A comparison of measurement error and between-pollutant relationships from alternative exposure assessment approaches for use in epidemiology studies

EPA United States Environmental Protection Agency

This research is funded by U.S. EPA - Science To Achieve Results (STAR) Program Grant # R834799

Kathie L. Dionisio[1], Lisa K. Baxter[1], Howard H. Chang[2]

[1]National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC, [2]Emory University, Atlanta, GA

Georgia Institute of Technology

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH

Background

Measurements from central site (CS) monitors are often used as exposure metrics in air pollution epidemiological studies. These measurements lack spatiotemporal coverage and do not account for exposures in different microenvironments (e.g. indoors and in-vehicle) where pollutant infiltration and indoor sources can substantially impact total exposures. All of these factors may contribute to an underestimation of exposure variability in the study population, and may lead to exposure error. In addition, CS monitors may not accurately capture multipollutant relationships, particularly for local pollutants with fine-scale spatial variability. This analysis is motivated by the hypothesis that more refined exposure estimates may more accurately represent between-pollutant relationships, and thus may result in health associations with potentially less bias and more precision as we move towards multipollutant epidemiological analyses. Analyzing the degree of measurement error in exposure estimates will allow for an assessment of the potential impact on epidemiologic model coefficients. We developed a set of alternative daily metrics of exposure for 169 ZIP codes in the Atlanta, GA metropolitan area, from 1999-2002, for $PM_{2.5}$ and its components (EC, SO_4), O_3, CO, and NO_x. Single pollutant analyses both within and between exposure metrics can be found in Dionisio et al., JESEE 2013, under review. Presented here are results of between-pollutant analyses.

Methods

Study design

- Study period: 1999-2002
- Study area: 169 ZIP codes in the 20 county Atlanta, GA metropolitan area
- Pollutants investigated: $PM_{2.5}$ and its components (EC, SO_4), CO, NO_x, O_3 (24-hr avg, 8-hr max for O_3)

- 3 tiers of exposure metrics
  1) Central site monitor measurements ("CS") ← No spatial variability
  2) Modeled ambient estimates ("Ambient") ← ZIP code level
     - Regulatory background from a statistical model
     - Local contribution from an emissions model (AERMOD)
  3) Modeled exposure estimates from the Stochastic Human Exposure and Dose Simulation model (SHEDS) ("Exposure") ← ZIP code level

Analytical methods

- Daily pollutant concentrations and measurement error were normalized by dividing by the annual average CS measurement for each pollutant, to allow for comparison across pollutants
- 3 types of measurement error (difference in concentration between metrics)
  - $\delta_{spatial}$ = Ambient - CS
  - $\delta_{personal}$ = Exposure – Ambient
  - $\delta_{spatial+personal}$ = Exposure - CS
- Between-pollutant correlations, across tiers of exposure metrics
- Between-pollutant correlations of the measurement error
- In boxplots, upper and lower limits are at the 25th and 75th percentiles, whiskers are at the 5th and 95th percentiles, and the line near the middle of the box represents the mean

Results

Exposure estimates and measurement error

- Mean ambient and exposure model estimates are lower than CS measurements

- Local pollutant concentrations have relatively more spatial variability (i.e. wider boxplots) compared to regional pollutants, for both ambient and exposure model estimates

- Measurement error for local pollutants
  - Spatial variability present to varying degrees across all types of error
  - Magnitude of measurement error varies by type of error
  - Spatial variability of measurement error implies differential measurement error is present
- Measurement error for regional pollutants
  - Little spatial variability; degree of spatial variability is consistent across types of error
  - Magnitude of measurement error varies by type of error
  - Total measurement error ($\delta_{spatial+personal}$) is comprised mostly of measurement error due to human exposure factors ($\delta_{personal}$, e.g. time-activity patterns, air exchange rate in the home, etc.)

