Critical Reviews in Toxicology

ISSN: 1040-8444 (Print) 1547-6898 (Online) Journal homepage: http://www.tandfonline.com/loi/itxc20

# Measurement error in environmental epidemiology and the shape of exposure-response curves

Lorenz R. Rhomberg, Juhi K. Chandalia, Christopher M. Long & Julie E. Goodman

**To cite this article:** Lorenz R. Rhomberg, Juhi K. Chandalia, Christopher M. Long & Julie E. Goodman (2011) Measurement error in environmental epidemiology and the shape of exposure-response curves, Critical Reviews in Toxicology, 41:8, 651-671, DOI: 10.3109/10408444.2011.563420

**To link to this article:** http://dx.doi.org/10.3109/10408444.2011.563420

Published online: 08 Aug 2011.

*Critical Reviews in Toxicology*, 2011; 41(8): 651–671

ISSN 1040-8444 print/ISSN 1547-6898 online
DOI: 10.3109/10408444.2011.563420

informa
healthcare

**REVIEW ARTICLE**

# Measurement error in environmental epidemiology and the shape of exposure-response curves

Lorenz R. Rhomberg, Juhi K. Chandalia, Christopher M. Long, and Julie E. Goodman

*Gradient, Cambridge, Massachusetts, USA*



**Abstract**

Both classical and Berkson exposure measurement errors as encountered in environmental epidemiology data can result in biases in fitted exposure-response relationships that are large enough to affect the interpretation and use of the apparent exposure-response shapes in risk assessment applications. A variety of sources of potential measurement error exist in the process of estimating individual exposures to environmental contaminants, and the authors review the evaluation in the literature of the magnitudes and patterns of exposure measurement errors that prevail in actual practice. It is well known among statisticians that random errors in the values of independent variables (such as exposure in exposure-response curves) may tend to bias regression results. For increasing curves, this effect tends to flatten and apparently linearize what is in truth a steeper and perhaps more curvilinear or even threshold-bearing relationship. The degree of bias is tied to the magnitude of the measurement error in the independent variables. It has been shown that the degree of bias known to apply to actual studies is sufficient to produce a false linear result, and that although nonparametric smoothing and other error-mitigating techniques may assist in identifying a threshold, they do not guarantee detection of a threshold. The consequences of this could be great, as it could lead to a misallocation of resources towards regulations that do not offer any benefit to public health.

**Keywords:** Epidemiology, exposure, exposure-response, measurement error, risk assessment



## Contents

Abstract ..... 651
1. Introduction ..... 652
2. Independent-variable measurement errors ..... 653
3. Sources of exposure measurement error ..... 654
4. Exposure measurement error in practice ..... 657
5. Independent-variable error and regression ..... 663
   5.1. Error in linear regression ..... 663
   5.2. Error in nonlinear regression ..... 664
6. Exposure measurement error and exposure-response relationships ..... 666
7. Conclusions ..... 669
Acknowledgment ..... 670
Declaration of interest ..... 670
References ..... 670

*Address for Correspondence*: Lorenz R. Rhomberg, Gradient, 20 University Road, Cambridge, MA 02138, USA. E-mail: lrhomberg@gradientcorp.com

*(Received 21 April 2010; revised 11 February 2011; accepted 12 February 2011)*

## 1. Introduction

Levels of chemical contaminants in the environment vary geographically as a consequence of varying distances from sources, patterns of dispersion and movements of environmental media, local sources of removal or sequestration, and other factors. In natural populations exposed to such chemicals in the environment, there is widespread person-to-person variation in the levels of exposure, depending on each person's habits and lifestyle, in addition to the effect of location and change in location. Uptake rates and pharmacokinetic factors that vary person to person can lead to a variation in the internal dose of chemicals, even for people exposed to identical concentrations in environmental media.

A challenge of environmental epidemiology, which seeks to identify and characterize any health impacts of such exposures, is that there are no discrete and preidentified "dose groups" in such natural populations. The degrees of exposure of the members of a study population must be estimated by biomonitoring, measuring of the agent in environmental media, or modeling, and the estimates so obtained will vary from one subject to the next. The investigation of potential associations of these varying exposures with different health states or probabilities of adverse effects is generally carried out using some variety of statistical regression, in which the degree of exposure serves as an independent variable and the outcome measures as dependent variables. (There may be further independent variables that capture demographic groupings or other factors that may need to be adjusted for in the evaluation.)

Because the exposure measures vary on a continuous scale among subjects, the results of such regressions lead to exposure-response relationships that can be expressed as curves relating increasing exposure to its effect on the magnitude or probability of response. Such curves are of great interest because they inform the process of establishing the degree of control of environmental emissions that is needed in order to protect public health. In particular, for the so-called criteria air pollutants in the United States, under the Clean Air Act, National Ambient Air Quality Standards (NAAQS) are set, and these are heavily influenced by the degree of impact estimated from such regression approaches.

The shapes of the exposure-response curves, and in particular their shapes at lower doses, are of particular interest, because they inform the levels of restriction of exposure needed to avoid undue impacts among members of the whole population. Many such curves for the widespread environmental pollutants appear to be more-or-less linear at low doses, even in cases for which animal testing data or mode of action considerations would suggest that there ought to be an exposure threshold below which exposures could be tolerated without harm, and adverse impacts would be unlikely. The true shape of exposure-response curves, and the true nature of the ability of low-level exposure to marginally increase the magnitude or probability of health impacts, is a controversial and much debated matter.

Our concern is that the levels of individual exposure, whether estimated by measurement or modeling, inevitably have some statistical error. This results for a number of reasons: the estimates may be based on measurements at central air monitoring stations that only approximate the personal exposures of the nearby population, people move about in varying ways and hence experience individual histories of varying exposure that are not easily related to fixed-site measurements, people may have local sources of exposures in their homes and workplaces, and others. The estimates of the biologically relevant exposures for each person need not be biased—the variation of estimated from true values can be unbiased random variation about true values, though in some cases, systematic errors in estimation are possible. Most discussion of exposure estimation has focused on the positive aspect—the usefulness of surrogate measures as estimates of individual exposures—but our concern is with the consequences of the inevitable residual degree of measurement error, larger or smaller depending on the methods of estimation used, and in particular with the effect of such measurement error on the apparent shape of empirically determined dose-response curves. Our aim is to show that, even when the measurement error is unbiased, its very existence and unavoidability affects the estimation of exposure-response relationships and can, in some identifiable cases, result in alteration of the apparent shape of exposure-response relationships.

It is well known among statisticians that random, unbiased errors in the values of independent variables may tend to bias regression results (Carroll et al., 2006). For monotonically increasing convex curves (such as those describing nonlinear dose-response relationships), this effect tends to flatten and apparently linearize what is in truth a steeper and perhaps more curvilinear or even threshold-bearing relationship. The issue is not one of statistical power to detect nonlinear relationships, it is one of biases that tend to make curves estimated from data with random independent-variable measurement error look more linear than they truly are. The degree of bias is tied to the magnitude and nature of the measurement error in the independent variables. Whether this is a trivial problem or a significant one for the interpretation of exposure-response curves and their low-dose shapes is a quantitative empirical question.

In the present paper, we investigate the potential for exposure measurement errors as could be encountered in environmental epidemiology data to result in biases in fitted exposure-response relationships that are large enough to affect interpretation and use of the apparent exposure-response shapes in risk assessment applications. We consider the identity and variety of sources of potential measurement error in the process of estimating individual exposures to environmental contaminants, and we review the evaluation in the literature of

the magnitudes and patterns of exposure measurement errors that prevail in actual practice. We then review the statistical literature on the effects of independent-variable measurement error on regression and discuss the practical impact of exposure measurement error on the interpretation of the shapes of exposure-response relationships actually measured in key environmental epidemiology applications.

## 2. Independent-variable measurement errors

To determine how exposure measurement error affects the shape of a specific exposure-response curve, one must first determine what type of error it is. Exposure measurement error is often described as either classical or Berkson, and can be additive and/or multiplicative (Carroll et al., 2006). It is assumed throughout this discussion that the error is random and homoscedastic, i.e., both independent of, and unbiased with respect to the independent variable.

Classical additive error occurs when a measured value varies around the true value. The standard mathematical formulation of classical error is that a measured value equals the true value plus additive error. In the equation below, $X$ is the true independent variable, $W$ is the measured independent variable, and $U$ is the classical error:

$$W = X + U \tag{1}$$

This additive error can be due to numerous factors, including instrument imprecision, operator error, and the use of surrogates as a substitute for parameters that cannot be measured (e.g., biomarkers). Typically, classical error is assumed to have a mean of zero and constant variance, but may have non-constant variance with respect to the magnitude of the independent variable. Classical measurement error variance is assumed to be independent of the true independent variable and is thus additive to the true value variance. The variance of the true values is less than that of the measured values. Hence, classical measurement error in an independent variable can affect any model that relies on sample variance in its formulation (such as the slope in a linear regression model).

In contrast with classical error, Berkson-type error refers to the case in which measurements have lower variability than the true values.

$$X = W + U \tag{2}$$

Again, $X$ is the true independent variable, $W$ is the measured independent variable, and $U$ is the Berkson additive error, which is assumed to be independent of the measured variable $W$. The canonical case of Berkson error is a controlled experiment in which different levels of treatment (e.g., doses in an animal bioassay) are set independently of measurement and applied to groups of subjects. There is actual inter-individual variation in the resulting exposure (owing to experimental imprecision and biological or behavioral differences among subjects), but it clusters around the nominal target values and those targets constitute mean effects that groups of subjects share. In environmental epidemiology, an example of the Berkson error model applies when measurements from a centralized point, such as an air quality monitoring station, are used to estimate individual exposures (the true values) to people in the vicinity of the central monitors. In this case, the numerous factors affecting individual exposure, and hence the true values, dictate that the true exposures have higher variance than the measured (estimated) exposures, yet each group's assigned exposure, being based for all its members on the same monitor reading, have in common the central estimate of what may in fact be varying ambient-level contributions to their personal exposure. Berkson error is formulated mathematically such that the true value equals the estimated (or measured) value plus measurement error. Unlike the classical error model, under the Berkson error model, the variance of the true values equals that of the measured values plus that due to Berkson error. As with the classical error model, the Berkson error model assumes that the error is unbiased, or has a mean of zero.

In practice, measurement error may have components of both classical and Berkson errors, and may also exhibit error characteristics not defined by these two error models. For example, error could be multiplicative (Equation 3) or a mixture of classical multiplicative and additive (Equation 4; Carroll et al., 2006):

$$W = X \times U \tag{3}$$

$$W = U_1 \times X + U_2 \tag{4}$$

Again, $X$ is the true independent variable, $W$ is the measured independent variable, and $U$, $U_1$, and $U_2$ are error terms.

In all of the models described above, error parameters are often modeled as normally distributed variables, but this assumption is not always necessary. Both the additive classical and Berkson error models assume that the error is unbiased (mean zero), but in practice this is not always true. When bias in the error is present, measurement error models must account for it. Although bias in measurement error can further complicate the estimation of dose-response curves, we focus below on the simpler case that the measurement of exposures is unbiased yet nonetheless (and inevitably) imperfect and show that even unbiased error can lead to a bias in regression parameters.

