



Published in final edited form as:

*Environmetrics*. 2013 December ; 24(8): 521–524. doi:10.1002/env.2249.

## Regression calibration in air pollution epidemiology with exposure estimated by spatio-temporal modeling

Donna Spiegelman\*

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

There has been substantial statistical research over the past 20 years on methods to adjust estimates and inference for exposure measurement error and misclassification, with several textbooks summarizing these developments Buonaccorsi (2010); Carroll *et al.* (2006); Gustafson (2003). Regression calibration Rosner *et al.* (1989); Carroll and Stefanski (1990) has been the most widely applied method for correcting for bias in estimation and inference due to exposure measurement error, with a growing number of examples of its use in environmental and occupational health, for example, Horick *et al.* (2006); Keshaviah *et al.* (2003); Li *et al.* (2006); Weller *et al.* (2007); Spiegelman and Casella (1997); Spiegelman and Valanis (1998); Van Roosbroeck *et al.* (2008); Spiegelman *et al.* (2001). In order for measurement error correction methods to be applied, the *main study*, which contains the usual data on the outcome, exposure, and confounders, is augmented by a second study, the *validation study*, in which the main study's surrogate exposure is validated against the gold standard for exposure assessment. In external validation studies as here, transportability must be assumed for valid estimation and inference. The standard transportability assumption is specified in Section 2.1 of this paper in the particular terms of this setting. Optimal design of main study/validation studies to minimize cost and maximize power has been discussed, for example, Spiegelman and Gray (1991); Tosteson and Ware (1990). Air pollution studies in which exposure is estimated by a spatio-temporal model, for example, Yanosky *et al.* (2009), diverge from this paradigm in that no gold standard is typically available. Rather, the first stage analysis involves building the model from environmental data external to the environmental health study. Point estimation in this particular setting is as usual for regression calibration: individual exposure estimates are plugged in to the main study (second stage) health effects model. Consistent estimates of the effect of the exposure on the health outcome under study are thus obtained.

### 1. CONSISTENCY VERSUS ASYMPTOTIC BIAS

The focus of this paper is on a rather unusual topic in statistics: the asymptotic bias of a *consistent* estimator. It should be noted that standard modern statistical estimation and inference procedures rely on consistency virtually universally, rarely if ever striving for asymptotic unbiasedness. See Pierce and Peters (1992) and the Discussion following for further details on the technical points of this debate. Readers should recall that maximum

\*Correspondence to: Donna Spiegelman. stdls@hsph.harvard.edu.

likelihood estimators, including the odds ratio estimator from the logistic regression model, are asymptotically biased. So is the hazard ratio estimated from the Cox model. Even the closed-form Mantel–Haenszel estimators of the odds ratio, rate ratio and risk ratio are asymptotically biased. Similarly, it is well established that the ‘two-stage’ regression calibration estimator that is the focus of this paper is not only consistent but also asymptotically biased, with asymptotic optimality characteristics similar to those of our other popular statistical tools Carroll *et al.* (2006); Carroll and Stefanski (1990).

On a minor technical level, early in Section 2.3, the authors state they are giving consistency conditions but subsequently speak only about asymptotic bias. I found this confusing. I do not believe the conditions given are those required for consistency of the regression calibration estimator, but rather, the conditions may be those needed for lack of asymptotic bias.

Because the regression calibration estimator in the context of air pollution epidemiology developed somewhat independently from parallel developments of regression calibration in statistics, biostatistics, nutritional epidemiology and other areas of environmental and occupational epidemiology (e.g. Horick *et al.*, 2006; Keshaviah *et al.*, 2003; Li *et al.*, 2006; Weller *et al.* (2007); Spiegelman and Casella (1997); Spiegelman and Valanis (1998); Van Roosbroeck *et al.* (2008)), the formal properties of the approach in this setting have not previously been linked to the well-established theoretical developments for measurement error correction in these other areas of inquiry. This disconnect led to an inappropriate use of multiple imputation in air pollution epidemiology. It had previously been proposed that a multiple imputation approach for  $\hat{x}$  from the model for  $E(x^*|\mathbf{z}; w(\mathbf{s}^*))^\dagger$  was needed in order to obtain the correct variance estimator to allow valid Wald-type confidence intervals and  $p$ -values. It was later pointed out by one author of this paper and colleagues Gryparis *et al.* (2009) that this approach was wrong because multiple imputation from the model  $E(x^*|\mathbf{z}; w(\mathbf{s}^*))$  will bias  $\hat{\beta}$  toward the null; the correct imputation must be from a conditional mean model which also depends on  $y$ , that is, from  $E(x^*|y; \mathbf{z}; w(\mathbf{s}^*))$ . It follows that if the point estimator is biased, inference will be biased, so this incorrect multiple imputation procedure will also produce invalid confidence intervals and  $p$ -values.

