Correction methods for measurement error in epidemiologic research

Nab, L.

4

Regression calibration for measurement error correction: The bias–variance trade off and finite sample performance

Correction of possible bias in exposure-outcome associations due to exposure measurement error using regression calibration may come at the cost of increased variance, referred to as the bias–variance trade off. Notably, in settings where measurement error is relatively large, the finite sample properties of regression calibration have not been investigated. We explore the bias–variance trade off for regression calibration and study the finite sample performance of regression calibration in settings where measurement error is relatively large using Monte Carlo simulation. The bias–variance trade off was of relevance in small samples (sample size <80) and was more pronounced in settings where measurement error was relatively large (reliability = 0.3) and residual error variance of the exposure-outcome association was relatively large (variance = 25). Particularly in settings where measurement error was relatively large (reliability <0.2) and sample size small (sample size <150), the performance of regression calibration was poor with percentage bias ranging from −99%–79% and mean squared error ranging from 6–25431. Application of regression calibration may not be useful in small sample size settings where measurement error is relatively large, because of the overall poor performance of the estimator in these settings.

This chapter has been submitted as: L. Nab and R.H.H. Groenwold, Regression calibration for measurement error correction: The bias–variance trade off and finite sample performance when measurement error is large

4.1. Introduction

Exposure measurement error is common in epidemiologic research but often neglected [1, 2]. When neglected, exposure measurement error can lead to bias in the exposure-outcome association [3], even when measurement error is random [4]. In the rare occasion of measurement error correction in epidemiologic research, regression calibration is among the methods used most often [1, 2]. Regression calibration relies on information about the measurement error model and its parameters, which can be estimated in extra data such as replicates data or internal validation data, or alternatively, informed by expert knowledge [5, 6].

When exposure measurement error is present, the estimator not correcting for this measurement error is typically biased. The application of regression calibration for measurement error correction is of particular interest when bias in the estimator not correcting for the exposure measurement error is relatively large. That is, when measurement error is relatively large, or equivalently, reliability of the error-prone measurement low. Regression calibration is a correction method designed to reduce this bias, at the price of an increased variance [7], a phenomenon referred to as the bias–variance trade off. We are unaware of reports of the finite sample performance of regression calibration in settings of highly unreliable measurements.

In this chapter we demonstrate settings in which the application of regression calibration can be useful, but importantly also when it may not. We report on settings where the estimator not correcting for exposure measurement error may be more efficient in terms of mean squared error than the regression calibration estimator (i.e., the bias–variance trade off). Additionally, we report on the performance of regression calibration in settings where the measurement error in the exposure is relatively large. Specific attention is paid to the performance of regression calibration in small samples. This is illustrated using an example of the association between active energy expenditure and lean body mass.

This chapter is organised as follows. In section 4.2, a study is introduced of the association between active energy expenditure and lean body mass. In section 4.3, the bias–variance trade off is illustrated for regression calibration. The finite sample performance of regression calibration when measurement error is relatively large is studied in section 4.4 by means of Monte Carlo simulation, focusing on settings where sample size is small. We conclude with a discussion of our results in section 4.5.

4.2. Example of lean body mass and energy expenditure

To motivate our study, we use an example of the association between energy expenditure and lean body mass. The association between active energy expenditure (mean active energy expenditure in kilo calories (kcal) per day) and percentage lean body mass (percentage of lean body mass of total body mass) was studied using publicly available data from the cross-sectional Karlsruhe Metabolomics and Nutrition study [8], detailed information on the study can be found here [9]. Body weight was measured in underwear and without shoes using a standardized scale. Lean body mass was measured in a standardized way by dual-energy X-ray absorptiometry and expressed as percentage of total body weight. The energy expenditure (in kcal per day) was measured by Actiheart® (CamNtech, Cambridge, United Kingdom). In addition, energy expenditure was measured using the international physical activity questionnaire (IPAQ). This questionnaire provides a substitute measure of energy expenditure, based on physical activity and expressed in metabolic equivalent of task (MET)-minutes. This measure was then transformed to approximate subject's energy expenditure in kilocalories per day [10].

Throughout this example, we consider energy expenditure measured by Actiheart® the reference measure and energy expenditure measured by the IPAQ the (error-prone) substitute measure. Figure 4.1 shows the agreement between energy expenditure in kcal per day measured by Actiheart® and the IPAQ in the Karlsruhe Metabolomics and Nutrition study. The correlation between the two measures of energy expenditure was 0.10 (95% confidence interval (CI): -0.02;0.21).

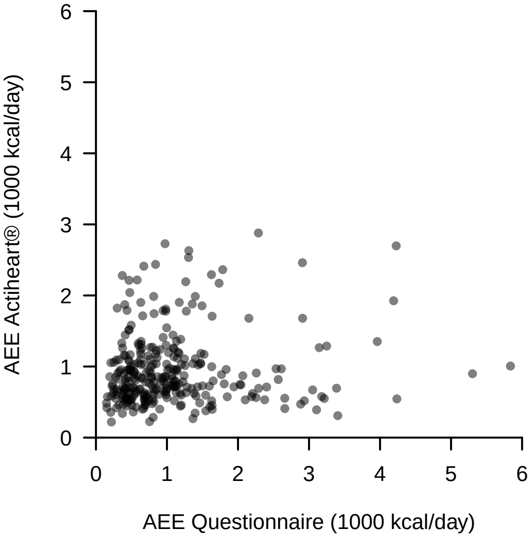

Figure 4.1: Agreement of mean active energy expenditure (AEE) measured by Actiheart® (reference measure) and AEE measured by the international physical activity questionnaire (substitute measure)

