

The importance of the exposure metric in air pollution epidemiology studies: When does it matter, and why?

Kathie L. Dionisio¹ · Lisa K. Baxter¹ · Janet Burke¹ · Halûk Özkaynak¹

Received: 16 January 2015 / Accepted: 10 June 2015 / Published online: 27 June 2015
© Springer Science+Business Media Dordrecht (outside the USA) 2015

Abstract Exposure error in ambient air pollution epidemiologic studies may introduce bias and/or attenuation of the health risk estimate, reduce statistical significance, and lower statistical power. Alternative exposure metrics are increasingly being used in place of central-site measurements, with the intent of reducing exposure error. Dependent on the study design, health outcome, and pollutant of interest, these metrics may provide a means of reducing error (leading to less bias and uncertainty in health risk estimates) if they capture variability in exposure which is not represented when central-site measurements are used. We examine the current evidence for answering the question of when the choice of exposure metric matters and why, focusing on studies which examined multiple exposure metrics in the same epidemiologic study. We conclude that for time-series and case-crossover studies, central-site measurements may be sufficient, especially for homogeneous pollutants, and in cohort and panel studies, approaches that increase spatial resolution of the exposure metrics do impact the health effect estimates. We note that while the current literature is widely dispersed across exposure metrics and health outcomes, the largest collective, common body of literature is focused on birth/pregnancy outcomes and traffic-related pollution. Also additional discussion and agreement is needed on how to classify metrics as “different” and “better.” Primary recommendations are to provide context for the reasons behind selection of exposure metrics and to encourage collaboration between exposure scientists and epidemiologists when designing and

implementing a study, as results can have important implications for the development of policies and regulations.

Keywords Air quality models · Exposure error · Exposure metric · Exposure models · Uncertainty · Epidemiology · Air pollution

Abbreviations

AER	Air exchange rate
APEX	Air Pollutants Exposure Model
CI	Confidence interval
CS	Central-site
EPA	Environmental Protection Agency
IDW	Inverse distance weighted
LBW	Low birth weight
LUR	Land-use regression
MESA	Multi-Ethnic Study of Atherosclerosis and Air Pollution
NAAQS	National Ambient Air Quality Standards
OR	Odds ratio
SHEDS	Stochastic Human Exposure and Dose Simulation Model
TEOM	Tapered element oscillating microbalance

Introduction

When central-site (CS) measurements are used as the exposure estimate in epidemiologic studies evaluating the health effects of exposure to ambient air pollution, the magnitude and impact of exposure error may vary. CS measurements may capture temporal variability well for most pollutants (dependent on the spatial and temporal scale of interest), as well as

✉ Kathie L. Dionisio
kathie.dionisio@gmail.com

¹ National Exposure Research Laboratory, U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, USA

spatial variability for homogeneous pollutants (e.g., PM_{2.5}, SO₄). However, CS measurements do not adequately capture the spatial variability for pollutants with local sources (e.g., NO₂, CO, EC) (Dionisio et al. 2013). Exposure error may be introduced when CS measurements do not reflect the spatial and/or temporal variability of ambient air pollutant concentrations and their relationships to true personal exposures in the study area (Dionisio et al. 2014). Exposure error may introduce bias and could lead to attenuation of the health risk estimate, reduce statistical significance, and lower statistical power (Goldman et al. 2010, 2011, 2012; Zeger et al. 2000). Further, the magnitude and resultant impact of exposure measurement error on health risk estimates vary by pollutant, and between single and multipollutant models (Dionisio et al. 2014; Goldman et al. 2010; Tolbert et al. 2007; Zeger et al. 2000). Many recent epidemiologic studies on the health effects of exposure to ambient air pollution move beyond CS measurements and incorporate alternative exposure estimates intended to reduce exposure error and/or to increase study power.

Alternative approaches for exposure estimation in air pollution epidemiologic studies have included GIS-based metrics to capture spatial variability (traffic metrics, land-use regression (LUR) models), satellite data and statistical techniques that combine measurements with air quality modeling of air concentrations using emissions and meteorological data, and human exposure models that also incorporate human activity data and infiltration of air pollutants indoors (Özkaynak et al. 2013). It is important to note that different metrics are often designed with different goals in mind. For example, CS measurements may approximate exposure to outdoor ambient pollution, while human exposure models may be used to approximate personal exposure to ambient generated pollution by accounting for pollutant infiltration indoors, or total personal exposure if exposure to indoor sources is also included in the model. Selecting an appropriate exposure estimation approach for a particular study may not be straightforward, or may require additional expertise not readily available, since the variety of alternative exposure metrics currently being used makes it hard to compare results across studies in a meaningful way, even for the same pollutant and health outcome. The latter also poses a challenge to the U.S. Environmental Protection Agency (EPA) when interpreting and synthesizing published literature during the periodic regulatory review of the science upon which National Ambient Air Quality Standards (NAAQS) are set, as well as in the emerging literature on accountability research which examines the causal relationship between air quality regulations and ensuring changes in health outcomes (Moore et al. 2008, 2010; Zigler et al. 2012).