It is puzzling to me why the authors are so concerned about asymptotic bias in the context of a consistent estimator. Although they are studying a standard regression calibration estimator, the title of the paper suggests that perhaps the authors do not realize that this is the case. It is evident from the results of Tables 1 and 2 that the standard regression calibration estimator is associated with quite a small amount of finite sample bias except in one extreme case considered in Table 1, Scenario 2, where the health effect,  $\beta$ , is very small and the exposure model is a poor fit to the exposure data. Similarly, the scenario studied in Figure 2 is also quite extreme, with  $n^*$ , the validation study size used to fit the exposure model, having only 200 data points, when the usual case is that these models are built in very large data sets orders of magnitude larger than that used to produce this Figure. For example, in Hart *et al.* (2009),  $n^* = 23,565$  over 16 years, with between 369 and 2722 monitors providing data per year!

---

<sup>†</sup>The notation used in this Commentary is that defined by Szpiro and Paciorek unless otherwise indicated

Nevertheless, although the concern about asymptotic bias in a consistent estimator may be misguided, some important points can be gleaned from this work: (1) the usual approach to regression calibration in air pollution epidemiology using spatio-temporal exposure models to estimate exposure in the main study under-estimates the variance of the effect estimate; and (2) under-smoothing of the exposure model in the regression calibration context in air pollution epidemiology leads to finite sample bias. I will discuss each of these in the text that follows.

## 2. THE USUAL APPROACH TO REGRESSION CALIBRATION IN AIR POLLUTION EPIDEMIOLOGY USING SPATIO-TEMPORAL EXPOSURE MODELS TO ESTIMATE EXPOSURE IN THE MAIN STUDY UNDERESTIMATES THE VARIANCE OF THE EFFECT ESTIMATE

Although the earlier proposals for multiple imputation in air pollution epidemiology were incorrect, the problem, which the authors of these proposals were, at least in part, attempting to address, is one that remains unsolved today and has again been demonstrated through the simulations included in this paper: the variance estimate of the regression calibration estimator used typically in air pollution epidemiology, where complex relatively high-dimensional spatio-temporal models are used to predict ambient air pollution concentrations, is incorrect and is, in fact, an underestimate.

This is apparent in Table 2 of this paper, where the standard error of the regression calibration estimator (misleadingly labeled ‘no correction’ when in fact this estimator is a theoretically consistent estimate of the true parameter and is corrected for measurement error) is underestimated. The correct variance is given by the bootstrap standard error in the line in the line below. It should be noted that the bootstrap is but one approach—a rather awkward and computationally intensive one not amenable, in general, to the setting of epidemiologic research where many models need to be run and data sets are typically quite large. Another approach in the big data setting of environmental epidemiology could use the asymptotic variance. Whenever the smoothing algorithm can be written as a richly-parameterized function of many parameters as is the case for many classes of smoothers used to estimate exposure including the ones considered in this paper, the derivation of the asymptotic variance should be straightforward. The estimating equations for  $\hat{\beta}$  from the primary regression model can be stacked up with the estimating equations for  $\hat{\gamma}$  from the exposure smoothing model, and the robust variance method can be used to obtain  $Var(\hat{\beta})$  as usual (See Appendix A.6 in Carroll *et al.*, 2006). If the spatio-temporal smoother used is such that it is not possible to stack up the estimating equations for  $\hat{\gamma}$ , which could be the case, perhaps if penalized regression or lowess smoothing was used, computational feasibility in practice could perhaps be improved by bootstrapping only to obtain  $Var(\hat{\gamma})$  in the first stage validation study; this variance estimate could be used repeatedly in all primary regression models that rely on the same estimated exposures, avoiding the need for many further bootstraps. So, yes, this paper makes an exceedingly important contribution to air pollution epidemiology by the following: (a) establishing that *the usual approach to account for measurement error using regression calibration underestimates the effect estimate*

variance when complex spatio-temporal exposure estimation models are used in the first stage of a two-stage analysis; (b) by proposing one option for addressing this shortcoming; and (c) establishing that the finite sample properties of this approach are good. I encourage the authors to pursue the asymptotic (big data) approach to valid variance estimation in future research.