Table 4.1 gives an overview of four different estimates of the association between energy expenditure and lean body mass. Using the reference measure of energy expenditure measured by Actiheart®, we found that an increase in energy expenditure of 1,000 kcal per day was associated with a 3.2 increase in lean body mass (95%CI: 1.8;4.7). Using the substitute measure of energy expenditure measured by IPAQ instead, it was found that an increase in energy expenditure of 1,000 kcal per day was associated with a 0.7 *decrease* in lean body mass (95%CI: −1.5;0.1). This estimate was considered biased due to the measurement error in the questionnaire-based energy expenditure level. When this estimate was corrected by means of regression calibration (RC), informed by the relation between the substitute measure and reference measure, we found an increase in energy expenditure of 1,000 kcal per day was associated with an 18.7 *decrease* in lean body mass. Notably, there is a large discrepancy between the point estimate obtained by regressing the reference measure of energy expenditure (measured by Actiheart®) and lean body mass (i.e., 3.2) and the measurement error corrected point estimate (i.e., −18.7). Two different methods for CI construction were available for the RC corrected estimate: the Delta method and the bootstrap percentile method, yielding 95% CIs of −56.6;19.2 and −270.0;217.2, respectively. The above estimates were all adjusted for sex. Covariate adjustment was restricted to sex for illustration purposes, the covariate adjustment set should potentially be expanded.

Table 4.1: Estimates of the association between an increase of 1,000 kcal/day in mean active energy expenditure and percentage lean body mass (adjusted for sex) and associated 95% confidence intervals (CIs) in the Karlsruhe Metabolomics and Nutrition study [8]

Method	**Point Estimate**	**95% CI**
Actiheart®	3.2	1.8;4.7
International Physical Activity Questionnaire	−0.7	−1.5;0.1
Regression Calibration and Delta for CI Construction	−18.7	−56.6;19.2
Regression Calibration and Bootstrap for CI Construction[a]	−18.7	−270.0;217.2

[a] Based on 999 replicates using percentiles

4.3. Bias–variance trade off for regression calibration

The estimator of the exposure-outcome association that does not account for measurement error in the exposure variable is typically biased. Nevertheless, a correction for this bias by means of RC may come at the price of an increased variance, sometimes referred to as the bias–variance trade off [7]. That is, the RC estimator is typically unbiased, yet it is more variable than the uncorrected estimator. Consequently, there may be circumstances where the uncorrected estimator is more efficient in terms of mean squared error (MSE) than the corrected estimator.

We illustrate this phenomenon here by graphical presentation of the MSE of the uncorrected estimator and the RC estimator in simple settings, by using the theoretical derivation of the MSE of the two estimators, described by Carroll et al. [7]. Since the theoretical derivation by Carroll et al. relies on the assumption that the correction factor used in RC is known, which is rare, we expand these results by means of Monte Carlo simulation to simple settings where the correction factor is not known and is estimated from the data.

The data generating mechanism used to generate sets of artificial data is described in Table 4.2. Parameters of the data generating mechanism were inspired by the motivating example of energy expenditure and body mass. For simplicity, we assume random measurement error in the error-prone AEE^* (i.e., AEE^* is distributed around AEE with independent error). In our artificial data, the reliability of AEE^* is equal to 0.25 / (0.25 + τ^2). This ratio is referred to as the ‘reliability’ in this chapter and in case of random measurement error as assumed here, the reliability is equal to the ‘correction factor’ mentioned above. There is an inverse relation between the measurement error variance (i.e., τ^2) and the reliability of the error-prone measure. When the measurement error variance is large, the reliability is low and vice versa.

Table 4.2: Data generating mechanism

Variable	Variable Name	Distribution
Active Energy Expenditure	AEE	$N(1, 0.25)$
Error-Prone Active Energy Expenditure	AEE^*	$N(AEE, \tau^2)$
Percentage Lean Body Mass	LBM	$N(80 + 3 \times AEE, \sigma^2)$

We refer to the estimator of the linear regression of *LBM* on the error-prone measurement of *AEE* (i.e., *AEE**) using ordinary least squares (OLS) as the OLS estimator. We refer to the corrected estimator by means of regression calibration (RC) as the ‘RC estimator’. In this chapter, the RC estimator available in the package `mecor` [6] is used. This package adopts the RC estimator described by Rosner et al. in [11], which is for linear regression equivalent to the method of moments estimator [3]. The RC estimator divides the OLS estimator by a ‘correction factor’ which can be estimated in extra data.

4.3.1. Correction factor known

From the results from Carroll et al. [7] and the data generating mechanism in Table 4.2, the bias in the OLS estimator is equal to 1 minus the correction factor times the effect of *AEE* on *LBM* (i.e., 3 in Table 4.2). The variance of the OLS estimator is equal to the variance of the residual errors (i.e., σ^2 in Table 4.2) divided by the number of observations (i.e., n in Table 4.2) times the variance of *AEE** (i.e., $0.25 + \tau^2$ in the above). The MSE of the OLS estimator is equal to its bias squared plus its variance. The RC estimator is assumed unbiased, and its variance is equal to the variance of the OLS estimator divided by the correction factor squared. Figure 4.2 shows the MSE of the OLS estimator and the RC estimator for different scenarios of variance of the residual errors, sample size and reliability. It illustrates that when the variance of the residual errors is relatively large (25) and sample size is small (≤60), the OLS estimator may be more efficient than the RC estimator in terms of MSE. This gain in efficiency becomes smaller and ultimately turns around in favour of the RC estimator as reliability increases, sample size increases, or the variance of the residual errors decreases. Note that we fixed the variance of *AEE* (i.e., 0.25) and the effect of *AEE* on *LBM* (i.e., 3) throughout this illustration. Varying these will impact the graphical illustrations of the bias–variance trade off in Figure 4.2, but the phenomenon would still exist.