Ideally, direct comparisons of alternative exposure metrics would be conducted within the epidemiologic studies themselves. However, few have implemented multiple exposure estimation approaches and reported the resultant health effect

estimates. In addition, the studies that have evaluated multiple exposure metrics have reported mixed findings, as the benefit of using a particular alternative exposure metric depends on the pollutant and health effect relationships being studied, the most appropriate type of exposure metric (e.g., a metric which evaluates ambient-generated exposures versus total personal exposure), and on the various study designs considered. Results from a group of studies investigating the impact of alternative exposure estimation approaches on health risk estimates associated with short-term exposure studies were summarized recently (Baxter et al. 2013b). A survey of more recently available literature revealed a relatively small number of additional published studies beyond those presented earlier. In this manuscript, we examine what evidence the available literature provides for answering the important question of when the choice of exposure metric matters and why. Here, we primarily focus on studies which examined multiple exposure estimates approximating exposure to ambient generated pollution in the same study and present our thoughts on areas where more research is needed.

Current evidence: when does it matter and why?

Study design

Study design has important implications for the interpretation of the results associated with the particular exposure metric(s) considered in epidemiologic studies. Alternative exposure metrics which reflect an increased level of spatial and/or temporal refinement compared to CS exposure metrics may reduce exposure error and thus increase our confidence in the resultant health risk estimate, dependent on the study design. We evaluate exposure metrics to determine in what instances (e.g., for a certain type of epidemiological data set) a more suitable exposure metric could be chosen which may result in more reliable and/or precise health effect estimates. In this regard, we discuss two aspects of study design—statistical analysis and spatial design considerations.

Case-crossover and time-series studies (single city)

Case-crossover and time-series studies by design utilize temporal contrasts in exposure. For these acute effects studies, the goal is to assess population-weighted average exposure over time. Exposure variability may or may not be captured well by a CS ambient monitor depending on a number of factors, some of which may also be time-varying (Baxter et al. 2013a; Dionisio et al. 2013). However, when temporal variability in exposure is captured well by the CS monitor, refinements in exposure estimation may have little impact for health risks obtained from these study designs. In addition, for a time-series study health outcome is often aggregated at the county level,

thus rendering the increased spatial resolution of the exposure estimate of little use. Nevertheless, analysis by SE Sarnat et al. in a time-series study at the finer ZIP code level showed some impact on effect estimates for cardiovascular and respiratory outcomes from exposure to traffic-related pollutants (EC, NO_x, CO) when dispersion/statistical models were used compared to CS measurements (Baxter et al. 2013b; Sarnat et al. 2013b). A case-crossover study of exposure to PM_{2.5}, using estimates from ambient monitors and from a semiempirical air exchange rate (AER)/mass balance model, showed little to no difference in odds ratios (ORs) or in confidence intervals (CIs) for transmural myocardial infarction outcomes (Hodas et al. 2013). However, this same study, in addition to a second case-crossover study (Jones et al. 2013) of respiratory hospitalizations in New York City, found evidence of effect modification by classifications of AER categories, which serve as indicators of infiltration rate, suggesting spatial consideration of housing factors could modify the health effects estimates for different population sub-groups.

Though the weight of evidence for time-series and case-crossover studies using a variety of statistical, dispersion modeling, and satellite-derived exposure estimates is limited and varied, typically little impact would be expected from the use of these more refined exposure estimation approaches as opposed to using CS-based measurement data, especially for pollutants that are distributed homogeneously (e.g., PM_{2.5}, O₃) as the CS measurements often adequately capture temporal variability of pollutant concentrations. This is particularly true when conventional time-series or case-crossover statistical methods are applied using aggregate health data. In addition, in the case-crossover design where each household is used as its own control and within-season AERs are less variable, effect modification because of AER or other enhanced exposure predictors is not likely to be detected unless prior stratification of the data is performed on key household, behavioral or locational factors (Hodas et al. 2013; Jones et al. 2013; Sarnat et al. 2013a).

While some factors that are not temporally varying (e.g., location of a subject's home) are likely to be less important for time-series studies (Hodas et al. 2013; Jones et al. 2013; Mannshardt et al. 2013; Sarnat et al. 2013b), other time-varying factors must still be taken into consideration. While certain time-varying factors such as behavioral changes due to weather are accounted for in human exposure models, these changes will only be reflected when human exposure models are used to obtain exposure estimates, and may not be as influential on exposure estimates as air concentration changes due to greater fluctuations in meteorology. In addition, the potential for correlation between behavioral changes and air concentration changes must be considered. We conclude that for time-series and case-crossover studies, the use of CS measurements may be appropriate for homogeneously distributed pollutants over the study areas.

