Advances in Modeling Exposure to Support Environmental Health Studies

Valerie Garcia, Vlad Isakov*, Tim Watkins

Office of Research and Development (ORD), National Exposure Research Laboratory (NERL), U. S. EPA, RTP, NC 27711

*Corresponding author address: Vlad Isakov, PhD, U.S. EPA, Mail Drop E243-04, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, Tel: (919) 541-2494, E-mail: Isakov.Vlad@epa.gov

ABSTRACT

Protecting public health from the harmful effects of exposure to air pollutants is a key concern for many environmental agencies. Air quality regulations designed to protect public health are supported by epidemiology studies, many of which have used measurements from a few central-site ambient monitors to characterize air pollution exposures. Relying solely on central-site ambient monitors does not account for the spatial-heterogeneity of ambient air pollution patterns or the influence of infiltration and indoor sources. Using this traditional approach is especially problematic for air pollutants that exhibit significant spatial-heterogeneity because misalignment between measurements and exposure can result in misclassification error. If this misclassification error is differential (different for the diseased and non-diseased groups), the health effect measure calculated from the epidemiology model may over- or under-state the risk, or may fail to detect a risk where risk occurs or vice versa. Thus, methods to refine exposure estimates are critical. This paper describes novel approaches to refine exposure estimates and then applies three of the techniques in an epidemiology study conducted in the New York City metropolitan area to illustrate the impact of misclassification on the study results.

Keywords: air pollution, epidemiologic studies, exposure, air quality modeling

1. INTRODUCTION

Concerned about the health impacts of air pollution, many environmental agencies in the world implement rules and regulations to reduce emissions from various sectors to maintain ambient air quality at acceptable levels. The United States Environmental Protection Agency (EPA) sets the National Ambient Air Quality Standards (NAAQS) for criteria pollutants, such as

ozone, nitrogen dioxide, and particulate matter to protect public health and the environment. However, because the NAAQS target pollutant concentrations in the ambient air the impact on reducing actual human exposure to harmful air contaminants may vary significantly and air quality management programs designed to meet the NAAQS may also vary in their effectiveness at protecting public health. Therefore, to fully assess the nature and magnitude of the air pollution problem, it is important to explicitly consider exposure to achieve a better understanding of how pollutants emitted into the environment ultimately impact human health.

Numerous health studies have used measurements from a few central-site ambient monitors to characterize air pollution exposures. Relying solely on central-site ambient monitors does not account for the spatial–heterogeneity of ambient air pollution patterns, or the influence of infiltration and indoor sources (Jerrett et. al. 2005, Sarnat et. al, 2006). Central-site monitoring becomes even more problematic for certain PM components (e.g., metals) or size fractions (e.g., coarse, ultrafine) that exhibit significant spatial-heterogeneity. This variation may be influenced by meteorology as well as emissions from both regional and local sources. In addition, using central site monitors does not reflect personal exposure contributions, such as time in microenvironments. Given that people spend the majority of their time indoors, the infiltration of outdoor air indoors and indoor sources can greatly affect the personal-ambient exposure relationship. Improving air pollution exposure characterization will result in more accurate risk estimates of associated health effects to inform future development of NAAQS and other air pollution regulations.

Not all analyses, however, will require the same level of exposure characterization depending on the goals of the epidemiologic assessments. More advanced methods to characterize exposure are needed in studies where intra-urban gradients of air pollutant concentrations are important. For example, while PM mass concentrations may be similar across locations, the composition may be very different, making the understanding of sources very important. In addition, health studies conducted across multiple geographic locations require an understanding of geographic specific factors that impact personal-ambient exposure relationships, such as climate, housing stock, and commuting patterns. Better refining exposure estimates even for those pollutants that are known to be spatially homogeneous may reveal associations not discernable before.

This paper illustrates the impact of this spatial misalignment on the health effect measure calculated from a study conducted in the New York City (NYC) metropolitan area to examine

the association between asthma and ozone. In this study, various definitions of exposure are applied in an epidemiology model to investigate the impact of the exposure metric on the inference of the health effect measure. Hence, the objectives of this paper are to (1) discuss the need for advanced approaches for estimating exposure and (2) illustrate how these advanced methods can impact epidemiologic analyses using an example case in New York City.