4.3.2. Correction factor not known

We compared the MSE of the OLS estimator and the RC estimator by means of Monte Carlo simulation investigating scenarios of finite samples where the correction factor is *not* known. The correction factor is estimated in an extra data set providing information about the reference measure *AEE* and the substitute measure *AEE**. We generated data using the data generating mechanism described in Table 4.2 and studied MSE for σ^2 equal to 5 or 25, reliability equal to 0.3, 0.6 or 0.9, and the number of observations 20, 40, 60, 80 or 100 in a full-factorial design (2 × 3 × 5 = 30 scenarios). We set the sample size of the set that is used to estimate the correction factor equal to the sample size of the study. For each scenario, 5000 datasets were generated. In each generated data set, the uncorrected effect was estimated by regressing the outcome percentage lean body mass on the error-prone active energy expenditure using standard software. Subsequently, the corrected effect was estimated by means of RC using the R package `mecor` [6]. The performance of these two estimators was evaluated in terms of MSE. Accompanying Monte Carlo standard errors (MCSE) were calculated [12], using the R package `rsimsum` [13]. All code used for the simulation study is publicly available via `https://github.com/LindaNab/woorc`. Figure 4.3 shows the MSE of the OLS estimator and the RC estimator. Overall, the same patterns were obtained as those described in section Correction factor known. An important difference is, however, that the MSE of the RC estimator was much larger than its theoretical derivation when sample size is small, which was most pronounced in the settings where reliability was 0.3 and the residual errors of the outcome model relatively high (i.e., σ^2 equal to 25) (Figure 4.3).

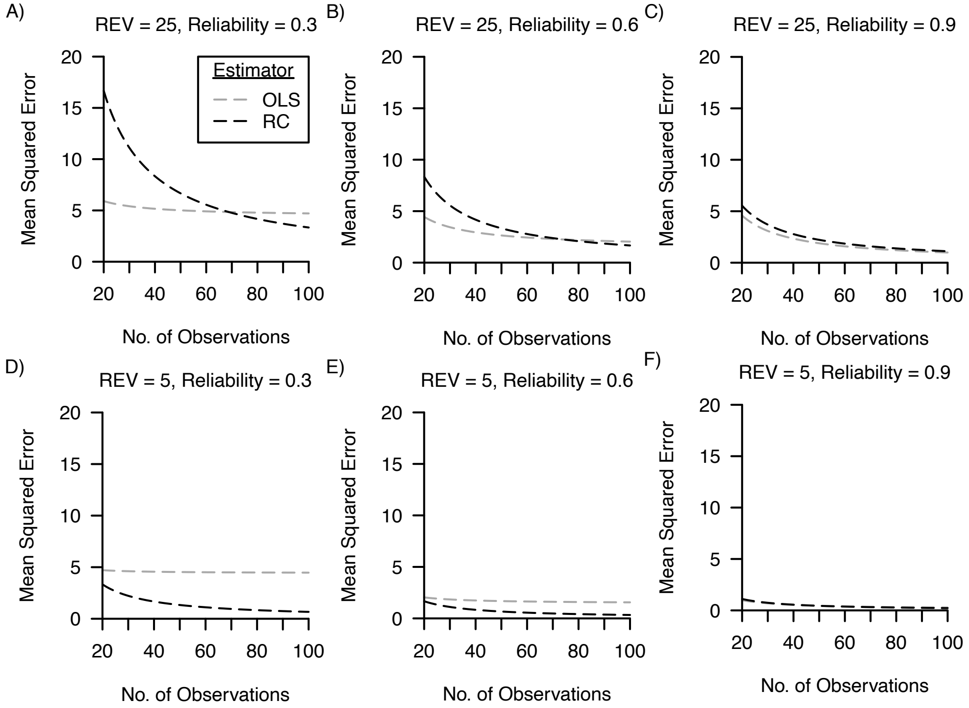

Figure 4.2: Theoretical mean squared error of the estimator not correcting for measurement error (OLS) (gray dashed line) and the regression calibration (RC) estimator (black dashed line), as derived by Carroll et al. [7], for varying sizes of the sample size (20-100, x axis) and for varying sizes of the residual error variance (REV) (25: panels A-C; 5: panels D-F) and for varying size of the reliability (0.3: panels A and D; 0.6: panels B and E; 0.9: panels C and F). In panel F, the lines overlap.