When human exposure modeling is used to obtain population exposure estimates, we gain the added benefit of increasing the between-individual variability of exposure to ambient pollution by factoring demographic differences, time-activity patterns, spatial gradients in air pollution, and AERs into exposure estimates. While these added model inputs may increase variability in the exposure estimates, they may also introduce additional uncertainty, depending on the overall applicability and accuracy of exposure model inputs or parameters. This leads to the outstanding question of whether the larger range of exposures predicted from more complex exposure models reflects the true exposure variability, thereby reducing exposure error.

There exist a number of time-series and case-control studies utilizing exposure modeling estimates, based on two EPA exposure models, the Stochastic Human Exposure and Dose Simulation (SHEDS) model and the Air Pollutants Exposure (APEX) model. Other human exposure models have not been explored in this context. Studies using SHEDS or APEX considered a variety of cardiovascular and respiratory outcomes, and multiple pollutants (PM_{2.5}, O₃, EC, SO₄, NO_x, CO) and reported results varying depending on the pollutant chosen. Health risk CIs utilizing human exposure model estimates which include only exposures to ambient generated pollution decreased with mixed impact on the health risk estimates compared to those obtained using ambient monitor measurements (Mannshardt et al. 2013; Sarnat et al. 2013b). However, posterior intervals increased in one study (Chang et al. 2012), and no change was seen in either the health risk estimate or the CIs in two studies (Hodas et al. 2013; Jones et al. 2013). The smaller confidence intervals are potentially due to less exposure error, which decreases the uncertainty in the health risk estimates and increases the chances for detecting their statistical significance. Though human exposure models may produce exposure estimates with increased variability, more evidence is needed to conclude whether the increased variability in the exposure estimates produces enough change in the resultant health risk estimate to overcome the additional uncertainty that may be present when human exposure modeling is used, as compared to results when often scarce measurement data alone are used in place of the known variations in human exposures to ambient pollution.

Cohort studies of pregnancy and birth outcomes

While typically cohort studies are long-term, prospective studies, and panel studies are often short-term in duration, cross-sectional studies, both are driven by temporal and spatial contrasts in exposure. Therefore, we expect exposure estimates which include refinements on the spatial scale to potentially influence health risk estimates when used in these types of epidemiologic studies. A number of such studies comparing health risk estimates obtained using different exposure metrics

were found. However, these studies were limited by the narrow range of health outcomes evaluated as all but one study were related to pregnancy complications or birth outcomes (preeclampsia, preterm birth, very preterm birth, small for gestational age, term low birth weight (LBW)). Making these studies even more difficult to evaluate is the fact that evidence is not clear on which exposure window is most appropriate to use in evaluating these outcomes, and studies still use a variety of different exposure windows. In addition, the majority of studies focused on local/traffic-related pollutants (CO, NO₂, NO, NO_x, O₃, PM_{2.5}, PM₁₀, BC, SO₂), adding to the limited nature of the evidence, and the conclusions that can be drawn from this set of studies. These studies predominantly used LUR and traffic-related metrics as their estimate of exposure, though a few used dispersion modeling and other types of models.

When comparing health risk estimates obtained from LUR estimates to CS measurements, we generally see decreased CIs for the health risk estimate when using LUR, though results sometimes differ by pollutant or study location (Basagana et al. 2012; Wang et al. 2014). Though mixed effects were seen for the impact on the OR, the CIs were smaller when LUR estimates were used compared to when inverse distance weighted (IDW) or direct CS measurements were used (Brauer et al. 2008; Laurent et al. 2013; Wu et al. 2011). Notably, there is a mix of both significance and lack of significance in the reported results. As discussed in Wu et al. (2011), an important consideration is whether CIs decreased because a larger spatial area could be included in the analysis allowing for an increased sample size (hence greater statistical power), or because the spatial resolution of the exposure estimates increased (Wu et al. 2011).

Evidence shows that specific traffic-related exposure metrics may result in different impacts on health risk estimates compared to CS measurements (Brauer et al. 2008; Laurent et al. 2013; Wu et al. 2011). The minimal cost and time typically associated with developing traffic-related exposure metrics makes them an advantageous tool. Though most studies collectively used a variety of traffic indicators (e.g., traffic density in various buffers of 50–300 m surrounding the subject's home, distance to nearest main road, residence <50 m from highway), we generally see evidence of smaller health risk CIs (Wu et al. 2011), and typically smaller CIs and a change from nonsignificant to significant ORs for two different areas (Laurent et al. 2013), when traffic-related exposure metrics are used compared to CS measurements. In contrast, distance to nearest road (freeway or major road) provided more mixed results, with evidence of OR increases along with CI increases compared to IDW for all pollutants (Brauer et al. 2008), and a mix of nonsignificant and sometimes protective results compared to nearest CS monitor (Laurent et al. 2013).

