

## Topics in Epidemiologically-Based Risk Assessment

Louise Ryan<sup>1</sup>, Brent Coull<sup>1</sup>, Chris Paciorek<sup>1</sup>, and Warren Strauss<sup>2</sup>. <sup>1</sup>*Harvard School of Public Health.*  
<sup>2</sup>*Battelle Memorial Laboratory.*

The first part of this project aimed to develop and apply improved statistical methods for the analysis of environmental exposure data and to quantify the uncertainty associated with those analyses.

Specific statistical issues that were addressed include:

- Development and application of latent variable methods for analyzing multiple exposure measurements;
- Use of source apportionment methods to assess health effect;
- Quantification of the impact of mistiming errors in exposure assessment; and
- Quantification of model uncertainty

The research is expected to advance statistical methods for handling the kinds of complex exposure data that arise in environmental health applications. Another benefit is the training of undergraduate and graduate students in environmental statistics.

The second part of this project builds on part 1 and addresses a series of statistical methodology issues associated with the design and analysis of spatially and temporally correlated biomonitoring and exposure data. Our goals are (1) to develop computationally efficient methods for fitting spatio-temporal models to biomonitoring data and (2) to provide guidelines for practical and efficient biomonitoring designs.

Our research will fill important gaps in terms of quantitative methods for analyzing and interpreting biomonitoring data. It will also provide practical and cost-saving tools for guiding researchers and policy makers on effective designs for collecting biomonitoring data.

**Implications:** Biomonitoring has emerged in recent years as powerful strategy for assessing human chemical exposures based on sampling and analysis of biological samples taken from fluids such as blood, urine, or breast milk, or tissues such as hair, nails, or bone. Despite widespread interest in biomonitoring, there are concerns that advances in measurement technologies can easily outpace the ability of scientists and policy-makers to interpret the resulting data. This project aims to address such concerns by applying rigorous statistical theory to develop sound design guidelines for biomonitoring studies. The challenges include taking account of spatial and temporal correlations, handling multiple endpoints, and accounting for measurement error and missing data.

**Start and end date:** August 2004 – December 2006 (part 1); August 2007 – July 2008 (part 2).

### Presentations:

Ryan, L. (2008). Computationally efficient fitting of generalized linear mixed models, with application to spatio-temporal modeling of diseases. Invited presentation at the Department of Biostatistics, University of Washington, April 2008.

Ryan, L. (2008). Computationally efficient fitting of generalized linear mixed models, with application to spatio-temporal modeling of diseases. Invited presentation at the National Cancer Institute, May 2008.

Ryan, L. (2006). The use of hierarchical models to combine data from multiple endpoints and multiple sources. Invited presentation at the American Statistical Association Joint Statistical Meeting, Seattle, Washington, August 2006.

Ryan, L.M. (2006). Statistics, the science of uncertainty. Buckingham Scholar Lecture Series, University of Miami. Oxford, Ohio,

Coull, B. (2006). Structural equation models for assessing source-specific health. Invited speaker at Institute for Prevention of Cardiovascular Disease at the Beth Israel Deaconess Medical Center, March 2006.

Gryparis, A. (2006). Semiparametric latent variable regression models for modeling the spatial and temporal variability of black carbon concentrations in the Greater Boston area. Invited talk at Spring Meeting of the Eastern North American Region (ENAR) of the International Biometrics Society, Tampa, Florida, March 2006.

Coull, B. (2006). Invited talk at an EPA-sponsored workshop Interpretation of Epidemiologic Studies of Multipollutant Exposure and Health Effects, Research Triangle Park, North Carolina, December 2006.

Schwartz, J. (2006). The association of black carbon (predicted from a latent variable model) with mortality in Boston. Presented at Annual Meeting of the ISEE, Paris, France. September 2006.

Tonne, C. (2006). The association of black carbon (predicted from a latent variable model) with myocardial infarction in a case-control study in Worcester. Presented at Annual Meeting of the ISEE meeting, Paris, France, September 2006.

Zeka, A. (2006). The association of black carbon (predicted from a latent variable model) with birth weight in eastern Massachusetts. Presented at Annual Meeting of the ISEE, Paris, France, September 2006.

#### **Peer-reviewed publications:**

Nikolov, M.C., Coull, B.A., Catalano, P.J., and Godleski, J.J. (2008). Statistical methods to evaluate health effects associated with major sources of air pollution: A case study of breathing patterns during exposure to concentrated Boston air particles. *J R Stat Soc Ser C (Applied Statistics)* 57(3): 357-378 (doi:10.1111/j.1467-9876.2008.00618.x).

Schwartz, J., Coull, B.A., Laden, F., and Ryan, L.M. (2008). The effect of dose and timing of dose on the association between airborne particles and survival: A Bayesian Model Averaging approach. *Env Health Persp.* 116(1): 64-69.

Maynard, D., Coull, B.A., Gryparis, A., and Schwartz, J. (2007) Mortality risk associated with short term exposure to traffic particles and sulfates. *Env Health Persp* 115(5): 751-755.

Morara, M., Ryan, L., Houseman, A., and Strauss, W. (2007). Optimal design for epidemiological studies subject to designed missingness. *Lifetime Data Analysis* 13: 583-605.

Gryparis, A., Coull, B.A., Schwartz, J., and Suh, H.H. (2007). Semiparametric latent variable regression models for spatio-temporal modeling of mobile source particles in the Greater Boston area. *J R Stat Soc Ser C. (Applied Statistics)* 56 (2): 183-209 (doi:10.1111/j.1467-9876.2007.00573.x).

Sánchez, B.N., Budtz-Jørgensen, E., Ryan, L.M. (2006). An estimating equations approach to fitting latent exposure models with longitudinal responses. *Biometrika*. (Submitted).

Sánchez, B.N., Budtz-Jørgensen, E., Ryan, L.M., and Hu, H. (2005). Structural equation models: A review with applications to environmental epidemiology. *JASA* 100(472): 1443-55.

Li, Y. and Ryan, L.M. (2005). Survival analysis with heterogeneous covariate measurement error. *JASA* 99: 724-35.

Morales, K.H., Ibrahim, J., Chen, C.-J., and Ryan, L.M. (2006). Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *JASA* 101(473): 9-17.

Nikolov, M.C., Coull, B.A., Catalano, P.J., and Godleski, J.J. (2006). An informative Bayesian structural equation model to assess source-specific health effects of air pollution. *Biostatistics* 8(3): 609-624 (doi:10.1093/biostatistics/kxl032).

**Other publications:**

Whitney M., et al. Bayesian model averaging techniques for covariate adjustment in environmental epidemiology. (In preparation for submission).

**Sponsors in addition to the LRI:** None.

**Abstract revision date:** May 2008