Reproduction of Simulation Study to Investigate Impact of Measurement Error

Deliverable 3

Sophie Hetche

2024-11-13

Introduction

This project aims to reproduce and adapt the simulation study from the paper "**Introduction to Statistical Simulations in Health Research**" by Boulesteix et al. (2020). The paper highlights how simulations can be used to evaluate statistical methods, especially when the true effects in real-world data are unknown due to sampling errors or measurement issues.

For illustration purposes, the paper comprises a small statistical simulation to investigate the impact of measurement error when conducting a linear regression analysis. Specifically, the study simulates the effects of measurement errors on the coefficients of a linear regression model that examines the relationship between glycated hemoglobin (HbA1c) and systolic blood pressure (bp). This simulation is reproduced within this project.

Regression Analysis

Firstly, two baseline regression models are run to predict blood pressure with HbA1c. The first regression (A) adjusts for the effects of age and sex in this relationship. The second regression (B) additionally adjust for BMI. The equations corresponding to both linear regressions are as follows:

$$\mathrm{bp} = \beta_0 + \beta_1 \cdot \mathrm{HbA1C} + \beta_2 \cdot \mathrm{age} + \beta_3 \cdot \mathrm{sex} + \epsilon \tag{A}$$

$$\mathrm{bp} = \beta_0 + \beta_1 \cdot \mathrm{BMI} + \beta_2 \cdot \mathrm{HbA1C} + \beta_3 \cdot \mathrm{age} + \beta_4 \cdot \mathrm{sex} + \epsilon \tag{B}$$

Where:

- β_0 is the intercept
- β_1, β_2 and β_3 are the regression coefficients (aka slopes) and
- ϵ is the residual error representing the difference between the observed bp values and the values predicted by the regression model.

Results

Disclaimer: For replication of this simulation, the data can be downloaded in xpt form from here.

The results of the first regression analysis (A) can be found in Table 1, and results of the second regression analysis (B) can be found in Table 2.

Table 1: Linear Regression Results for Blood Pressure (Model A)

Variable	Estimate	Std. Error	Statistic	p-value	Conf. Low	Conf. High
(Intercept)	98.751	1.214	81.332	0	96.371	101.132
Hemoglobin A1C	1.126	0.203	5.551	0	0.729	1.524
Age	0.445	0.013	34.648	0	0.420	0.470
Sex (Male vs Female)	-3.248	0.452	-7.191	0	-4.133	-2.363

Table 2: Linear Regression Results for Blood Pressure (Model B)

Variable	Estimate	Std. Error	Statistic	p-value	Conf. Low	Conf. High
(Intercept)	92.656	1.393	66.506	0	89.925	95.387
Hemoglobin A1C	0.752	0.206	3.650	0	0.348	1.156
Body Mass Index (BMI)	0.286	0.033	8.724	0	0.222	0.351
Age	0.446	0.013	34.979	0	0.421	0.471
Sex (Male vs Female)	-3.631	0.450	-8.060	0	-4.514	-2.748

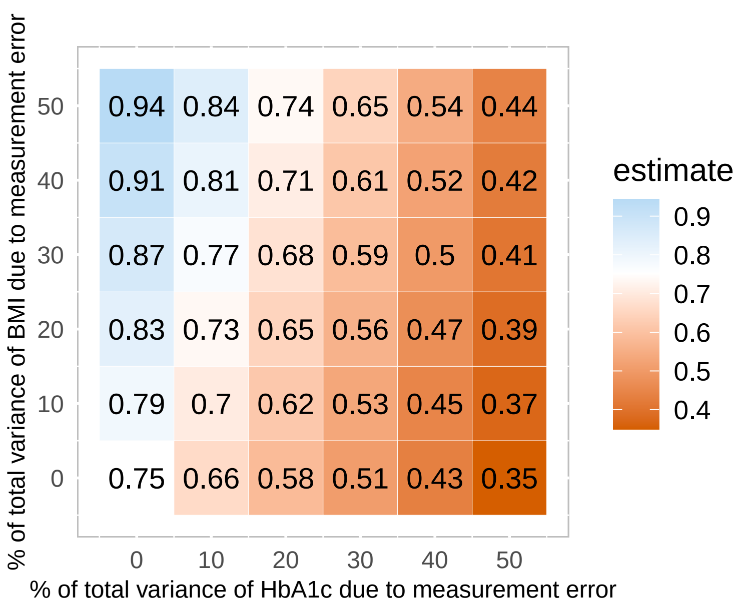
The Simulation

The simulation introduces random errors into the key predictors—HbA1c and BMI—to assess how measurement inaccuracies influence the estimated regression coefficients. These errors were drawn from a normal distribution with varying magnitudes, ranging from no error to up to 50% error in HbA1c and BMI. The simulation was conducted across 1,000 repetitions for each scenario to account for potential variability and assess the impact of measurement error on the model's results.

Results of Simulation

Figure 1 displays the impact of measurement error on the estimated association between HbA1c levels and systolic blood pressure, after adjusting for BMI, across various simulation scenarios.

Figure 1: Impact of Measurement Error



Conclusion

In conclusion, this simulation study demonstrates the significant impact that measurement error can have on the estimation of relationships in linear regression analysis. By simulating different levels of measurement error in key variables such as HbA1c and BMI, we observed how such errors could attenuate or strengthen effect estimates, depending on whether the error occurred in the exposure or confounding variable. This study highlights the importance of accounting for measurement error in epidemiological research, as failure to do so can lead to biased results

References

Boulesteix, A.-L., Groenwold, R. H., Abrahamowicz, M., Binder, H., Briel, M., Hornung, R., Morris, T. P., Rahnenführer, J., & Sauerbrei, W. (2020). Introduction to statistical simulations in health research. *BMJ Open*, *10*(12), e039921. https://doi.org/10.1136/bmjopen-2020-039921