A limited number of dispersion models were employed in cohort studies. Two studies investigated the use of CALINE

model predictions for their exposure estimate, with one finding a decrease in health risk CIs for CO and PM_{2.5} for CALINE compared to nearest CS monitor data; however, OR were protective and nonsignificant for term LBW, making these results difficult to interpret (Laurent et al. 2013). The other found decreases in the CI with mixed impact on the OR for NO_x and PM_{2.5} for all outcomes (preeclampsia, preterm birth, very preterm birth) when using CALINE in two adjacent geographic areas (Wu et al. 2011). A third study investigated the CALPUFF model in a cohort study exploring the impact of exposure to forest fires on respiratory and cardiovascular health outcomes and found similar results of a significant association between exposure metrics and respiratory-related outcomes when using the CALPUFF model or tapered element oscillating microbalance (TEOM) measurements of PM₁₀ as the exposure metric (Henderson et al. 2011), though it is difficult to compare these results to those previously described due to the unique and specific source of the exposure. One study compared risk estimates for various birth outcomes derived from satellite or monitor-based measurements. The study had mixed results, with both OR and CI increasing for satellite-based estimates compared to monitor measures, potentially because satellite-based measures may reduce exposure error by capturing more of the spatial variability in health and exposures compared to monitor measurements (Hyder et al. 2014). Because the available evidence encompasses a large number of different exposure metrics and health outcomes, it is difficult to draw firm conclusions. However, evidence does suggest that LUR and other traffic-related exposure metrics may reduce exposure error and increase our confidence in related health risk estimates compared to when CS measurements are used instead.

Studies of chronic health effects associated with long-term exposure

All studies discussed thus far focus on evaluating health effects associated with short-term exposures (with the exception of the studies on birth outcomes, described above). In studies of health effects associated with long-term exposure to ambient pollution, short-term (e.g., hourly) temporal refinement is not as important, and refinements in spatial resolution due to factors such as traffic sources may not be as influential when looking at average concentrations over a longer time period within a single city, or across multiple locations with similar trends in pollutant concentrations. CS monitors may capture the detail needed for these longer term measures; however, there may be some benefit to alternative exposure metrics if intra-city variations in pollutant concentrations are large enough to maintain their gradients over long-term averages (Jerrett et al. 2005; Wilson et al. 2005).

Though many studies of the chronic health effects associated with long-term exposure to air pollution (often from

large, well-known cohorts) examine the use of an alternative exposure metric in an epidemiologic study (Adar et al. 2013; Brauer et al. 2002; Jerrett et al. 2008; Puett et al. 2011; Young et al. 2014; Zmirou et al. 2004), only the seven studies found by the authors and described below examine multiple approaches to exposure assessment, and include comparison of the epidemiologic results obtained when using each exposure metric. However, many manuscripts lacked definitive conclusions about which exposure metric was most beneficial, and the wide range of health outcomes, exposure metrics, and exposure durations of interest used make it difficult to draw larger conclusions from these studies. In addition, previous work has shown that ambient concentrations of gaseous pollutants may be more strongly associated with personal exposure to $PM_{2.5}$ than with their respective personal exposures, and that these associations may differ by location, which may also contribute to the lack of conclusive results (Koutrakis et al. 2005). It is also important to note that the alternative exposure metrics used in these studies aimed to increase spatial resolution but did not incorporate the influence of infiltration or time-activity patterns on exposure. These studies provide a basis for initial evidence on the impact of alternative exposure approaches in long-term exposure studies and will perhaps lead to more conclusive evidence when combined with results of future studies.

The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air) produced three manuscripts thus far providing evidence on the impact of alternative exposure metrics in chronic health effects studies. Two studies using the MESA Air cohort examined annual average exposure to $PM_{2.5}$ mass and components, and associations with subclinical atherosclerosis (right common carotid intima-media thickness and coronary artery calcium) (Kim et al. 2013; Sun et al. 2013). Sun et al. used three exposure metrics: measurements from the nearest monitor, inverse distance weighting of monitor measures, and a city-wide average, while Kim et al. utilized city-specific spatiotemporal models, and a national spatial model. Findings in Sun et al. were generally consistent across the three approaches, and were not highly sensitive to the approach (Sun et al. 2013). Kim et al. showed that findings from the two models were similar, but were often sensitive to adjustment for city (Kim et al. 2014). Across the two studies, findings were consistent for some $PM_{2.5}$ components (though still with differences in magnitude and precision), and different for others. Kim et al. concluded that increased measurement error and lack of fine-scale spatial variability in exposure estimates may have driven the difference in findings with Sun et al., and that overall, when simpler exposure metrics are used for $PM_{2.5}$ components dominated by fine-scale spatial variability, measures of association may be inconclusive. Also originating from the MESA Air cohort, O'Neill et al. examined the association between arterial stiffness and 20-year exposure to $PM_{2.5}$ and PM_{10} , with exposure metrics calculated

from PM_{10} nearest monitor measurements, and imputed PM_{10} and $PM_{2.5}$ both derived from a space-time model (O'Neill et al. 2011). The study found that increases in long-term particle mass exposure were not associated with greater arterial stiffness; further, there was no consistent pattern across exposure metrics for the few associations which were seen.