The simplest example of random (i.e., unbiased) error in the independent variable causing a bias in regression parameters occurs in the linear, univariate model (more complex models are discussed in Section 5). (Zeger et al., 2000) demonstrated that random error produces a bias towards the null in the slope parameter, and we will outline their analysis below. For the linear, univariate model, they described the following equation:

$$Y = \alpha + \beta_x X + \epsilon \quad (5)$$

where $X$ is the independent variable with variance $\sigma_x^2$, $Y$ is the dependent variable, $\alpha$ and $\beta$ are regression parameters, and $\varepsilon$ is the dependent-variable error term. Error is introduced into the model when a surrogate (designated as $W$ here) is used for the independent variable $X$. As described above, for classical additive error, $W = X + U$, where $U$ is generally taken as normally distributed with mean of zero and variance, $\sigma_u^2$. The naïve estimate of $Y$ based on the measured variable $W$,

$$Y = \alpha^* + \beta_w W + \epsilon^* \quad (6)$$

produces a coefficient $\beta_w$ distinct from the actual regression parameter $\beta_x$, such that

$$\beta_w = \beta_x \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \quad (7)$$

where $\sigma_x^2$ is the variance in the true variable and $\sigma_x^2 + \sigma_u^2$ is the variance in the measured variable. The ratio of these variances is known as the regression attenuation factor, $\gamma$ (this definition of attenuation factor differs from that described below, which relates the difference between personal and ambient exposures to air pollutants):

$$\gamma = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \quad (8)$$

Because $\gamma$ is always less than 1, $\beta_w$ consistently *underestimates* $\beta_x$ for a linear model as in Equation 5. The larger the error in the measured variable $W$ (i.e., the larger the variance of $U$), the more biased the regression slope.

In contrast to classical error, unbiased Berkson error in the independent variable in univariate models leads to the measured variable, $W$, being an unbiased estimate of $X$. As a result, the regression slope parameter is unbiased because the measured variable is less error-prone than the actual variable (Zeger et al., 2000).

After a discussion of the sources of exposure measurement error and exposure measurement error in practice in Sections 3 and 4, we discuss more specific error models in the context of exposure measurement error in Section 5.

## 3. Sources of exposure measurement error

Exposure measurement error is a well-recognized limitation of environmental and occupational epidemiology studies. Exposure measurement error is commonly defined as the difference between the "measured exposure" and the "true exposure," with the potential for error being a function of the relationship between the exposure predictor of interest in a given disease model and the measurable exposure surrogate chosen as the operationally defined predictor (Sarnat et al., 2007; Heid et al., 2004). The exposure predictor of interest in a disease model (e.g., personal exposure to a pollutant) may often be several steps removed from a biologically effective dose, and this separation is typically further amplified for an exposure surrogate (e.g., community-average ambient air concentrations). As described below, exposure measurement error is inherent in any instrument used to assess exposure, including personal exposure monitors, biomonitoring tools, exposure models, and questionnaires. This section provides a brief overview of some of the common sources of exposure measurement error, highlighting in particular sources of exposure measurement error in air pollution studies.

Exposure measurement error is widely regarded as an unavoidable problem in air pollution epidemiology studies where ambient air measurements made at one or a few fixed, central-site monitoring stations are frequently used to represent personal exposures to such criteria air pollutants as fine particulate matter ($PM_{2.5}$), nitrogen dioxide ($NO_2$), and ozone ($O_3$). In particular, the National Research Council (NRC) Committee on Research Priorities for Airborne Particulate Matter identified exposure measurement error and its impacts on particulate matter (PM) health effects studies as a key research priority in its seminal report "Research Priorities for Airborne Particulate Matter: I. Immediate Priorities for Long-Range Research Portfolio" (NRC, 1998). NRC (1998) identified several sources of exposure measurement error associated with the use of central-site ambient monitoring data to represent individual exposures to a pollutant, including "errors in the accuracy and precision of the monitoring instrument; differences in exposure due to the placement of the ambient monitor (related to the zones of representation for a monitor or to the spatial homogeneity of the environmental agent measured); differences between ambient concentrations used to characterize a pollutant exposure and the average personal exposure to that pollutant or, for particulate matter, its mass or the size fractions and chemicals of biological significance; and the differences between average personal exposure levels and the exposure of a given individual."

Since the 1998 NRC report, additional studies have elaborated on the nature of the exposure measurement error for air pollution epidemiology studies, and proposed additional sources of exposure measurement error. (Zeger et al., 2000) summarized three principal components of exposure measurement error for time-series air pollution epidemiology studies relying upon central-site ambient air concentrations as exposure surrogates: (1) the difference between an individual's personal exposure and the population-average personal exposure; (2) the difference between the population-average personal exposure and the true ambient level (as influenced by differences in microenvironmental exposures, including those due to indoor source contributions); and (3) the difference between the measured and true ambient levels (as influenced by both instrument error and spatial variation). (Zeger et al., 2000) classified the first and third error components as Berkson-type errors (discussed in

detail below), but hypothesized that indoor source contributions could be associated with ambient concentrations for at least short-term periods. They proposed that the second error component could be a source of "substantial bias" to regression coefficients.

Subsequent studies have restricted their assessment of exposure measurement errors in ambient air $PM_{2.5}$ epidemiology analyses to those associated with the use of centrally monitored ambient concentrations to represent personal exposures to $PM_{2.5}$ of ambient origin only (e.g., Sheppard et al., 2005; Sarnat et al., 2009b; Avery et al., 2010a). This is consistent with the idea that it is the fraction of personal exposure to ambient-derived pollution, rather than total personal exposure, that is the relevant exposure measure for application to situations in which health effects are to be avoided by controlling ambient pollutant levels. Recent studies have thus downplayed the potential role of indoor and personal air pollutant exposures as a source of exposure measurement error for ambient pollutant exposures, citing study findings indicating that non-ambient (i.e., indoor-generated and personal-generated) source terms are not generally correlated with ambient source terms.

Although variability in non-ambient source terms is unlikely to be an important source of exposure measurement error in ambient air pollution studies (i.e., studies using central-site ambient monitoring data as an exposure surrogate for ambient pollutant exposures), it is important to emphasize that variation in indoor and personal source contributions can still be a significant source of exposure measurement error for total pollutant exposure, which ultimately is the causative variable in dose-response analyses and hence important for understanding the true nature of the dose-response relationship. In short, there is an abundance of common indoor and personal sources of both gaseous and particulate airborne species, including building materials, household furnishings, indoor combustion sources, consumer products, occupant activities (e.g., cooking, cleaning, smoking, spraying), and indoor appliances. Given that people spend approximately 90% of their time indoors, it is well accepted that total exposures to both gaseous and particulate airborne species are intricately tied to individual lifestyles and activities, with indoor and personal sources providing major and often dominant contributions (United States Environmental Protection Agency [US EPA], 2008, 2009). Variability among individual indoor and personal exposures can thus be a source of exposure measurement error affecting dose-response analyses, in addition to the error associated with person-to-person differences in ambient-source exposures.

For air pollution studies as well as other studies using measurement data as exposure surrogates, exposure measurement error can also arise due to the failure of measurements to adequately represent exposures during the etiologically relevant period. In particular, this can be an issue for exposure measurements that are grab samples reflecting concentrations at a specific location and time. For example, epidemiology studies of drinking water disinfection by-products (DBPs) typically rely upon average concentrations of drinking water disinfection by-products at a municipal treatment plant or within a distribution system as exposure surrogates (Wright and Bateson, 2005; Arbuckle et al., 2002). Given that US drinking water utilities are only required to report DBP concentrations on a quarterly basis, however, averages of these snapshot measurements may poorly reflect actual annual average DBP concentrations in water systems due to known temporal variability in DBP concentrations (Arbuckle et al., 2002). In addition, peak DBP occurrences may be entirely missed. Although there remains uncertainty regarding the critical exposure window for DBPs, it is clear that infrequent DBP grab samples may not adequately represent either chronic or short-term peak population DBP exposures.

In their review of exposure measurement error in air pollution epidemiology studies, (Sarnat et al., 2007) addressed this idea, proposing that differences between the averaging time of an exposure surrogate (i.e., the period over which ambient measurements are integrated, e.g., 24 hours for daily $PM_{2.5}$ measurements) and the biologically relevant exposure interval can introduce exposure measurement error. This is likely not an uncommon source of error given that the biologically relevant exposure intervals for particular health outcomes are often not well known, and there is thus uncertainty associated with the selection of particular averaging times to represent health outcomes in epidemiology studies. (Sarnat et al., 2007) cited the work of (Adar et al., 2007), who reported differences in patterns of exposure measured using continuous personal monitors versus daily-average central monitors among a panel of active seniors in St. Louis. Specifically, for continuous monitoring of elderly subjects for $PM_{2.5}$, black carbon, and ultrafine particle counts over 24-hour periods that included commuting events, (Adar et al., 2007) observed short-term peak pollutant exposure levels that were not evident in daily-average central-site monitoring data. (Adar et al., 2007) showed that these short-term pollution exposure events were independently associated with measures of heart rate variability, suggesting that daily-averaged $PM_{2.5}$ central-site measurements may poorly represent a critical exposure window for short-term cardiovascular health effects.

Focusing on DBP epidemiology studies, the common use of town-average DBP concentrations in municipal water samples as a surrogate for population DBP exposure is widely regarded to contribute to significant exposure measurement error (Wright and Bateson, 2005; Arbuckle et al., 2002). As discussed extensively by (Arbuckle et al., 2002) and (Wright and Bateson, 2005), other sources of exposure measurement error in DBP studies involve individual heterogeneity in water usage, temporal and spatial variability in DBP formation, and inter-individual differences in various factors that can modify individual-level DBP exposures, including point-of-use filtration or

Table 1. Examples of components of exposure measurement error for a German epidemiology study of residential radon exposure and lung cancer where indoor radon gas measurements were used to derive a surrogate exposure metric for residential radon exposure (adapted from Heid et al., 2004).[†1]

| Error component | Type |
|---|---|
| (1)Estimating average radon gas concentration in home during exposure of detector | |
| (a)Error from between-measurement variability (i.e., instrument error) | Classical |
| (b)Error from between-laboratory variability | Classical |
| (c)Error from between-detector-placement variability (i.e., detector location in a room) | Classical |
| (d)Error from between-room variability | Classical |
| (2)Using (1) to estimate average radon gas concentration in home over the year of the measurement | |
| (a)Error from between-season variability (for < one-year of measurements) | Classical |
| (a*)Error from mis-specifying seasonal correction factor (sampling error) | Classical |
| (a**)Error from assigning a group-matched correction factor for seasonal variation | Berkson |
| (3)Using (2) to estimate radon gas exposure of an individual over a certain period of years prior to measurement | |
| (a)Error from between-year variability | Classical |
| (b)Error from between-subphase variability (i.e., between radon-relevant alterations to a home) | Classical |
| (b*)Error from mis-specifying the correction factor for house phases | Classical |
| (b**)Error from applying group-matched correction factor for house phases | Berkson |
| (c)Error from between-owner variability | Classical |
| (c*)Error from mis-specifying the correction factor for previous owners | Berkson |
| (c**)Error from applying group-matched correction factor for previous owners | Classical |
| (4)Using (3) as surrogate exposure metric for residential radon exposure | |
| (a)Error from differences in room- and daytime-dependent ventilation habits | Classical |
| (b)Error from between-environment variability | Classical |
| (c)Error from between-home variability | Classical |
| (d)Error from recall of residency time | Classical |
| (e)Error from false recall of relative occupancy of bedroom | Classical |
| (f)Error from false recall of absolute occupancy of home | Classical |
| (g)Error from mis-specifying the relevant exposure-window | Classical |
| (h)Error from ignoring the absolute occupancy time of home | Berkson |

[†1] Does not include measurement error associated with the use of residential radon exposure as a proxy for the true biological predictor in the disease-causing process, namely the effective organ dose (i.e., alpha dose). Components of this measurement error include uncertainty related to the equilibrium factor for radon gas and error from between-person variability due to differences between individuals in factors that affect inhaled dose, such as inhalation rate (both classified by Heid et al. as Berkson-type errors).



boiling prior to consumption and ventilation use during showering/bathing.