### 3. UNDER-SMOOTHING OF THE EXPOSURE MODEL IN THE REGRESSION CALIBRATION CONTEXT IN AIR POLLUTION EPIDEMIOLOGY LEADS TO FINITE SAMPLE BIAS

In fact, optimal modeling guidelines for the measurement error model in regression calibration have not been well developed to date. It is well-known in the prediction literature that over-smoothing leads to bias in predictions; these results likely apply directly to the regression calibration setting in air pollution epidemiology when complex, highly parameterized spatio-temporal smoothers are used to obtain  $\hat{x}$ . This is apparent in Table 2, where the regression calibration estimator from the over-smoothed model with 10 degrees of freedom and a cross-validated  $R^2$  of 0.41 was considerably different, and presumably biased, relative to the much more optimally smoothed 5 degree of freedom model with a substantially higher cross-validated  $R^2$  of 0.68. It is not clear if the confounders from the primary regression model were included in the exposure model in the data shown in Table 2, as they should have been; if not, the estimates of  $\hat{\beta}$  would remain biased to the extent that these risk factors for the outcome were also correlated with the underlying true exposure of the MESA-Air study participants. A careful study of optimal exposure modeling in the context of regression calibration is an important direction for future research. I might speculate that standard approaches for choosing the optimal degree of smoothing via cross-validation may work here as well, although it could be of interest to investigate theoretically, if possible, and by simulation, if not, whether this approach will always result in the minimum variance regression calibration estimator.

I would like to make a few additional points raised by my reading of this paper before closing.

### 4. THE DICHOTOMY BETWEEN BERKSON OR BERKSON-LIKE ERRORS AND CLASSICAL OR CLASSICAL-LIKE ERRORS IS NOT USEFUL

To apply regression calibration here, the parameters  $\gamma$  of the model  $E(x^*|\mathbf{z}; w(s^*); \gamma)$  need to be estimated. There is no particular advantage if it turns out that this conditional mean model follows the form expected under Berkson or classical assumptions. Although standard regression calibration results assume homoscedasticity, for example,

$Var(x^*|\mathbf{z}, w(s^*); \gamma) = \sigma_\eta^2$ , it has been found that the estimator is remarkably robust to departures from this homoscedasticity assumption and that second order corrections for heteroscedasticity do not improve performance in finite samples, e.g. Spiegelman *et al.* (2011). In addition, if the sample size of the exposure validation study is small, it may appear that the  $\hat{x}$  in the main study are correlated through their common dependency on  $\hat{\gamma}$ .

Although by standard asymptotic estimating equations theory, these correlations should not cause inconsistency of the estimator, they may need to be accounted for in variance estimation to obtain valid inference. On the other hand, because it has been found that the bias correction version of regression calibration of Rosner and colleagues is algebraically identical to the substitution version of regression calibration originally proposed by Carroll and colleagues under a wide class of generalized linear models Thurston *et al.* (2003), these correlations may vanish asymptotically because it has been shown that the asymptotic variance of the bias correction version of regression calibration does not include any covariance between the uncorrected main study estimate of  $\beta$  and the validation study estimate of  $\gamma$  Spiegelman *et al.* (2001).

## 5. THE MOST IMPORTANT SOURCE OF EXPOSURE MEASUREMENT ERROR IS THAT BETWEEN PERSONAL EXPOSURE AND ESTIMATED AMBIENT EXPOSURE