Figure 1. Normalized exposure estimates from 3 methods

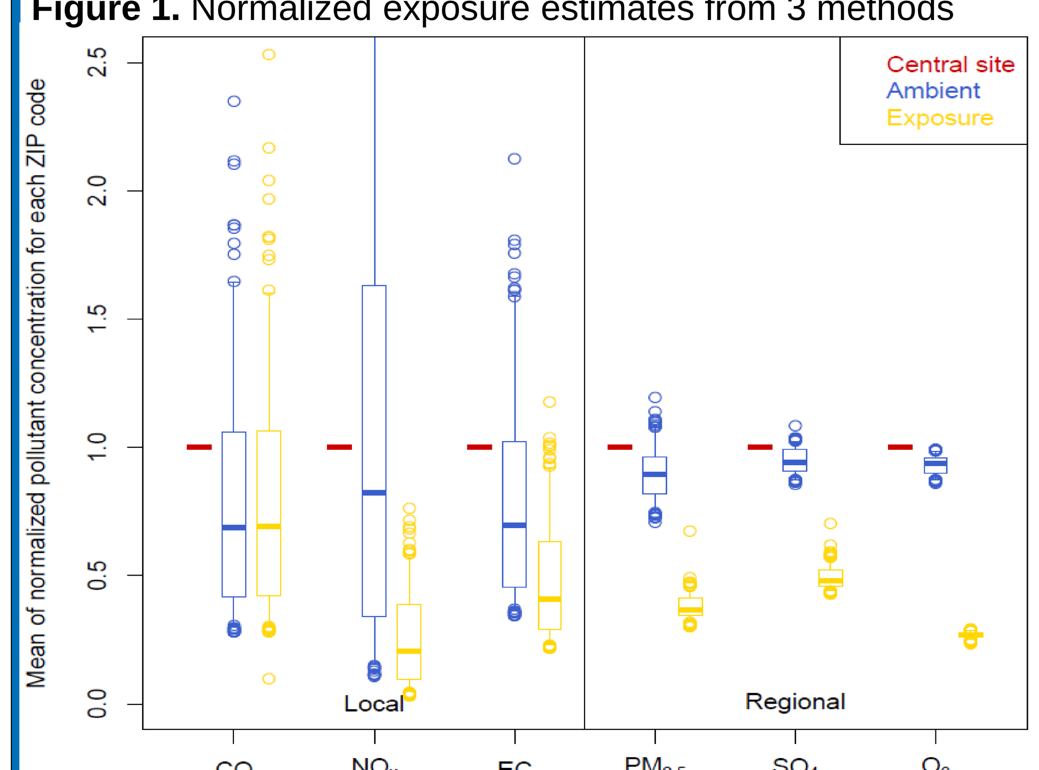
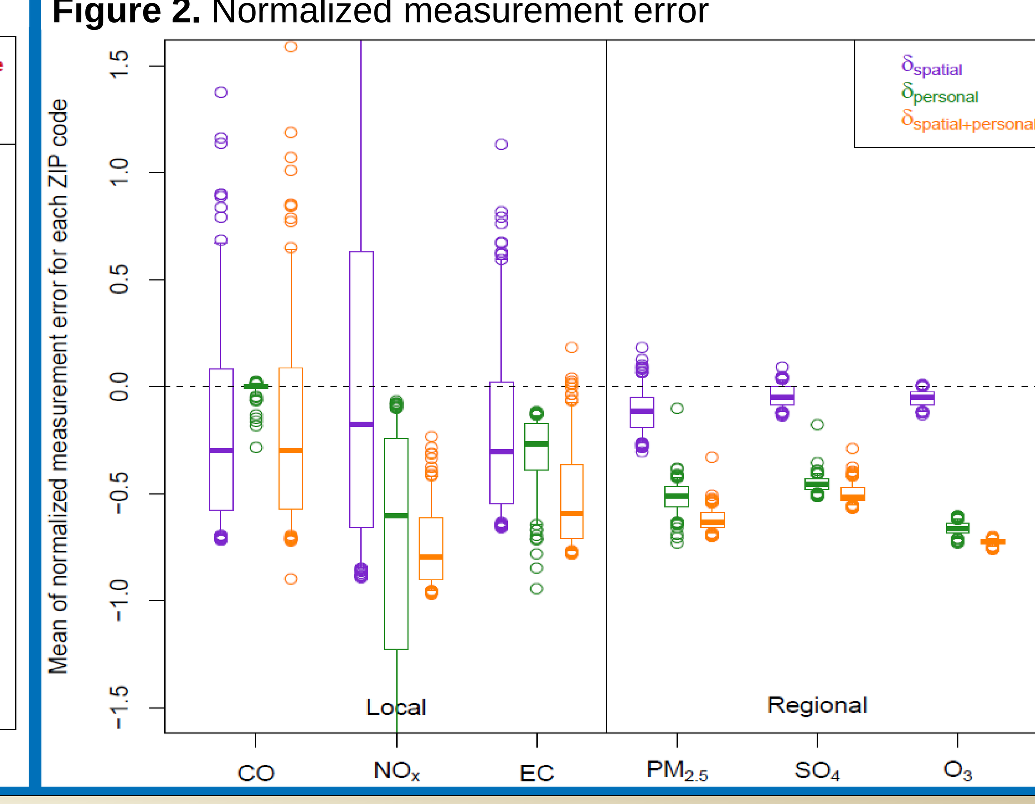