A simulation study examining the impact of exposure to $PM_{2.5}$ on cardiovascular events presents some insight to add to evidence from the MESA Air studies described above. Kim et al. examined nearest monitor measures versus a kriging interpolation metric for exposure assessment. Findings from this study showed that when the underlying exposure distribution has a high degree of spatial dependence, both kriging and nearest monitor metrics provided reasonable health effect estimates. However, when exposures had little spatial dependence, kriging was preferable, yet gave very uncertain estimates (Kim et al. 2009).

Two studies examined the association between long-term exposure to air pollution and cancer outcomes—one for associations of ambient air pollution with lung cancer, and another for associations of traffic-related pollution with childhood cancer. Hystad et al. used satellite data to develop spatiotemporal models for all of Canada to estimate annual residential exposure to $PM_{2.5}$, NO_2 , and O_3 over 20 years. Fixed-site monitor measurements from urban centers were used to conduct a sub-analysis. Results show positive associations of lung cancer incidence with both exposure metrics, with varying degrees of magnitude and significance (Hystad et al. 2013). Heck et al. examined associations between childhood cancers and traffic-related exposures to $PM_{2.5}$ and CO, both in utero, and during the child's first year. The CALINE model, traffic density metrics, and monitor measurements were all used as exposure metrics, with weak associations found between exposure metrics and several childhood cancers (Heck et al. 2013).

Lastly, one manuscript on the Children's Health Study examined the association between traffic-related pollution and childhood asthma using four exposure metrics: NO_2 measured outside of the subjects' homes, distance from residence to nearest freeway, traffic volume on nearby roads, and model-based estimates of pollution at the home originating from nearby roads (Gauderman et al. 2005). Though overall the study found robust associations of several traffic-related exposure metrics with asthma-related outcomes, some metrics showed no association with asthma outcomes. More validation work must be done to determine the most appropriate indicator of traffic exposure to use in such a study. The authors concluded that failure to do such validation work may explain previous inconsistent findings on the association between traffic-related exposure and asthma outcomes.

As described, the seven studies the authors found which describe the application of multiple approaches to exposure assessment in studies of the health effects of long-term

exposure to air pollution cover a variety of health outcomes, exposure assessment approaches, and exposure durations. This, coupled with the lack of definitive conclusions on which exposure assessment approach is most beneficial shows that more studies in this area are needed before firm recommendations can be made.

Implications of spatial considerations: single versus multiple geographic areas

Researchers have commonly conducted time-series or case-crossover studies in single cities, counties or regions. Often a single daily exposure value (CS-based measurements or more refined exposure metrics) is used to represent daily aggregate exposure for the study area. When multi-geographic epidemiologic studies are conducted, separate exposure estimates may represent different census tracts, cities, counties, or ZIP codes. The statistical methodologies applied when CS measurements are used are more prone to estimation biases and/or higher imprecision in these multi-geographic epidemiologic applications, if the spatial distribution of pollutants or household characteristics (e.g., AER, infiltration rates, other exposure predictors) vary between the study sites and exhibit a different pattern from the ambient pollution concentrations over time and space. These types of studies often benefit from using more refined exposure metrics than the traditionally used ambient monitoring data (Baxter et al. 2013b; Dai et al. 2014; Hodas et al. 2013; Özkaynak et al. 2013).

Challenges in comparing exposure metrics

In evaluating the impact of the use of alternative exposure metrics on epidemiologic study results, one must answer the question of how to compare health risk estimates obtained using alternative exposure metrics. A “better” exposure estimate is not necessarily one that increases the magnitude of the health effect estimate or decreases the CI of the risk estimate. We prefer considering a better exposure estimate as the one which includes less exposure error and thus increases our confidence in the health risk estimate. Thus, there is a need for validation studies to determine if and when alternative exposure metrics are truly reducing or eliminating exposure error. We must also consider the various components of error introduced in exposure models, namely the “Berkson-like” component and the “classical-like” components of error (Szpiro et al. 2011). Studies have shown that error in air pollution epidemiology studies is due to a mixture of these two types of errors, which may introduce both bias and a loss of precision in results (Cefalu and Dominici 2014; Kioumourtoglou et al. 2014; Sheppard et al. 2012; Szpiro et al. 2011). While these methodological studies address theoretical properties and analyze performance through

simulation studies, there is a need for these methods to be applied in actual health studies. Currently, the wide range of results makes drawing definitive conclusions about the impact of an alternative exposure metric for a specific epidemiological study difficult.