2. APPROACHES FOR ESTIMATING EXPOSURES

Estimates of ambient concentrations have been enhanced by utilizing both measurements and modeling tools. Statistical interpolation techniques and passive monitoring methods can provide additional spatial resolution in ambient concentration estimates. In addition, spatio-temporal models, which integrate GIS data and other factors, such as meteorology, have been developed to produce more resolved estimates of ambient concentrations. Models, such as the Community Multi-Scale Air Quality (CMAQ) model, estimate ambient concentrations by combining information on meteorology, source emissions, and chemical-fate and transport (Byun and Schere, 2006). In addition, hybrid modeling approaches, which integrate regional scale models (such as CMAQ) with local scale dispersion models, provide new alternatives for characterizing ambient concentrations.

Publically available data on housing characteristics and commuting patterns can be utilized to understand the personal-ambient exposure relationships. The age and size of the home will affect the proportion of personal exposure due to ambient air, and commuting patterns will influence how representative a central site ambient monitor is to ambient exposure. However, since publically available data are limited, modeling approaches, such as the Stochastic Human Exposure and Dose Simulation Model (SHEDS), to estimate personal exposure are being developed (Burke et. al. 2001, Özkaynak et. al. 2007). The SHEDS model is a population exposure model that calculates distribution of exposures within the study population. An integrated air quality and exposure modeling system provides the means to predict the distribution of exposures for the population of interest in various microenvironments by linking air quality modeling information to SHEDS. An integrated modeling system could also be operationally applied in air quality management practices such as standard setting, standard implementation, risk mitigation and accountability (Isakov et. al. 2006, Isakov et. al. 2009).

3. APPLICATION OF ADVANCED APPROACHES IN A HEALTH STUDY

In this illustrative study, we apply three of the techniques discussed above to characterize ozone exposure in the New York City (NYC) metropolitan area: 1) ozone concentration surfaces over the study area derived from monitoring data using statistical interpolation techniques; 2) ozone concentration estimates based on emissions and meteorological data using the chemical transport model CMAQ; and 3) ozone exposure estimates using the SHEDS model with inputs from CMAQ. The study investigated the association between ozone exposure and respiratory-related hospital admissions for five summers (June – August; 2001 – 2005) in four counties of the NYC metropolitan area (Bronx, New York, Queens and Kings). First, we examine the key features in exposure metrics produced using these different approaches. Figure 1 compares spatial aspects of two exposure metrics: ozone concentrations averaged over the study period modeled exposures from SHEDS. The figure illustrates that the exposure concentration levels estimated with SHEDS is much lower than the interpolated observations since individuals spend a large amount of time indoors and also that the spatial heterogeneity in exposure estimates that is achieved by incorporating exposure factors such as housing type, infiltration rates, or individual's locations and activities. One question to be addressed in this paper is whether the more spatially resolved exposure estimate generated from the SHEDS model will result in a more significant association with the health endpoints, despite the fact that the exposure levels are lower than the spatially interpolated estimates.