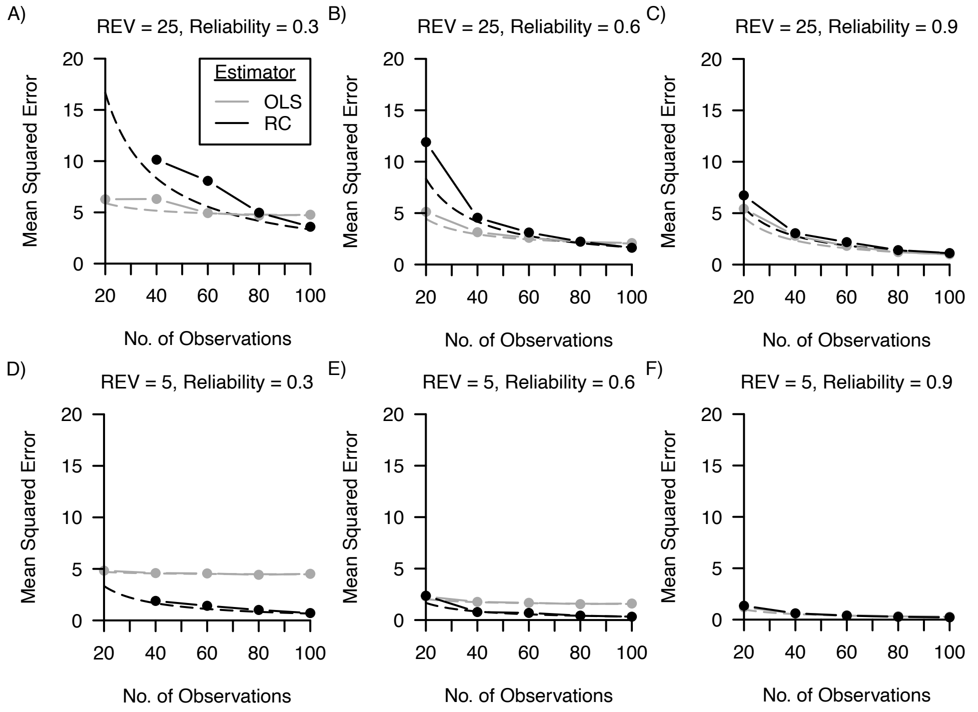

Figure 4.3: Results of a Monte Carlo simulation study of the mean squared error of the estimator not correcting for measurement error (OLS) (gray solid line with dots indicating the estimates) and the regression calibration (RC) estimator (black solid line with dots indicating the estimates) for varying sizes of the sample size (20-100, x-axis) and for varying sizes of the residual error variance (REV) (25: panels A-C; 5: panels D-F) and for varying size of the reliability (0.3: panels A and D; 0.6: panels B and E; 0.9: panels C and F). The dashed gray and black lines represent the theoretical mean squared error of the estimator not correcting for measurement error and the regression calibration estimator, respectively, as derived by Carroll et al. [7]. In panel A and D, the mean squared errors of the regression calibration estimator in the Monte Carlo simulation study fell outside the range of the graph when number of observations was 20, and were 211 (Monte Carlo standard error (MCSE) 142) and 133 (MCSE 89), respectively. In panel F, all lines overlap.

4.4. Finite sample properties of regression calibration

In case of exposure measurement error in a linear regression, RC provides consistent estimates if the correction factor is estimated consistently [3]. A consistent estimate of the correction factor can be obtained in extra data such as internal validation data or replicates data. However, earlier studies (e.g., [14]) suggested that the RC estimator is not necessarily unbiased, specifically in settings where the reliability of the error-prone measurement is low (i.e., 0.2). In addition, in our investigation of the efficiency of the RC estimator described in the previous section, we found that when the reliability was equal to 0.3 and sample size was 20, the MSE of the RC estimator was extremely large compared to the MSE of the OLS estimator (i.e., 211 vs 6 and 133 vs 5, for residual error variance equal to 25 and 5, respectively). Here, we aim to extend these results and investigate the finite sample performance of RC in settings where the measurement error is relatively large (i.e., reliability low), thereby focusing on small samples.

4.4.1. Data generating mechanism

Again, we used the generating mechanism described in Table 4.2. The number of observations (25, 50, 150, 300 and 600) and the reliability of the error-prone exposure AEE^* (0.99, 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05 and 0.01) were varied in a full factorial design (5 × 13 = 65 scenarios). Although our specific interest was the performance of RC for an error-prone measure with low reliability, we studied the full range of the reliability for illustration purposes. We set the sample size of the set that is used to estimate the correction factor equal to the sample size of the study. For each scenario, 5,000 datasets were generated.

4.4.2. Assessment of performance

In each generated data set, the uncorrected effect was estimated by regressing the outcome variable *LBM* on the error-prone AEE^* using standard software. Subsequently, the corrected effect was estimated by means of RC using the R package `mecor` [6]. Ninety-five percent CIs of the uncorrected analysis were constructed using standard software, and for the RC analysis these were constructed using the Delta method and bootstrap resampling using 999 replicates and taking the 2.5% and 97.5% percentiles, both available in the R package `mecor`. Performance of the two different analyses was evaluated in terms of bias, MSE, confidence interval coverage (the proportion of 95% CIs that contained the true value of the true effect), empirical SE, and model based SE. Model based SE was estimated using the standard errors for the uncorrected analysis from standard software, and using the standard errors estimated by the Delta method or the standard deviation of the 999 replicates of the bootstrap samples for the RC analysis. Monte Carlo standard errors (MCSE) were calculated for all performance measures [12], using the R package `rsimsum` [13]. All code used for the simulation study is publicly available via `https://github.com/LindaNab/woorc`.

4.4.3. Results

Figures 4.4 and 4.5 show percentage bias, MSE and confidence interval coverage for varying levels of the reliability of the error-prone measure and number of observations. The OLS estimator was biased, with decreasing bias for increasing levels of reliability. Bias in the

OLS estimator was independent of sample size. Generally, the RC estimator was unbiased, except when reliability was 0.01 for all levels of the sample size. Specifically, for a sample size of 150 and reliability 0.01, the percentage bias was 659% (Monte Carlo SE (MCSE) of bias 23). In addition, the RC estimator was severely biased for a sample size of 50 and reliability equal to 0.01 and 0.05 (percentage bias was 78.7% (MCSE of bias 0.624) and -73.9% (MCSE of bias 2.255), respectively) and for a sample size of 25 and reliability 0.01, 0.05 and 0.1 (percentage bias was -9.4% (MCSE of bias 1.91), -83.7% (MCSE 0.514) and -34.2% (MCSE 0.273), respectively). MSE of the OLS estimator and the RC estimator decreased when reliability increased (Figures 4.4 and 4.5). Generally, the RC estimator was more efficient in terms of MSE than the OLS estimator, except for reliability equal to 0.01 for all sample sizes; reliability ≤ 0.2 for a sample size of 50; or reliability ≤ 0.3 for a sample size of 25 (Table 4.3). In addition, the RC estimator and OLS estimator show similar efficiency for high reliability (i.e., reliability ≥0.9).

Confidence interval coverage was around the nominal level of 95% for the CIs constructed using bootstrap resampling, independent of sample size or reliability. CI coverage was slightly above the nominal level of 95% for the CI constructed using the Delta method for reliability ≤ 0.8 (i.e., ranging between 96%–100%, MCSE <0.05) and at the nominal level for reliability greater or equal to 0.9, independent of sample size. Generally, the coverage of the CIs of the OLS estimator was lower than the nominal level of 95% and moved closer to the nominal level for increasing values of the reliability (ranging between 0%–97%, MCSE < 0.05).