Further, we must also decide what is “different” when comparing health risk estimates. With risk estimates often presented on different scales, obtained using metrics assessing different types of exposure, or presented for slightly different health outcomes, it is difficult to assign an absolute measure of relevant difference between the alternative exposure metrics considered. Further complicating matters, there is currently no consensus in the field on how to measure or classify relevant “difference.” For instance, should we consider the impact of alternative exposure metrics on the effect estimates directly, or in conjunction with their estimate of significance level/CI widths? Also, how does one determine if there is a difference in CI widths? We note that risk estimates with overlapping CIs may not be considered statistically different; however, they may still be desirable in this type of evaluation if one set of CIs are smaller in width than the other. Further, though comparison of confidence interval widths can provide some empirical insight, without knowing what the “true” exposure is in a given study, it is impossible to assess true bias. In summary, the available published literature which has included comparisons of alternative exposure metrics has produced a broad range of results, making it difficult to draw definitive conclusions for most epidemiologic study scenarios about the expected impact of applying an alternative exposure metric.

Implications for policy or regulatory decision making

The U.S. EPA has the responsibility for setting NAAQS for pollutants emitted from numerous and diverse sources considered harmful to public health and the environment. Periodically, EPA reviews the science upon which the NAAQS are based and the standards themselves. For many of the NAAQS reviews, epidemiologic evidence plays a key role. Because of the variations in the multitude of factors controlling air pollution health effects, it is important for these epidemiologic studies to understand the spatial and temporal distributions of the pollutants being evaluated and the complex interactions affecting these air pollutants. This understanding can reduce uncertainties in the epidemiologic analysis and provide further insight into interpreting the results. This can become especially important in better understanding the associations between any observed health effects and exposure to lower concentrations of these air pollutants, particularly for at risk populations that may not be included in clinical study analyses.

Conclusions and recommendations

Alternative approaches to exposure estimation for use in studies of the health effects of exposure to air pollution have the potential to reduce exposure error. However, studies summarized here show that current evidence may provide case-by-case conclusions, but the ability to draw general conclusions on the impact of alternative exposure metrics on resultant health risk estimates is still limited. We describe evidence showing that for time-series and case-crossover studies, CS measurements (versus statistical or dispersion models, or satellite-derived estimates) may be an adequate exposure metric for pollutants distributed homogeneously over the study area. More evidence is needed to conclude whether increased variability in estimates of exposure to ambient pollution from human exposure models produces enough change in the resultant health risk estimate to overcome the additional uncertainty introduced. Evidence for cohort and panel studies in this area is focused on pregnancy and birth outcomes, and evidence does suggest that LUR and other traffic-related exposure metrics reduce exposure error and increase confidence in related health risk estimates compared to results when CS measurements are used. Lastly, evidence from long-term studies of the chronic effects of exposure to air pollution lacks definitive conclusions due to the limited number of studies, broad range of health outcomes, exposure metrics, and exposure durations of interest.

The ability to evaluate and compare exposure metrics is becoming increasingly important as new exposure estimation methods (e.g., satellite-based measures) continue to be developed. Air pollution epidemiologic studies must build on existing evidence regarding the use of alternative exposure metrics to ensure the most efficient use of and derive greatest benefit from increasingly limited resources available for research projects. We suggest that in addition to analyzing and publishing results of epidemiologic studies with multiple exposure estimates applied separately to the same health data set, future effort should be put into providing context for the selection of the exposure metrics themselves, including details of why such exposure metrics were chosen, and what additional information or benefit they provide. In addition, collaboration between exposure scientists and epidemiologists should be encouraged from the beginning to the end of the study (Cefalu and Dominici 2014), so that exposure scientists can provide input on the feasibility of using various exposure metrics (e.g., indicators of AER or indoor infiltration rates derived using mechanistic or semiempirical methods), while at the same time epidemiologists can provide insight on the types of sources of variability most important to the epidemiologic analysis chosen. The ability to effectively evaluate exposure error and evaluate a larger body of literature on this topic may allow us to draw more definitive conclusions about which classes of exposure metrics perform better for specific

study designs, pollutants, or health outcomes. Ultimately, we may be able to improve utility of our analysis if we can make more educated a priori decisions on which exposure metrics are most beneficial in comparison to the traditional use of CS ambient monitoring data and the best ways to utilize available CS measurements in air pollution epidemiology studies.

Acknowledgments The authors would like to acknowledge Beth Hassett-Sipple for her feedback on the manuscript.