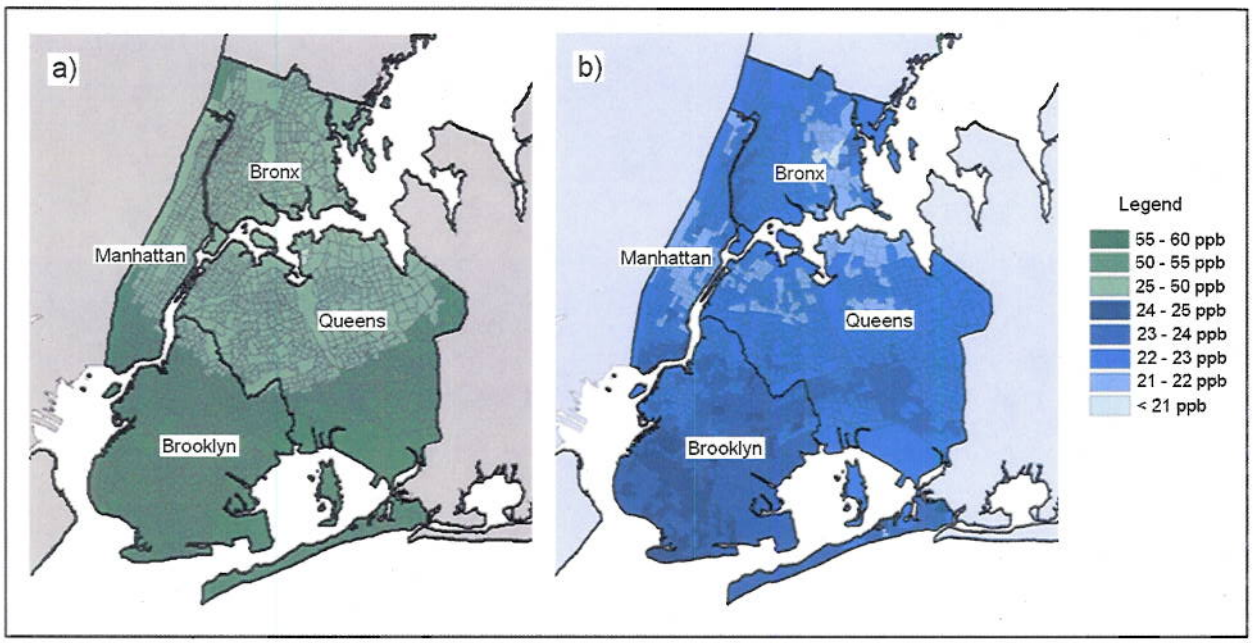

Figure 1. Spatial maps of interpolated observations of ambient ozone concentrations (a) and modeled personal exposures (b) in the greater New York City metropolitan area.

The 8-hour maximum daily averaged ozone concentrations were calculated from the hourly measurements archived at the Environmental Protection Agency's (EPA) Air Quality System (AQS) database (http://www.epa.gov/oar/data/aqsdb.html). Available ozone observations (a total of 6 stations) were interpolated to 1 km horizontal grid resolution and then averaged for each of the four counties (Bronx, New York, Kings and Queens). In addition, 8-hr maximum daily averaged ozone concentrations were calculated from the hourly concentration values simulated by the CMAQ model, version 4.5 (Byun and Schere, 2006; Appel et al., 2007). The CMAQ estimates were used to generate a combined observed and modeled surface using a multiplicative adjusted bias approach (Garcia et al., 2010) referred to as "bias-corrected CMAQ" throughout the rest of the paper. Finally, output from the SHEDS model was used to estimate individual exposure by accounting for infiltration of pollutants into buildings and daily activity patterns (Burke et al, 2001).

Hospital admission information was obtained from the NYS Department of Health Statewide Planning & Research Cooperative (SPARCS), which collects inpatient information for all NYS hospitals, excluding psychiatric and federal hospitals. SPARCS is a legislatively mandated discharge database that is known to include at least 95% of acute care hospitalizations.

136 Respiratory diseases were based on the International Classification of Diseases, 9th Revision
137 (ICD-9 code; US Department of Health and Human Services, 1991), and included: asthma
138 (ICD-9 code 493), chronic bronchitis (491), emphysema (492), and chronic obstructive
139 pulmonary disease (COPD; 496). There were a total of 1,840 daily respiratory-related hospital
140 admissions across the four NYC counties during the 460 days of the study time period.

141 A Generalized Additive Model (GAM; Wood, 2010) was used to investigate potential
142 associations between ozone and respiratory-related hospital admissions in the greater NYC
143 metropolitan area. The model relates number of hospital admissions to the 3-day moving
144 average of the daily 8-hr maximum ozone concentration for each day. This time-series model
145 accounts for inter- and intra-annual variability as well as holidays and weekend/weekday effects.
146 Other indicator variables include average maximum temperature and average dew point. The
147 GAM model was run iteratively, with one of the three definitions of exposure applied in the
148 model as the main health effects variable each time.

149 Using the GAM described above, relative risk was calculated for same-day, lag 1 day, lag
150 2 days, lag 3 days and 3-day SMA for each exposure metric (averaged observations, bias-
151 corrected CMAQ, and exposure model output). All other variables in the GAM remained the
152 same. In addition, a sensitivity analysis was conducted by randomly selecting and removing data
153 points and then examining the repeatability of the results to determine the stability of the
154 epidemiology model. In addition, residuals were examined for collinearity and autocorrelation.