Model based standard errors were equal to empirical standard error of the analysis ignoring measurement error for all studied settings (Figures 4.6 and 4.7 and Table 4.4). Generally, model based standard errors obtained by bootstrap resampling better approximated empirical standard errors of the RC analysis (Figures 4.6 and 4.7). Model based standard error of the RC analysis were equal to empirical standard error for reliability ranging between 0.1–1 for a sample size of 600; reliability ranging between 0.2–1 for a sample size of 300 or 150; and reliability ranging between 0.9–1 for a sample size of 50 or 25 (Figures 4.6 and 4.7 and Table 4.5). For all other studied simulation settings, model based standard errors differed from the empirical standard errors of the RC analysis (Table 4.5).

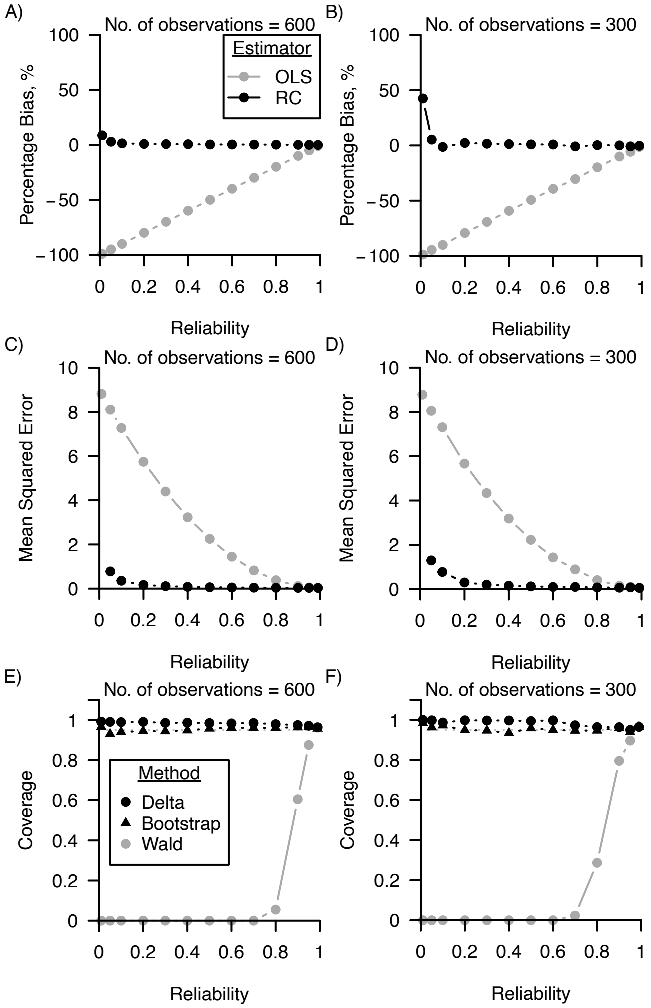

Figure 4.4: Performance of the analysis ignoring measurement error (OLS) and regression calibration (RC) in a setting with 600 (first column) and 300 (second column) observations, in terms of percentage bias (panels A and B); mean squared error (panels C and D) and coverage (panels E and F) for varying values of reliability of the error-prone exposure (x-axis). In panel C and D, the mean squared errors of the regression calibration estimator fell outside the range of the graph when reliability was 0.01, and were 12 (Monte Carlo standard error (MCSE) 2) and 80 (MCSE 29), respectively.