It is reasonable to assume that given sufficient validation data and transportability of the measurement error model from the exposure validation study to the main study, which seems to be roughly equivalent to conditions 1 and 2, although those conditions may be sufficient but not necessary,  $E(\hat{x}) = x$  for all main study participants. However,  $x$  as defined in this paper is not equal to the main study participant's true exposure, the quantity of interest for health effect estimation, because the participant may not be at their residence all day but traveling to and from other destinations as well as spending significant amounts of time at other locations. In addition, study participants likely spend much time indoors, and indoor penetration of ambient exposure varies and needs to be accounted for. Personal exposure validation studies of exposure to PM<sub>2.5</sub>, PM<sub>10</sub>, black carbon, and other air pollution constituents show that the correlation between the nearest monitor exposure concentration and spatio-temporal predicted exposure with personal exposure concentrations range between 0.3 and 0.6. Liu *et al.* (2003); Meng *et al.* (2005); Sarnat *et al.* (2000); Williams *et al.* (2003a); Williams *et al.* (2003b); Kioumourtzoglou *et al.* (2014). Ideally, we would want to estimate personal exposure from environmental monitor measurements and primary regression model variables—the health risk factor model confounders—to perform a regression calibration analysis that can truly be interpreted as estimating the effect of personal exposure (ambient or total, as is of interest) on the health outcomes.

Unfortunately, personal exposure monitoring is extremely expensive, and these studies are relatively small and conducted in a limited number of locations. This likely makes it infeasible to regress personal exposure,  $v$ , directly on the full array of available ambient exposure data available from routine environmental monitoring, as might be optimal if larger personal exposure validation studies were available. Thus, measurement error correction through regression calibration in an air pollution epidemiology setting may best be considered as a three-stage process:

1. fit the ambient exposure model using the full array of available ambient exposure data from routine environmental monitoring over space and time following one

of the existing approaches, for example, Szpiro *et al.* (2010) or Yanosky *et al.* (2009);

2. using the model developed in step 1 to estimate  $\hat{x}$ , fit the measurement error model for personal exposure,  $v$  as  $\hat{E}(v|\hat{x}, z) = \hat{v}$ , in the personal exposure validation study;
3. using the model developed in step 2 to estimate  $\hat{v}$ , plug  $\hat{E}(v|\hat{v}, z)$  into the primary regression model for the health outcome in the main study, to obtain the regression calibration estimate of  $\beta$ .

It is not clear to me whether the variance from fitting the model in step 1 to estimate  $\hat{x}$  needs to be included in the variance estimation for  $\hat{\beta}$  in step 3. I suspect not, although further research would be needed to prove this, as well as to compare the efficiency of the three-stage approach described above to the standard two-stage approach where  $v$  is regressed directly on  $w(s)$  in the personal exposure validation study.

## Acknowledgments

Support for this research provided by NIH R01ES009411.

## References

- Buonaccorsi, JP. Measurement Error: Models, Methods, and Applications. CRC Press; Boca Raton, FL: 2010.
- Carroll, RJ., Ruppert, D., Stefanski, LA., Crainiceanu, CM. Measurement Error in Nonlinear Models. 2. Chapman & Hall; London: 2006.
- Carroll RJ, Stefanski LA. Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*. 1990; 85:652–663.
- Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*. 2009; 10(2):258–274. [PubMed: 18927119]
- Gustafson, P. Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments. Chapman and Hall CRC; Boca Raton, FL: 2003.
- Hart JE, Yanosky JD, Puett RC, Ryan L, Dockery DW, Smith TJ, Garshick E, Laden F. Spatial modeling of PM10 and NO2 in the Continental United States, 1985-2000. *Environmental Health Perspectives*. 2009; 117(11):1690–1696. [PubMed: 20049118]
- Horick N, Weller E, Milton DK, Gold DR, Li RF, Spiegelman D. Home endotoxin exposure wheeze in infants: correction for bias due to exposure measurement error. *Environmental Health Perspectives*. 2006; 114(1):135–140. [PubMed: 16393671]
- Keshaviah AP, Weller E, Spiegelman D. Occupational exposure to methyl tertiary butyl ether in relation to key health symptom prevalence: the effect of measurement error correction. *Environmetrics*. 2003; 14(6):573–582.
- Kioumourtzoglou M-A, Spiegelman D, Szpiro AA, Sheppard L, Kaufman JD, Yanosky JD, Williams R, Laden F, Hong B, Suh H. Exposure Measurement Error in PM2.5 Health Effects Studies: A Pooled Analysis of Eight Personal Exposure Validation Studies. *Environmental Health*. 2014 in press.
- Li RF, Weller E, Dockery DW, Neas LM, Spiegelman D. Association of indoor nitrogen dioxide with respiratory symptoms in children: application of measurement error correction techniques to utilize data from multiple surrogates. *Journal of Exposure Science and Environmental Epidemiology*. 2006; 16(4):342–350. [PubMed: 16519415]

