Teaching regression calibration to correct for measurement error to develop statistical thinking

Anarina L. Murillo and Avery Maytin
Brown University, U.S.A.
anarina_murillo@brown.edu

Introductory statistics and biostatistics courses are essential for cultivating statistical thinking. This work explores the effectiveness of teaching measurement error (ME) methods in an introductory biostatistics course to teach three statistical concepts: bias, uncertainty, and decision-making. A tutorial was developed to illustrate the application of ME methods for a specific case study using a sample of U.S. adults (n=600) from the National Health and Nutrition Examination Survey (NHANES). Data were analyzed using logistic regression models to predict diabetes status based on blood pressure levels adjusted for age, race, calories, alcohol intake, and body mass index (BMI). To imitate real world applications of ME, simulated random noise was introduced based on reliability levels. Regression calibration (RC) was applied and improved the estimate of the noisy data. Students were taught that RC may reduce the effect of ME in statistical modeling to reduce bias, and to also improve uncertainty and decision-making. This work demonstrated the utility of an ME tutorial to develop understanding of statistical bias, uncertainty, and decision-making.

INTRODUCTION

A major goal in many introductory courses in statistics and biostatistics is to cultivate statistical thinking. Data literacy is essential for advancing public health and medicine using data-driven and evidence-based approaches. Educators in statistics, mathematics, and data science have formulated learning objectives related to statistical thinking. For examples, Rumsey et al. (2002) states the learning outcomes of statistical literacy should allow students to be good statistical citizens and to develop good research scientist skills to effectively explain, decide, judge, evaluate, and make decisions. This involves basic statistical communication skills such as reading, writing, demonstrating and exchanging statistical information. Similarly, Varlamis (2025) proposed that statistical thinking should lead to improved problem-solving abilities. This requires students to acquire hands-on experience with messy data and exploring real-world data scenarios to cultivate critical thinking skills, data wrangling, and evaluation of data quality in the context of data science. More specifically, gaining experience with the complexities of data science such as inconsistencies, missing values, and noise may support the development of practical data skills and critical thinking in the context of data science by fostering problem-solving skills through the challenge of navigating and managing imperfect data.

Real-world scenarios are expected in introductory courses to motivate statistical methods. In biostatistics courses, a problem-first approach allows students to draw connections in statistics and data science to their respective fields in STEAM, medicine, or public health. A typical curriculum for an introductory statistics course includes basic concepts of statistics, such as, descriptive statistics, types of data variables, common probability models, and basic methods of statistical inference such as parameter estimation, hypothesis testing, and confidence intervals. Biostatistics courses introduce basic statistical methods commonly used in public health, biology, medicine, and the social sciences. This may include exploratory data analysis, as well as understanding and modeling bivariate and multivariate relationships between variables, emphasizing regression analysis for modeling and drawing inference about these relationships. Generally, many introductory courses target undergraduates and graduate students with no prior exposure to college-level statistics. Hence, courses must be designed to

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*.

introduce statistics required for both quantitative fields (statistics, math or other technical subjects) or non-quantitative fields (public health, social science or medicine).

In order to balance methods and applications, these courses are generally designed to be a hands-on and introductory-level overview of commonly-used statistical methods. Many statistical methods are illustrated using real-world data from public health and biomedical research. Courses will often incorporate statistical software including R, STATA, SAS, etc to emphasize practical applications of these methods. The courses also serve as a platform to inspire interest in statistics and recruit the next generation of statisticians. However, rarely do introductory courses include advanced topics beyond regression analysis to deepen statistical thinking and data literacy. Moreover, measurement error methods, such as regression calibration (RC), is an advanced topic that can be presented in introductory courses as it builds off of basic concepts in statistics, probability, calculus, and linear algebra. Through a discussion of measurement error, students can engage in: 1) critical thinking of bias due to errors in data collection based on instruments used, data quality, or missingness; 2) problem-solving of quantifying and describing uncertainty; and 3) evaluation of decision-making when bias is or is not corrected using measurement error methods. This paper presents the utility of teaching measurement error in an introductory biostatistics course to teach three statistical concepts: bias, uncertainty, and decision-making.