As discussed by (Heid et al., 2004), exposure measurement error is also unavoidable when individual-specific measurements, rather than community-based measurements, are used to represent personal pollutant exposures. Specifically, (Heid et al., 2004) provides one of the more exhaustive reviews of sources of exposure measurement error for the case of residential radon exposure assessment. Based on a German radon study where alpha track detectors were used to measure radon gas concentrations in the bedrooms and living rooms of the current residences of study participants over approximate 1-year periods, (Heid et al., 2004) documented over 20 components of exposure measurement error associated with the use of these radon gas concentration measurements to derive a proxy measure of cumulative residential radon exposure (Table 1). These error components included between-measurement variability (i.e., instrument error), between-room variability (associated with use of radon concentration measurements in two rooms of the house as a proxy for concentrations in all other rooms), between-year variability (associated with the use of measurement data from one year to represent other years, with differences in weather and occupant habits contributing to concentration variability), and between-home variability (associated with the use of radon concentration measurements in a current residence as a proxy for concentrations in prior residences). As shown in Table 1, (Heid et al., 2004) classified most error components as classical-type errors, and only a few as Berkson-type errors. Importantly, (Heid et al., 2004) discussed how there would be additional Berkson-type error components if the predictor of interest was an effective organ dose (i.e., alpha dose) rather than the residential radon exposure, including components associated with assigning an equilibrium factor and between-person variability.

We have thus far discussed effects of measurement error of external exposures (e.g., average concentrations of drinking water disinfection by-products at a municipal treatment plant or within a distribution system) on risk estimates. Yet, to elucidate a true exposure-response relationship, one must consider the internal dose (e.g., chloroform levels in target organ issues, such as the liver). Pharmacokinetic processes can contribute to exposure

measurement error because there are generally several toxicokinetic steps separating an external exposure surrogate from an internal dose. In other words, owing to the impacts of pharmacokinetic processes (e.g., absorption, distribution, metabolism, and excretion)—and to variation among individuals in these factors—a dose estimated based on external exposure measurements will fail to account for variations in the biologically effective dose as experienced by different identically exposed individuals, and it is this biologically effective dose that is ultimately functionally tied to any observed effects.

One cannot necessarily avoid exposure measurement error by using biomarkers of exposure in place of external exposure measurements, as errors in these measurements can also occur, stemming both from imprecision of measurement and from the degree to which such biomarkers are surrogates for still more fundamental internal dose measures. For example, although the availability of biomarkers of lead exposure is generally viewed as an improvement over air lead measurements for predicting lead health impacts, both of the commonly used biomarkers of lead exposure, namely blood lead and bone lead measurements, are considered to have significant exposure measurement error (US EPA, 2006a). In part this is because, with the exception of health outcomes related to hematopoiesis and bone health, lead in these peripheral tissues is typically several toxicokinetic steps removed from lead in critical targets organs, such as the brain (US EPA, 2006a). US EPA (2006a) identified several sources of exposure measurement error associated with the use of blood lead levels as an exposure surrogate for brain lead levels, including (1) from the use of a single blood lead concentration as a proxy of the lead total body burden in an individual; (2) from differences between individuals in the relationship between blood lead levels and lead total body burdens; and (3) from differences in the blood lead concentrations at the time of measurement versus levels at the etiologically relevant period. Overall, it concluded that the use of blood lead levels as a proxy of brain lead levels results principally in classical-type measurement error. Bone lead levels are considered to be a more valid surrogate of long-term lead exposure than blood lead levels, but US EPA (2006a) highlighted several components of exposure measurement error associated with their use, including (1) instrument error associated with the current X-ray fluorescence (XRF) methods (e.g., their lack of sensitivity for low-dose community exposures); (2) between-laboratory variability in XRF results, due in part to the lack of standard reference materials; (3) differences in XRF results for different types of bones; and (4) differences between individuals in release of bone-deposited lead.

In summary, there are numerous common sources of exposure measurement error in environmental and occupational epidemiology studies. As discussed above, exposure measurement error is a pervasive problem not only for indirect exposure assessment metrics, such as central-site ambient air pollutant measurements, but also for biomarkers of exposure. Importantly, as illustrated by some of the examples provided above, the structure of exposure measurement error in air pollution studies is complex and is generally considered to reflect the combined influences of both Berkson and classical errors, rather than a single error type. The next section addresses exposure measurement error in practice, focusing in particular on air pollution epidemiology studies where recent studies have attempted to characterize exposure measurement errors.

## 4. Exposure measurement error in practice

Despite widespread recognition of exposure measurement error in environmental and occupational epidemiology studies, relatively few studies have attempted to quantify the degree of exposure measurement error for different study designs, health predictors, and exposure surrogates. Among the studies that have attempted to quantify exposure measurement errors is a body of recent air pollution studies, with perhaps the greatest number of studies focusing on exposure errors for $PM_{2.5}$. As described below, two dominant metrics have emerged in the $PM_{2.5}$ literature to quantify exposure errors related to the use of central-site ambient monitoring data in epidemiology studies, namely longitudinal, within-person correlation coefficients between central-site ambient concentrations and either total personal $PM_{2.5}$ exposures or personal exposures to ambient $PM_{2.5}$, and a factor that has been termed either an "attenuation factor" ($\alpha$; see Sheppard et al., 2005; Sarnat et al., 2007) or an "outdoor personal exposure factor" ($F_{pex}$; see Wallace and Williams, 2005). Focusing on $PM_{2.5}$ but also drawing upon the more limited information available for other air pollutants such as $NO_2$ and $O_3$, this section discusses analyses conducted in the air pollution literature to quantify exposure measurement errors and highlights scenarios where the potential for bias exists. Because the studies provided indirect evidence of the size of exposure measurement errors, this section concludes with a brief discussion of some recent air pollution health effect studies that have attempted to examine the impact of such errors on relative risk estimates. This discussion of the impact of exposure measurement error on exposure-response relationships is continued in the next section. It is important to note that this section is focused on findings from the air pollution literature owing to the growing number of studies addressing exposure measurement error in this field, not because exposure measurement error is any less of a potential problem for other types of health effect studies. The focus of this literature is on the positive aspect—gauging the usefulness of practically accessible measures of exposure for use as surrogates of inaccessible yet fundamental personal toxicologically effective dose—but our interest in reviewing these studies is for gaining insights into the magnitude and nature of the inevitable imperfections of such accessible measures.

Regulatory documents, such as the various US EPA Air Quality Criteria Documents (e.g., US EPA, 2006b), and more recently the US EPA Integrated Science Assessments (e.g., US EPA, 2008, 2009), have used data on daily correlations between central-site ambient concentrations and personal exposure concentrations to support the validity of using central-site measurements as surrogate exposure measures in air pollution epidemiology studies. For example, there is the opinion among the scientific and regulatory communities that these longitudinal correlations are sufficiently strong to support the use of central-site measurements as a proxy for ambient $PM_{2.5}$ exposures (Sarnat et al., 2007, 2009a). Although in general high, these correlations do not indicate an absence of significant exposure measurement error or potential bias in the case of $PM_{2.5}$. In fact, a pair of recent publications (Avery et al., 2010a, 2010b) cautioned of the potential for substantial measurement error based on the high degree of variability in reported ambient-personal $PM_{2.5}$ correlations.

(Avery et al., 2010a) conducted a meta-analysis of published exposure studies reporting within-person correlations ($r$) between total personal $PM_{2.5}$ levels and central-site ambient $PM_{2.5}$ concentrations. Figure 1 summarizes estimates of median (with 95% confidence intervals) within-participant personal $PM_{2.5}$–ambient $PM_{2.5}$ correlation coefficients reported by (Avery et al., 2010a) for the 18 eligible studies they identified using comprehensive literature searches. As shown in Figure 1, estimated summary correlation coefficients (median .54; range .09 to .83) and their standard deviations (median .12; range .04 to .31) show substantial variability. Based on their findings, (Avery et al., 2010a) concluded, "The wide range in estimated correlations between personal and ambient $PM_{2.5}$, as well as the associations with participant, study and environment characteristics, suggest that the potential for exposure misclassification can be substantial." In addition, for their companion study where they estimated summary correlations between total personal $PM_{2.5}$ levels and outdoor residential $PM_{2.5}$ concentrations rather than central-site ambient $PM_{2.5}$ concentrations, (Avery et al., 2010b) also reported a wide range in correlation coefficients (median .53; range: .25 to .79) and their standard deviations (median 53.6; range 9.4 to 548.1) for the nine eligible studies identified in their systematic literature review. Based on these and their prior findings, (Avery et al., 2010b) concluded, "Collectively, the meta-analyses suggest that residential outdoor-personal and ambient-personal $PM_{2.5}$ correlations merit greater consideration when evaluating the potential for bias in studies of $PM_{2.5}$-mediated health effects."

As discussed previously, exposures to non-ambient PM, such as indoor-generated PM, are not expected to be correlated with ambient concentrations, meaning that higher, more uniform personal-ambient correlations would be anticipated for studies using novel approaches to separate out ambient $PM_{2.5}$ exposures from total personal $PM_{2.5}$ exposures. However, studies evaluating the correlations between time-series of personal exposures

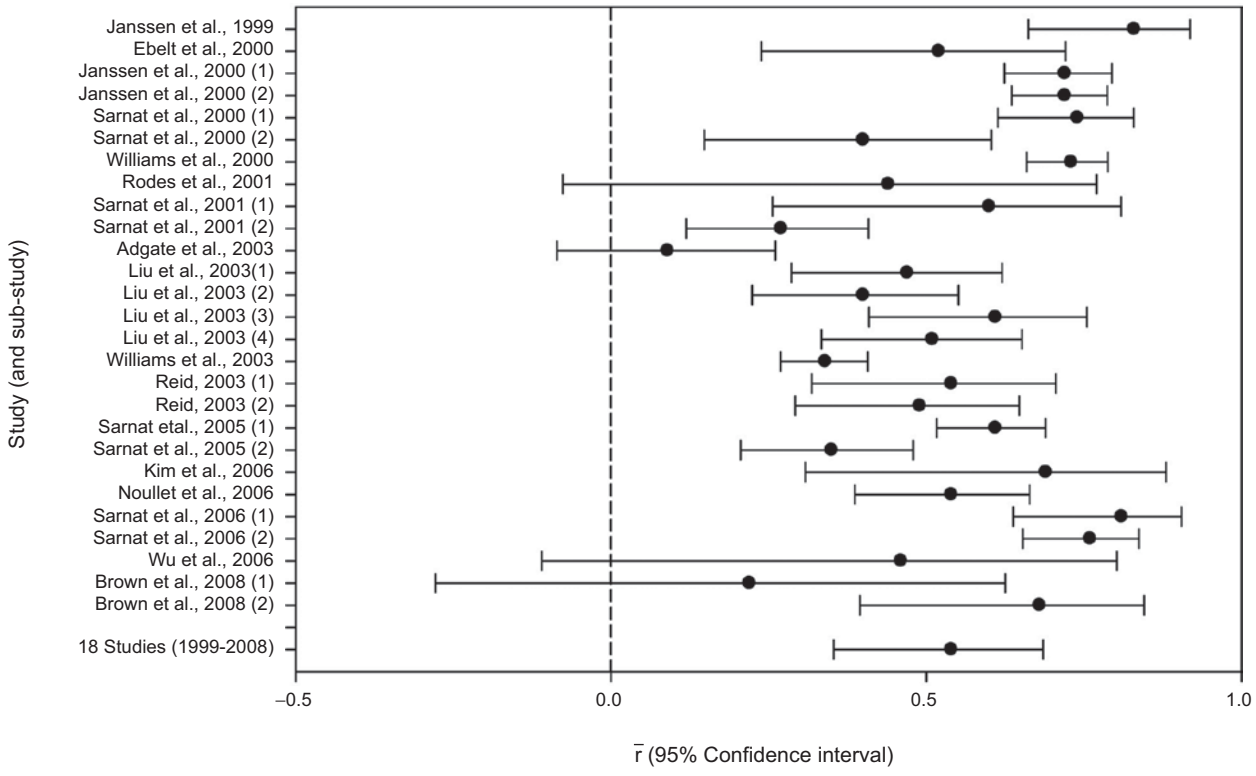


Figure 1. Estimates of median correlation coefficients (with 95% CI) from 18 studies of the within-participant correlation between ambient and personal $PM_{2.5}$ exposure (data from Avery et al., 2010a).

