Published in final edited form as:

Epidemiology. 2010 July ; 21(Suppl 4): S1–S3. doi:10.1097/EDE.0b013e3181dda514.

Opening the Black Box of Biomarker Measurement Error

Enrique F. Schisterman and Roderick J. Little

This special issue of *Epidemiology* concerns two very challenging issues that are commonly confronted in environmental epidemiology, and in particular exposure analysis: the assessment of chemicals found in mixtures and measurements subject to a limit of detection (LOD). When exposures occur in mixtures, collinearity and high dimensionality become pressing issues which make it difficult to distinguish the influences of individual chemicals on the response variable. When exposure levels are low, measurement of chemicals is subject to inadequate instrument sensitivity resulting in a large percentage of measurements falling below the LOD. Often the measurement process is treated as a black box, leading to distortions of the statistical analysis. The following papers describe statistical methods for handling mixtures and LOD, separately and in combination. We hope that these techniques will be applied and will motivate further study into this area of research.

To introduce the special issue, we give a short overview of the papers which include background information regarding the LOD and measurement error; applied data analysis techniques for data subject to a LOD using linear regression, longitudinal models, regression calibration, and Kaplan-Meier estimators; multiple imputation techniques for handling data with a LOD; and approaches for estimating associations between health outcomes and complex exposure mixtures.

In an expository article, Browne and Whitcomb define and compare the LOD, limit of quantification (LOQ), and limit of blank (LOB) thresholds, which commonly arise in epidemiologic studies using biomarkers.¹ They highlight that the choice of an appropriate strategy for dealing with data affected by such limits requires an understanding of the standard experimental and statistical procedures generally used for estimating these different detection limits. These issues are described in the context of analysis of fat-soluble vitamins and micronutrients in human serum.

Assay measurement error leads to the thresholds discussed by Browne and Whitcomb. Guo, Harel and Little use raw calibration data for fat-soluble vitamins to analyze the measurement error throughout the range of measurement.² Using a Bayesian model that allows changes in the variance of the measurement error with the level of the true value to be estimated, they develop prediction intervals for the true value of serum vitamin levels for different observed values. Prediction intervals for values above the LOQ are wider than values below the LOQ, and the width increases with the measured value. Prediction intervals below the LOQ provide more information than simply noting that the value is less than the LOQ. They conclude that the current paradigm of transmitting data from calibration assays provides a distorted picture of the actual measurement error and new methods for communicating measurement error to users are needed.

Other articles in the issue discuss data analysis when variables are subject to values below the LOD or LOQ. Nie et al. study various approaches for linear regression with an independent variable X subject to a LOD, from the statistical viewpoint of the LOD representing an example of left censoring.³ Deletion of cases with levels below the LOD and simple substitution methods are compared with more sophisticated maximum likelihood methods based on normality assumptions. Simulations are conducted to compare the

performance for normal and non-normal data, indicating improved performance of likelihood-based methods when the LOD is a serious problem.

The effects of a LOD on longitudinal data were also explored. Chu et al. apply a segmental Bernoulli/lognormal random effects model to assess and adjust for the effects of left-censored viral loads.⁴ Their methods account for within-subject correlation and accommodate a high degree of censoring. The work is motivated by data from HIV viral load trajectories over 8 years following HAART initiation, in the Multicenter AIDS Cohort Study and the Women's Interagency HIV Study.

Albert et al. evaluate ways of combining information from multiple assays to assess an environmental exposure.⁵ The ideas are motivated by the varying sensitivities and costs of the assays. The authors focus on maximizing efficiency for the case of two assays with different degrees of measurement error, and values below the LOD for subsets of individuals.

Whitcomb et al. consider the analysis of a calibration experiment that includes data from multiple batches performed within the main experiment.⁶ Conventionally, the calibration experiment from each batch is used to calibrate each batch independently. This approach incorporates batch variability but is subject to limitations given the small number of calibration measurements in each batch. The authors compare this approach with mixed effects models and simple pooling of data across batches. Using a real data example with biomarker and outcome information, they show that risk estimates may vary depending on the calibration approach utilized. Under minimal interbatch variability, as shown in the data, conventional batch-specific calibration is not the best use of available data and results in attenuated risk estimates.

Several authors consider use of multiple imputation for data affected by detection limits. LOD issues can be viewed as a missing data problem, where values below the LOD or LOQ are known to lie within an interval, but the precise value is missing. A popular modern technique of missing data analysis is multiple imputation, which creates multiple data sets with different imputed values that are subsequently analyzed using simple multiple imputation combining rules (e.g. Rubin, 1987, Little and Rubin, 2002). Multiple imputation is suggested as a promising method for analysis of LOD in this special issue.

Chen et al. describe the use of multiple imputation to address LOD issues with serum dioxin concentrations.⁷ The methods are used to quantify the population-based background concentrations of dioxin in serum using data from the University of Michigan Dioxin Exposure Study (UMDES) and the National Health and Nutrition Examination Survey (NHANES) 2001-2002. Linear and quantile regression methods for complex survey data are used to estimate the mean and percentiles of background serum dioxin concentrations for females and males aged 20-85 years. These methods and results have wide application for studies focusing on the concentrations of chemicals in human serum and in environmental samples.

Kang considers the presence of artificial zero values in datasets.⁸ Artificial zero values may result due to rounding error, replacement of observations below the LOD, or a variety of other reasons. Kang proposes and examines parametric and distribution-free methods for comparing such data sets, specifically extending the empirical likelihood technique for estimating confidence intervals in datasets that contain artificial zeros due to a LOD, while allowing for robust comparisons of different populations of interest.