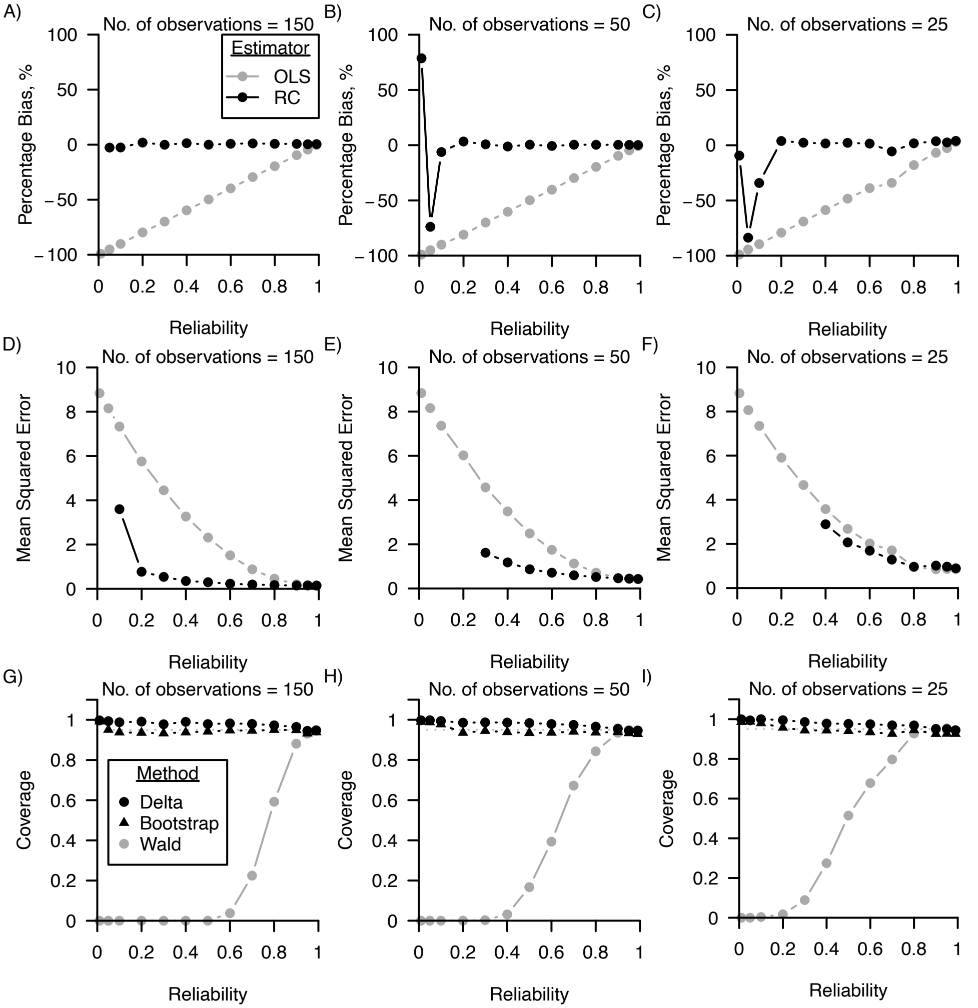

Figure 4.5: Performance of the analysis ignoring measurement error (OLS) and regression calibration (RC) in a setting with 150 (first column), 150 (second column) and 25 (third column) observations, in terms of percentage bias (panels A-C); mean squared error (panels D-F) and coverage (panels G-I) for varying values of reliability of the error-prone exposure (x-axis). In panel A, the percentage bias in the regression calibration estimator fell outside the range of the graph when reliablity was 0.01, and was 659% (Monte Carlo standard error (MCSE) of bias 23). The values that fell outside the range of panels D-F, can be found in Table 4.3

Table 4.3: Mean Squared error (MSE) of the regression calibration estimator in the settings which fell outside the plot range of the graphs in Panel D-F in Figure 4.5

n	Reliability	MSE	MCSE	Panel
150	0.01	2 536 515	2 307 534	D
	0.05	20	3	
50	0.01	1950	446	E
	0.05	25 431	9740	
	0.10	1025	325	
	0.20	20	5	
25	0.01	7092	1062	F
	0.05	1328	231	
	0.10	374	59	
	0.20	630	540	
	0.30	57	37	

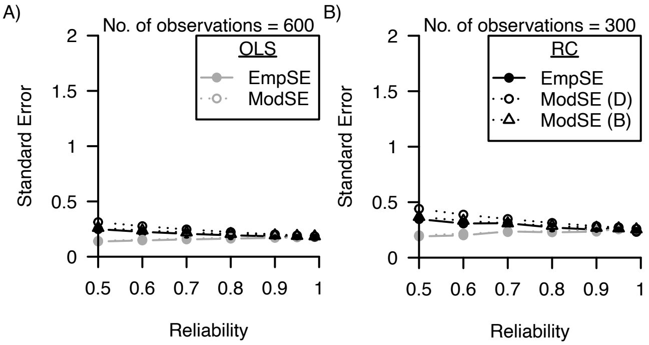

Figure 4.6: Empirical standard error (EmpSE) of the analysis ignoring measurement error (OLS) (solid gray lines with dots indicating the estimates) and regression calibration (RC) (solid black lines with dots indicating the estimates); and model based standard error (ModSE) of the analysis ignoring measurement error (OLS) (dotted gray lines with open dots indicating the estimates) and regression calibration (RC) using the Delta method (D) (dotted black lines with open dots indicating the estimates) or bootstrap resampling (B) (dotted black lines with open triangles indicating the estimates) in a setting with 600 (first column) and 300 (second column) observations for varying values of the reliability of the error-prone exposure (x-axis). The lines of the OLS estimator for the empirical standard error and model based standard error overlap. The lines of the RC estimator for the empirical standard error and model based standard error using bootstrap resampling overlap.

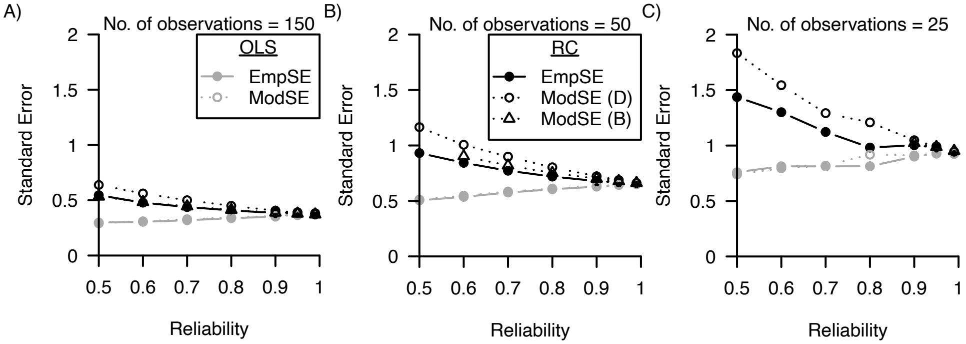

Figure 4.7: Empirical standard error (EmpSE) of the analysis ignoring measurement error (OLS) (solid gray lines with dots indicating the estimates) and regression calibration (RC) (solid black lines with dots indicating the estimates); and model based standard error (ModSE) of the analysis ignoring measurement error (OLS) (dotted gray lines with open dots indicating the estimates) and regression calibration (RC) using the Delta method (D) (dotted black lines with open dots indicating the estimates) or bootstrap resampling (B) (dotted black lines with open triangles indicating the estimates) in a setting with 150 (first column), 50 (second column) and 25 (third column) observations for varying values of the reliability of the error-prone exposure (x-axis). The lines of the OLS estimator for the empirical standard error and model based standard error overlap. The lines of the RC estimator for the empirical standard error and model based standard error using bootstrap resampling overlap in panel A. In panel B, the model based standard error of the RC estimator using bootstrap resampling fell outside the range of the graph for reliability equal to 0.5 and was 2.5 (Monte Carlo SE (MCSE) 0.411). In panel E, the model based standard error of the RC estimator using bootstrap resampling fell outside the range of the graph for reliability equal to 0.5, 0.6, 0.7, and 0.8, and was 211.9 (MCSE 25.489), 27.8 (MCSE 9.575), 6.4 (MCSE 0.551), 2.4 (MCSE 0.241), respectively.

