main chapter can be found in Table S5.1.

Table S5.1: Values of the parameters R-squared and skewness varied in the simulation study in a full factorial design.The values for τ and λ present the values for that parameter in the data generating mechanism that corresponds to the given R-squared and skewness, respectively.

(a) R-squared and corresponding τ

R-Squared	τ
0.2	1.8
0.4	1.1
0.6	0.7
0.8	0.4
0.9	0.3

(b) Skewness and corresponding λ

Skewness	λ
0.1	65.6
1.0	0.7
1.5	0.3
3.0	0.1

S5.3. Simulation study results

The results of the simulation study that were left out the main chapter for brevity are shown in the following subsections. Full results of the simulation study can also be found on the online repository at `https://github.com/LindaNab/me_neo`. Specifically, `Rds` summary files are available at `https://github.com/LindaNab/me_neo/results/summaries`. These summary files contain more detailed information on e.g. model based standard errors, empirical standard errors and Monte Carlo standard errors. Additionally, output of each single run of the simulation study can be found at `https://github.com/LindaNab/me_neo/data/output` and subsequent folders.

S5.3.1. Internal validation restricted analysis

The main results of the internal validation restricted analysis were shown in the main chapter. Panels A and B in Figure S5.1 show the mean squared error of the association between visceral adipose tissue and insulin resistance using an internal validation sample of 25% of the main study's sample size. Table S5.2 shows the mean squared error of the association under study in the scenarios where R-squared was equal to 0.9 or skewness was equal to 1.0, that were left out the main chapter for brevity. Tables S5.3 and S5.4 show the percentage bias and coverage, respectively, of the association under study in the scenarios

where R-squared was equal to 0.9 or skewness was equal to 1.0.

S5.3.2. Validation regression calibration

The main results of validation regression calibration were shown in the main chapter. Panels C and D in Figure S5.1 show the mean squared error of the association between visceral adipose tissue and insulin resistance using an internal validation sample of 25% of the main study's sample size. Table S5.5 shows the mean squared error of the association under study in the scenarios where R-squared was equal to 0.9 or skewness was equal to 1.0, that were left out the main chapter for brevity. Tables S5.6 and S5.7 show the percentage bias and coverage, respectively, show the percentage bias and coverage of the association under study in the scenarios where R-squared was equal to 0.9 or skewness was equal to 1.0.

S5.3.3. Efficient regression calibration

The results of the application of efficient regression calibration for measurement error correction were as follows. Figure S5.2 shows the mean squared error of the association between visceral adipose tissue and insulin resistance using an internal validation sample of 10%, or 40% of the main study's sample size. Figure S5.3 shows the mean squared error of the association between visceral adipose tissue and insulin resistance using an internal validation sample of 25% of the main study's sample size. Table S5.8 shows the mean squared error of the association under study in the scenarios where R-squared was equal to 0.9 or skewness was equal to 1.0, that were left out Figure S5.2 and S5.3 for comparability with Figure 5.5 and 5.6 in the main chapter. Table S5.9 and S5.10 show the percentage bias in the association between visceral adipose tissue and insulin resistance using an internal validation sample of 10%, 25% or 40% of the main study's sample size for a linear and non-linear measurement error model, respectively. Table S5.11 and S5.12 show the coverage of the association between visceral adipose tissue and insulin resistance using an internal validation sample of 10%, 25% or 40% of the main study's sample size for a linear and non-linear measurement error model, respectively.

S5.3.4. Standard regression calibration

The results of the application of standard regression calibration for measurement error correction were as follows. Table S5.13 and S5.14 show the mean squared error of the association between visceral adipose tissue and insulin resistance using an internal validation sample of 10%, 25% or 40% of the main study's sample size for a linear and non-linear measurement error model, respectively. Table S5.15 and S5.16 show the percentage bias in the association between visceral adipose tissue and insulin resistance using an internal validation sample of 10%, 25% or 40% of the main study's sample size for a linear and non-linear measurement error model, respectively. Table S5.17 and S5.18 show the coverage of the association between visceral adipose tissue and insulin resistance using an internal validation sample of 10%, 25% or 40% of the main study's sample size for a linear and non-linear measurement error model, respectively.

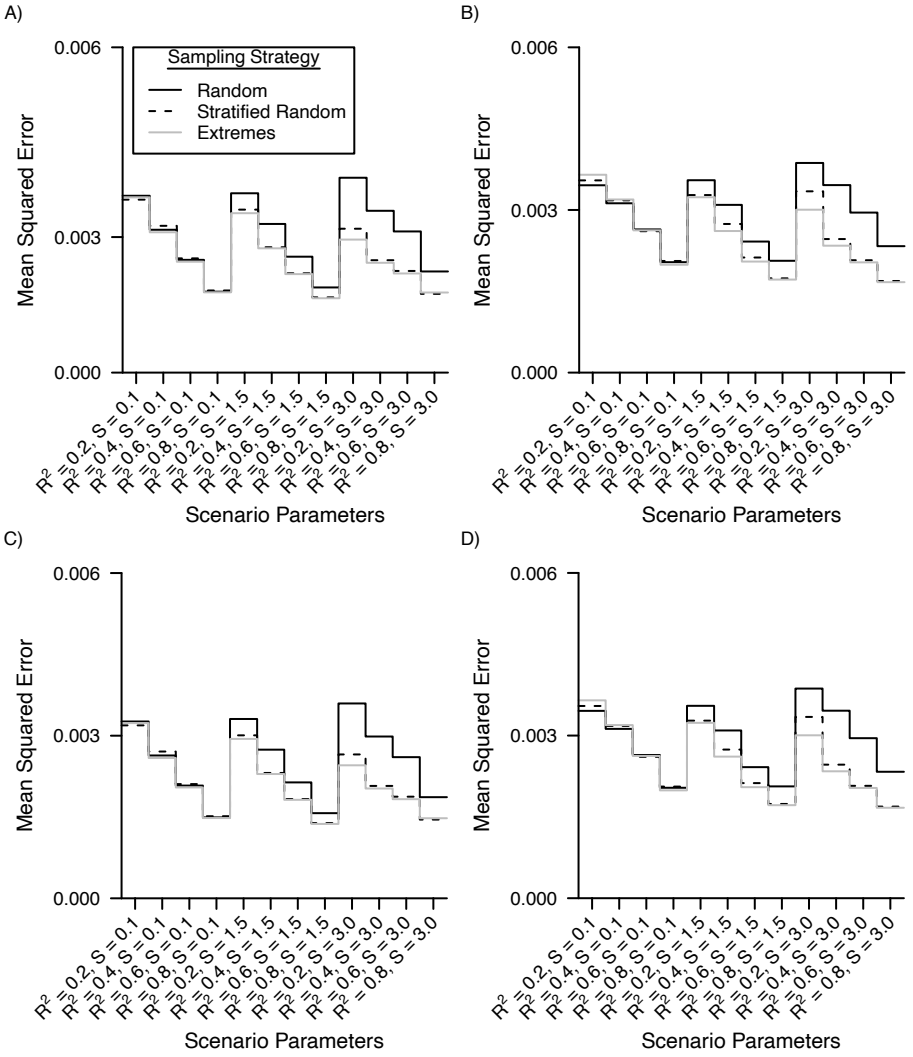

Figure S5.1: Nested loop plot of the mean squared errors in the analysis restricted to the internal validation sample (panels A and B) and the validation regression analysis (panels C and D) for the three different sampling strategies. A and C) Linear measurement error model and an internal validation sample of 25% of the main study; and B and D) Non-linear measurement error model and an internal validation sample of 25% of the main study. Order from outer to inner loops: Skewness of the residual errors of the gold standard measure (S, 3 levels, increasing); R-squared of the measurement error model (R^2, 4 levels, increasing).

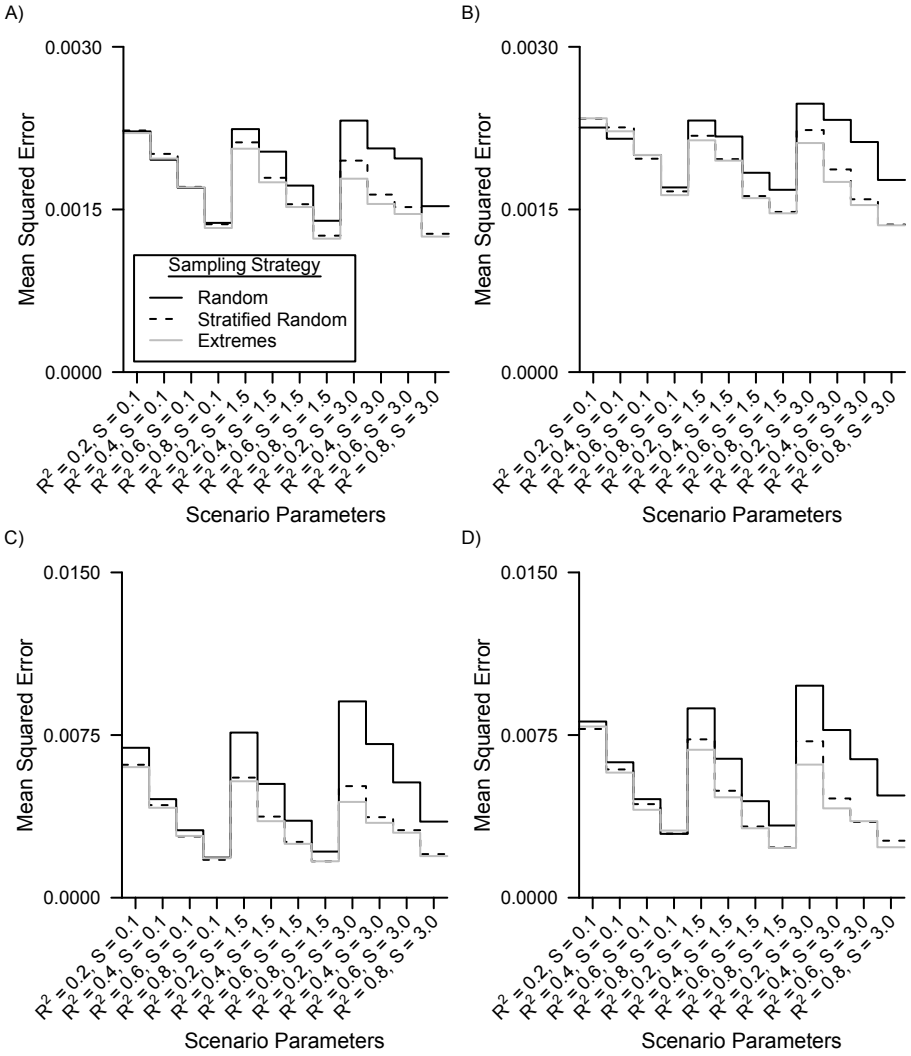

Figure S5.2: Nested loop plot of the mean squared errors in the analysis using efficient regression calibration to correct for the measurement error for the three different sampling strategies. A) Linear measurement error model and an internal validation sample of 40% of the main study; B) Non-linear measurement error model and an internal validation sample of 40% of the main study; C) Linear measurement error model and an internal validation sample of 10% of the main study; and D) Non-linear measurement error model and an internal validation sample of 10% of the main study. Order from outer to inner loops: Skewness of the residual errors of the gold standard measure (S, 3 levels, increasing); R-squared of the measurement error model (R^2, 4 levels, increasing).