155 **4. RESULTS AND DISCUSSION**

156 The refined exposure estimates yielded a relative risk greater in magnitude and with
157 greater precision (smaller confidence intervals) (Fig. 2) from just under 2% for observations, to
158 3% (bias-corrected CMAQ) and over 2% (exposure model output). Additionally, the lag 1 day,
159 lag 2 days and 3-day SMA all produce confidence intervals that exclude 1.0 whereas using
160 observations alone produced confidence intervals that included 1.0. This result is consistent
161 with the findings reported in Garcia et al., 2010 that the air quality model provides additional
162 information with regard to spatial texture, which may account for the greater precision. Jerrett et
163 al. (2005) also demonstrate the intra-urban variability and improved estimates from using

approaches that consider infiltration and time in microenvironments. The results of this study are consistent with these previous studies in that the estimated health effect measure was greater in magnitude for both the biased-corrected CMAQ and exposure model metrics as compared to observations alone.

In examining the results above, it is important to consider several issues relevant to the inference of the health effect measure apart from the relative risk and confidence intervals and how these values change between approaches. Such issues include misclassification and selection bias, differing scales and intervals at which health and air pollution data are available and aggregated, sampling bias due to data limitations, and confounding that exists between the model variables and the health outcome. In this study, selection bias (association between exposure and disease differs for those who participate and those who do not) is moderated by the study design since the entire NYC population is part of the study. The impact of exposure misclassification (also known as information bias or Berkson error) on the inference of the health effect measure is determined by whether the error is differential (bias is different for diseased and non-diseased groups) and nondifferential (bias is unrelated to disease occurrence). In this study, any information bias would likely be nondifferential because the error associated with the main health effect (ozone) should be the same in both the respiratory illness and healthy groups. Thus, any bias would likely attenuate the findings toward the null making the study results conservative (under-estimation of effect).

Confounding is an issue in this study as temperature is correlated with ozone and with the health outcome. For this study, Garcia et al. 2010 provides evidence that temperature alone cannot account for the association seen between ozone and respiratory-related hospital admissions in the New York City metropolitan area. Other analyses examining the homogeneity of the health effect modifier for temperature versus ozone concentrations indicated that after accounting for temperature, the variability remaining was sufficiently explained by ozone for two groups stratify by the mean concentration. Finally, the large sample size of the study (n = 1,840) minimizes the existence of sampling bias. Thus, the change in the health effect measure resulting from the application of the more refined exposure estimates can be attributed to the improvement in matching the daily change in exposures with the appropriate population.

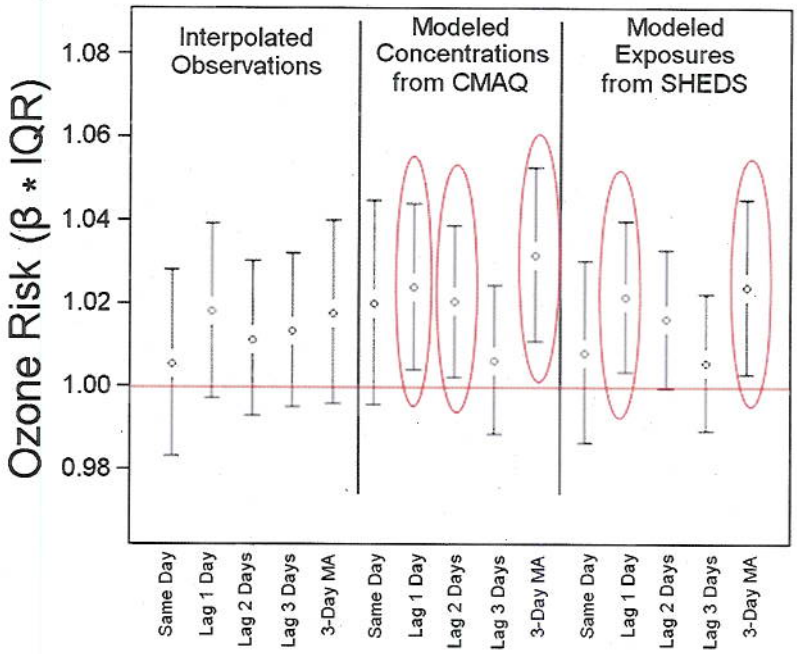

Figure 2. Risk and 95th percentile confidence intervals calculated by applying averaged observations, bias-adjusted CMAQ predictions, and exposure model output in the GAM. Coefficients are multiplied by the inter-quartile range to account for the varying distributions of the three datasets. Ovals highlight significant findings.