to $PM_{2.5}$ of ambient origin and either ambient $PM_{2.5}$ central-site measurements or residential outdoor $PM_{2.5}$ measurements have reported correlation coefficients less than 1 that show some interpersonal heterogeneity, providing evidence for a possible role of exposure measurement error. For example, for a panel of 37 residents in the Research Triangle Park area in North Carolina monitored for week-long periods in each of four seasons, Wallace and Williams (2005) reported a median $R^2$ value of .73, but a range of .19 to .88, for within-participant correlations of estimated outdoor contributions to personal exposure versus central-site $PM_{2.5}$ correlations. Importantly, Wallace and Williams (2005) noted that their $R^2$ values likely represent upper bounds on the true correlations between personal exposures to $PM_{2.5}$ of ambient origin and central-site ambient $PM_{2.5}$ measurements given the assumptions used in their model to estimate exposures to particles of ambient origin. In addition, for a series of recent panel-based exposure assessment studies conducted by researchers at the Harvard School of Public Health where over 200 study participants in four US cities (Baltimore, Boston, Atlanta, and Steubenville, Ohio) were monitored for 7- to 12-day periods, (Sarnat et al., 2007, 2009a) reported higher personal-ambient associations for sulfates ($SO_4^{2-}$), a $PM_{2.5}$ constituent that is known to have few indoor sources and that has gained use as an exposure surrogate for $PM_{2.5}$ of ambient origin, than for $PM_{2.5}$. However, although generally high (over 70% of subjects had correlations greater than .80), there remained some interpersonal heterogeneity in these sulfate personal-ambient correlations, including negative correlations.

The attenuation factor ($\alpha$), also sometimes referred to as the "outdoor personal exposure factor" ($F_{pex}$), is another measure of the exposure-surrogate difference that has been used to quantify exposure measurement errors resulting from the use of central-site ambient monitoring data in place of personal $PM_{2.5}$ data. The attenuation factor ($\alpha$) is simply the ratio of the personal exposure to ambient $PM_{2.5}$ and the ambient $PM_{2.5}$ concentration, as generally measured by a central-site ambient monitor. It is thus an index of the fraction of ambient particles that contribute to personal exposure (Wallace and Williams, 2005). As discussed below, several recent studies have postulated that systematic differences in the relationship between personal exposures and ambient concentrations (i.e., in ambient attenuation factors) may be a key source of bias in $PM_{2.5}$ concentration-response relationships (Sarnat et al., 2009a; Meng et al., 2005; Sheppard et al., 2005). In support of this idea, there is a growing body of data indicating that values of $\alpha$ vary with season, geographic region, individual, and spatiotemporal averaging times (Sarnat et al., 2007; Wallace and Williams, 2005; Allen et al., 2003; Janssen et al., 2002). For example, based on personal and outdoor sulfate measurements, Wallace and Williams (2005) estimated 24-hour values of $\alpha$ that were well below unity (overall average 0.54; range 0.33 to 0.77) and that exhibited substantial interpersonal and day-to-day variation (Figure 2). However, it is important to note that ambient attenuation is unlikely to be a source of bias in all studies. For example, in the case of a perfect personal-ambient correlation, the ambient attenuation factor could be less than 1, but not a source of potential bias to regression coefficients (i.e., a constant proportionality factor) or arguably even a component of the measurement error structure. In addition, random variation in ambient attenuation factors, rather than systematic variation that is correlated with ambient concentration, can be a source of Berkson-type error rather than classical-type error and thus not a source of potential bias to regression coefficients.

More recently, for the Harvard four-city panel study data, (Sarnat et al., 2009a) calculated within-subject personal-ambient attenuation factors using personal sulfate concentrations and ambient sulfate concentrations averaged over complete sampling sessions (245 total sampling sessions, each consisting of 7 to 12 days of monitoring, including sessions in multiple seasons). As with the Wallace and Williams (2005) daily values of $\alpha$, these time-averaged attenuation factors were generally well less than 1, and were also found to vary by individual, city, and season. For example, for analyses where data were stratified by city across seasons, (Sarnat et al., 2009a) observed mean values of $\alpha$ that ranged from 0.53 in Baltimore to 0.76 in Steubenville. An even larger range (0.47 in Baltimore in winter to 0.83 in Boston in summer) was observed when data were stratified by city and season. Seasonal fluctuations were highly variable across the four cities, with negligible seasonal differences for Steubenville and Atlanta, but considerable seasonal differences for Boston (0.59 in winter and 0.83 in summer) and Baltimore (0.43 in winter and 0.60 in summer).

Based on their findings, (Sarnat et al., 2009a) proposed that situations could arise where exposure measurement error from the use of time-averaged ambient $PM_{2.5}$ concentrations as surrogates for average personal exposures could contribute to bias in $PM_{2.5}$ concentration-response relationships, in particular for epidemiology studies of long-term $PM_{2.5}$ exposures and chronic health effects. Such bias in concentration-response relationships could result from independent-variable measurement error attributable to systematic variation in the relationship between personal exposures and ambient concentrations in different cities and locations. This bias can arise if individuals in different locations/cities or even within the same location exposed to the same ambient concentration may not receive the same personal exposure, owing to differences in building construction, ventilation practices (e.g., use of air conditioning versus opening windows), spatial variability in $PM_{2.5}$ concentrations, time-activity patterns (e.g., amount of time spent outdoors), and particle properties (e.g., particle size distributions, chemistry, thermodynamics). In support of this idea, (Sarnat et al., 2009a) pointed to their findings of different ordinal rankings for the four cities for mean ambient concentrations versus averaged personal

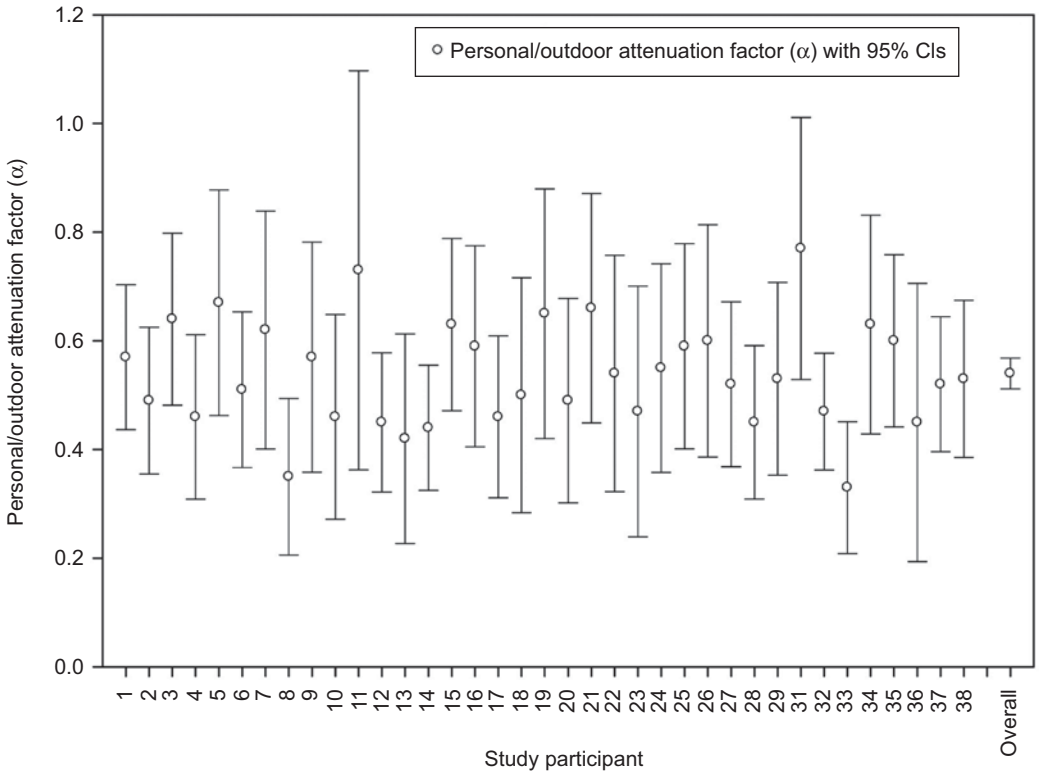


Figure 2. Estimates of $PM_{2.5}$ personal/outdoor attenuation factors (α) from the RTP Particulate Matter Panel Study (data from Wallace and Williams, 2005).

exposures (e.g., Baltimore had the highest mean ambient concentrations, but was second to Steubenville in average personal sulfate exposures). (Sarnat et al., 2009a) concluded, "In these cases, systematic differences between average personal exposure and ambient concentrations by city, location or within a given location can lead to concentration-response relationships that do not reflect the true relationships between exposure and response, thus leading to biased estimated health risks."

Based on findings from an earlier study of three US cities (Houston, Los Angeles, and Elizabeth, NJ), (Meng et al., 2005) also suggested potential situations where exposure-response bias could result from the use of central-site ambient measurements as an exposure surrogate for personal exposures to ambient $PM_{2.5}$. Using a series of models with increasingly accurate assumptions for predicting the infiltration of outdoor particles into indoor environments, (Meng et al.,) demonstrated the broadening in the distribution of exposures to $PM_{2.5}$ of ambient origin as models improved in their ability to account for home-to-home and day-to-day variations in particle infiltration behavior (which are caused by variations in building construction, ventilation practices, and particle properties). They showed that the simplest model, which assumed a constant infiltration factor as would be case if central-site PM was a "perfect surrogate" for ambient PM exposure, significantly underestimated the variance of exposures to $PM_{2.5}$ of ambient origin. Meng et al. (2005) thus concluded that home-to-home and day-to-day variations in particle infiltration behavior can introduce significant exposure measurement error in air pollution studies relying on central-site ambient measurements. They classified this error as predominantly Berkson error, but hypothesized that in some cases, namely for seasonal variations in particle infiltration behavior, potential bias to regression coefficients could arise, in particular for time-series epidemiology analyses. As demonstrated in their data, seasonal differences in infiltration behavior can not only coincide with fluctuations in ambient particle concentrations, but can also vary with location. In particular, in summer when $PM_{2.5}$ concentrations are generally higher, (Meng et al., 2005) observed an increase in infiltration factors in New Jersey homes due to the opening of windows to increase home ventilation, but a reduction in infiltration factors in Texas homes as people instead relied on air conditioners to cool their homes. As with the (Sarnat et al., 2009a) analysis, (Meng et al., 2005) further suggested that variation in particle infiltration could also contribute to bias in chronic health studies, since differences in particle infiltration behavior between communities could differentially impact the personal-ambient relationships (e.g., mean ambient $PM_{2.5}$ concentrations could be higher in City A versus City B, but due to differences in particle infiltration behavior in the two cities, mean exposures to ambient $PM_{2.5}$ could instead be higher in City B versus City A).