METHODS

To introduce measurement error methods in courses, students first learn about the mechanisms generating error in clinical research. Measurement error is calculated by examining the difference between the unknown true value and the measured value of a quantity. Errors can arise from multiple sources and are described as either systematic or random error. Random error is defined as error in measurement which occurs unpredictably and by chance. For example, blood pressure readings vary due to within-individual variability based on the instrument or physiology. Another example is self-reported diet behavior, which may be inaccurate due to participants' misremembering their food consumption. Systematic error is repeatable error that consistently occurs in the same direction, either over- or under-estimating the true values. For example, an incorrectly calibrated blood pressure device will consistently overestimate the true blood pressure of a patient. Hence, measurement error can manifest in different patterns. For example, with self-reported data there may be consistent under-reporting or over-reporting of variables. This is common in nutritional studies, when surveys are used to collect diet behaviors and may be subject to inaccurate responses due to participants' inability to accurately recall food consumption. These concepts give students insight into study design, data collection instruments, and promotes critical thinking of noise versus error in data.

Students then learn about the consequence of measurement error in statistical analysis which is that it leads to bias in statistical inference and incorrect conclusions. Measurement errors can lead to biased estimates of the effects of error-prone measures on the outcomes of interest, loss of statistical power for detecting associations due to potential excess variability, and can obscure true features of the data (e.g., of linear and nonlinear trends and associations between data variables). Several statistical methods are available to reduce bias due to measurement error, such as regression calibration (RC). RC is one of the most widely used methods in epidemiological studies to correct for measurement error. RC methods have shown to reduce bias in analyses involving self-reported data such as diet measures, alcohol patterns, smoking behavior, and physical activity. Moreover, prior studies showed improvements in statistical inference of nutrition data following RC. Prentice et al. (6, 20) showed that unadjusted estimates of food intake were not significantly associated with cancer; however, after bias-adjustment of error using RC, energy and protein density were indeed positively associated with cancer incidence (20). This example illustrates the utility of RC to improve the estimation of the association between diet and cancer. Students can then learn about how to quantify uncertainty before and after measurement error correction. Lastly, students learn how to draw conclusions and make decisions after adjusting for bias in statistical analysis using RC.

Not only is RC effective but it is an accessible statistical topic for students enrolled in an introductory statistics or biostatistics course. RC requires basic understanding of statistics, calculus, and linear algebra making it a suitable topic for an interdisciplinary audience with less statistical training. The purpose of this class lesson was to illustrate the use and value of RC to reduce potential biases due to measurement error in a real-world scenario. Students studied a tutorial with public health data to explore a case study on the association between blood pressure and diabetes status requiring correction for measurement error.

Data Description

To illustrate measurement error methods, a sample of U.S. adults (n=600) from the National Health and Nutrition Examination Survey (NHANES) were analyzed. The NHANES is conducted by the CDC National Center for Health Statistics over 2 year cycles and describes the overall population health of U.S. adults. This study uses curated data from the cardioStatsUSA package (Jaeger *et al*., 2023) in R statistical software which provides data from NHANES 1999-2020. The goal of the real-world scenario is to predict diabetes status based on blood pressure levels adjusted for age, race, calories, alcohol intake, and body mass index (BMI).

Simulating Noise in the Dataset

First, to imitate real world applications of measurement error, simulated random noise was introduced to the dataset based on reliability levels. The artificial random noise is added to imitate random error in the dataset. The data now has the true blood pressure measures from NHANES denoted, *X*, and the simulated blood pressure with error defined as $X^* = X + error$, where $error$ is random noise drawn defined as $error \sim N(0, \sigma^2)$. Variance is based on reliability defined as $\sigma^2 = \frac{1}{reliability - 1}$. When variance is low, this implies that *X** and *X* are similar and thus, there is higher reliability. When variance is high, this implies that *X** and *X* are different and thus, there is lower reliability. Figure 1 below demonstrates different scenarios of additive noise introduced to the dataset ranging from low reliability, moderate reliability, and high reliability comparing the true blood pressure values, *X*, and the simulated blood pressure with error, *X**. As noise increases, the reliability between the two measures decreases.