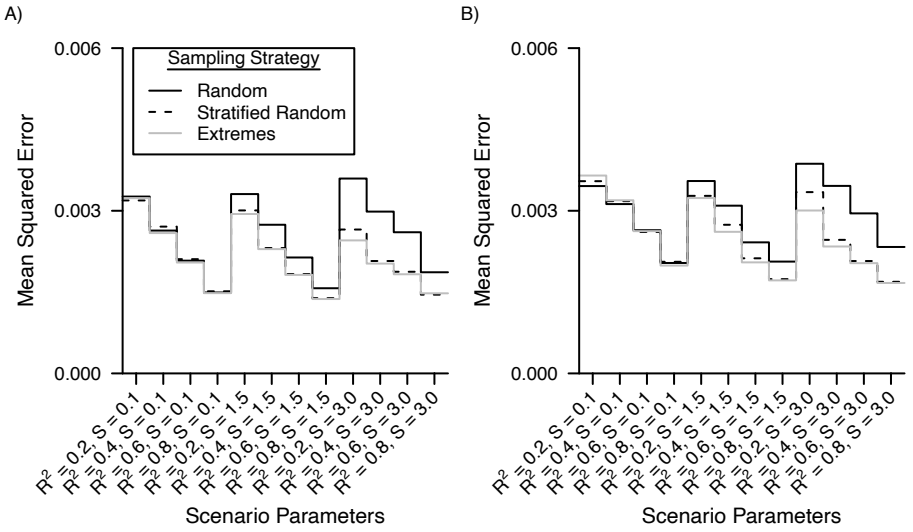

Figure S5.3: Nested loop plot of the mean squared errors in the analysis using efficient regression calibration to correct for the measurement error for the three different sampling strategies. A) Linear measurement error model and an internal validation sample of 25% of the main study; and B) Non-linear measurement error model and an internal validation sample of 25% of the main study. Order from outer to inner loops: Skewness of the residual errors of the gold standard measure (S, 3 levels, increasing); R-squared of the measurement error model (R^2, 4 levels, increasing).

Table S5.2: Mean squared error of the estimated association between visceral adipose tissue and insulin resistance in the analysis restricted to the internal validation sample

Scenario			IVS 40% of Main Study[a]			IVS 25% of Main Study[a]			IVS 10% of Main Study[a]		
Linear	Skew-ness	R^2	Mean Squared Error[b]			Mean Squared Error[b]			Mean Squared Error[b]		
			R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.9	0.0023	0.0020	0.0019	0.0039	0.0031	0.0031	0.0100	0.0080	0.0084
	1.0	0.2	0.0024	0.0022	0.0022	0.0038	0.0034	0.0033	0.0104	0.0086	0.0080
		0.4	0.0024	0.0021	0.0020	0.0038	0.0031	0.0031	0.0103	0.0077	0.0069
		0.6	0.0024	0.0019	0.0018	0.0039	0.0029	0.0027	0.0100	0.0068	0.0067
		0.8	0.0023	0.0019	0.0018	0.0038	0.0029	0.0027	0.0106	0.0066	0.0070
		0.9	0.0024	0.0019	0.0018	0.0039	0.0028	0.0029	0.0103	0.0064	0.0071
	1.5	0.9	0.0024	0.0018	0.0017	0.0039	0.0025	0.0026	0.0109	0.0052	0.0059
	3.0	0.9	0.0024	0.0015	0.0013	0.0040	0.0020	0.0019	0.0120	0.0036	0.0040
No	0.1	0.9	0.0023	0.0020	0.0019	0.0038	0.0032	0.0031	0.0100	0.0089	0.0081
	1.0	0.2	0.0024	0.0022	0.0022	0.0040	0.0036	0.0034	0.0104	0.0094	0.0085
		0.4	0.0024	0.0021	0.0020	0.0038	0.0033	0.0031	0.0107	0.0084	0.0074
		0.6	0.0024	0.0019	0.0019	0.0038	0.0029	0.0028	0.0102	0.0074	0.0068
		0.8	0.0024	0.0019	0.0018	0.0039	0.0029	0.0027	0.0103	0.0068	0.0064
		0.9	0.0023	0.0018	0.0017	0.0039	0.0027	0.0026	0.0103	0.0069	0.0064
	1.5	0.9	0.0024	0.0016	0.0016	0.0039	0.0024	0.0023	0.0108	0.0060	0.0053
	3.0	0.9	0.0026	0.0015	0.0014	0.0042	0.0019	0.0019	0.0123	0.0038	0.0036

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.3: Percentage bias in the estimated association between visceral adipose tissue and insulin resistance in the analysis restricted to the internal validation sample

Scenario			IVS 40% of Main Study[a]			IVS 25% of Main Study[a]			IVS 10% of Main Study[a]		
Linear	Skewness	R^2	Percentage Bias (%)[b]			Percentage Bias (%)[b]			Percentage Bias (%)[b]		
			R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.9	−0.3	0.0	0.5	−0.5	−0.3	0.5	−0.7	−0.3	0.1
	1.0	0.2	0.3	−0.1	0.1	0.1	−0.4	0.0	0.9	−0.9	0.2
		0.4	−0.1	−0.6	−0.3	0.2	−0.7	−0.7	−0.6	−0.4	−0.3
		0.6	0.0	−0.2	−0.4	−0.2	−0.6	−0.2	0.8	−0.3	−0.3
		0.8	0.5	−0.2	0.1	0.2	−0.3	0.1	0.3	−0.4	0.3
		0.9	−0.2	−0.2	−0.2	−0.3	−0.4	−0.3	0.2	−0.6	−0.8
	1.5	0.9	−0.6	−0.5	−0.3	−1.1	−0.8	−0.4	−1.8	0.1	−0.5
	3.0	0.9	0.4	−0.2	−0.2	1.0	−0.1	−0.3	2.4	−0.1	0.0
No	0.1	0.9	−0.4	−0.1	−0.3	−0.4	−0.5	0.0	−0.7	−0.2	−0.7
	1.0	0.2	0.0	0.3	0.0	−0.6	0.0	−0.4	0.3	0.1	−0.5
		0.4	−0.4	−0.1	0.2	−0.5	0.1	−0.1	−0.7	0.0	0.4
		0.6	0.4	0.2	0.1	0.1	0.5	0.0	−0.1	0.5	0.8
		0.8	0.3	0.2	−0.1	−0.3	0.2	0.0	−0.3	−0.3	0.1
		0.9	0.2	0.2	0.5	0.3	0.4	0.2	0.7	1.0	0.4
	1.5	0.9	−0.5	0.2	0.2	−0.3	−0.2	0.3	0.4	−0.2	0.6
	3.0	0.9	0.1	0.2	0.0	0.7	0.2	0.2	1.7	0.2	0.6

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.4: Coverage of the estimated association between visceral adipose tissue and insulin resistance in the analysis restricted to the internal validation sample

Scenario			IVS 40% of Main Study[a] Coverage (%)[b]			IVS 25% of Main Study[a] Coverage (%)[b]			IVS 10% of Main Study[a] Coverage (%)[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.9	95.1	94.9	95.1	94.9	94.8	94.9	94.6	93.8	94.7
	1.0	0.2	94.9	94.6	94.8	95.2	94.2	94.2	94.5	94.7	94.3
		0.4	94.7	94.9	94.5	94.7	94.9	94.1	94.7	94.5	94.3
		0.6	95.1	94.6	95.1	94.6	94.4	95.3	94.8	94.7	94.3
		0.8	94.9	95.3	95.0	94.9	94.7	95.4	94.5	94.4	94.5
		0.9	94.9	94.5	95.1	94.5	94.6	94.7	94.4	94.1	94.5
	1.5	0.9	95.3	94.5	94.9	95.3	94.4	94.6	94.3	94.6	94.5
	3.0	0.9	95.2	95.5	95.3	95.2	94.6	95.1	94.7	93.9	94.3
No	0.1	0.9	95.3	95.1	94.9	94.9	94.9	94.6	95.4	94.1	94.3
	1.0	0.2	95.0	95.3	94.6	94.2	95.1	94.6	94.5	94.4	94.1
		0.4	94.8	94.7	94.7	95.1	94.4	94.7	94.3	94.7	94.7
		0.6	94.7	95.1	95.0	94.9	95.0	94.7	94.8	95.2	94.7
		0.8	94.7	94.6	95.2	95.1	94.2	94.9	94.5	95.0	94.9
		0.9	94.9	95.4	95.4	94.7	94.7	95.3	94.5	94.7	94.7
	1.5	0.9	95.0	95.2	94.8	94.6	95.3	94.7	94.5	94.3	95.1
	3.0	0.9	94.3	94.8	94.7	94.6	94.5	94.9	94.0	94.6	94.7

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.5: Mean squared error of the estimated association between visceral adipose tissue and insulin resistance in the validation regression calibration analysis

Scenario			IVS 40% of Main Study[a] Mean Squared Error[b]			IVS 25% of Main Study[a] Mean Squared Error[b]			IVS 10% of Main Study[a] Mean Squared Error[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.9	0.0011	0.0011	0.0011	0.0012	0.0012	0.0012	0.0014	0.0013	0.0013
	1.0	0.2	0.0022	0.0021	0.0020	0.0033	0.0030	0.0029	0.0073	0.0058	0.0057
		0.4	0.0019	0.0018	0.0017	0.0025	0.0023	0.0023	0.0052	0.0038	0.0037
		0.6	0.0016	0.0014	0.0014	0.0021	0.0017	0.0017	0.0035	0.0025	0.0025
		0.8	0.0012	0.0012	0.0012	0.0014	0.0013	0.0013	0.0019	0.0016	0.0016
		0.9	0.0011	0.0011	0.0011	0.0012	0.0011	0.0011	0.0014	0.0012	0.0013
	1.5	0.9	0.0011	0.0010	0.0010	0.0012	0.0011	0.0011	0.0015	0.0013	0.0013
	3.0	0.9	0.0012	0.0011	0.0011	0.0014	0.0013	0.0013	0.0022	0.0016	0.0015
No	0.1	0.9	0.0014	0.0013	0.0013	0.0016	0.0015	0.0016	0.0021	0.0020	0.0024
	1.0	0.2	0.0022	0.0021	0.0022	0.0034	0.0032	0.0032	0.0081	0.0070	0.0066
		0.4	0.0021	0.0019	0.0019	0.0030	0.0027	0.0027	0.0067	0.0052	0.0047
		0.6	0.0018	0.0016	0.0016	0.0025	0.0021	0.0021	0.0048	0.0036	0.0032
		0.8	0.0015	0.0014	0.0014	0.0019	0.0016	0.0016	0.0031	0.0023	0.0023
		0.9	0.0013	0.0012	0.0012	0.0015	0.0014	0.0014	0.0021	0.0017	0.0018
	1.5	0.9	0.0013	0.0013	0.0012	0.0016	0.0014	0.0013	0.0024	0.0019	0.0017
	3.0	0.9	0.0016	0.0015	0.0014	0.0020	0.0020	0.0016	0.0043	0.0031	0.0024