Less is known about the size and nature of the exposure measurement errors associated with the use of central-site ambient monitoring data as exposure surrogates for the gaseous criteria air pollutants (e.g., $NO_2$, $O_3$,

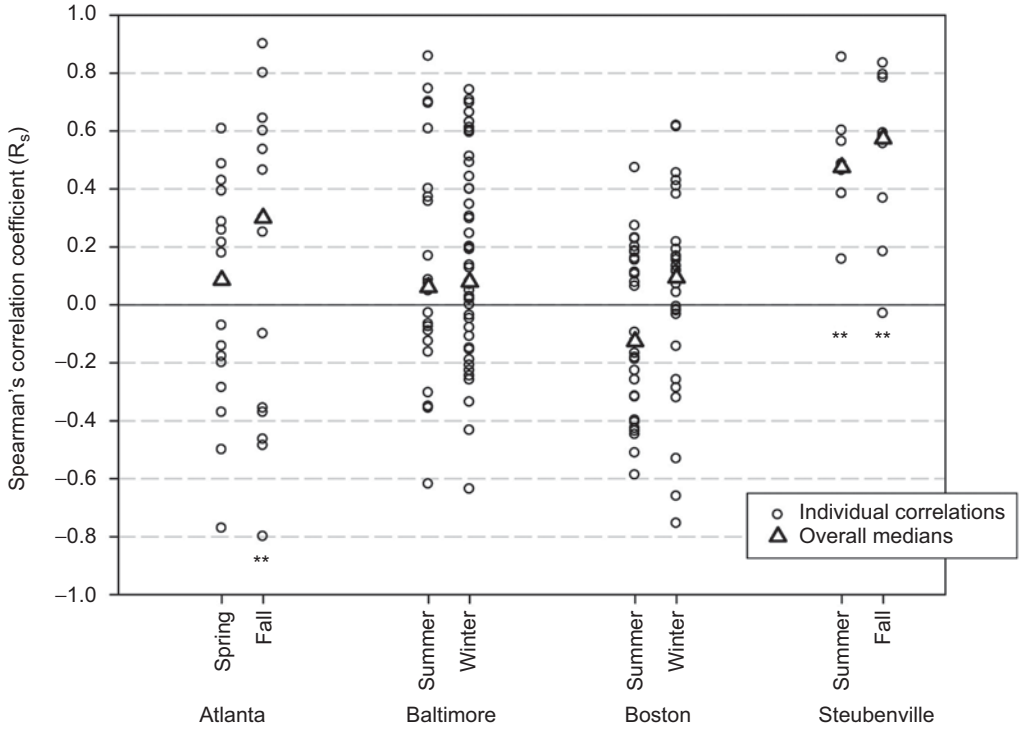


Figure 3. Distribution of subject-specific Spearman's Spearman's correlation coefficients for personal-to-ambient $NO_2$ exposure by city in the Harvard panel studies (data from Sarnat et al., 2007). Triangles represent the median correlation coefficient. $^{**}p < .0001$.

$SO_2$, CO). This is because fewer studies have characterized personal-ambient correlations or personal-ambient attenuation factors for the gaseous criteria air pollutants. The available data for such pollutants as $NO_2$ and $O_3$ suggest poorer correlations, greater attenuation factors, and thus larger exposure measurement errors than for $PM_{2.5}$. For example, as summarized in (Sarnat et al., (2007), the Harvard investigators generally observed weak and statistically insignificant (and, in some cases, negative) personal-ambient $NO_2$ associations for subjects in three of the US cities included in their studies, namely Baltimore, Boston, and Atlanta (see Figure 3). Importantly, these panel studies generally observed stronger associations between ambient $NO_2$ concentrations and personal $PM_{2.5}$ exposures than between ambient $NO_2$ concentrations and personal $NO_2$ exposures. Only for their Steubenville (Ohio) panel, where subjects were clustered around the ambient monitoring site, did these investigators observe stronger personal-ambient $NO_2$ associations (i.e., subject-specific Spearman's correlation coefficients that were generally above .50). US EPA (2008) reported a range in $NO_2$ personal-ambient attenuation factors of ~0.3 to ~0.6 for the limited study data available.

$NO_2$ personal-ambient correlations are affected by such factors as spatial variability, indoor source contributions, seasonal and geographic variability in ambient $NO_2$ infiltration into indoor environments, and time-activity patterns in various microenvironments. Greater within-city spatial variability for this locally generated pollutant is considered to be one of the key factors underlying the lower personal-ambient correlations (US EPA, 2008). For example, localized traffic-related emissions are a major source of $NO_2$ exposures, and $NO_2$ exposures received during commuting may be poorly reflected in measurements of urban background levels at central-site monitors. For the same reason, weaker correlations, and thus larger exposure measurement errors, are also expected for ultrafine particles (i.e., particles with diameters <100 nm).

For $O_3$, (Sarnat et al., 2001, 2005) investigated the association between personal $O_3$ exposures and central-site $O_3$ measurements for two of the Harvard panels (Baltimore and Boston) using mixed-model regression. For a Baltimore-based cohort of 56 subjects recruited from among populations considered to be more susceptible to air pollution (senior citizens, children, and individuals with chronic obstructive pulmonary disease [COPD]), (Sarnat et al., (2001) reported no correlation between 24-hour personal and ambient ozone measurements for both winter and summer sampling periods (mixed-model regression slopes of .00 and .01, respectively). In a similarly designed study conducted in Boston among a cohort of 20 healthy senior citizens and 23 school children, (Sarnat et al., 2005) reported similar results, finding no correlation between personal and ambient ozone concentrations for winter sampling data (statistically insignificant slope of .04) and only a moderate correlation between personal and ambient ozone concentrations for summer sampling data (statistically significant slope of .27). Furthermore, due in part to its high reactivity that results in substantially diminished indoor ozone concentrations compared to ambient levels, (Sarnat et al., 2005) observed that ambient $O_3$ levels overestimated personal exposures 3- to 4-fold in the summer and 25-fold in the winter.

For both $NO_2$ and $O_3$, relatively weak personal-ambient correlations and low personal-ambient attenuation factors are a function of the interplay of a number of individual-, season-, and city-specific factors, including time-activity patterns, building characteristics, and ventilation practices, that influence the relationship between ambient concentrations of these pollutants and personal exposures. The use of central-site ambient measurements as an exposure surrogate for $NO_2$ and $O_3$ results in exposure measurement error because people do not spend 100% of their time outdoors, and large microenvironmental differences in levels of these pollutants can result from differences in air exchange rates, outdoor infiltration, indoor circulation rates, and indoor removal processes (US EPA, 2006a, 2008). For $O_3$ in particular, its high reactivity results in more complex interactions with indoor microenvironments, including greater deposition and reaction with surfaces, than for less reactive pollutants such as $PM_{2.5}$. Complicating matters for $NO_2$, it can have common indoor sources, including gas stoves and heaters. As for $PM_{2.5}$, home characteristics and ventilation are thus key factors determining $NO_2$ and $O_3$ exposure variability within and between locations and seasons (US EPA, 2006b, 2008). Similar to the bias postulated by (Sarnat et al., 2009a) and (Meng et al., 2005) for $PM_{2.5}$, varying relationships between personal exposures and ambient concentrations for different seasons and in different cities are possible sources of bias for $NO_2$ and $O_3$.

In addition, for $O_3$ in particular, researchers have hypothesized an additional source of potential bias involving pollution-induced behavioral modification of activity patterns (Bateson et al., 2007). As described by (Bateson et al., 2007), it is hypothesized that some people modify their activity levels and elect to stay inside on high air pollution days, thus resulting in personal exposures for these high pollution days that are less than exposures for lower pollution days. This is considered to be a more likely source of bias for $O_3$ and $NO_2$ than for $PM_{2.5}$ because indoor $O_3$ concentrations show greater reductions compared to ambient concentrations, meaning that the averting behavior would result in a greater difference between ambient and personal $O_3$ concentrations.

In concluding this discussion of exposure measurement error in practice, it is important to note the emergence of a body of air pollution studies addressing the impacts of exposure measurement error (e.g., Sheppard et al., 2005; Schwartz et al., 2007; Sarnat et al., 2009b; Baxter et al., 2010). Rather than quantifying the magnitude of exposure measurement error itself, these studies have attempted to quantify its impacts on health effect estimates. They thus provide an indirect measure of the magnitude of exposure measurement error from using central-site ambient monitoring data as an exposure surrogate for personal exposure. These studies are not discussed in great detail in the next section because they have generally assumed a linear regression model and thus examine bias to the slope of this model, rather than bias to the shape of the exposure-response relationship. However, as discussed briefly below, these studies provide additional evidence that the magnitude of exposure measurement error differs across the various criteria air pollutants, and they also demonstrate the potential bias that can arise from seasonal variation in attenuation factors.

In particular, (Schwartz et al., 2007) conducted a simulation study to investigate changes in assumed true health risk estimates when exposure surrogates (e.g., central-site ambient concentrations) were used in place of actual personal exposures. They relied upon empirical data (from the Harvard Baltimore panel study) to simulate particulate and gaseous pollutant correlations and covariances, and to estimate the distributions of personal exposures associated with a corresponding set of ambient concentrations. For a time-series study design and modeling scenarios where a true health association was assumed with personal exposures to $PM_{2.5}$, $O_3$, or $NO_2$, (Schwartz et al., 2007) reported a statistically significant health association for only the corresponding ambient concentration of $PM_{2.5}$. Although statistically significant, the size of the observed $PM_{2.5}$ risk estimate was diminished by approximately 70% compared to the assumed true health risk estimate. These findings thus suggest that exposure measurement error is of greater magnitude for $O_3$ and $NO_2$ than for $PM_{2.5}$, and (Schwartz et al., 2007) concluded that their results demonstrated the inadequate performance of central-site ambient monitors as exposure surrogates for the gaseous criteria air pollutants. Similarly, based on a time-series analysis of Atlanta emergency room visits data, (Sarnat et al., 2009b) reported findings demonstrating the effects of exposure measurement error from spatial variations in ambient concentrations for two spatially heterogeneous criteria air pollutions (CO and $NO_2$).