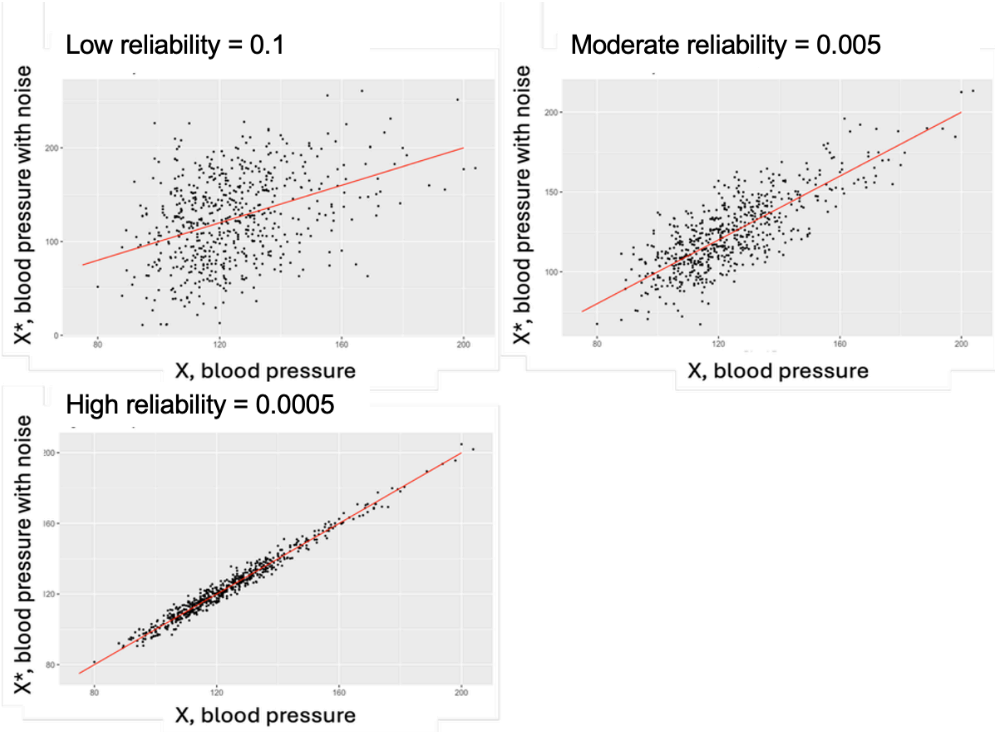

Figure 1. Simulated noise comparing the true blood pressure values, X, and the simulated blood pressure with error, X*. As noise increases, reliability decreases.

Data Analysis

Once data are simulated and noise with one corresponding reliability score is selected, students proceed with conducting the data analysis. This includes:

- Exploratory data analysis and data visualization (Lesson 2)
- Examining bivariate associations between diabetes status, blood pressure, and other demographic or clinical variables (Lesson 3).
- Conducting regression analysis before and after correcting for measurement error (Lesson 4-6)

Logistic regression models were used to predict diabetes status based on blood pressure levels adjusted for age, race, calories, alcohol intake, and body mass index (BMI). Regression calibration (RC) was applied and improved the estimate of the noisy data. A tutorial was developed to demonstrate this lesson.

RESULTS

A publicly available online tutorial was developed and presented to two introductory biostatistics classes for undergraduate and graduate students in public health and medicine (Murillo and Maytin, 2024). The courses have multiple sections with class sizes ranging from 22 to 68 students. Class activities were developed using the online tutorial. The learning goals of the class activities included:

1. Students will understand the main concepts of measurement error, including random and systematic error, as well as the data generating mechanisms underlying measuring error.
2. Students will understand the effects of measurement error on bias, uncertainty and decision-making.

3. Students will be able to detect and assess bias in exploratory data analysis using data visualization and descriptive statistics.
4. Students can apply and analyze statistical methods to correct for error using software.
5. Students can correctly interpret results and draw conclusions of analysis.

The online tutorial includes several lesson chapters: data visualization (Lesson 2), bivariate associations and initial model fitting without error correction (Lesson 3), evaluating modeling diagnostics to examine effects of measurement error (Lesson 4), simulating additive noise in the dataset (Lesson 5), and applying RC methods to regression analysis (Lesson 5). A preview of the tutorial is shown in Figures 2 and 3 below.