Figure 2. Normalized measurement error

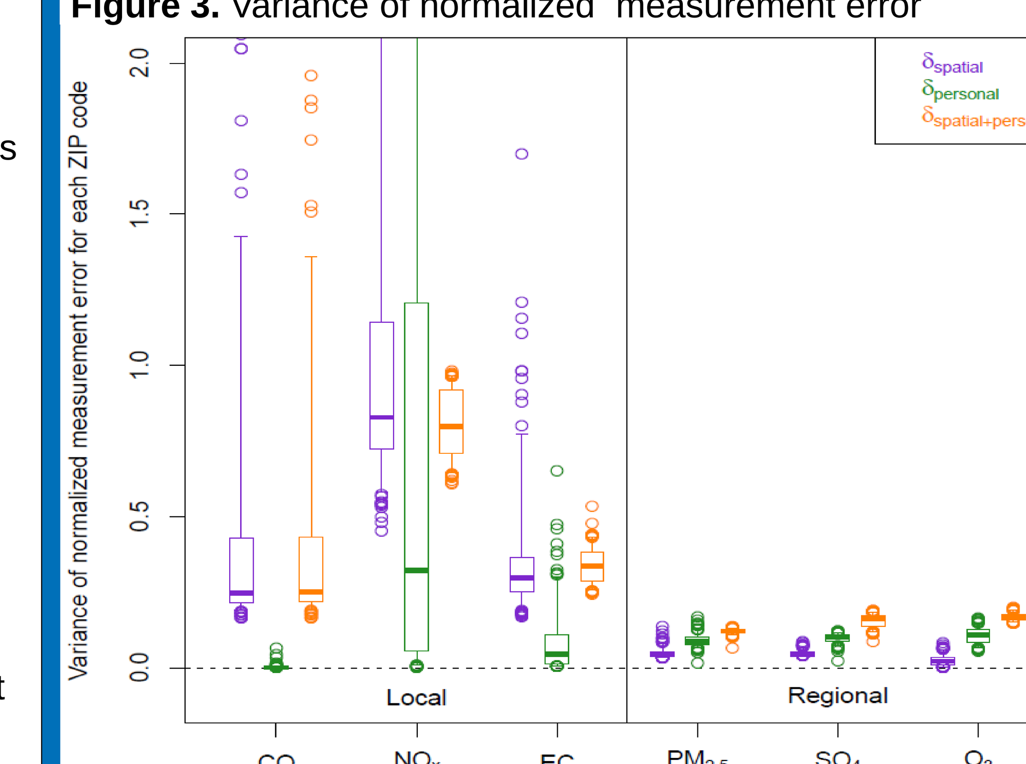


Variance of measurement error

- Regional pollutants
  - Variance of measurement error is low (near zero) and has little spatial variability