(Sheppard et al., 2005) conducted an analysis in which they simulated population variation in the personal-ambient attenuation factor and examined the resulting impact on $PM_{2.5}$ health effect estimates. They included several different cases differing in the strength of the temporal correlation (weak, moderate, strong) between the variation in the attenuation factor and the ambient central-site concentrations used as the exposure metric. As with the (Schwartz et al., 2007) study, (Sheppard et al., 2005) used a time-series study design, with local outdoor, ambient central-site, and personal $PM_{2.5}$ measurement data from a Seattle panel study informing the error structure. Ambient central-site concentrations used as the exposure metric in the simulations were characterized by higher values in the winter and lower values in the summer. Importantly, (Sheppard et al., 2005) demonstrated that random population variation in the personal-ambient attenuation factor (i.e., variation that has little or no temporal correlation with ambient concentration) does not cause significant bias to effect estimates. For the cases of seasonal variation in the attenuation factor that was highly correlated with ambient concentrations, however, they demonstrated substantial bias in health effect

estimates. Interestingly, they observed that the direction of this bias depended on the nature of the association of the attenuation factor with the ambient concentrations. Specifically, they observed a significant downward bias in the effect estimate for the case where low attenuation factors were assumed for the winter, but a significant upward bias in the effect estimate when high attenuation factors were assumed for the winter.

Sheppard et al. (2005) also conducted two additional simulation analyses, including one in which they examined the impact of random variation in non-ambient exposure (i.e., variation in non-ambient exposure that was not temporally correlated with ambient concentration) on the ambient $PM_{2.5}$ effect estimates. In this case, they observed no bias in the effect estimate for ambient $PM_{2.5}$ concentrations. In the third simulation study, Sheppard et al. (2005) examined the impact of spatial variation in ambient monitors, modeling cases where only one monitor or the average of several monitors was used to estimate population ambient concentration. In their simulation, where spatial variation was randomly distributed and small, they demonstrated a slight bias in the effect estimate, reflecting some classical-type error, for the use of a single monitor. This bias rapidly disappeared with an increased number of monitors. Although bias was small in this case, owing to the low degree of spatial variability in ambient $PM_{2.5}$ concentrations, this simulation shows the potential for classical-type measurement error, and resulting bias to regression coefficients, associated with spatial variability in pollutant concentrations.

In summary, recent studies show relatively strong longitudinal correlations between central-site ambient $PM_{2.5}$ concentrations and personal exposures to $PM_{2.5}$ of ambient origin. These correlations have been interpreted as supporting the validity of using central-site measurements as surrogate exposure measures in $PM_{2.5}$ epidemiology studies. Their usefulness, or even their lack of bias, does not indicate the absence of exposure measurement error, however. As discussed above, exposure measurement error is a common feature of $PM_{2.5}$ epidemiology studies, as well as health studies of other air pollutants, which rely upon central-site ambient measurements as exposure surrogates, with the structure of this measurement error generally consisting of a mix of both Berkson-type and classical-type error components. For example, as discussed above, measurement error associated with ambient attenuation can have both types of error components attributable to random population variation (a source of Berkson-type error) and to systematic variability in attenuation that is correlated with ambient concentration (a source of classical-type error). Berkson-type error, such as that associated with random population variation in ambient attenuation, is associated with reduced information on the true exposure variability, but is not considered to be a significant source of bias in health effect parameters. Classical-type error, however, is an established source of bias to measurements of relationships between health effect response and the measured exposures, with studies such as Sheppard et al. (2005), Meng et al. (2005), and Sarnat et al. (2009a) identifying systematic variability in ambient attenuation by season and/or by region as a possible source of bias to $PM_{2.5}$ health effect estimates. Although a greater number of studies have addressed the sources and impacts of exposure measurement error for $PM_{2.5}$, we have also discussed how the potential bias from exposure measurement error may be larger for gaseous criteria air pollutants (e.g., $O_3$, $NO_2$) than for $PM_{2.5}$. Overall, exposure measurement error is pervasive and inescapable; it constitutes a challenge of interpretation of patterns in environmental and occupational epidemiology, and a potential source of bias in estimating associations of exposure and effect, adding to the importance of better understanding how such errors may affect the shape of concentration-response and exposure-response curves (see next section).

## 5. Independent-variable error and regression

To determine the effects of exposure measurement error on an exposure-response curve, one must first understand the more general case of error in independent variables on regression. Independent variable error results in the loss of statistical power, parameter estimate biases, and the masking of features in the true regression curve. Both biased parameter estimates and masked features are distinct from and not by-products of the loss of statistical power. Although we do not aim at a comprehensive review, below we discuss several types of error and how they bias parameter estimates and mask features in both linear and nonlinear regression curves that often describe exposure-response relationships.

It is important to note that the results discussed below are dependent upon specific assumptions and simulation parameter choices, so not all of the results will be applicable to all situations with independent variable error. In practice, generalizing results to real-world situations should be done with caution because we do not know the exact form of the exposure response model or the error characteristics of the exposure variables.

### 5.1. Error in linear regression

In Section 2, we demonstrated that random error in the independent variable in linear regression generally biases the slope parameter towards the null. That is, for the linear, univariate model, classical error causes the slope parameter to be underestimated by a factor $\gamma$, which is less than 1 (Equation 8). In contrast to classical error, unbiased Berkson error in the independent variable results in an unbiased regression slope parameter (Zeger et al., 2000).

With regard to multivariate regression models in which there are classical errors in multiple predictor variables, the situation is not clear cut. Classical error in each measured covariate biases the regression slope, but the direction (towards or away from the null) and magnitude

of the bias depends on the correlation among the independent variables as well as among the error variances (Carroll and Galindo, 1998). Zeger et al. (2000) found that in the presence of negative correlations, the bias tends to be away from the null.

In contrast to classical error, unbiased Berkson error in the independent variable in univariate models leads to the measured variable, *W*, being an unbiased estimate of *X*. As a result, the regression slope parameter is unbiased because the measured variable is less error-prone than the actual variable. This is also true when there is Berkson error in several independent variables in multivariate models (Zeger et al., 2000).

In addition to contributing to bias in parameter estimates, error also increases the variability of estimates, which results in the loss of statistical power. In the presence of measurement error, there can be a tradeoff between bias and variability (Carroll et al., 2006). In particular, though the regression attenuation factor, $\gamma$, can reduce the bias in a univariate regression slope estimate, it also results in an increase in the variance proportional to the inverse of the square of the regression attenuation factor, $1/\gamma^2$.

Overall, it is important to note that Berkson error and classical error affect regression curves differently. It is well known and has been shown above that classical error results in bias in the regression relationship. Pure Berkson error results in less bias in the regression curve but at the cost of reduced power in estimates of regression parameters. In practice, most error is a combination of classical and Berkson.

### 5.2. Error in nonlinear regression

There are many types of nonlinear curves (e.g., logistic, exponential, polynomial, power and segmented linear ["hockey stick"] curves), and error in independent variables tends to smooth out and flatten, essentially linearize, these curves. This phenomenon has been noted in a variety of fields and can be seen analytically and computationally for a variety of nonlinear curves, under different assumptions, and over broad parameter choices. This is demonstrated below for both polynomial and logistic regression models.

The quadratic model is the polynomial curve most similar to linear regression. Kuha and Temple (2003) assessed the effects of additive errors in independent variables on quadratic regression curves and found that classical error flattens the curve. They noted that if the true relationship is

$$Y = \beta_0 + \beta_1 \times X + \beta_2 \times X^2 \tag{9}$$

then naïvely regressing *Y* against the measured variable, $W = X + U$, leads to

$$Y \approx \beta_0^* + \beta_1^* \times X + \beta_2^* \times X^2 \tag{10}$$

where the new regression coefficients are biased with respect to the true coefficients. For a monotonically increasing regression curve with normally distributed *X* and *U*, the constant $\beta_0^*$ is biased upwards and the quadratic coefficient is biased downwards by the square of the attenuation factor, $\gamma$:

$$\beta_2^* = \beta_2 * \gamma^2 \tag{11}$$

Although the direction of the bias of the linear coefficient, $\beta_1^*$, depends upon the particular parameters of the regression curve, the biases in the constant and quadratic regression parameters are sufficient to flatten and linearize the curve, which means that for large values of the independent variable, the response is smaller than it would be with unbiased parameters, but for small values, the response is larger than it would be without error.

This also extends to logistic regression. Using the following logistic regression model:

$$Y = \frac{1}{[1 + exp(-\beta_0 * X)]} \tag{12}$$

where *Y* is the response variable, *X* is the true independent variable, and $\beta_0$ is the estimated regression parameter. Stefanski and Carroll (1985) investigated the effect of independent variable error and reported:

"When covariates are measured with error the usual logistic regression estimator of βo is asymptotically biased [see Clark (1982) and Tripathi (1980)]. As a consequence of bias there is generally a tendency to underestimate the disease probability for high-risk cases and overestimate for low-risk cases; it will be said that measurement error attenuates predicted probabilities."

Complementing the work of Kuha and Temple (2003), Crump (2005) assessed the effects of classical multiplicative errors in independent variables in power regression curves. Although Crump (2005) was attentive to the parameter estimate biases caused by measurement error, he also addressed the issue of masking of features caused by independent variable error. Thus, unlike the previous analyses, he analyzed whether the form of the curve could change in the presence of error. He concluded that unbiased errors in the independent variable can convert a sublinear curve (power function with exponent >1) to a supralinear curve (exponent <1), essentially changing the shape of the regression curve. This result is especially important in the context of exposure-response curves because a threshold function is a special case of a sublinear curve. The transition from a sublinear towards a supralinear curve corresponds to a masking of a threshold or linearization or flattening of the curve.

Crump (2005) used the classical multiplicative error model ($W = U \times X$) and assumed that both the covariate, *X*, and error term, *U*, are lognormal such that ln(*X*) has mean μ and standard deviation, $\sigma_x$, and ln(*U*) has mean, υ, and standard deviation, $\sigma_u$. A true power exposure response of order *r* appears to be of order *K* where $K < r$ in the presence of measurement error. Specifically, the true response:

$$Y = \alpha + \beta * X^{r} \tag{13}$$

appears to be

$$Y = \alpha + \beta * A * W^{K} \tag{14}$$

where $K = r * \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$ and $\sigma_x$ and $\sigma_u$ characterize the variability in the exposure and error distributions, respectively. The power $r$ is reduced in the presence of error by a factor that is similar to the regression attenuation factor encountered in linear regression. The parameter $A$ depends on parameters characterizing the exposure and error distributions.

Although quadratic and more general polynomial and power regression curves exhibit nonlinearities, segmented models are the most definitive threshold models. There are many types of segmented models that describe functions with thresholds. For example, Equations 15 and 16 describe linear and exponential models, respectively:

$$Y = \alpha_0 + \beta_0 (X - \tau_0)_+ \text{ for } (X - \tau)_+ = \begin{cases} 0 & X < \tau \\ (X - \tau) & X \geq \tau \end{cases} \tag{15}$$

$$Y = \exp\{\alpha_0 + \beta_0 (X - \tau_0)_+\} \text{ for } (X - \tau)_+ = \begin{cases} 0 & X < \tau \\ (X - \tau) & X \geq \tau \end{cases} \tag{16}$$

$Y$ is constant at a level $\alpha_0$ in the linear model and $\exp(\alpha_0)$ in the exponential model for all $X$ less than the threshold, $\tau_0$. For all $X$ greater than or equal to $\tau_0$, $Y$ increases linearly in the linear model and exponentially with the exponential model. A graph of the linear function is shown in Figure 4.