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.6: Percentage bias in the estimated association between visceral adipose tissue and insulin resistance in the validation regression calibration analysis

Scenario			IVS 40% of Main Study[a]			IVS 25% of Main Study[a]			IVS 10% of Main Study[a]		
Linear	Skewness	R^2	Percentage Bias (%)[b]			Percentage Bias (%)[b]			Percentage Bias (%)[b]		
			R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.9	−0.1	0.0	0.0	0.0	0.0	0.0	0.3	0.3	0.3
	1.0	0.2	0.0	−0.4	−0.2	−0.5	−1.0	−0.6	−1.2	−2.2	−1.6
		0.4	−0.1	−1.5	−1.2	0.3	−2.5	−1.9	2.2	−3.9	−3.7
		0.6	0.0	−2.1	−2.3	0.2	−3.7	−3.1	2.3	−6.2	−6.4
		0.8	0.1	−2.2	−2.3	0.3	−3.5	−3.0	1.4	−5.6	−4.9
		0.9	0.0	−1.5	−1.6	0.1	−2.3	−1.9	0.6	−3.7	−2.9
	1.5	0.9	0.0	−3.1	−3.4	0.2	−4.6	−4.1	1.0	−6.8	−5.7
	3.0	0.9	0.6	−5.5	−6.8	1.3	−8.2	−8.3	4.1	−11.9	−10.6
No	0.1	0.9	−0.1	−1.6	0.1	−0.1	−1.1	2.3	0.7	0.0	6.4
	1.0	0.2	−0.5	−0.1	−0.3	−1.5	−0.7	−0.9	−3.8	−1.7	−2.7
		0.4	−0.6	−1.0	−0.9	−0.9	−1.5	−2.4	−0.2	−2.6	−6.1
		0.6	0.5	−1.2	−2.0	0.8	−1.8	−4.4	3.1	−1.4	−6.2
		0.8	0.3	−3.2	−2.7	0.7	−3.5	−3.0	2.5	−3.1	−0.4
		0.9	0.2	−4.8	−2.1	0.5	−5.4	−1.3	1.8	−6.5	−0.6
	1.5	0.9	0.2	−6.9	−4.3	0.7	−8.6	−4.4	2.5	−11.3	−5.6
	3.0	0.9	0.7	−11.6	−9.6	1.8	−15.6	−11.6	6.7	−22.1	−17.4

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.7: Coverage of the estimated association between visceral adipose tissue and insulin resistance in the validation regression calibration analysis

Scenario			IVS 40% of Main Study[a] Coverage (%)[b]			IVS 25% of Main Study[a] Coverage (%)[b]			IVS 10% of Main Study[a] Coverage (%)[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.9	94.4	95.0	94.9	94.2	94.6	94.4	93.2	93.4	93.4
	1.0	0.2	95.1	94.7	95.1	94.6	94.5	94.9	91.5	93.5	94.4
		0.4	94.9	94.6	94.9	94.6	94.0	95.0	90.3	91.8	92.6
		0.6	94.4	94.9	95.0	93.9	94.1	94.6	90.6	90.4	91.2
		0.8	94.8	95.0	95.1	94.5	94.4	94.6	92.4	92.3	92.7
		0.9	94.5	94.5	94.3	93.9	94.5	94.1	92.7	93.4	93.5
	1.5	0.9	95.3	94.6	94.7	94.6	94.2	94.4	92.9	91.9	92.4
	3.0	0.9	94.5	93.2	92.4	93.4	90.8	91.2	90.2	86.4	88.1
No	0.1	0.9	94.8	94.7	95.3	94.1	94.4	94.8	92.3	93.0	92.3
	1.0	0.2	95.3	95.7	94.8	94.6	95.4	95.5	92.3	94.5	94.8
		0.4	94.9	94.9	95.2	94.6	94.9	94.9	91.0	93.1	93.5
		0.6	94.7	95.2	94.8	94.2	94.8	94.6	91.2	92.5	93.4
		0.8	94.7	94.5	94.2	94.1	94.3	94.1	90.2	91.8	93.2
		0.9	95.0	94.1	94.5	94.4	93.4	94.9	91.3	90.9	92.9
	1.5	0.9	94.1	92.9	93.8	93.7	91.3	93.7	90.6	86.5	91.1
	3.0	0.9	93.1	87.4	89.7	91.6	81.4	87.4	84.6	64.2	76.9

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.8: Mean squared error of the estimated association between visceral adipose tissue and insulin resistance in the efficient regression calibration analysis

Scenario			IVS 40% of Main Study[a] Mean Squared Error[b]			IVS 25% of Main Study[a] Mean Squared Error[b]			IVS 10% of Main Study[a] Mean Squared Error[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.9	0.0012	0.0012	0.0012	0.0013	0.0013	0.0013	0.0014	0.0014	0.0014
	1.0	0.2	0.0023	0.0022	0.0022	0.0033	0.0032	0.0032	0.0072	0.0059	0.0060
		0.4	0.0020	0.0019	0.0019	0.0026	0.0025	0.0025	0.0049	0.0038	0.0038
		0.6	0.0018	0.0016	0.0016	0.0022	0.0019	0.0019	0.0033	0.0026	0.0026
		0.8	0.0014	0.0013	0.0012	0.0015	0.0014	0.0014	0.0020	0.0017	0.0016
		0.9	0.0012	0.0012	0.0012	0.0013	0.0012	0.0012	0.0015	0.0013	0.0014
	1.5	0.9	0.0012	0.0011	0.0011	0.0013	0.0012	0.0012	0.0015	0.0013	0.0013
	3.0	0.9	0.0013	0.0011	0.0011	0.0014	0.0012	0.0012	0.0021	0.0014	0.0013
No	0.1	0.9	0.0015	0.0014	0.0014	0.0017	0.0017	0.0017	0.0022	0.0021	0.0024
	1.0	0.2	0.0023	0.0022	0.0022	0.0035	0.0033	0.0033	0.0082	0.0073	0.0069
		0.4	0.0022	0.0021	0.0020	0.0031	0.0029	0.0029	0.0064	0.0053	0.0050
		0.6	0.0019	0.0018	0.0018	0.0025	0.0023	0.0023	0.0043	0.0037	0.0034
		0.8	0.0017	0.0015	0.0015	0.0020	0.0018	0.0018	0.0030	0.0024	0.0024
		0.9	0.0014	0.0013	0.0013	0.0017	0.0015	0.0015	0.0022	0.0018	0.0019
	1.5	0.9	0.0015	0.0013	0.0013	0.0017	0.0014	0.0014	0.0023	0.0019	0.0017
	3.0	0.9	0.0016	0.0014	0.0013	0.0000	0.0016	0.0015	0.0037	0.0025	0.0020

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.9: Percentage bias in the estimated association between visceral adipose tissue and insulin resistance in the efficient regression calibration analysis for a linear measurement error model

Scenario			IVS 40% of Main Study[a] Percentage Bias (%)[b]			IVS 25% of Main Study[a] Percentage Bias (%)[b]			IVS 10% of Main Study[a] Percentage Bias (%)[b]		
Linear	Skew-ness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.2	−1.2	−0.3	−0.4	−2.2	−0.6	−0.7	−6.3	−2.5	−2.3
		0.4	−0.6	0.0	−0.1	−0.9	−0.2	−0.3	−1.9	0.0	0.0
		0.6	0.6	0.8	0.5	0.6	0.9	0.7	1.4	1.3	1.2
		0.8	0.0	0.1	0.2	0.0	0.1	0.4	0.8	0.3	0.7
		0.9	−0.1	0.0	0.1	−0.1	−0.1	0.2	0.2	0.2	0.3
	1.0	0.2	−0.6	−0.8	−0.5	−1.9	−1.7	−1.1	−5.8	−4.5	−3.1
		0.4	−0.5	−1.5	−1.2	−0.6	−2.4	−2.2	−1.7	−4.5	−4.2
		0.6	−0.2	−1.6	−1.9	−0.3	−3.0	−2.7	0.7	−5.3	−5.6
		0.8	0.1	−1.6	−1.6	0.1	−2.6	−2.2	1.0	−4.7	−3.9
		0.9	−0.1	−1.1	−1.1	0.0	−1.7	−1.4	0.6	−3.1	−2.3
	1.5	0.2	−1.1	−1.1	−0.7	−2.2	−2.0	−1.7	−5.8	−5.6	−4.7
		0.4	−0.2	−1.5	−1.6	−0.4	−3.1	−2.7	−1.3	−7.4	−7.4
		0.6	0.3	−2.4	−2.4	0.7	−4.5	−3.9	2.1	−8.9	−8.9
		0.8	0.3	−2.4	−2.9	0.4	−4.3	−3.9	1.9	−7.9	−7.0
		0.9	−0.2	−2.2	−2.3	−0.2	−3.4	−3.1	0.6	−5.5	−4.4
	3.0	0.2	−1.0	−1.4	−1.5	−2.1	−3.2	−2.6	−5.4	−9.5	−8.0
		0.4	0.2	−3.3	−3.7	0.4	−7.1	−5.9	0.8	−16.2	−16.0
		0.6	0.7	−5.3	−6.5	1.5	−9.9	−9.7	5.3	−19.2	−19.2
		0.8	0.9	−5.1	−6.4	1.8	−8.4	−8.9	6.2	−14.3	−13.3
		0.9	0.6	−3.6	−4.4	1.3	−5.5	−5.7	4.0	−8.9	−7.5

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.10: Percentage bias in the estimated association between visceral adipose tissue and insulin resistance in the efficient regression calibration analysis for a non-linear measurement error model