Table 4.4: Empirical standard error (EmpSE) and model based standard error (ModSE) of the analysis ignoring measurement error for varying values of the sample size and reliability of the error-prone exposure measure

n	Reliability	EmpSE	MCSE	ModSE	MCSE
600	0.01	0.02	<0.001	0.02	<0.001
	0.05	0.05	<0.001	0.05	<0.001
	0.10	0.07	0.001	0.07	<0.001
	0.20	0.09	0.001	0.10	<0.001
	0.30	0.11	0.001	0.11	<0.001
	0.40	0.13	0.001	0.13	<0.001
	0.50	0.14	0.001	0.14	<0.001
300	0.01	0.03	<0.001	0.03	<0.001
	0.05	0.07	0.001	0.07	<0.001
	0.10	0.09	0.001	0.10	<0.001
	0.20	0.13	0.001	0.13	<0.001
	0.30	0.15	0.002	0.16	<0.001
	0.40	0.17	0.002	0.18	<0.001
	0.50	0.19	0.002	0.20	<0.001
150	0.01	0.04	<0.001	0.04	<0.001
	0.05	0.10	0.001	0.10	<0.001
	0.10	0.14	0.001	0.14	<0.001
	0.20	0.19	0.002	0.19	<0.001
	0.30	0.24	0.002	0.23	<0.001
	0.40	0.26	0.003	0.26	<0.001
	0.50	0.30	0.003	0.29	<0.001
50	0.01	0.08	0.001	0.08	<0.001
	0.05	0.17	0.002	0.17	<0.001
	0.10	0.24	0.002	0.25	0.001
	0.20	0.33	0.003	0.34	0.001
	0.30	0.40	0.004	0.41	0.001
	0.40	0.46	0.005	0.46	0.001
	0.50	0.51	0.005	0.51	0.001
25	0.01	0.12	0.001	0.12	<0.001
	0.05	0.22	0.002	0.25	0.001
	0.10	0.37	0.004	0.36	0.001
	0.20	0.51	0.005	0.50	0.002
	0.30	0.61	0.006	0.60	0.002
	0.40	0.70	0.007	0.68	0.002
	0.50	0.76	0.008	0.74	0.002

Table 4.5: Empirical standard error and model based standard error using the Delta method or bootstrap (btstrp) resampling of regression calibration and associated Monte Carlo standard errors (MCSE) for varying values of the sample size and reliability

n	Relia-bility	EmpSE	MCSE	ModSE Delta	MCSE	ModSE Btstrp	MCSE
600	0.01	3.44	0.034	28.23	10.792	719.18	73.283
	0.05	0.88	0.009	1.20	0.004	1.42	0.076
	0.10	0.60	0.006	0.81	0.002	0.62	0.001
	0.20	0.41	0.004	0.54	0.001	0.42	0.001
	0.30	0.33	0.003	0.43	< 0.001	0.34	< 0.001
	0.40	0.28	0.003	0.36	< 0.001	0.29	< 0.001
	0.50	0.25	0.002	0.31	< 0.001	0.26	< 0.001
300	0.01	8.84	0.088	17 882.74	8938.773	1006.92	173.947
	0.05	1.13	0.011	1.78	0.012	72.07	28.696
	0.10	0.88	0.009	1.17	0.004	1.08	0.131
	0.20	0.54	0.005	0.77	0.002	0.60	0.001
	0.30	0.45	0.004	0.61	0.001	0.48	0.001
	0.40	0.38	0.004	0.51	0.001	0.41	0.001
	0.50	0.35	0.003	0.44	0.001	0.37	0.001
150	0.01	1592.68	15.928	140 226.41	65 175.844	3727.84	1670.360
	0.05	4.49	0.045	102.07	17.746	1132.07	123.372
	0.10	1.89	0.019	3.35	0.438	97.39	14.841
	0.20	0.88	0.009	1.15	0.004	0.99	0.008
	0.30	0.73	0.007	0.89	0.002	0.72	0.003
	0.40	0.59	0.006	0.74	0.002	0.61	0.002
	0.50	0.55	0.005	0.64	0.001	0.53	0.001
50	0.01	44.09	0.441	12 448.39	2456.747	4439.06	756.806
	0.05	159.47	1.595	306 333.55	65 894.251	4958.66	631.884
	0.10	32.02	0.320	5841.30	1090.653	864.77	225.879
	0.20	4.42	0.044	140.83	22.227	178.76	13.801
	0.30	1.27	0.013	1.73	0.011	47.74	5.814
	0.40	1.09	0.011	1.37	0.007	13.04	1.355
	0.50	0.93	0.009	1.17	0.004	2.47	0.411
25	0.01	84.22	0.842	155 586.92	46 169.38	6699.71	1731.772
	0.05	36.35	0.364	2866.83	263.583	1173.37	48.557
	0.10	19.32	0.193	1048.99	125.869	2287.16	234.035
	0.20	25.09	0.251	4762.48	2369.783	527.13	100.104
	0.30	7.58	0.076	384.43	153.234	628.15	257.196
	0.40	1.70	0.017	2.44	0.112	207.84	78.895
	0.50	1.44	0.014	1.83	0.014	211.88	25.489

4.5. Discussion

This chapter studied settings in which application of regression calibration (RC) may not be appropriate for correcting bias induced by exposure measurement error. Particularly in small samples, the RC estimator may be less efficient in terms of MSE than an estimator not correcting for the exposure measurement error. This bias–variance trade off was most pronounced in settings where reliability was low and residual error variance high. In an investigation of the finite sample properties of RC, we showed that particularly when the measurement error is relatively large and sample size small, RC provided biased estimates, large MSEs and large empirical standard errors. Particularly, in these settings, the model based standard errors did not agree with the empirical standard errors and the RC estimator was instable as shown by large Monte Carlo standard errors.