5. SUMMARY

Since environmental managers and health scientists rely on the ambient concentrations in managing air quality, it is important that we have a better understanding of human exposure to air contaminants to adequately protect human health. Improving characterization of air pollution exposures involves novel approaches using measurements and modeling tools to enhance estimates of ambient air concentrations and to gain a better understanding of the personal-ambient relationships. These approaches include statistical interpolation techniques, passive monitoring methods, and spatio-temporal models which produce more resolved estimates of ambient air concentrations. In addition, models such as the Community Multi-Scale Air Quality (CMAQ) model and hybrid modeling approaches provide new alternatives for characterizing ambient air concentrations. Finally, creative use of publically available data and the application of personal exposure modeling tools, such as SHEDS, alone and in combination with other air quality models provide opportunities to better define and predict personal-ambient associations. Many of these exposure characterization approaches are currently being applied and evaluated in several epidemiological investigations supported by the U.S. EPA.

In this study, we used advanced techniques to characterize exposure in an epidemiologic analysis conducted in New York. The study investigated the association between ozone exposure and respiratory-related hospital admissions for five summers (June – August; 2001 – 2005) in the NYC metropolitan area using various definitions of exposure: concentration surfaces derived form monitoring data; 2) bias-adjusted modeled concentrations from CMAQ; and 3) modeled exposure estimates from the SHEDS model. The results of this study support that refined exposure estimates (estimates supplemented with air quality and exposure modeling data) result in more precise effect estimates than those produced when using observations alone. Regardless of these findings, however, care needs to be taken when using only the health effect measure as the indicator for improving exposure estimates. A full range of metrics need to be identified for gauging whether refined exposure estimates are truly improving our ability to discern a health signal or whether we are simply introducing uncertainty and bias. Such metrics include examining misclassification (including colinearity and autocorrelation), confounding and sampling bias.

REFERENCES

Appel, K. W., A. B. Gilliland, G. Sarwar, and R. C. Gilliam (2007), Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance: Part I--Ozone, Atmospheric Environment, 41(40), 9603-9615.

Burke, J., M. Zufall, and H. ¨Ozkaynak (2001). A population exposure model for particulate matter: Case study results for PM2.5 in Philadelphia, PA. *Journal of Exposure Analysis and Environmental Epidemiology* 11, 470 – 489.

Byun, D. W., and K. L. Schere (2006), Review of the governing equations, computational algorithms, and other componenets of the MOdels-3 Community Multiscale Air Qualtiy (CMAQ) modeling system, Applied Mechanics Reviews, 59, 51-77.

Garcia, V., K. Foley, E. Gego, D. Holland, and S.T. Rao (2010), A comparison of statistical techniques for combining modeled and observed concentrations to create high-resolution air quality surfaces. Air & Waste Manage. Assoc. 60: 586-595.

Wood, S. (2010), Package 'mgcv' (v. 1.6-2): GAMs with GCV/AIC/REML smoothness estimation and GAMMs by PQLRep. 19:09:58.

Isakov, V., J. Touma, J. Burke, D. Lobdell, T. Palma, A., Rosenbaum, H., Özkaynak (2009). Combining Regional and Local Scale Air Quality Models with Exposure Models for Use in Environmental Health Studies. *J. A&WMA*. 59: 461-472.

US Department of Health and Human Services, International Classification of Diseases, 9th Revision (ICD-9), (1991).

Jerrett, M., Burnett, R.T., Ma, R., Pope, C.A., 3rd, Krewski, D., Newbold, K.B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E.E., Thun, M.J. (2005). Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology*, 16, 727-736.

Sarnat, S.E., Klein, M, Peel, J.L., Mulholland, J., Sarnat, J.A., Flanders, W.D., Waller, L.A., and Tolbert, P.E. (2006). Spatial considerations in a study of ambient air pollution and cardiorespiratory emergency department visits. *Epidemiology*; 17(6) Suppl:S242-S243.

Özkaynak, H.; Palma, T.; Touma, J.; Thurman, J. (2008). Modeling population exposures to outdoor sources of hazardous air pollutants. *Journal of Exposure Science and Environmental Epidemiology*, 8, 45-58.