Gillespie investigates the reverse Kaplan-Meier (KM) estimator for use in estimating the distribution function, and thus population percentiles, for left-censored data.⁹ This method

leads to efficient estimation of the distribution and population percentiles. The author also provides guidance on how to utilize built-in Turnbull estimators to achieve the not often built-in reverse KM estimator.

Analysis of associations between health outcomes and complex mixtures is complicated by the lack of knowledge regarding causal components of the mixture, highly correlated mixture components, potential synergistic effects of mixture components, and measurement difficulties. Herring extends recently proposed nonparametric Bayes shrinkage priors for model selection to these settings by developing a formal hierarchical modeling framework to allow for different degrees of shrinkage for main effects and interactions, and to handle truncation of exposures with a LOD.¹⁰

Gennings et al. evaluated the relation between polychlorinated biphenyl (PCB) mixtures exposure and risk of endometriosis in women in response to varying selection of congeners in the literature.¹¹ An optimization algorithm was developed to determine the weights in a linear combination of scaled PCB levels that lead to the strongest possible association with risk of endometriosis. Integrating toxicological and biologic interpretation with refined estimation procedures can create testable hypotheses that might not otherwise be explored.

In conclusion, while much progress has been made in the measurement and analysis of environmental exposures, further research is needed in this challenging area of epidemiology. We hope that the papers contained in this special issue will not only be applied to current research involving problems of mixtures and LOD, but will also stimulate further research leading to new analytical methods. The heightened focus on epigenomics makes the evaluation of gene-environment interactions with data on mixtures of exposures even more important, along with the development of new study designs that emphasize cost and analytical efficiency. Good solutions for such problems require an integrated approach, combining the best of epidemiologic, basic science, and statistical research. Some current statistical assumptions in the analysis of biomarkers are not always true, and the only way to correct such assumptions is through statistical models that incorporate more realistic scientific assumptions. We hope that further advances in the design and analysis of studies of environmental exposures will enable us to assess the effects of multiple small exposures on human health outcomes.

Acknowledgments

This project was funded by a grant to Enrique F. Schisterman by the long range initiative of the American Chemistry Council (ACC). Through their support, two working groups were formed to independently tackle the issues of environmental mixtures and environmental exposures subject to a LOD. The LOD group was composed of Drs. Paul Albert, Rick Browne, Haitao Chu, Steve Cole, Ying Guo, Ofer Harel, Rod Little, Aiyi Liu, Lei Nie, Neil J. Perkins, Enrique F. Schisterman, Albert Vexler, and Brian Whitcomb. The mixtures group consisted of Germaine Buck Louis, Ed Carney, Chris Gennings, Patrick Heagerty, Amy Herring, Neil J. Perkins, Anindya Roy, Enrique F. Schisterman, Rajeshwari Sundaram, and Albert Vexler.

This work was also funded in part by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health.

References

1. Browne R, Whitcomb BW. Procedures for determination of detection limits: application to high-performance liquid chromatography analysis of fat soluble vitamins in human serum. *Epidemiology* 2010;21:S4–S9. [PubMed: 20220523]
2. Guo Y, Harel O, Little RJ. How well quantified is the limit of quantification? *Epidemiology* 2010;21:S10–S16. [PubMed: 20526201]

3. Nie L, Chu H, Liu C, Cole SR, Vexler A, Schisterman EF. Linear regression with an independent variable subject to a detection limit. *Epidemiology* 2010;21:S17–S24.
4. Chu H, Gange SJ, Li X, et al. The effect of HAART on HIV RNA trajectory among treatment-naïve men and women: a segmental Bernoulli/lognormal random effects model with left censoring. *Epidemiology* 2010;21:S25–S34. [PubMed: 20386106]
5. Albert P, Harel O, Perkins N, Browne R. Use of multiple assays subject to detection limits with regression modeling in assessing the relationship between exposure and outcome. *Epidemiology* 2010;21:S35–S43. [PubMed: 20386105]
6. Whitcomb BW, Perkins NJ, Albert PS, Schisterman EF. Treatment of batch in the detection, calibration, and quantification of immunoassays in large-scale epidemiologic studies. *Epidemiology* 2010;21:S44–S50.
7. Chen Q, Garabrant DH, Hedegeman E, et al. Estimation of background serum 2,3,7,8-TCDD concentrations by using quantile regression in the UMDES and NHANES populations. *Epidemiology* 2010;21:S51–S57. [PubMed: 20220524]
8. Kang L, Vexler A, Tian L, Cooney M, Buck GM. Empirical and parametric likelihood interval estimation for populations with many zero values: application for assessing environmental chemical concentrations and reproductive health. *Epidemiology* 2010;21:S58–S63.
9. Gillespie BW, Chen Q, Reichert H, et al. Estimating population distributions when some data are below a limit of detection by using a reverse Kaplan-Meier estimator. *Epidemiology* 2010;21:S64–S70. [PubMed: 20386104]
10. Herring A. Nonparametric Bayes shrinkage for assessing exposures to mixtures subject to limits of detection. *Epidemiology* 2010;21:S71–S76. [PubMed: 20526202]
11. Gennings C, Sabo R, Carney E. Identifying subsets of complex mixtures most associated with complex diseases: polychlorinated biphenyls and endometriosis as a case study. *Epidemiology* 2010;21:S77–S84.