Scenario			IVS 40% of Main Study[a] Percentage Bias (%)[b]			IVS 25% of Main Study[a] Percentage Bias (%)[b]			IVS 10% of Main Study[a] Percentage Bias (%)[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
No	0.1	0.2	−0.9	−0.4	−0.3	−2.0	−1.2	−0.7	−7.9	−4.8	−2.8
		0.4	−0.3	−0.3	−0.7	−0.9	−0.6	−1.4	−4.2	−1.9	−3.4
		0.6	−0.3	−0.7	−1.2	−0.4	−0.4	−2.5	−2.2	0.3	−3.0
		0.8	0.2	−0.1	0.3	0.4	0.6	1.0	0.7	2.7	4.9
		0.9	−0.2	−1.5	−0.1	−0.2	−1.3	1.3	0.3	−0.3	4.5
	1.0	0.2	−1.1	−0.4	−0.6	−2.8	−1.4	−1.6	−6.9	−4.2	−4.5
		0.4	−1.4	−1.2	−1.2	−2.3	−1.9	−2.9	−5.4	−4.1	−6.8
		0.6	0.2	−0.9	−1.7	−0.1	−1.4	−3.8	−0.7	−1.9	−5.9
		0.8	0.3	−2.2	−2.1	0.3	−2.6	−2.5	1.0	−3.0	−0.9
		0.9	0.3	−3.4	−1.4	0.5	−4.1	−1.2	1.4	−5.4	−0.5
	1.5	0.2	−1.3	−0.6	−0.9	−2.3	−1.6	−1.6	−7.3	−4.8	−4.5
		0.4	−1.4	−1.5	−1.8	−2.1	−2.5	−3.5	−5.5	−5.0	−9.0
		0.6	−0.1	−1.9	−2.6	−0.2	−3.2	−5.5	−0.4	−5.5	−10.0
		0.8	−0.1	−4.0	−3.7	0.2	−5.3	−5.1	1.5	−7.3	−5.6
		0.9	0.0	−4.9	−3.1	0.4	−6.5	−3.4	1.9	−9.5	−4.3
	3.0	0.2	−1.3	−0.8	−0.7	−2.7	−1.6	−1.5	−8.1	−4.6	−5.6
		0.4	−0.9	−2.5	−2.7	−1.7	−4.4	−5.0	−4.2	−9.5	−12.8
		0.6	−0.2	−4.4	−4.8	−0.1	−7.5	−9.3	1.1	−13.3	−18.0
		0.8	0.8	−6.6	−6.4	1.4	−9.8	−9.5	5.3	−15.8	−13.7
		0.9	0.6	−8.0	−6.8	1.6	−11.3	−8.6	5.6	−17.6	−12.8

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.11: Coverage of the estimated association between visceral adipose tissue and insulin resistance in the efficient regression calibration analysis for a linear measurement error model

Scenario			IVS 40% of Main Study[a] Coverage (%)[b]			IVS 25% of Main Study[a] Coverage (%)[b]			IVS 10% of Main Study[a] Coverage (%)[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.2	93.4	92.7	92.6	93.9	93.0	92.4	92.9	93.2	92.2
		0.4	92.1	91.2	90.3	93.4	90.8	91.1	93.6	92.0	91.5
		0.6	90.9	89.7	88.7	92.7	90.4	89.6	93.8	92.3	91.5
		0.8	89.6	88.8	88.5	91.9	89.5	89.6	94.2	92.6	91.5
		0.9	88.5	87.1	87.0	90.6	89.0	88.1	93.2	91.4	91.1
	1.0	0.2	93.5	92.3	91.9	93.5	92.1	91.2	92.3	91.9	91.1
		0.4	92.0	90.2	90.0	92.9	90.7	89.7	92.3	91.5	90.2
		0.6	91.0	89.0	88.5	91.6	89.7	89.4	93.4	89.6	88.7
		0.8	90.5	88.8	88.7	91.7	89.6	90.2	93.5	90.4	90.2
		0.9	88.7	87.5	86.7	90.4	88.8	88.5	92.7	90.2	90.1
	1.5	0.2	93.7	92.7	92.5	93.4	92.5	92.0	91.9	92.5	91.4
		0.4	91.9	91.0	91.0	92.4	90.9	90.4	92.3	88.9	88.8
		0.6	90.8	89.4	88.2	92.1	88.9	88.4	92.7	86.5	85.7
		0.8	89.3	87.7	87.1	91.7	88.2	87.6	92.7	87.3	86.8
		0.9	89.4	86.7	86.6	90.6	87.9	87.6	92.5	89.0	88.5
	3.0	0.2	93.0	92.1	92.5	93.0	92.2	92.4	90.7	88.5	88.8
		0.4	92.0	90.9	89.7	92.3	88.4	88.3	90.9	77.8	79.0
		0.6	89.8	86.3	85.3	90.2	83.0	83.5	90.0	69.2	70.2
		0.8	89.2	85.2	83.4	90.2	83.1	81.9	90.2	76.2	77.4
		0.9	89.1	85.8	84.8	89.8	85.3	84.9	91.4	83.5	84.6

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.12: Coverage of the estimated association between visceral adipose tissue and insulin resistance in the efficient regression calibration analysis for a non-linear measurement error model

Scenario			IVS 40% of Main Study[a] Coverage (%)[b]			IVS 25% of Main Study[a] Coverage (%)[b]			IVS 10% of Main Study[a] Coverage (%)[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
No	0.1	0.2	94.2	93.7	93.4	93.6	93.3	92.7	93.0	92.7	92.9
		0.4	93.6	91.8	91.4	93	92.2	91.6	93.2	92.7	90.8
		0.6	92.2	91.3	90.0	92.5	91.6	89.8	93.2	92.5	90.7
		0.8	91.3	89.4	89.6	92.4	90.5	90.3	93.8	92.6	92.2
		0.9	90.4	89.1	88.7	91.3	89.6	89.1	93.7	92.2	92.3
	1.0	0.2	94.3	93.8	92.8	93.6	93.4	93.1	92.0	93.2	92.6
		0.4	93.3	92.2	91.7	93.7	91.9	91.1	92.1	91.8	89.9
		0.6	92.1	90.7	89.9	92.9	91.1	89.3	93.3	91.5	89.5
		0.8	90.3	88.5	88.0	91.8	89.4	89.2	93.0	91.0	90.8
		0.9	90.0	88.1	87.7	91.3	88.4	89.1	93.2	90.1	91.0
	1.5	0.2	94.1	93.5	93.6	94.3	93.9	93.3	92.4	92.9	92.3
		0.4	93.1	91.9	91.5	93.0	91.7	91.2	92.5	91.1	88.8
		0.6	92.7	91.4	90.6	93.5	90.9	89.1	93.3	91.1	87.3
		0.8	90.6	87.3	87.0	91.4	88.1	87.2	91.9	88.9	88.2
		0.9	89.7	86.5	86.9	90.8	87.0	87.7	92.8	85.5	88.9
	3.0	0.2	93.5	92.7	92.8	93.7	93.0	92.9	91.8	92.1	92.2
		0.4	92.6	91.1	90.9	92.5	91.0	89.8	91.1	88.0	85.1
		0.6	91.4	89.3	88.8	91.7	87.4	85.3	90.5	82.1	76.0
		0.8	90.1	85.8	86.0	90.8	83.5	83.2	90.2	76.1	79.0
		0.9	88.6	82.1	83.5	89.6	78.8	82.1	89.6	69.9	78.9

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.13: Mean squared error of the estimated association between visceral adipose tissue and insulin resistance in the standard regression calibration analysis for a linear measurement error model

Scenario			IVS 40% of Main Study[a] Mean Squared Error[b]			IVS 25% of Main Study[a] Mean Squared Error[b]			IVS 10% of Main Study[a] Mean Squared Error[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.2	0.013	0.011	0.011	0.020	0.012	0.011	2.144	0.042	0.015
		0.4	0.005	0.005	0.005	0.006	0.005	0.005	0.013	0.007	0.006
		0.6	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.003	0.003
		0.8	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		0.9	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	1.0	0.2	0.014	0.011	0.011	0.094	0.012	0.011	5.839	0.107	0.016
		0.4	0.005	0.005	0.004	0.006	0.005	0.005	0.013	0.006	0.006
		0.6	0.003	0.003	0.002	0.003	0.003	0.003	0.005	0.003	0.003
		0.8	0.002	0.001	0.001	0.002	0.002	0.001	0.002	0.002	0.002
		0.9	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	1.5	0.2	0.013	0.011	0.011	0.019	0.012	0.011	2.667	0.067	0.013
		0.4	0.005	0.004	0.004	0.006	0.004	0.004	0.641	0.006	0.005
		0.6	0.003	0.002	0.002	0.003	0.003	0.002	0.006	0.003	0.003
		0.8	0.002	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002
		0.9	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001
	3.0	0.2	0.017	0.011	0.010	0.030	0.011	0.011	20.259	1.602	0.016
		0.4	0.007	0.005	0.004	0.011	0.005	0.005	11.778	0.007	0.006
		0.6	0.004	0.003	0.003	0.005	0.003	0.003	0.012	0.005	0.005
		0.8	0.002	0.002	0.002	0.002	0.002	0.002	0.005	0.003	0.003
		0.9	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.14: Mean squared error of the estimated association between visceral adipose tissue and insulin resistance in the standard regression calibration analysis for a non linear measurement error model

Scenario			IVS 40% of Main Study[a] Mean Squared Error[b]			IVS 25% of Main Study[a] Mean Squared Error[b]			IVS 10% of Main Study[a] Mean Squared Error[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
No	0.1	0.2	0.167	0.025	0.023	0.355	0.030	0.025	6.808	0.247	0.045
		0.4	0.010	0.009	0.008	0.012	0.009	0.009	0.143	0.014	0.010
		0.6	0.005	0.005	0.004	0.005	0.005	0.004	0.011	0.007	0.005
		0.8	0.003	0.002	0.002	0.003	0.003	0.003	0.004	0.003	0.004
		0.9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003
	1.0	0.2	0.036	0.020	0.019	46.762	0.023	0.021	19.508	8.612	1.576
		0.4	0.009	0.008	0.007	0.012	0.008	0.008	1.074	0.012	0.008
		0.6	0.004	0.004	0.004	0.005	0.004	0.004	0.014	0.005	0.004
		0.8	0.002	0.002	0.002	0.003	0.002	0.002	0.004	0.003	0.003
		0.9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	1.5	0.2	0.244	0.021	0.020	1.050	0.023	0.021	106.246	9.679	0.116
		0.4	0.009	0.008	0.007	0.011	0.008	0.007	10.060	0.012	0.008
		0.6	0.004	0.004	0.004	0.005	0.004	0.004	0.012	0.005	0.004
		0.8	0.003	0.002	0.002	0.003	0.002	0.002	0.005	0.003	0.003
		0.9	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.002
	3.0	0.2	0.036	0.022	0.020	0.402	0.029	0.022	66.440	0.601	0.085
		0.4	0.011	0.007	0.007	0.019	0.008	0.007	2.702	0.021	0.008
		0.6	0.006	0.004	0.004	0.007	0.004	0.005	0.082	0.006	0.006
		0.8	0.003	0.003	0.003	0.004	0.003	0.003	0.012	0.004	0.003
		0.9	0.002	0.002	0.002	0.002	0.003	0.002	0.005	0.004	0.003

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.15: Percentage bias in the estimated association between visceral adipose tissue and insulin resistance in the standard regression calibration analysis for a linear measurement error model