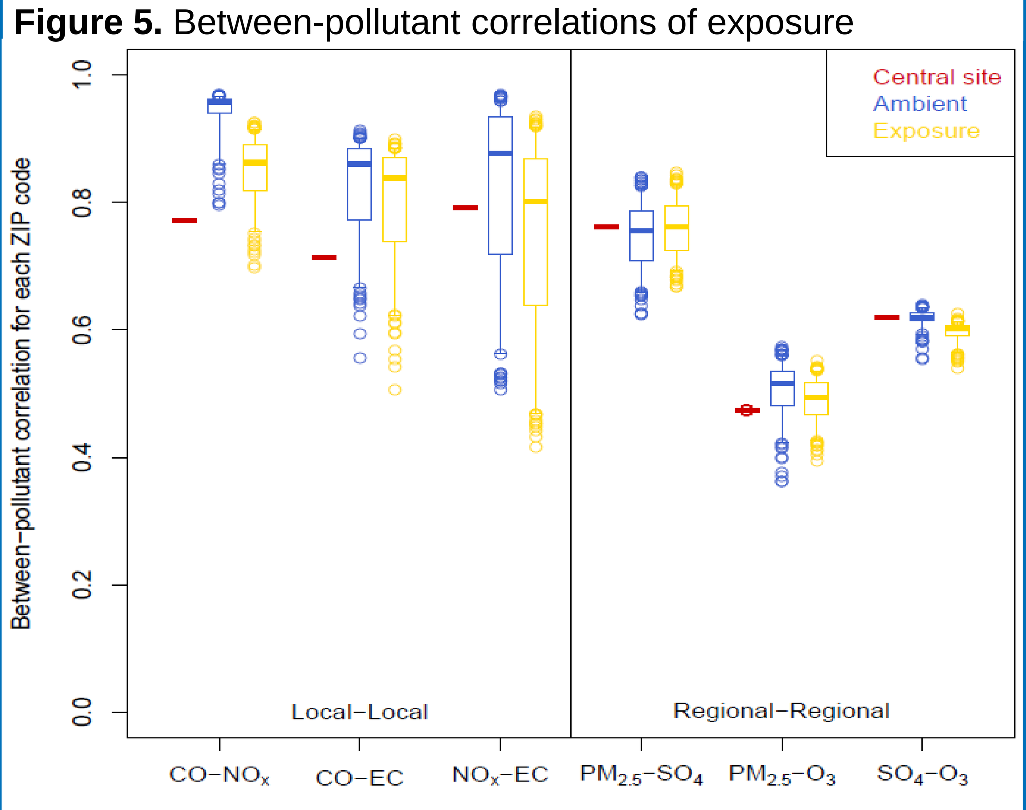
- Local pollutants
  - Variance of measurement error is greater than zero
  - Spatial variability is present
  - Magnitude is dependent on pollutant and on type of error

Figure 3. Variance of normalized measurement error



Between-pollutant correlations of exposure

Figure 5. Between-pollutant correlations of exposure



- Moderate-strong, positive, correlations for all pairs, with stronger correlations for local-local pollutant pairs

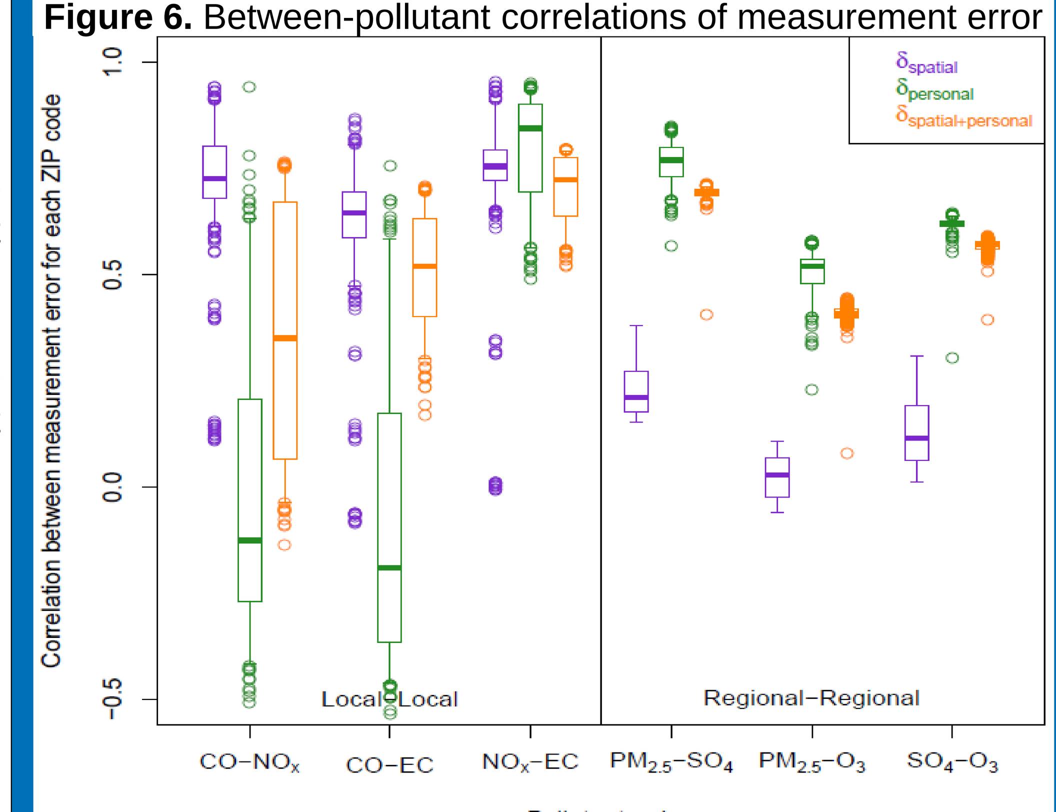
- Correlations of local-regional pollutant pairs are more varied and typically weaker (not shown)