Küchenoff and Carroll (1997) investigated bias and masking of features in a segmented linear threshold model and found that independent variable error can mask a true threshold. They used a classical additive error model such that $W = X + U$, with $U$ characterized by a normal distribution with mean zero and variance, $\sigma_u^2$. They determined analytically the result of naïvely regressing $Y$ against the measured variable $W$. The resulting analytical expression is the following:

$$E(Y \mid W) = \alpha_0 + \beta_0 \gamma^{1/2} \sigma_u \{a\Phi(a) + \varphi(a)\} \tag{17}$$

where $\varphi$ and $\Phi$ are the normal density and cumulative distribution functions, respectively, and

$$a = \frac{\{\mu(1-\gamma) + \gamma W - \tau_0\}}{(\gamma \cdot \sigma_u^2)^{1/2}} \tag{18}$$

where $\mu$ is the mean of $X$, $\gamma$ is attenuation factor, $\sigma_u^2$ is the variance in the measurement error, and $\tau_0$ is the true threshold. As a result of the measurement error in $W$, what was originally a segmented threshold function became a smooth function of W. The function is curved, and a clear threshold is no longer evident. For large $W$, the slope asymptotically approaches $\beta_0 \times \gamma$, which is the same bias found in linear regression. A graph of the true threshold function and those with varying levels of error is shown in Figure 5.

In other simulations, when fitting data with independent variable error to the segmented linear model, Küchenoff and Carroll (1997) found that the threshold parameter was more strongly underestimated with increasing error. Plots of the naïve estimate of threshold as a function of error variance showed a decrease of about 40% as the error variance increased from zero to twice the variance of the independent variable $X$. Figure 6

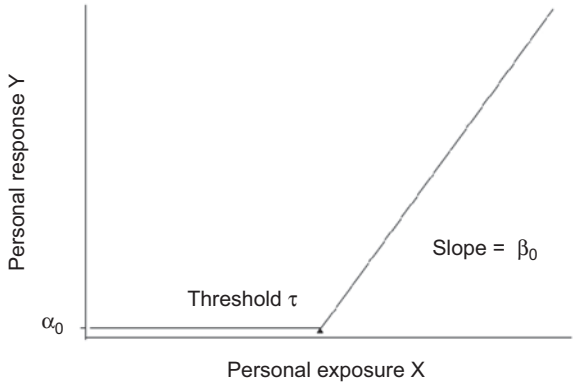


Figure 4. Example of a "hockey-stick" individual threshold risk function with a linear link (adapted from Brauer et al., 2002).

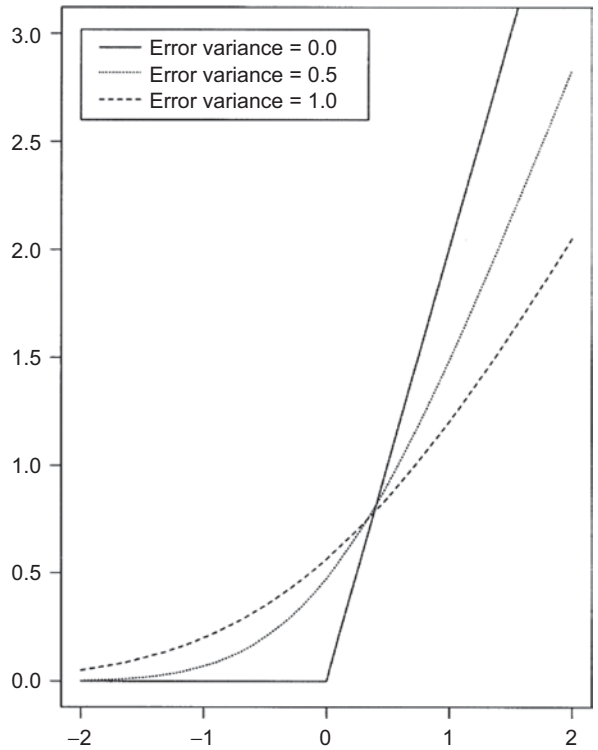


Figure 5. The true segmented linear ("hockey-stick") model and observed curves with increasing measurement error (reproduced from Kuchenhoff and Carroll, 1997). For these curves, $\alpha_0 = \tau_0 = 0$, $\beta_0 = 2$, and $\sigma_x^2 = 1$. Error variance increases from 0 to 0.5 and 1.

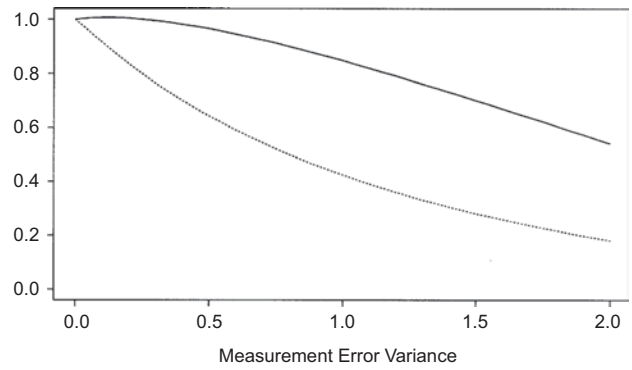


Figure 6. Naïve (solid line) and error-mitigated (dashed line) estimates of a threshold (designated threshold limiting value or TLV) with increasing error. For these estimates, $\alpha_0 = 0$, $\tau_0 = 1$, $\beta_0 = 2$, and $\sigma_x^2 = 1$. Figure reproduced from Kuchenoff and Carroll, 1997.

illustrates the results of increasing underestimation of the threshold parameter with increasing error.

Overall, these studies show that a naïve regression of outcome against an independent variable measured with error can bias regression parameters, linearize a true nonlinear response, and mask a threshold. Effects of independent variable error in linear regression generally result in bias of parameter estimates towards the null assuming random error, though this result can be complicated by associations between multiple independent variables measured with error. The notion of "flattening" and bias towards the null in linear curves can be extended naturally to polynomial and power curves. As the curves become more complicated, however, the effect of error is likewise more complicated: bias in parameter estimates becomes intertwined with masking of features in the true regression curve. For the definitive segmented threshold model, independent variable error can alter the shape of the response function and mask or bias the true threshold.

## 6. Exposure measurement error and exposure-response relationships

Exposure-response relationships summarize our knowledge of the human health risks of a pollutant. Because policy and regulations depend upon the information contained in the exposure-response relationship, any potential uncertainty in the relationship has far-reaching consequences. To this end, it is important to examine all possible sources of bias or error in these relationships, especially with respect to exposure measurement error and the problem of ecological fallacy, which is the assumption that all individuals in a group have the average characteristics (e.g., exposure) of the entire group. Below we relate the results from the previous sections about bias and masking of features for nonlinear regression to simulations of exposure-response curves using experimentally based parameters.

In epidemiology studies of exposure-response patterns, the independent variable is the exposure measure and the dependent variable is the human health outcome. Numerous simulation papers using data for the air pollutants $PM_{10}$, $PM_{2.5}$, and $SO_4^{2-}$ have addressed exposure measurement error and its consequences with applications to real-world situations. Not only are these papers informative to all types of exposure-response relationships, they also quantify the actual error and effects for criteria pollutants of interest and illustrate that exposure measurement error could result in artificial linearity in the exposure-response curve based on actual experimental data when they meet the assumptions of the simulation. Though there are techniques that could help mitigate the effects of error, these techniques are imperfect and cannot be reliably applied to all situations. Overall, because of the prevalence of exposure measurement error in epidemiology data and lack of reliable error-mitigating techniques, conclusions about the linearity of the exposure-response curve must be examined carefully and treated with some skepticism.

Before launching into a discussion of bias in exposure-response curves, we would like to point out that we have confined our investigations to non-differential error, in that error in exposure is independent of the response. An example of differential error would be when subjects with a smaller response have a measured exposure larger than the true exposure and subjects with a greater response have a measured exposure smaller than the true exposure. Examination of differential error and the possibility of a resulting systematic bias in the exposure-response relationship is much more complicated and beyond the scope of this paper, though, as a general statement, if there is a bias in the measurement error itself, it is only a further opportunity for bias in the exposure-response relationship. Overall, we will show that in the best-case scenario of non-differential bias/unbiased error, a bias results in the dose-response fit. The assumptions for the error structure (normal or lognormal) are stated for each regression model.

For the past two decades, studies have acknowledged the importance of exposure measurement error in the distortion of the true exposure-response curve. Many studies have examined air pollutants, specifically PM (Lipfert and Wyzga, 1996; Watt et al., 1995) and $NO_2$ (Yoshimura, 1990) and found linearization of the true threshold exposure-response curve for the pollutants using experimentally-derived parameters. Lipfert and Wyzga (1996) and Watt et al. (1995) explored the problem of measurement error in environmental epidemiology studies, specifically with respect to airborne particles using data from Lioy et al. (1990) (specifically the correlations between outdoor and personal $PM_{10}$). Lipfert and Wyzga (1996) found that for a true $PM_{10}$ threshold of up to 150 μg/m$^3$, an underlying "hockey-stick" risk model would appear consistent with linear (no-threshold) models in the presence of independent variable error. They concluded that it is inappropriate to estimate health risks from agents with large amounts of exposure measurement error and that exposure error could mask nonlinearities

in true exposure-response functions, especially curvature and thresholds. They also found that estimated underlying thresholds are biased low. They wrote:

"If the variables that we are forced to work with (from fixed ambient monitors) already contain a lot of exposure error, no amount of analysis of this type can provide a remedy since the error cannot be removed. In such situations (which may include most of the PM studies), even sophisticated statistical analysis cannot impart real meaning to the data."

Motivated by the differences between personal and ambient $PM_{10}$ measurements for traffic wardens, Watt et al., (1995) also conducted simulations estimating human health risks from $PM_{10}$ in the presence of measurement error. Using the same data and parameters from Lioy et al., (1990) as Lipfert and Wyzga (1996) but a slightly different computational approach in which individual exposures were assumed to be lognormally distributed around the central/ambient exposure, they also showed that error can mask a true threshold function.

A more recent study best illustrates the effects of measurement error on assessing thresholds. (Brauer et al., 2002) conducted simulations in which a common individual threshold was varied and assessed the effect of measurement error on the ability to detect that threshold, or even a nonlinear exposure-response pattern (Figure 7). (Brauer et al., 2002) looked specifically at the magnitude of exposure measurement error of $PM_{2.5}$ and sulfate concentrations—that is, the magnitude of uncertainty in individual exposure measurements was drawn from actual cases of agents for which linear exposure-response patterns have been claimed.

Brauer et al., (2002) used a fairly general way to capture the relationship between the measured and actual independent variables:

$$X_{ij} = A_i + B_i W_j + \varepsilon_{ij} \tag{19}$$

where the subscripts $i$ and $j$ refer to subject $i$ on day $j$, $X$ is true exposure to either $PM_{2.5}$ or $SO_4^{2-}$, $W$ is measured exposure to $PM_{2.5}$ or $SO_4^{2-}$, and the parameters $A$, $B$, and $\varepsilon$ each have their own distributions based on regressing personal exposure data of $PM_{2.5}$ and sulfate of 16 subjects against corresponding ambient monitor data from Vancouver, Canada. This error model generalizes from the classical additive and multiplicative error models (Equations 1 and 2). (Brauer et al., 2002) used the segmented linear function with a linear link to represent individual risk, where each individual shared a common risk function and risk-related parameters were estimated from Vancouver-area data. Referring to Equation 15, $Y$ is the mortality rate, $\alpha$ is the local baseline mortality rate, and $\beta$ was determined using mortality risk estimate from the World Health Organization (WHO) Air Quality Guidelines.