Scenario			IVS 40% of Main Study[a] Percentage Bias (%)[b]			IVS 25% of Main Study[a] Percentage Bias (%)[b]			IVS 10% of Main Study[a] Percentage Bias (%)[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.2	3.5	1.2	0.9	7.4	2.1	1.6	26.9	7.3	4.4
		0.4	0.7	0.4	0.2	1.8	0.8	0.7	6.9	3.4	2.5
		0.6	1.5	1.4	1.3	2.0	1.8	1.6	4.3	2.8	2.6
		0.8	0.3	0.2	0.3	0.4	0.3	0.5	1.2	0.6	1.1
		0.9	0.1	0.0	0.0	0.1	0.1	0.1	0.4	0.3	0.3
	1.0	0.2	3.4	−0.2	−0.7	9.1	−0.1	0.1	10.6	2.9	1.9
		0.4	1.0	−2.2	−2.4	2.3	−3.2	−2.5	9.2	−3.5	−3.6
		0.6	0.1	−3.3	−3.6	0.7	−4.8	−4.3	3.5	−6.7	−7.1
		0.8	−0.1	−2.7	−3.0	0.2	−4.0	−3.5	1.5	−5.8	−5.2
		0.9	0.0	−1.7	−1.8	0.1	−2.5	−2.1	0.7	−3.7	−2.9
	1.5	0.2	2.2	−2.2	−2.3	6.7	−2.6	−1.9	30.8	1.2	−2.5
		0.4	1.7	−4.4	−4.5	3.5	−6.8	−5.3	20.5	−9.6	−9.8
		0.6	1.4	−4.9	−5.5	2.3	−7.8	−6.9	6.8	−12.0	−12.3
		0.8	0.5	−4.4	−5.0	1.1	−6.7	−6.1	2.9	−10.2	−9.3
		0.9	−0.1	−3.4	−3.7	0.1	−4.8	−4.3	1.0	−7.0	−5.8
	3.0	0.2	6.4	−5.6	−6.0	14.1	−9.2	−5.3	108.3	−22.6	−12.2
		0.4	4.1	−9.8	−11.4	8.6	−16.4	−13.2	51.9	−25.6	−25.6
		0.6	2.6	−11.6	−14.2	5.2	−18.1	−17.3	15.4	−27.5	−27.7
		0.8	1.6	−9.5	−11.7	2.8	−13.9	−14.0	8.6	−20.0	−19.0
		0.9	0.7	−6.1	−7.5	1.4	−8.7	−8.8	4.2	−12.1	−10.9

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.16: Percentage bias in the estimated association between visceral adipose tissue and insulin resistance in the standard regression calibration analysis for a non-linear measurement error model

Scenario			IVS 40% of Main Study[a]			IVS 25% of Main Study[a]			IVS 10% of Main Study[a]		
Linear	Skewness	R^2	Percentage Bias (%)[b]			Percentage Bias (%)[b]			Percentage Bias (%)[b]		
			R	SR	E	R	SR	E	R	SR	E
No	0.1	0.2	5.6	1.9	0.0	16.4	4.1	1.3	45.8	16.7	7.1
		0.4	3.0	1.2	−0.4	5.3	2.1	−0.9	16.4	6.7	−0.4
		0.6	0.3	−0.2	−2.1	1.3	1.0	−3.3	6.5	5.2	−1.1
		0.8	0.9	0.3	0.5	1.5	1.9	2.2	3.2	5.2	8.9
		0.9	0.1	−2.0	0.4	0.2	−1.2	2.9	1.1	0.3	7.3
	1.0	0.2	7.9	0.7	−0.3	−29.6	2.3	0.6	2.3	−5.5	−4.9
		0.4	1.0	−2.8	−4.2	3.4	−2.9	−5.8	21.2	0.7	−7.9
		0.6	1.9	−2.4	−3.9	2.9	−2.8	−6.8	9.0	−0.6	−7.3
		0.8	0.8	−4.7	−4.0	1.5	−4.5	−4.0	3.7	−3.2	−0.2
		0.9	0.6	−6.0	−2.7	0.8	−6.3	−1.5	2.1	−6.9	−0.5
	1.5	0.2	9.8	0.3	−0.5	14.3	1.9	0.4	−79.8	39.5	4.5
		0.4	1.0	−4.4	−5.4	3.7	−5.1	−7.8	−16.0	−3.0	−11.3
		0.6	1.2	−5.7	−7.2	2.4	−7.1	−10.9	9.5	−6.9	−13.5
		0.8	0.2	−8.3	−7.5	1.2	−9.1	−8.5	4.4	−9.5	−7.3
		0.9	0.5	−8.8	−5.6	0.9	−10.0	−5.3	2.9	−11.9	−6.1
	3.0	0.2	8.9	−1.5	−2.9	26.9	−0.5	−2.9	115.6	1.6	−1.5
		0.4	3.3	−9.1	−9.2	8.8	−12.6	−13.4	29.2	−13.6	−22.0
		0.6	2.3	−13.1	−13.9	5.1	−17.6	−20.0	20.1	−21.1	−28.0
		0.8	1.9	−14.8	−14.1	3.7	−18.6	−17.5	12.3	−22.8	−21.2
		0.9	1.0	−14.9	−12.7	2.3	−18.5	−14.1	7.6	−23.7	−19.0

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.17: Coverage of the estimated association between visceral adipose tissue and insulin resistance in the standard regression calibration analysis for a linear measurement error model

Scenario			IVS 40% of Main Study[a] Coverage (%)[b]			IVS 25% of Main Study[a] Coverage (%)[b]			IVS 10% of Main Study[a] Coverage (%)[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
Yes	0.1	0.2	96.9	96.9	96.8	96.8	97.0	97.0	94.9	96.7	97.0
		0.4	95.9	95.9	95.8	96.4	95.9	95.9	95.6	96.0	95.7
		0.6	95.9	96.0	96.1	95.9	96.0	96.2	96.3	96.3	96.7
		0.8	95.8	95.5	95.7	95.6	95.6	95.9	95.9	95.6	95.8
		0.9	94.7	94.7	94.9	94.9	95.0	95.2	95.3	94.9	95.2
	1.0	0.2	96.5	96.4	96.3	96.5	96.3	96.5	94.5	95.6	95.9
		0.4	95.9	95.8	96.1	96.2	95.3	95.7	94.7	94.2	94.1
		0.6	95.7	95.3	95.4	95.3	94.7	95.1	95.3	92.2	93.0
		0.8	95.7	95.5	95.6	95.8	95.0	95.4	95.7	93.8	94.3
		0.9	95.2	95.0	95.0	94.9	94.8	94.9	94.8	94.6	94.9
	1.5	0.2	96.9	96.7	96.9	96.3	96.4	97.0	94.0	94.6	95.5
		0.4	96.2	95.8	96.2	96.2	95.1	95.7	94.6	90.5	92.1
		0.6	96.1	95.2	95.1	95.6	93.5	94.3	94.7	88.5	89.7
		0.8	95.9	95.0	94.9	95.8	93.7	93.9	95.6	90.2	90.8
		0.9	95.5	95.1	95.2	95.2	94.8	95.0	95.4	93.0	93.5
	3.0	0.2	96.4	95.9	96.6	95.8	94.7	96.2	92.5	88.3	91.7
		0.4	95.2	93.5	93.8	95.0	89.6	92.3	92.4	74.3	77.7
		0.6	94.3	91.3	90.0	93.4	84.2	86.3	92.0	64.6	66.8
		0.8	94.4	91.8	90.0	93.2	86.7	87.1	91.3	74.8	78.1
		0.9	94.6	93.2	92.5	94.2	91.0	90.9	93.4	86.9	88.7

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

Table S5.18: Coverage of the estimated association between visceral adipose tissue and insulin resistance in the standard regression calibration analysis for a non-linear measurement error model

Scenario			IVS 40% of Main Study[a] Coverage (%)[b]			IVS 25% of Main Study[a] Coverage (%)[b]			IVS 10% of Main Study[a] Coverage (%)[b]		
Linear	Skewness	R^2	R	SR	E	R	SR	E	R	SR	E
No	0.1	0.2	97.3	97.2	97.2	96.9	97.3	97.3	94.7	96.6	97.0
		0.4	97.2	96.9	96.6	97.4	97.2	96.9	95.6	97.1	96.5
		0.6	96.7	96.3	96.1	96.6	96.3	96.0	95.4	96.8	96.3
		0.8	95.8	95.9	96.1	95.9	96.1	96.1	96.4	96.4	96.6
		0.9	95.6	95.1	95.5	95.9	95.1	95.5	96.1	96.0	96.3
	1.0	0.2	97.4	97.3	97.2	97.0	97.3	97.1	94.8	96.7	96.7
		0.4	96.4	96.3	96.1	96.5	96.2	95.7	94.3	95.4	94.3
		0.6	96.5	96.2	96.0	96.5	95.9	95.0	95.7	95.4	94.1
		0.8	95.7	94.8	95.1	95.5	95.2	94.9	95.3	94.4	95.5
		0.9	95.6	94.1	95.1	95.7	94.2	95.5	95.9	92.9	95.5
	1.5	0.2	97.4	97.3	97.5	96.4	97.0	97.3	94.4	95.9	96.4
		0.4	96.5	96.2	96.2	96.0	95.9	95.5	94.1	94.0	92.9
		0.6	96.4	95.7	95.4	96.5	95.3	94.2	95.0	93.9	91.5
		0.8	94.9	93.6	93.7	94.9	92.7	93.4	94.2	91.1	93.2
		0.9	95.2	92.5	94.6	95.1	91.5	94.5	94.8	88.4	93.1
	3.0	0.2	97.2	97.2	97.1	96.6	96.6	97.0	93.9	94.9	95.4
		0.4	96.5	95.1	95.7	95.9	93.6	94.1	92.9	88.9	86.1
		0.6	95.2	92.9	92.5	94.5	88.9	87.4	92.8	80.5	75.1
		0.8	94.6	89.2	90.1	93.6	84.1	85.9	91.6	73.5	79.5
		0.9	94.1	86.1	89.3	93.3	80.0	87.1	90.4	66.1	78.7

[a] Internal validation sample
[b] For varying sampling strategies of the internal validation sample, R: random, SR: stratified random, E: extremes

References

[1] 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.

[2] R. H. Keogh, P. A. Shaw, P. Gustafson, R. J. Carroll, V. Deffner, K. W. Dodd, H. Küchenhoff, J. A. Tooze, M. P. Wallace, V. Kipnis, L. S. Freedman, STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1—Basic theory and simple methods of adjustment, Statistics in Medicine 39 (16) (2020) 2197–2231. doi:10.1002/sim.8532.

[3] 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.

[4] B. Rosner, D. Spiegelman, W. C. Willett, Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error, American Journal of Epidemiology 132 (4) (1990) 734–745. doi:10.1093/oxfordjournals.aje.a115715.

[5] D. Spiegelman, R. J. Carroll, V. Kipnis, Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument, Statistics in Medicine 20 (1) (2001) 139–160. doi:10.1002/1097-0258(20010115)20:1<139::AID-SIM644>3.0.CO;2-K.