- Liu LJ, Box M, Kalman D, Kaufman J, Koenig J, Larson T, Lumley T, Sheppard L, Wallace L. Exposure assessment of particulate matter for susceptible populations in Seattle. *Environmental Health Perspectives*. 2003; 111(7):909–918. [PubMed: 12782491]
- Meng QY, Turpin BJ, Korn L, Weisel CP, Morandi M, Colome S, Zhang JJ, Stock T, Spektor D, Winer A, Zhang L, Lee JH, Giovanetti R, Cui W, Kwon J, Alimokhtari S, Shendell D, Jones J, Farrar C, Maberti S. Influence of ambient (outdoor) sources on residential indoor and personal PM<sub>2.5</sub> concentrations: analyses of RIOPA data. *Journal of Exposure Analysis and Environmental Epidemiology*. 2005; 15(1):17–28. [PubMed: 15138449]
- Pierce DA, Peters D. Practical use of higher order asymptotics for multiparameter exponential families. *Journal of the Royal Statistical Society Series B (Methodological)*. 1992; 54(3):701–737.
- Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*. 1989; 8(9):1051–1069. discussion 1071–1053. [PubMed: 2799131]
- Sarnat JA, Koutrakis P, Suh HH. Assessing the relationship between personal particulate and gaseous exposures of senior citizens living in Baltimore, MD. *Journal of the Air & Waste Management Association*. 2000; 50(7):1184–1198. [PubMed: 10939211]
- Spiegelman D, Casella M. Fully parametric and semi-parametric regression models for common events with covariate measurement error in main study/validation study designs. *Biometrics*. 1997; 53(2): 395–409. [PubMed: 9192443]
- Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Statistics In Medicine*. 2001; 20(1):139–160. [PubMed: 11135353]
- Spiegelman D, Gray R. Cost-efficient study designs for binary response data with gaussian covariate measurement error. *Biometrics*. 1991; 47(3):851–869. [PubMed: 1789885]
- Spiegelman D, Logan R, Grove D. Regression calibration with heteroscedastic error variance. *International Journal of Biostatistics*. 2011; 7(1):4. [PubMed: 22848187]
- Spiegelman D, Valanis B. Correcting for bias in relative risk estimates due to exposure measurement error: a case study of occupational exposure to antineoplastics in pharmacists. *American Journal of Public Health*. 1998; 88(3):406–412. [PubMed: 9518972]
- Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman JD. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics*. 2010; 21(6):606–631.
- Thurston SW, Spiegelman D, Ruppert D. Equivalence of regression calibration methods for main study/external validation study designs. *Journal of Statistical Planning and Inference*. 2003; 113:527–539.
- Tosteson TD, Ware JH. Designing a logistic regression study using surrogate measures for exposure and outcome. *Biometrika*. 1990; 77:11–21.
- Van Roosbroeck S, Li R, Hoek G, Lebret E, Brunekreef B, Spiegelman D. Traffic-related outdoor air pollution and respiratory symptoms in children: the impact of adjustment for exposure measurement error. *Epidemiology*. 2008; 19:409–416. [PubMed: 18379431]
- Weller EA, Milton DK, Eisen EA, Spiegelman D. Regression calibration for logistic regression with multiple surrogates for one exposure. *Journal of Statistical Planning and Inference*. 2007; 137(2): 449–461.
- Williams R, Suggs J, Rea A, Leovic K, Vette A, Croghan C, Sheldon L, Rodes C, Thornburg J, Ejire A, Herbst M, Sanders W. The Research Triangle Park particulate matter panel study: PM mass concentration relationships. *Atmospheric Environment*. 2003a; 37(38):5349–5363.
- Williams R, Suggs J, Rea A, Sheldon L, Rodes C, Thornburg J. The Research Triangle Park particulate matter panel study: modeling ambient source contribution to personal and residential PM mass concentrations. *Atmospheric Environment*. 2003b; 37(38):5365–5378.
- Yanosky JD, Paciorek CJ, Suh HH. Predicting chronic fine and coarse particulate exposures using spatiotemporal models for the northeastern and midwestern United States. *Environmental Health Perspectives*. 2009; 117(4):522–529. [PubMed: 19440489]