In settings where the reliability of the error-prone measure was low (i.e., reliability <0.2) and sample size small (i.e., sample size <150), the performance of RC was poor. This is explained by the fact that by application of RC, the uncorrected estimate was divided by an estimate of the correction factor. This correction factor was equal to the reliability of the error-prone measurement in our study. In settings in which the correction factor was close to zero, it was more likely that in one of the replications in the simulation study the correction factor approached zero. Specifically when sample size was small. Consequently, the corrected estimate in that specific replication was large, affecting mean percentage bias, MSE, and the empirical standard error of the setting under study, since outliers affect these summary estimates. Bootstrapped confidence intervals were sensitive to this property as well. That is, independent of the original artificial data, one of the 999 replicates could provide a correction factor approaching zero, affecting the distribution of the estimates in the different bootstrap samples, and thus standard errors based on the standard deviation of that distribution. Taking the 2.5% and 97.5% bootstrap percentiles for CI construction was less sensitive to outliers, but when many of the bootstrap resamples provided a correction factor approaching zero, clearly the percentile-based CIs were affected too.

In our motivating example of active energy expenditure and lean body mass, RC provided an effect estimate that was large compared to the uncorrected estimate (-17.8 versus -0.7) accompanied with wide confidence intervals (-56.6;19.2 (Delta) and -270.0;217.2 (bootstrap)). The large width of the bootstrap confidence intervals can be explained by the fact that the correction factor was small and approached zero in some of the bootstrap resamples.

We only studied relatively simple settings, i.e., random measurement error and univariable models. However, the two phenomena explained here can be extended to settings where the measurement error is not random (e.g., in case of systematic measurement error) and in multivariable models. When differential measurement error is expected, the use of RC for measurement error correction is inappropriate [3, 15].

RC is not only suited for exposure measurement error correction in linear regression models but serves as a fair approximation in logistic regression and survival models as well [3]. In case of logistic regression and survival models, RC is only approximately consistent if ‘measurement error is small’ and the odds ratio or hazard ratio ‘small to moderate’. See for a detailed discussion of RC for logistic regression, Kuha et al. [16] and for a detailed discussion of RC, Carroll et al. [17]. An investigation of the bias–variance trade off and finite sample performance of RC in logistic regression or survival models when measurement error is large is a topic for future research.

RC provides a valuable tool for exposure measurement error correction in epidemiologic studies but may not be particularly useful in settings where sample size is small and reliability of the error-prone exposure low. In those settings, it is advised to replace the substitute error-prone exposure by a more reliable measure of exposure and/or the collection of more data is needed.

References

[1] T. B. Brakenhoff, M. Mitroiu, R. H. Keogh, K. G. M. Moons, R. H. H. Groenwold, M. van Smeden, Measurement error is often neglected in medical literature: A systematic review, Journal of Clinical Epidemiology 98 (2018) 89–97. doi:10.1016/j.jclinepi.2018.02.023.

[2] P. A. Shaw, V. Deffner, R. H. Keogh, J. A. Tooze, K. W. Dodd, H. Küchenhoff, V. Kipnis, L. S. Freedman, Epidemiologic analyses with error-prone exposures: Review of current practice and recommendations, Annals of Epidemiology 28 (11) (2018) 821–828. doi:10.1016/j.annepidem.2018.09.001.

[3] 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.

[4] J. A. Hutcheon, A. Chiolero, J. A. Hanley, Random measurement error and regression dilution bias, BMJ 340 (c2289) (2010). doi:10.1136/bmj.c2289.

[5] R. H. Keogh, J. W. Bartlett, Measurement error as a missing data problem, in: G. Yi, A. Delaigle, P. Gustafson (Eds.), Handbook of measurement error models, 1st Edition, CRC Press, Boca Raton, FL, 2021, Ch. 20, pp. 429–452.

[6] L. Nab, M. van Smeden, R. H. Keogh, R. H. H. Groenwold, Mecor: An R package for measurement error correction in linear regression models with a continuous outcome, Computer Methods and Programs in Biomedicine 208 (2021) 106238. doi:10.1016/j.cmpb.2021.106238.

[7] R. J. Carroll, D. Ruppert, L. A. Stefanski, C. M. Crainiceanu, Bias versus variance, in: Measurement error in nonlinear models, 2nd Edition, Chapman & Halll/CRC, Boca Raton, FL, 2006, Ch. 3, pp. 60–63.

[8] N. Biniaminov, Data from: Irisin, physical activity and fitness status in healthy humans: no association under resting conditions in a cross-sectional study, Dryad, Dataset (2019). doi:10.5061/dryad.ck501.

[9] N. Biniaminov, S. Bandt, A. Roth, S. Haertel, R. Neumann, A. Bub, Irisin, physical activity and fitness status in healthy humans: No association under resting conditions in a cross-sectional study, PLOS ONE 13 (1) (2018) e0189254. doi:10.1371/journal.pone.0189254.

[10] B. E. Ainsworth, W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Basset, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, A. S. Leon, 2011 Compendium of physical activities, Medicine & Science in Sports & Exercise 43 (8) (2011) 1575–1581. doi:10.1249/MSS.0b013e31821ece12.

[11] 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.

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

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

[14] L. Nab, R. H. H. Groenwold, Sensitivity analysis for random measurement error using regression calibration and simulation-extrapolation, Global Epidemiology 3 (2021) 100067. doi:10.1016/j.gloepi.2021.100067.

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

[16] J. Kuha, Corrections for exposure measurement error in logistic regression models with an application to nutritional data, Statistics in Medicine 13 (11) (1994) 1135–1148. doi:10.1002/sim.4780131105.

[17] R. J. Carroll, D. Ruppert, L. A. Stefanski, C. M. Crainiceanu, Regression calibration for survival analysis, in: Measurement error in nonlinear models, 2nd Edition, Chapman & Hall/CRC, Boca Raton, FL, 2006, Ch. 14, pp. 321–323.