- Spatial variability (i.e. width of the boxplot) is present for some correlations in the ambient and exposure tiers, especially for CO-EC and NO_x-EC pairs

- Between-pollutant correlations of local-local pollutant pairs are typically higher when ambient or exposure model estimates are used compared to CS measurements

Between-pollutant correlations of measurement error

Figure 6. Between-pollutant correlations of measurement error



- Correlation of error for pollutant pairs is dependent on the pollutant pair and the type of error

- Measurement error between pollutants can be moderately-highly correlated

- Spatial variability (i.e. width of the boxplot) is present in the correlation of measurement error

Shown in Figure 4 is the 20 country Atlanta metropolitan area.

Filled areas represent ZIP codes, colored lines represent major roads.

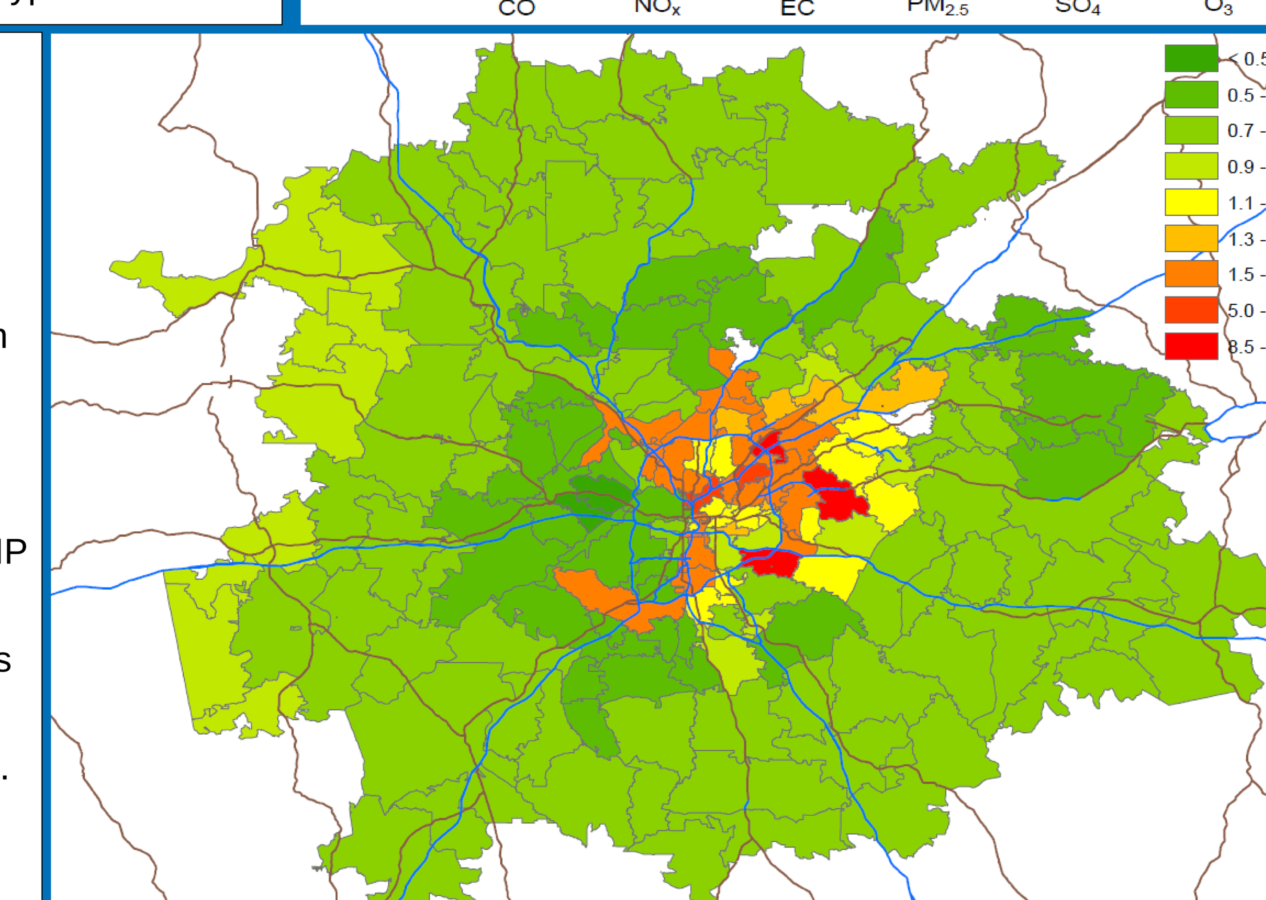


Figure 4. Variance of normalized spatial measurement error ($\delta_{spatial}$) for NO_x

Epidemiologic implications and conclusions

- Pollutant concentrations are correlated, and sometimes highly so
- Correlation of measurement error between pollutants can be moderate-strong
  - Dependent on pollutant pair and type of measurement error
  - May cause bias and attenuation of coefficients in related bivariate pollutant epidemiologic models
- Variance of measurement error
  - Often non-zero for local pollutants, must be aware of the potential for bias in model coefficients
  - Close to zero for regional pollutants, bias may be minimal in epidemiologic model

- Magnitude of the variance of measurement error is dependent on the type of measurement error
  - Magnitude of the bias will differ depending on the metric used to represent exposure