S8

Supplementary material Chapter 8

These are the supplementary materials accompanying Chapter 8. The supplementary materials are structured as follows. In section S8.1, the bias formulas of a conditional model and marginal structural model estimated using inverse probability weighting are derived. Section S8.2 illustrates the application of the bias formulas in a quantitative bias analysis.

S8.1. Quantification of bias due to classification error in a confounding variable

S8.1.1. Conditional model

Under the assumptions and notation described in section 8.2 of the main chapter and by the law of total expectation, the expected value of the outcome Y given the covariates A and L^* is,

$$\begin{aligned}
\mathsf{E}[Y|A,L^*] = \mathsf{E}_{L|A,L^*}[\mathsf{E}[Y|A,L^*,L]] &= \mathsf{E}_{L|A,L^*}[\alpha + \beta A + \gamma L] \\
&= \alpha + \beta A + \gamma \mathsf{E}[L|A,L^*] \\
&= \alpha + \beta A + \gamma \phi_{aL^*} \\
&= \{\alpha + \gamma\phi_{00}\} + \{\beta + \gamma(\phi_{10} - \phi_{00})\}A \\
&+ \{\gamma(\phi_{01} - \phi_{00})\}L^* \\
&+ \gamma(\phi_{11} - \phi_{10} - \phi_{01} + \phi_{00})AL^*,
\end{aligned}$$

which relies on the assumption that L^* is non-differentially misclassified with respect to the outcome (i.e., $L^* \perp\!\!\!\perp Y|L$) and includes an interaction between A and L^*. Further, ϕ_{al^*} is the probability that confounding variable L is one, given that treatment A is a and that misclassified confounding variable L^* is l^*, or,

$$\begin{aligned}
\phi_{al^*} &= P(L = 1|A = a, L^* = l^*) \\
&= \frac{P(A|L = 1, L^* = l^*)P(L = 1|L^* = l^*)}{P(A = a|L^* = l^*)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{P(A=a|L=1)P(L=1|L^*=l^*)}{P(A=a|L^*=l^*)} \\
&= \frac{P(A=a|L=1)\frac{P(L^*=l^*|L=1)P(L=1)}{P(L^*=l^*)}}{P(A=a|L^*=l^*)} \\
&= \frac{P(A=a|L=1)P(L^*=l^*|L=1)P(L=1)}{P(A=a|L^*=l^*)P(L^*=l^*)} \\
&= \frac{\lambda(1-\pi_1)^{(1-a)}\pi_1^a(1-p_1)^{(1-l^*)}p_1^{l^*}}{(1-\pi_{l^*}^*)^{(1-a)}\pi_{l^*}^{*\,a}(1-\ell)^{(1-l^*)}\ell^{l^*}}.
\end{aligned}
$$

Here $\ell = P(L^* = l^*) = p_0(1-\lambda) + p_1\lambda$ and $\pi_{l^*}^*$ is the probability of receiving treatment A given that the misclassified confounding variable $L^* = l^*$. Note that the above is only defined if $0 < \ell < 1$ and $0 < \pi_{l^*}^* < 1$. To satisfy that $0 < \ell < 1$, we use our assumption that $0 < \lambda < 1$, and additionally, we assume that if $p_0 = 1$ then $p_1 \neq 1$, and if $p_0 = 0$ then $p_1 \neq 0$ (and vice versa). Under the assumption that $0 < \ell < 1$, it follows that,

$$
\begin{aligned}
\pi_{l^*}^* &= P(A=1|L^*=l^*) \\
&= \Sigma_l P(A=1|L^*=l^*, L=l)P(L=l|L^*=l^*) \\
&= \Sigma_l P(A=1|L=l)P(L=l|L^*=l^*) \\
&= \Sigma_l P(A=1|L=l)\frac{P(L^*=l^*|L=l)P(L=l)}{P(L^*=l^*)} \\
&= \Sigma_l \pi_l \frac{(1-p_l)^{(1-l^*)}p_l^{l^*}(1-\lambda)^{1-l}\lambda^l}{(1-\ell)^{1-l^*}\ell^{l^*}},
\end{aligned}
$$

we find that $0 < \pi_{l^*}^* < 1$, if, again, $0 < \lambda < 1$, and if $p_0 = 1$ then $p_1 \neq 1$, and if $p_0 = 0$ then $p_1 \neq 0$ (and vice versa) and $0 < \pi_l < 1$ (positivity assumption).

S8

The bias in the regression based estimator of the effect of A is $\gamma(\phi_{10} - \phi_{00})$ if the interaction between A and L^* is included in the model. However, in this model, the coefficient for A now represents the treatment effect given that L^* is null. Typically, only main effects of A and L^* are included in a regression model of Y conditional on A and L^*:

$$
\begin{aligned}
\mathsf{E}_{AL^*|A,L^*}\{\mathsf{E}[Y|A,L^*]\} &= \{\alpha + \gamma\phi_{00}\} + \{\beta + \gamma(\phi_{10} - \phi_{00})\}A + \{\gamma(\phi_{01} - \phi_{00})\}L^* \\
&+ \gamma(\phi_{11} - \phi_{10} - \phi_{01} + \phi_{00})\mathsf{E}[AL^*|A,L] \\
&= \{\alpha + \gamma\phi_{00} + \delta u_0\} + \{\beta + \gamma(\phi_{10} - \phi_{00}) + \delta u_A\}A \\
&+ \{\gamma(\phi_{01} - \phi_{00}) + \delta u_{L^*}\}L^*,
\end{aligned}
$$

where u_0, u_A, and u_{L^*} are the coefficients of the linear model $\mathsf{E}[AL^*|A,L^*] = u_0 + u_A A + u_{L^*}L^*$ and $\delta = \gamma(\phi_{11} - \phi_{10} - \phi_{01} + \phi_{00})$. Here,

$$
\begin{aligned}
u_A &= \frac{\mathrm{Var}(L^*)\mathrm{Cov}(A,AL^*) - \mathrm{Cov}(A,L^*)\mathrm{Cov}(L^*,AL^*)}{\mathrm{Var}(L^*)\mathrm{Var}(A) - \mathrm{Cov}(A,L^*)^2}, \\
u_{L^*} &= \frac{\mathrm{Var}(A)\mathrm{Cov}(L^*,AL^*) - \mathrm{Cov}(A,L^*)\mathrm{Cov}(A,AL^*)}{\mathrm{Var}(L^*)\mathrm{Var}(A) - \mathrm{Cov}(A,L^*)^2}, \\
u_0 &= \overline{AL^*} - u_A\overline{A} - u_{L^*}\overline{L^*},
\end{aligned}
$$

where $\overline{AL^*}$, $\overline{A}$, and $\overline{L^*}$ denote the mean of A times L^*, A, and L^*, respectively.

If we want to express u_A and u_{L^*} in terms of λ, π_0, π_1, p_0, and p_1, we can write a linear model for A conditional on L^* denoting that $P(A = 1|L^* = l^*) = \pi^*_{l^*}$ and using standard regression theory to get an expression for $\text{Cov}(A, L^*)$:

$$\mathsf{E}[A|L^*] = \pi^*_0 + (\pi^*_1 - \pi^*_0)L^*, \quad \pi^*_1 - \pi^*_0 = \frac{\text{Cov}(A, L^*)}{\text{Var}(L^*)},$$
$$\text{thus} \quad \text{Cov}(A, L^*) = (\pi^*_1 - \pi^*_0)\text{Var}(L^*),$$

where $\text{Var}(L^*) = \ell(1 - \ell)$. Since $\mathsf{E}[AL^*|L^* = 0] = 0$ and $\mathsf{E}[AL^*|L^* = 1] = \mathsf{E}[A|L^* = 1] = \pi^*_1$, it follows,

$$\mathsf{E}[AL^*|L^*] = \pi^*_1 L^*, \quad \pi^*_1 = \frac{\text{Cov}(AL^*, L^*)}{\text{Var}(L^*)}, \quad \text{thus} \quad \text{Cov}(AL^*, L^*) = \pi^*_1\text{Var}(L^*).$$

Equivalently, since $\mathsf{E}[AL^*|A = 0] = 0$ and $\mathsf{E}[AL^*|A = 1] = \mathsf{E}[L^*|A = 1]$, it follows that,

$$\mathsf{E}[AL^*|A] = \mathsf{E}[L^*|A = 1]A = \frac{P(A = 1|L^* = 1)P(L^* = 1)}{P(A = 1)}A,$$
$$\mathsf{E}[L^*|A = 1] = \frac{\pi^*_1\ell}{\omega}, \quad \frac{\pi^*_1\ell}{a} = \frac{\text{Cov}(AL^*, A)}{\text{Var}(A)}, \quad \text{thus} \quad \text{Cov}(AL^*, A) = \frac{\pi^*_1\ell}{\omega}\text{Var}(A).$$

Here, $\text{Var}(A) = \omega(1-\omega)$, and $\text{Var}(L^*) = \ell(1-\ell)$. Denoting that $\omega = P(A = 1) = \pi^*_0(1-\ell)+\pi^*_1\ell$. Combining the different expressions gives,

$$\begin{aligned}
u_A &= \frac{\pi^*_1\ell/\omega\text{Var}(A)\text{Var}(L^*) - \pi^*_1(\pi^*_1 - \pi^*_0)\text{Var}(L^*)^2}{\text{Var}(A)\text{Var}(L^*) - (\pi_1 - \pi_0)^2\text{Var}(L^*)^2} \\
&= \frac{\pi^*_1\ell/\omega\text{Var}(A) - \pi^*_1(\pi^*_1 - \pi^*_0)\text{Var}(L^*)}{\text{Var}(A) - (\pi_1 - \pi_0)^2\text{Var}(L^*)} \\
&= \ell \times \frac{\pi^*_1(1 - \omega) - \pi^*_1(\pi^*_1 - \pi^*_0)(1 - \ell)}{\omega(1 - \omega) - (\pi^*_1 - \pi^*_0)^2\ell(1 - \ell)} \\
&= \ell \times \frac{\pi^*_1 - \pi^{*^2}_1}{(\pi^*_1 - \pi^{*^2}_1)\ell + (\pi^*_0 - \pi^{*^2}_0)(1 - \ell)}, \\
u_{L^*} &= \frac{\pi^*_1\text{Var}(A)\text{Var}(L^*) - \pi^*_1\ell/\omega(\pi^*_1 - \pi^*_0)\text{Var}(A)\text{Var}(L^*)}{\text{Var}(L^*)\text{Var}(A) - ((\pi^*_1 - \pi^*_0)\text{Var}(L^*))^2} \\
&= \frac{\pi^*_1\omega - \pi^*_1\ell(\pi^*_1 - \pi^*_0)}{\omega - (\pi^*_1 - \pi^*_0)^2\text{Var}(L^*)/(1 - \omega)} \\
&= \frac{\pi^*_1\pi^*_0(1 - \pi^{*^2}_1)\ell + \pi^*_1\pi^*_0(1 - \pi^{*^2}_0)(1 - \ell)}{(\pi^*_1 - \pi^{*^2}_1)\ell + (\pi^*_0 - \pi^{*^2}_0)(1 - \ell)}, \\
u_0 &= \overline{AL^*} - u_A\overline{A} - u_{L^*}\overline{L^*}.
\end{aligned}$$