Conflict of interest Although this work was reviewed by EPA and approved for publication, it may not necessarily reflect official Agency policy. EPA does not endorse the purchase of any commercial products or services mentioned in this publication.

References

- Adar SD et al (2013) Fine particulate air pollution and the progression of carotid intima-medial thickness: a prospective cohort study from the Multi-Ethnic Study of Atherosclerosis and Air Pollution. *PLoS Med* 10:1–9
- Basagana X et al (2012) Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmos Environ* 54:634–642
- Baxter LK et al (2013a) Influence of human activity patterns, particle composition, and residential air exchange rates on modeled distributions of PM_{2.5} exposure compared with central-site monitoring data. *J Expo Sci Environ Epidemiol* 23:241–247
- Baxter LK et al (2013b) Exposure prediction approaches used in air pollution epidemiology studies: Key findings and future recommendations. *J Expo Sci Environ Epidemiol* 23:654–659
- Brauer M et al (2002) Air pollution from traffic and the development of respiratory infections and asthmatic and allergic symptoms in children. *Am J Respir Crit Care Med* 166:1092–1098
- Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C (2008) A cohort study of traffic-related air pollution impacts on birth outcomes. *Environ Health Perspect* 116:680–686
- Cefalu M, Dominici F (2014) Does exposure prediction bias health-effect estimation? The relationship between confounding adjustment and exposure prediction. *Epidemiology* 25:583–590
- Chang HH, Fuentes M, Frey HC (2012) Time series analysis of personal exposure to ambient air pollution and mortality using an exposure simulator. *J Expo Sci Environ Epidemiol* 22:483–488
- Dai L, Zanobetti A, Koutrakis P, Schwartz JD (2014) Associations of fine particulate matter species with mortality in the united states: a multicity time-series. *Anal Environ Health Perspect* 122:837–842
- Dionisio KL et al (2013) Development and evaluation of alternative approaches for exposure assessment of multiple air pollutants in Atlanta, Georgia. *J Expo Sci Environ Epidemiol* 23:581–592
- Dionisio KL, Baxter LK, Chang HH (2014) An empirical assessment of exposure measurement error and effect attenuation in bipollutant epidemiologic models. *Environ Health Perspect*. doi:10.1289/ehp.130772
- Gauderman WJ, Avol E, Lurmann F, Kuenzli N, Gilliland F, Peters J, McConnell R (2005) Childhood asthma and exposure to traffic and nitrogen dioxide. *Epidemiology* 16:737–743
- Goldman GT et al (2010) Ambient air pollutant measurement error: characterization and impacts in a time-series epidemiologic study in Atlanta. *Environ Sci Technol* 44:7692–7698
- Goldman GT, Mulholland JA, Russell AG, Strickland MJ, Klein M, Waller LA, Tolbert PE (2011) Impact of exposure measurement