- Must do a better job of characterizing intraurban variation in exposure to local pollutants in order to avoid differential measurement error due to the spatial variability of measurement errors and their between-pollutant correlations, which may lead to differential bias and attenuation of epidemiologic model coefficients

- Stronger between-pollutant correlations of exposure for ambient and exposure model estimates compared to CS measurements
  - Modeling may provide additional information on between-pollutant relationships missed when CS measurements are used

- A simulation (Table 1) shows the possible impact on model coefficients for a bi-pollutant model under the conditions shown here for local pollutants: positively correlated exposures ($Corr(x_1,x_2)$), unequal variance of measurement error ($Var(\delta_1)$, $Var(\delta_2)$), and varying degrees of measurement error correlation ($Corr(\delta_1, \delta_2)$)

Table 1. Predicted bias in bivariate regression coefficients under different correlations (corr) between the true exposures and measurement errors with indicated variances (var) when both variables have a true effect: $\beta_{x_1} = \beta_{x_2} = 1.0$.

$Corr(x_1,x_2)$	$Var(\delta_1)$	$Var(\delta_2)$	$Corr(\delta_1, \delta_2)$	$E(\hat{\beta}_{z_1})$	$E(\hat{\beta}_{z_2})$
0.0	1.0	1.0	0.0	0.50	0.50
0.5	1.0	1.0	0.0	0.60	0.60
-0.5	1.0	1.0	0.0	0.33	0.33
0.0	1.0	1.0	0.5	0.40	0.40
0.0	1.0	1.0	-0.5	0.67	0.67
0.0	0.5	2.0	0.0	0.67	0.33
0.5	0.5	2.0	0.0	0.71	0.53
0.5	0.5	2.0	0.3	0.66	0.27
0.5	0.5	2.0	0.5	0.64	0.21
0.5	0.5	2.0	0.7	0.64	0.14
0.5	0.5	2.0	-0.5	0.83	0.50
0.5	0.5	2.0	-0.7	0.91	0.57
0.5	0.5	2.0	-0.9	1.00	0.66

We assume $var(x_1) = var(x_2) = 1$.

The authors would like to acknowledge Janet Burke, Vlad Isakov, Jim Mulholland, Jeremy Sarnat, Stefanie Sarnat, and Haluk Ozkaynak for their contributions to the development of exposure metrics used in this analysis.

Table 1 from: Zeger, S. L., et al. (2000). " Environ Health Perspect **108**(5): 419-426.