The intercept, the coefficient for A and the coefficient for L^* of the conditional regression model for Y given A and L^* which includes only main effects of A and L^* are, respectively:

$$\alpha + \gamma\phi_{00} + \delta u_0,$$

$$\beta + \gamma(\phi_{10} - \phi_{00})\Big(1 - \ell \times \{\frac{\pi_1^* - \pi_1^{*^2}}{(\pi_1^* - \pi_1^{*^2})\ell + (\pi_0^* - \pi_0^{*^2})(1-\ell)}\}\Big)$$
$$+\gamma(\phi_{11} - \phi_{01})\Big(\ell \times \{\frac{\pi_1^* - \pi_1^{*^2}}{(\pi_1^* - \pi_1^{*^2})\ell + (\pi_0^* - \pi_0^{*^2})(1-\ell)}\}\Big),$$
$$\text{and} \quad \gamma(\phi_{01} - \phi_{00}) + \delta u_{L^*}.$$

S8.1.2. Marginal structural model estimated using inverse probability weighting

Under the assumptions described in section 8.2 of the main chapter, an MSM-IPW under model (8.2) is estimated by fitting a linear regression model for A on Y, where each subject i is weighted by 1 over the probability of that subject's observed exposure given the misclassified confounding variable L^*. Hence, an MSM-IPW proceeds by solving the weighted regression model,

$$\sum_{i=1}^{n} \frac{1}{P(A_i|L_i^*)}(Y_i - \alpha_{\text{msm}} - \beta A_i) = 0 \quad \text{and} \quad \sum_{i=1}^{n} \frac{A_i}{P(A_i|L_i^*)}(Y_i - \alpha_{\text{msm}} - \beta A_i) = 0.$$

Solving these equations for α_{msm} and β result in the following estimators:

$$\hat{\alpha}_{\text{msm}} = \overline{Y}_{w^*} - \hat{\beta}_{\text{msm}}\overline{A}_{w^*} \quad \text{and} \quad \hat{\beta} = \frac{\sum_{i=1}^{n} \frac{1}{P(A_i|L_i)}(Y_i - \overline{Y}_{w^*})(A_i - \overline{A}_{w^*})}{\sum_{i=1}^{n} \frac{1}{P(A_i|L_i)}(A_i - \overline{A}_{w^*})^2},$$

where,

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

S8

Let n_{al}^* be the number of subjects with $A = a$ and $L^* = l^*$ and n_{al} be the number of subjects with $A = a$ and $L = l$. In a population of n subjects,

$$\begin{aligned}
n_{00}^* &= nP(A = 0, L^* = 0) = nP(A = 0|L^* = 0)P(L^* = 0) \\
&= n\sum^{l} P(A = 0|L = l, L^* = 0)P(L = l|L^* = 0)P(L^* = 0) \\
&= n\sum^{l} P(A = 0|L = l)P(L = l|L^* = 0)P(L^* = 0) \\
&= n\sum^{l} P(A = 0|L = l)P(L = l)P(L^* = 0|L = l) \\
&= n_{00}(1 - p_0) + n_{01}(1 - p_1),
\end{aligned}$$

which relies on the assumption that L^* is non-differentially misclassified with respect tot the exposure (i.e., $L^* \perp\!\!\!\perp A|L$). Equivalently,

$$n_{01}^* = n_{00}p_0 + n_{01}p_1, \quad n_{10}^* = n_{10}(1 - p_0) + n_{11}(1 - p_1),$$

$$\text{and} \quad n_{11}^* = n_{10}p_0 + n_{11}p_1.$$

Hence,

$$\begin{aligned}
\sum_{i=1}^{n} 1/P(A_i|L_i^*) &= \sum_{i=1}^{n} \frac{1}{\sum_l [P(A_i|L_i^*, L=l)P(L=l|L_i^*)]} \\
&= \sum_{i=1}^{n} \frac{1}{\sum_l [P(A_i|L=l)P(L=l|L_i^*)]} \\
&= \sum^{n_{00}^*} \frac{1}{\sum_l [(1-\pi_l)P(L=l|L^*=0)]} \\
&+ \sum^{n_{01}^*} \frac{1}{\sum_l [(1-\pi_l)P(L=l|L^*=1)]} \\
&+ \sum^{n_{10}^*} \frac{1}{\sum_l [\pi_l P(L=l|L^*=0)]} \\
&+ \sum^{n_{11}^*} \frac{1}{\sum_l [\pi_l P(L=l|L^*=1)]}.
\end{aligned}$$

Here,

$$\begin{aligned}
\sum^{n_{00}^*} \frac{1}{\sum_l [(1-\pi_l)P(L=l|L^*=0)]} &= \\
\frac{n_{00}(1-p_0) + n_{01}(1-p_1)}{(1-\pi_0)P(L=0|L^*=0) + (1-\pi_1)P(L=1|L^*=0)} &= \\
\frac{n_{00}(1-p_0) + n_{01}(1-p_1)}{(1-\pi_0)\frac{P(L^*=0|L=0)(1-\lambda)}{P(L^*=0)} + (1-\pi_1)\frac{P(L^*=0|L=1)\lambda}{P(L^*=0)}} &= \\
\frac{n_{00}(1-p_0) + n_{01}(1-p_1)}{\frac{n_{00}}{nP(L^*=0)}(1-p_0) + \frac{n_{01}}{nP(L^*=0)}(1-p_1)} &= \\
\frac{1}{1/(nP(L^*=0))} &= \\
nP(L^*=0) &= n(1-\ell), \\
\sum^{n_{01}^*} \frac{1}{\sum_l [(1-\pi_l)P(L=l|L^*=1)]} &= \\
nP(L^*=1) &= n\ell, \\
\sum^{n_{10}^*} \frac{1}{\sum_l [\pi_l P(L=l|L^*=0)]} &= \\
nP(L^*=0) &= n(1-\ell), \\
\sum^{n_{11}^*} \frac{1}{\sum_l [\pi_l P(L=l|L^*=1)]} &=
\end{aligned}$$

$$nP(L^* = 1) = n\ell.$$

From these expressions it follows that,

$$\sum_{i=1}^{n} 1/P(A_i|L_i^*) = 2n(1-\ell) + 2n\ell = 2n.$$

Further,

$$\begin{aligned}
\sum_{i=1}^{n} \mathsf{E}[Y_i]/P(A_i|L_i^*) &= \sum^{n_{00}^*} \frac{\mathsf{E}[Y_i]}{\sum_l[(1-\pi_l)P(L=l|L^*=0)]} \\
&+ \sum^{n_{01}^*} \frac{\mathsf{E}[Y_i]}{\sum_l[(1-\pi_l)P(L=l|L^*=1)]} \\
&+ \sum^{n_{10}^*} \frac{\mathsf{E}[Y_i]}{\sum_l[\pi_l P(L=l|L^*=0)]} \\
&+ \sum^{n_{11}^*} \frac{\mathsf{E}[Y_i]}{\sum_l[\pi_l P(L=l|L^*=1)]} \\
&= \sum^{n_{00}^*} \frac{\alpha + \gamma P(L=1|A=0,L^*=0)}{\sum_l[(1-\pi_l)P(L=l|L^*=0)]} \\
&+ \sum^{n_{01}^*} \frac{\alpha + \gamma P(L=1|A=0,L^*=1)}{\sum_l[(1-\pi_l)P(L=l|L^*=1)]} \\
&+ \sum^{n_{10}^*} \frac{\alpha + \beta + \gamma P(L=1|A=1,L^*=0)}{\sum_l[\pi_l P(L=l|L^*=0)]} \\
&+ \sum^{n_{11}^*} \frac{\alpha + \beta + \gamma P(L=1|A=1,L^*=1)}{\sum_l[\pi_l P(L=l|L^*=1)]} \\
&= n\alpha(1-\ell) + n\gamma(1-\ell)\phi_{00} + n\alpha\ell + n\gamma\phi_{01} \\
&+ n(\alpha+\beta)(1-\ell) + n\gamma(1-\ell)\phi_{10} \\
&+ n(\alpha+\beta)\ell + n\gamma\phi_{11} \\
&= 2n\alpha + n\beta + n\gamma(1-\ell)(\phi_{00}+\phi_{10}) + n\gamma\ell(\phi_{01}+\phi_{11}),
\end{aligned}$$

and,

$$\begin{aligned}
\sum_{i=1}^{n} A_i/P(A_i|L_i) &= \sum^{n_{10}^*} \frac{1}{\sum_l[\pi_l P(L=l|L^*=0)]} + \sum^{n_{11}^*} \frac{1}{\sum_l[\pi_l P(L=l|L^*=1)]} \\
&= n(1-p_0)(1-\lambda) + n(1-p_1)\lambda + np_0(1-\lambda) + np_1\lambda = n.
\end{aligned}$$

Combining these expressions leads to,

$$\mathsf{E}[\overline{Y}_{w^*}] = \alpha + \beta/2 + \gamma/2(1-\ell)(\phi_{00}+\phi_{10}) + \gamma/2\ell(\phi_{01}+\phi_{11})$$

$$\text{and} \quad \overline{A}_{w^*} = n/2n = 1/2, \quad \text{and}$$

$$\begin{aligned}
\sum_{i=1}^{n} \frac{(A_i - \overline{A}_{w^*})^2}{P(A_i|L_i^*)} &= \sum^{n_{00}^*} \frac{(-1/2)^2}{\sum_l[(1-\pi_l)P(L=l|L^*=0)]} \\
&+ \sum^{n_{01}^*} \frac{(-1/2)^2}{\sum_l[(1-\pi_l)P(L=l|L^*=1)]} \\
&+ \sum^{n_{10}^*} \frac{(1-1/2)^2}{\sum_l[\pi_l P(L=l|L^*=0)]} \\
&+ \sum^{n_{11}^*} \frac{(1-1/2)^2}{\sum_l[\pi_l P(L=l|L^*=1)]} \\
&= 1/4 \times \sum_{i=1}^{n} 1/P(A_i|L_i^*) = n/2.
\end{aligned}$$

