Published in final edited form as:

Epidemiology. 2013 May ; 24(3): 466–467. doi:10.1097/EDE.0b013e31828b284b.

Regression Calibration Is Valid When Properly Applied

Xiaomei Liao,

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA

Donna Spiegelman, and

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA

Raymond J. Carroll

Department of Statistics, Texas A&M University, College Station, TX

Xiaomei Liao: stxia@channing.harvard.edu

To the Editors

Guo et al¹ address an important, previously unsolved problem: how to adjust for measurement error when a confounder is unavailable in the validation study. They propose a multiple imputation approach to solve this problem, under the assumption that $f(Y, X, Z|W)$ follows a multivariate normal distribution with linear means and a constant variance-covariance matrix.

The authors compare their approach to two alternative estimators. The “regression prediction” estimator is what is usually called the “regression calibration” estimator.² If Z is a confounder, that is, a risk factor for the outcome and a correlate of the exposure, X , regression calibration is not valid when Z is not available in the validation study.^{3,4} This was clearly stated in Discussion section of Rosner et al³: “To implement this method, a validation study must be performed to relate true exposure for the covariates measured with error to observed exposure for the covariates measured with error and to the covariates measured without error.” It has also been long known that if Z is a risk factor for the outcome and not a correlate of exposure (ie, not a confounder), then it drops out of the measurement error model, $f(X|W, Z) = f(X|W)$, and regression calibration will be unbiased even when Z is missing from the validation study.⁵ Guo et al¹ focus on the case when Z is a confounder but missing from the validation study. The authors show through extensive simulations that a method known from basic theory to be biased is in fact biased.

Guo et al¹ consider an additional estimator, which they call the “classical calibration” estimator. It is easy to show that this estimator is valid only in the trivial case when there is no measurement error (see eAppendix 1, <http://links.lww.com/EDE/A670>). In the first set of simulations by Guo et al¹, $\beta_0 = 0$ and $\beta_1 = 1.1$. Because this scenario is so close to the classical measurement error case, theory tells us that a large bias should be expected, as found in the simulations presented in Table 1 (see eAppendix 1, case 1, <http://>

links.lww.com/EDE/A670). Even if the mismeasured variable is generated by the linear measurement error model $W = \beta_0 + \beta_1 X + e$, what we need for calibration is $E(X|W) = \alpha_0 + \alpha_1 W$. Using $\hat{X}_{CC} = (W - \hat{\beta}_0)/\hat{\beta}_1$ to estimate $E(X|W)$ as Guo et al¹ did results in a biased estimator (see eAppendix 1, case 2, <http://links.lww.com/EDE/A670>). Because the classical calibration estimator is biased and will typically approach the naive estimator, just about any alternative approach is likely to have better performance. Thus, Guo et al¹ applied two methods to a setting where the methods are theoretically biased and then showed that their method, designed for the case considered, was valid.

An additional point—the multiple imputation method can be viewed as Monte Carlo integration of the likelihood for $f(Y, X, Z|W)$ over the observed data, with $f(Y, Z|W)$ identified in the main study and $f(X|W)$ identified in the validation study. It is clear that the more restrictive surrogacy assumption Guo et al¹ imposed is required for identifiability of their method (see eAppendix 2, <http://links.lww.com/EDE/A670>); it is not clear if the linear, homoscedastic multivariate normality assumption for $f(Y, X, Z|W)$ is also needed. The simulations suggest that possibly not; however, then a misspecified likelihood is being used that will therefore deliver an inconsistent estimator.

Finally, it is worth mentioning that, to maintain some uniformity of terminology, in the measurement error literature, an external validation study is one in which Y is missing but Z is present if it is a confounder, and an internal validation study is one in which Y is present along with Z .^{6–8}

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by NIH/NCI R01CA50597 and NIH/NIEHS R01 ES09411.

References

1. Guo Y, Little RJ, McConnell DS. On using summary statistics from an external calibration sample to correct for covariate measurement error. *Epidemiology*. 2012; 23:165–174. [PubMed: 22157312]
2. Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. *Measurement Error in Nonlinear Models*. 2. London: Chapman & Hall; 2006.
3. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol*. 1990; 132:734–745. [PubMed: 2403114]
4. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr*. 1997; 65(4 suppl):1179S–1186S. [PubMed: 9094918]
5. Armstrong BG, Whittemore AS, Howe GR. Analysis of case-control data with covariate measurement error—application to diet and colon cancer. *Stat Med*. 1989; 8:1151–1163. [PubMed: 2799135]
6. Spiegelman D, Gray R. Cost-efficient study designs for binary response data with Gaussian covariate measurement error. *Biometrics*. 1991; 47:851–869. [PubMed: 1789885]
7. Spiegelman, D. Reliability studies. In: Colton, T.; Armitage, P., editors. *Encyclopedia of Biostatistics*. Sussex, UK: Wiley; 1998. p. 3771–3775.

8. Spiegelman, D. Validation studies. In: Colton, T.; Armitage, P., editors. Encyclopedia of Biostatistics. Sussex, UK: Wiley; 1998. p. 4694-4700.