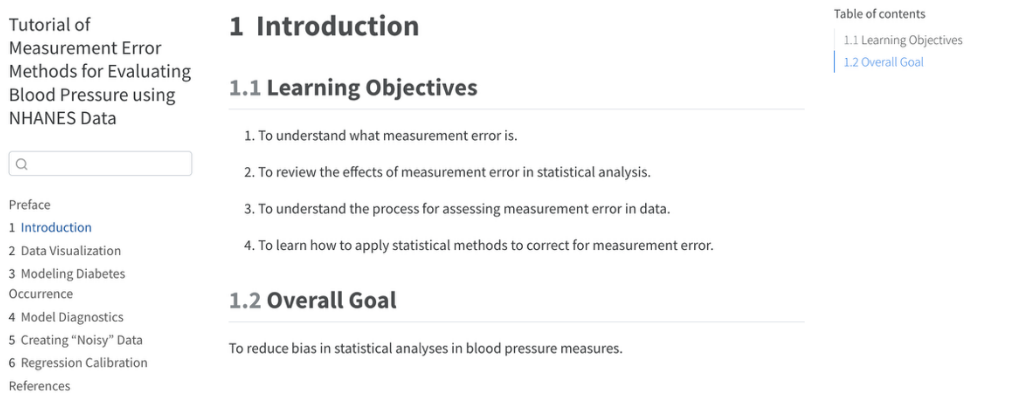

Figure 2. Brief description of learning objectives of the measurement error tutorial.

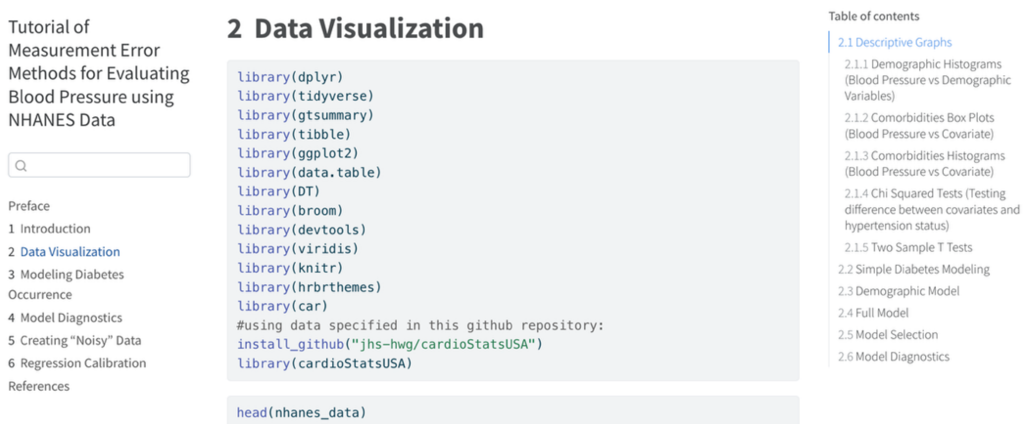

Figure 3. Sample of data visualization and exploratory data analysis.

Students were given a lecture, engaged in class discussion, and instructed to complete a hands-on activity with the tutorial. Several of the learning outcomes were achieved. Students reported improved understanding of statistical inference, statistical reasoning (understanding noise, randomness, variability, and uncertainty). They also developed a clear explanation of the research question and issues with measurement error in the context of the real-world scenario. Finally, they understood the application of RC and were able to interpret results to make decisions.

CONCLUSION

Interdisciplinary approaches are essential for motivating statistics education in public health and medicine, as well as to cultivate statistical thinking and data literacy. This work demonstrated the utility of an online tutorial on ME methods to introduce advanced statistical concepts in introductory courses to reinforce understanding of bias, uncertainty, and decision-making.

REFERENCES

Jaeger, B. C., Chen, L., Foti, K., Hardy, S. T., Bress, A. P., Kane, S. P., Huang, L., Herrick, J. S., Derington, C. G., Poudel, B., Christenson, A., Colantonio, L. D., & Muntner, P. (2023). Hypertension Statistics for US Adults: An Open-Source Web Application for Analysis and Visualization of National Health and Nutrition Examination Survey Data. *Hypertension*, *80*(6), 1311–1320. https://doi.org/10.1161/hypertensionaha.123.20900

Murillo A. L. & Maytin, A. (2024). *Tutorial of Measurement Error Methods for Evaluating Blood Pressure using NHANES Data* [R Software]. GitHub. https://anarinalm.github.io/BP_MeasurementError/visualization.html

Prentice R. L. (1996). Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *Journal of the National Cancer Institute, 88*(23), 1738–1747. https://doi.org/10.1093/jnci/88.23.1738

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of statistics education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910678

Varlamis, I. (2025). Messy Data in Education: Enhancing Data Science Literacy Through Real-World Datasets in a Master's Program. *Education Sciences*, *15*(4), Article 500. https://doi.org/10.3390/educsci15040500