Further,

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\mathsf{E}[(Y_i - \overline{Y}_{w^*})](A_i - \overline{A}_{\tilde{w}})}{P(A_i|L_i^*)} &= \\
\sum^{n_{00}^*} \frac{\beta/4 - \gamma/2\phi_{00} + \gamma/4(1-\ell)(\phi_{00}+\phi_{10}) + \gamma/4\ell(\phi_{01}+\phi_{11})}{\sum_l[(1-\pi_l)P(L=l|L^*=0)]} &+ \\
\sum^{n_{01}^*} \frac{\beta/4 - \gamma/2\phi_{01} + \gamma/4(1-\ell)(\phi_{00}+\phi_{10}) + \gamma/4\ell(\phi_{01}+\phi_{11})}{\sum_l[(1-\pi_l)P(L=l|L^*=1)]} &+ \\
\sum^{n_{10}^*} \frac{\beta/4 + \gamma/2\phi_{10} - \gamma/4(1-\ell)(\phi_{00}+\phi_{10}) - \gamma/4\ell(\phi_{01}+\phi_{11})}{\sum_l[\pi_l P(L=l|L^*=0)]} &+ \\
\sum^{n_{11}^*} \frac{\beta/4 + \gamma/2\phi_{11} - \gamma/4(1-\ell)(\phi_{00}+\phi_{10}) - \gamma/4\ell(\phi_{01}+\phi_{11})}{\sum_l[\pi_l P(L=l|L^*=0)]} &= \\
n(1-\ell)(\beta/4 - \gamma/2\phi_{00} + \gamma/4(1-\ell)(\phi_{00}+\phi_{10}) + \gamma/4\ell(\phi_{01}+\phi_{11})) &+ \\
n\ell(\beta/4 - \gamma/2\phi_{01} + \gamma/4(1-\ell)(\phi_{00}+\phi_{10}) + \gamma/4\ell(\phi_{01}+\phi_{11})) &+ \\
n(1-\ell)(\beta/4 + \gamma/2\phi_{10} - \gamma/4(1-\ell)(\phi_{00}+\phi_{10}) - \gamma/4\ell(\phi_{01}+\phi_{11})) &+ \\
n\ell(\beta/4 + \gamma/2\phi_{11} - \gamma/4(1-\ell)(\phi_{00}+\phi_{10}) - \gamma/4\ell(\phi_{01}+\phi_{11})) &= \\
n/2(\beta(1-\ell) + \beta\ell - \gamma(1-\ell)\phi_{00} - \gamma\ell\phi_{01} + \gamma(1-\ell)\phi_{10} + \gamma\ell\phi_{11}) &= \\
n/2(\beta + \gamma(1-\ell)(\phi_{10}-\phi_{00}) + \gamma\ell(\phi_{11}-\phi_{01}).
\end{aligned}$$

The above mentioned leads to the following expression for the expected estimated value of the effect of A, based on the MSM-IPW,

$$\mathsf{E}[\hat{\beta}] = \beta + \gamma(\phi_{10}-\phi_{00})(1-\ell) + \gamma(\phi_{11}-\phi_{01})\ell \quad \text{and}$$
$$\mathsf{E}[\hat{\alpha}_{\text{msm}}] = \alpha + \gamma/2 \times [2(1-\ell)\phi_{00} + 2\ell\phi_{01})] = \alpha + \gamma\phi_{00}(1-\ell) + \gamma\phi_{01}\ell.$$

S8.2. Illustration: quantitative bias analysis

Using an example study of blood pressure lowering therapy, we illustrate how the bias expressions in section 8.3 of the main chapter can be used to perform a quantitative bias analysis for misclassification of a confounding variable. For our illustration we use data of the National Health And Nutritional Examination Survey (NHANES) [1, 2]. Specifically, we study the average treatment 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. This supplement comprises background material that complements the motivating example in the main chapter. Additionally, equations are derived to inform the quantitative bias analysis.

NHANES. The NHANES survey consists of questionnaires, followed by a standardized health examination in specially equipped mobile examination centers. In the 2011-2014 sample 19,151 participants were physically examined. Of the 19,151 physically examined people, 12,185 participants aged over 16 were asked to fill out a questionnaire, including questions on self-reported weight and height, used to calculate self-reported BMI. For this illustration, we used complete data on 585 users of diuretics and 824 users of beta blockers (excluding non-users and people using both).

Parameters estimated in NHANES. In the NHANES data, it was found that the prevalence of self-reported overweight/obese was 0.77 (ℓ), the probability of receiving treatment given that one self-reports to be underweight/normal weight is 0.32 (π_0^*), the probability of receiving treatment given that one self-reports to be overweight/obese is 0.44 (π_1^*). Finally, the association between L^* and Y, given that $A = 0$ estimated in a conditional regression model including an interaction between A and L* was -6.63.

BMI measured by trained technicians. In the NHANES, anthropometric measures were also taken by trained health technicians. By using these measures to calculate BMI category, we found that the specificity of self-reported BMI category was 0.94 (p_1), and the sensitivity was 0.92 ($p_0 = 0.08$). The average treatment effect (95 % CI) of diuretics use in comparison to beta blocker use on mean blood pressure was -3.59 (-5.84; -1.35) estimated using MSM-IPW (by inverse weighting with the propensity for diuretic or beta blocker use given categorical BMI). Given that a subject is not overweight/obese, the fitted weights were 1.48 and 3.09 for beta blocker and diuretics use, respectively. Given that a subject is overweight/obese, the fitted weights were 1.77 and 2.30, respectively. In comparison, if self-reported categorical BMI was used, the fitted weights slightly differed: 1.46, 3.17, 1.79 and 2.26, respectively. Consequently, estimates of the average treatment effect differed, depending on the BMI measure that was used to calculate the inverse probability weights (-3.59 using categorical BMI versus -3.52 using categorical self-reported BMI (Table 8.3, main chapter)).

Performing a quantitative bias analysis. To inform a quantitative bias analysis, one needs to specify the bias parameters for sensitivity (p_1) and specificity ($1-p_0$) using external validation data, internal validation data, or an educated guess. From the data, one can estimate the prevalence of misclassified confounding variable L^* (i.e., ℓ), the probability of receiving treatment given that L^* is null (i.e., π_0^*) and the probability of receiving treatment given that L^* is one (i.e., π_1^*). We calculate the probability of receiving treatment given that L is null or one (i..e, π_0, and π_1, respectively) using the data and the assumed values of p_0

and p_1. Since,

$$\pi_0^* = \frac{\pi_0(1-p_0)(1-\lambda) + \pi_1(1-p_1)\lambda}{(1-\ell)}, \quad \text{and} \quad \pi_1^* = \frac{\pi_0 p_0(1-\lambda) + \pi_1 p_1 \lambda}{\ell},$$

it follows that if $p_0 = 1$, $\pi_1 = \pi_0^*$ and if $p_1 = 0$, $\pi_0 = \pi_1^*$ (using that $0 < \ell < 1$, as used in S8.1 section S8.1.1). Further, if $p_0 = 1$ and $0 < p_1 < 1$, we obtain,

$$\pi_1 = \pi_0^*, \quad \text{and} \quad \pi_0 = \frac{\pi_0^* p_1 \lambda - \pi_1^* \ell}{(1-\lambda)}.$$

Additionally, if $p_1 = 0$ and $0 < p_0 < 1$, we obtain

$$\pi_0 = \pi_1^*, \quad \text{and} \quad \pi_1 = \frac{\pi_0^*(1-\ell) - \pi_1^*(1-p_0)(1-\lambda)}{\lambda}.$$

If we assume that $p_0 \neq 1$ and $p_1 \neq 0$ and use our assumption that $0 < \lambda < 1$, it follows that,

$$\pi_0 = \frac{\pi_0^*(1-\ell) - \pi_1(1-p_1)\lambda}{(1-p_0)(1-\lambda)}, \quad \pi_1 = \frac{\pi_1^* \ell - \pi_0 p_0(1-\lambda)}{p_1 \lambda}. \tag{S8.1}$$

By rewriting the expression for π_1 using the expression for π_0, it follows that,

$$\begin{aligned}
\pi_1 &= \frac{\pi_1^* \ell - \pi_0 p_0(1-\lambda)}{p_1 \lambda} \\
&= \frac{\pi_1^* \ell - \frac{\pi_0^*(1-\ell) - \pi_1(1-p_1)\lambda}{(1-p_0)(1-\lambda)} p_0(1-\lambda)}{p_1 \lambda} \\
&= \frac{\pi_1^* \ell - (\pi_0^*(1-\ell) - \pi_1(1-p_1)\lambda)\frac{p_0}{(1-p_0)}}{p_1 \lambda} \\
&= \frac{\pi_1^* \ell - \pi_0^*(1-\ell)\frac{p_0}{(1-p_0)} + \frac{(1-p_1)p_0}{(1-p_0)}\lambda \pi_1}{p_1 \lambda} \\
&= \frac{\pi_1^* \ell - \pi_0^*(1-\ell)\frac{p_0}{(1-p_0)}}{p_1 \lambda} + \frac{(1-p_1)p_0}{(1-p_0)p_1}\pi_1 \\
&= \frac{\pi_1^* \ell - \pi_0^*(1-\ell)\frac{p_0}{(1-p_0)}}{p_1 \lambda} + \frac{(1-p_1)p_0}{(1-p_0)p_1}\pi_1.
\end{aligned}$$

Consequently,

$$\begin{aligned}
(1 - \frac{(1-p_1)p_0}{(1-p_0)p_1})\pi_1 &= \frac{\pi_1^* \ell - \pi_0^*(1-\ell)\frac{p_0}{(1-p_0)}}{p_1 \lambda}, \\
\pi_1 &= \frac{\frac{\pi_1^* \ell - \pi_0^*(1-\ell)\frac{p_0}{(1-p_0)}}{p_1 \lambda}}{\frac{(1-p_0)p_1 - (1-p_1)p_0}{(1-p_0)p_1}}
\end{aligned}$$

$$= \frac{\pi_1^*\ell - \pi_0^*(1-\ell)\frac{p_0}{(1-p_0)}}{p_1\lambda} \times \frac{(1-p_0)p_1}{(1-p_0)p_1 - (1-p_1)p_0}. \tag{S8.2}$$

From expression (S8.2) we now obtain a value for π_1, which we use to obtain a value for π_0 from expression (S8.1). We calculate the prevalence of L (i.e., λ) by,

$$\lambda = p_0, \quad \text{if } p_0 = p_1 \quad \text{and} \quad \lambda = \frac{\ell - p_0}{p_1 - p_0} \quad \text{otherwise.}$$

Subsequently, the expressions for π_0, π_1 and λ can be used to obtain estimates for ϕ_{al^*} using the expression in section Conditional model. Lastly, an estimate for γ can be obtained by fitting a conditional regression model on Y given A and L^*, including an interaction between A and L^*. The coefficient for L^* from this model is then divided by $(\phi_{01} - \phi_{00})$ to get an estimate for γ, holding that $\phi_{01} \neq \phi_{00}$. The inequality $\phi_{01} \neq \phi_{00}$ holds if $p_0 \neq p_1$, in the case that $p_0 = p_1$, γ is not identifiable from the data (and thus, bias is not identifiable). The bias expressions (8.3) and (8.4) in the main chapter of the article can subsequently be used to calculate bias in the average treatment effect estimator.

References

[1] 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

[2] 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