- error in air pollution epidemiology: effect of error type in time-series studies. *Environ Health* 10:1–11
- Goldman GT, Mulholland JA, Russell AG, Gass K, Strickland MJ, Tolbert PE (2012) Characterization of ambient air pollution measurement error in a time-series health study using a geostatistical simulation approach. *Atmos Environ* 57:101–108
- Heck JE et al (2013) Childhood cancer and traffic-related air pollution exposure in pregnancy and early life. *Environ Health Perspect* 121:1385–1391
- Henderson SB, Brauer M, MacNab YC, Kennedy SM (2011) Three measures of forest fire smoke exposure and their associations with respiratory and cardiovascular health outcomes in a population-based cohort. *Environ Health Perspect* 119:1266–1271
- Hodas N et al (2013) Refined ambient PM_{2.5} exposure surrogates and the risk of myocardial infarction. *J Expo Sci Environ Epidemiol* 23:573–580
- Hyder A, Lee HJ, Ebisu K, Koutrakis P, Belanger K, Bell ML (2014) PM_{2.5} exposure and birth outcomes: use of satellite- and monitor-based data. *Epidemiology* 25:58–67
- Hystad P, Demers PA, Johnson KC, Carpiano RM, Brauer M (2013) Long-term residential exposure to air pollution and lung cancer risk. *Epidemiology* 24:762–772
- Jerrett M et al (2005) A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol* 15:185–204
- Jerrett M et al (2008) Traffic-related air pollution and asthma onset in children: a prospective cohort study with individual exposure measurement. *Environ Health Perspect* 116:1433–1438
- Jones RR, Özkaynak H, Nayak S, Garcia V, Hwang S-A, Linn S (2013) Associations between summertime ambient pollutants and respiratory morbidity in NYC: comparison of ambient concentrations versus predicted exposures. *J Expo Sci Environ Epidemiol* 23:616–626
- Kim S-Y, Sheppard L, Kim H (2009) Health effects of long-term air pollution: influence of exposure prediction methods. *Epidemiology* 20:442–450
- Kim S-Y, Sheppard L, Hannigan MP, Dutton SJ, Peel JL, Clark ML, Vedal S (2013) The sensitivity of health effect estimates from time-series studies to fine particulate matter component sampling schedule. *J Expo Sci Environ Epidemiol* 23:481–486
- Kim S-Y et al (2014) Individual-level concentrations of fine particulate matter chemical components and subclinical atherosclerosis: a cross-sectional analysis based on 2 advanced exposure prediction models in the multi-ethnic study of atherosclerosis. *Am J Epidemiol* 180:1–11
- Kioumourtoglou M-A et al (2014) Exposure measurement error in PM_{2.5} health effects studies: a pooled analysis of eight personal exposure validation studies. *Environ Health* 13:1–11
- Koutrakis P, Suh HH, Sarnat JA, Brown KW, Coull BA, Schwartz J (2005) Characterization of particulate and gas exposures of sensitive subpopulations living in Baltimore and Boston vol Report 131. Health Effects Institute
- Laurent O, Wu J, Li L, Chung J, Bartell S (2013) Investigating the association between birth weight and complementary air pollution metrics: a cohort study. *Environ Health* 12
- Mannshardt E, Suci K, Fuentes M, Reich B, Frey C, Jiao W (2013) Comparing exposure metrics for the effects of speciated particulate matter on population health outcomes. *J Expo Sci Environ Epidemiol* 23:627–636
- Moore K, Neugebauer R, Lurmann F, Hall J, Brajer V, Alcom S, Tager I (2008) Ambient ozone Concentrations cause increased hospitalizations for asthma in children: an 18-year study in Southern California. *Environ Health Perspect* 116:1063–1070
- Moore K, Neugebauer R, Lurmann F, Hall J, Brajer V, Alcom S, Tager I (2010) Ambient ozone concentrations and cardiac mortality in Southern California 1983–2000: application of a new marginal structural model approach. *A J Epidemiol* 171:1233–1243
- O'Neill MS et al (2011) Long-term exposure to airborne particles and arterial stiffness: the Multi-Ethnic Study of Atherosclerosis (MESA). *Environ Health Perspect* 119:844–851
- Özkaynak H, Baxter LK, Dionisio KL, Burke J (2013) Air pollution exposure prediction approaches used in air pollution epidemiology studies. *J Expo Sci Environ Epidemiol* 23:566–572
- Puett RC, Hart JE, Suh H, Mittleman M, Laden F (2011) Particulate matter exposures, mortality, and cardiovascular disease in the health professionals follow-up study. *Environ Health Perspect* 119:1130–1135
- Sarnat JA et al (2013a) Spatiotemporally-resolved air exchange rate as a modifier of acute air pollution related morbidity in Atlanta. *J Expo Sci Environ Epidemiol* 23:606–615
- Sarnat SE et al (2013b) Application of alternative spatiotemporal metrics of ambient air pollution exposure in a time-series epidemiological study in Atlanta. *J Expo Sci Environ Epidemiol* 23:593–605
- Sheppard L, Burnett RT, Szpiro AA, Kim S-Y, Jerrett M, Pope CA III, Brunekreef B (2012) Confounding and exposure measurement error in air pollution epidemiology. *Air Qual Atmos Health* 5:203–216
- Sun M et al (2013) Particulate matter components and subclinical atherosclerosis: common approaches to estimating exposure in a Multi-Ethnic Study of Atherosclerosis cross-sectional study. *Environ Health* 12:1–14
- Szpiro AA, Paciorek CJ, Sheppard L (2011) Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology* 22:680–685
- Tolbert PE, Klein M, Peel JL, Sarnat SE, Sarnat JA (2007) Multipollutant modeling issues in a study of ambient air quality and emergency department visits in Atlanta. *J Expo Sci Environ Epidemiol* 17:S29–S35
- Wang M et al (2014) Performance of multi-city land use regression models for nitrogen dioxide and fine particles. *Environ Health Perspect* 122:843–849
- Wilson JG, Kingham S, Pearce J, Sturman AP (2005) A review of intraurban variations in particulate air pollution: implications for epidemiological research. *Atmos Environ* 39:6444–6462
- Wu J, Wilhelm M, Chung J, Ritz B (2011) Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. *Environ Res* 111:685–692
- Young MT, Sandler DP, DeRoo LA, Vedal S, Kaufman JD, London SJ (2014) Ambient air pollution exposure and incident adult asthma in a nationwide cohort of U.S. Women. *Am J Respir Crit Care Med* 190:914–921
- Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, Cohen A (2000) Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect* 108:419–426
- Zigler CM, Dominici F, Wang Y (2012) Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics* 13:289–302
- Zmirou D et al (2004) Traffic related air pollution and incidence of childhood asthma: results of the Vesta case-control study. *J Epidemiol Community Health* 58:18–23