For each possible level of ambient pollution and three different levels of the threshold parameter $\tau_0$, the true exposure was modeled and corresponding risks calculated and summed for the total simulation population. Importantly, the derived exposure-response relationship is what could be found by a study in which each true curve was affected by exposure measurement error of the magnitude actually observed to occur in real studies. In this study, personal exposure was more highly correlated with ambient exposure for sulfate than for $PM_{2.5}$, so comparison between the sulfate and $PM_{2.5}$ results illustrates the effect of error in the independent variable. Figure 7 shows the simulation results for $PM_{2.5}$ and $SO_4^{2-}$. The results for $PM_{2.5}$ show that the threshold is completely obscured for the lowest value of the threshold parameter (20 μg/m³). As the threshold parameter increases to 40 and 60 μg/m³, there is some evidence of nonlinearity but not at the correct threshold values. In contrast, the simulation results for $SO_4^{2-}$ do show smoothing of the underlying individual risk function but still offer evidence of nonlinearity and a fairly clear threshold response for all three levels of the threshold parameter (5, 10, and 15 μg/ m³). From both sets of figures, it is evident that when surrogate measures (ambient concentrations) are not highly correlated with

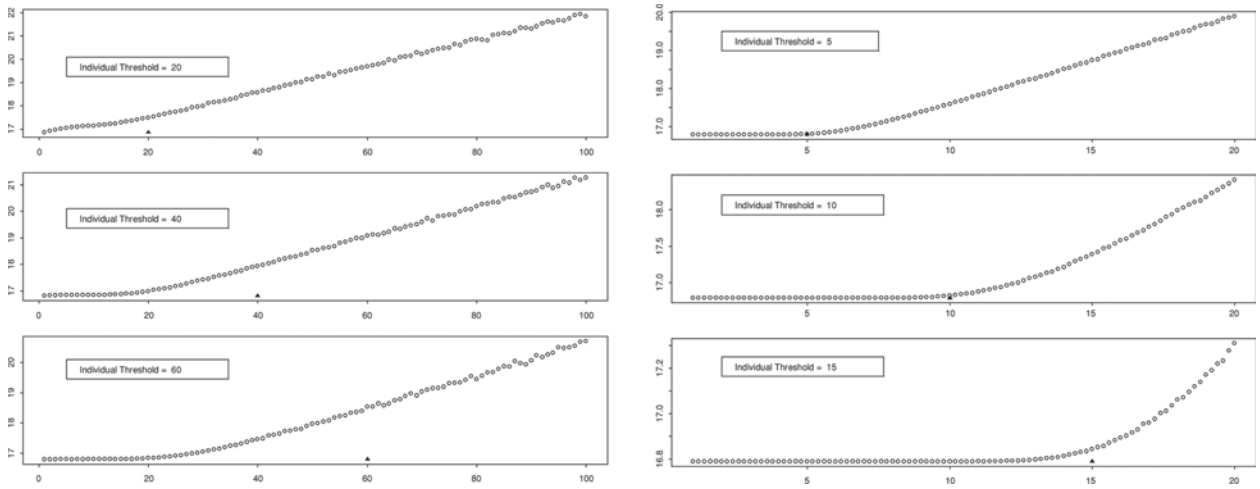


Figure 7. Simulation results reproduced from Brauer et al. (2002). The graphs show ambient concentration of $PM_{2.5}$ (left) or $SO_2$ (right) in μg/m³ on the $x$-axis and the expected number of deaths per 1,000,000 on the $y$-axis. The true underlying individual thresholds are indicated on the graphs as triangles.

personal exposures, a threshold can be masked at the population level even if there is a clear, common threshold at the individual level (individual risk function). Furthermore, even if the threshold is not completely masked, it is likely to be biased.

The simulations of (Brauer et al., 2002) as well as those in the previously mentioned papers, strongly suggest that the inability to detect a threshold in many epidemiology studies does not mean that no threshold exists. The apparent linearity of observed exposure-response relationships in studies with prevailing levels of exposure measurement error could be explicable as an artifact that arises from the effect of exposure measurement imprecision, and this has been recognized for at least two decades.

Methods for mitigating the effects of exposure measurement error have become an active area of research. Many techniques exist for possible correction, though they have varying measures of success for different error models and exposure distributions, as well as the shape of the underlying exposure-response curve. Techniques include a nonparametric smoothing technique in which a low-order polynomial is locally fit to the data (LOESS), simulation extrapolation (simex), regression calibration, and weighted nonlinear regression. These techniques perform differently under different assumptions and the literature detailing their strengths and limitations is substantive.

An example of the difficulties associated with error mitigating techniques has been provided by Cakmak et al., (1999). A thorough examination of this study is warranted because, in responding to a critique of (White et al., 2009), (Burke et al., 2009) cited the (Cakmak et al., 1999) study as evidence of modeling techniques that can identify potential thresholds, stating, "Although [a small range of exposures and measurement error] need to be considered in evaluating epidemiologic study results, modeling techniques such as nonparametric smoothing methods have demonstrated the capacity to identify potential threshold relationships even in the context of relatively extreme measurement error (Cakmak et al., 1999; Schwartz and Zanobetti, 2000)." It is our opinion that Cakmak's detection of thresholds is not totally reliable even with error mitigating techniques, which are not even clearly reproducible across different assumptions and parameter estimates.

Like Brauer, (Cakmak et al., 1999) evaluated the simulated association between air pollution and mortality and tried to distinguish thresholds in the presence of measurement error for air pollutants. They used a risk function that is an exponential transformation of the segmented linear threshold model and assumed multiplicative error in the measured covariate W such that $W = X \times U$ where U has mean 1 and constant variance $\sigma_u^2$. (Cakmak et al., 1999) characterized the error in the multiplicative factor, U, by varying $\sigma_u$ to achieve desired correlation coefficients between W and X, $\rho_{xw}$. The approach of Cakmak et al., (1999) was to simulate results from a study in which varying daily pollution levels, as measured by central monitoring stations, are compared to total non-accidental mortality numbers in a city on the same (or on the previous) day—a design that is often used in practice. Because no serial autocorrelation, seasonal components, or other reasons for non-independence from day to day are included, each day's result in essence corresponds to a simulated observation of a response rate at an exposure level measured with error, a large array of which yields observations that can be used to characterize the dose-response pattern. Cakmak et al., (1999) used the nonparametric smoothing technique, called LOESS, on the raw simulation output and evaluated the ability of such techniques to detect and characterize assumed thresholds.

Despite our close reading, we were unable fully to discern the exact nature of the exposure measurement error that (Cakmak et al., 1999) simulated; it could be Berkson, classical, or some mix of the two. The error was estimated from the degree to which single air monitoring stations in Toronto, Canada, differed from the readings at other monitors in the city on the same day—that is, it is a measure of geographic patchiness of daily pollutant concentrations (plus some effect of imprecision in the measuring instrument), rather than a direct measure of error in exposure to individuals. Although such patchiness would lead to variation among individual exposures in inhabitants of a city, it is not clear that such geographic variation was modeled, since the measurement error appears to have been applied as an error to the daily monitor reading to give the average ambient level to the subject population. These daily average ambient levels were then associated with daily total non-accidental mortality simulated according to a Poisson variation around a threshold dose-response curve. The simulated mortality and ambient level data were then compared to threshold and non-threshold models both visually and computationally to determine if error could mask the true underlying threshold dose-response curve. That is, (Cakmak et al., 1999) simulated the results of an ecologic time-series study of the sort often used to investigate short-term impact on mortality of daily fluctuations in air pollution as measured by central monitors, but in doing so, there appears to be no simulation of non-ambient sources or any other source of differences between average exposure concentration and average ambient levels. As is pointed out by (Zeger et al., 2000), such factors do come into play in actual studies, and they are the primary concern for classical (as opposed to Berksonian) error, and hence for the biasing of regression parameters and the threshold-disguising effect that we focus on in the present paper. That is, (Cakmak et al., 1999) appear to have simulated pure Berksonian error, which should enhance their ability to detect and characterize thresholds using used the nonparametric smoothing technique, LOESS, on the raw simulation output compared to real studies and simulations assuming classical error.

Despite this, the ability to detect and characterize exposure thresholds for added daily mortality was marginal. The authors showed that with simulations of threshold

models, although a threshold model was selected over a linear model based on the Akaike's information criterion (AIC) the majority of the time, this "majority" ranged from 52% to 90%, depending on the exposure error and threshold concentration in the simulated data. Thus, between 10% and 48% of the time, a linear model was incorrectly chosen over a threshold model. Even when the threshold model was chosen, the parameter estimates for the model were biased from the true parameters and generally biased upwards. Estimates of the threshold ranged anywhere from a few percent to 30% higher than the actual threshold, whereas estimates of the slope ranged from unbiased at zero error to almost 70% lower than the true slope for the most extreme amount of error. In addition, the standard errors for the estimated threshold parameters increased 5-fold for the scenarios with the most extreme error. Under the assumption of Berkson error (less bias yet more loss of power), one would expect even more error-prone results for classical error, and it is important to remember that, in practice, error is a mixture of classical and Berkson. The graphical results from simulations conducted by (Cakmak et al., 1999) are shown in Figure 8.

(Cakmak et al., 1999) illustrated that although certain techniques may aid in the detection of bias, they are imperfect. Specifically, the nonparametric algorithm they used requires a span parameter to determine how far away to look for points to contribute to the local regression. For this reason, specification of the weight function and span parameter are very important, and results can vary widely with different parameter choices. To determine the general utility of this technique, it would be useful to see the results of a "sensitivity" analysis by adjusting the span parameters. Furthermore, they stated that for their technique to work, the threshold of the exposure-response function must be at least as large as median of exposure distribution; the parameter estimates for the lower threshold models were not as consistent as those for the higher threshold models. A technique robust to parameter specification and exposure distributions could be a useful tool for assisting in threshold determination, though the reliability and reproducibility of the LOESS technique across different platforms, software, and algorithms would be important. Perhaps the overarching issue, however, is that the (Cakmak et al., 1999) results apply to their particular simulation of exposure measurement error, which they defined in a way that appears to comprise pure Berksonian error. This is in contrast to the more complex sources of error that apply to real studies, which (by including classical error as well as Berksonian error) would be expected to lead to a more pronounced masking of thresholds and a larger tendency to appear linear at low exposure levels despite actual biological thresholds.

Interestingly, though White et al. (2009) cited the (Cakmak et al., 1999) paper as a method for detecting thresholds in the presence of bias, they did not advocate use of error-mitigating techniques in determination of exposure-response curves. Model averaging has been suggested as another method of low-dose extrapolation, but this technique does not mitigate the presence of exposure measurement error.

The majority of the literature, both theoretical and experimentally based, indicates that measurement error results in the masking of a threshold, and that even when exposure measurement error is not large enough to completely linearize a truly threshold exposure-response relationship, bias still exists. Importantly, it has been shown that the degree of bias known to apply to actual studies is sufficient to produce a false linear result, and that although nonparametric smoothing and other error-mitigating techniques may assist in identifying a threshold, they do not guarantee detection of a threshold. Theoretical and computational literature are in agreement; the smoothing of the underlying segmented linear risk model is seen in the simulations of (Brauer et al., 2002) (Figure 7), as well as the analytical results from Kuchenoff and Carroll (1997) (Figure 5). In addition, because the risk function in (Cakmak et al., 1999) was well approximated by a first-order Taylor expansion and was essentially a linear link function in the vicinity of the thresholds, it allows for more direct comparison to the results of Brauer et al. (2002). The graphs from (Brauer et al., 2002) and (Cakmak et al., 1999) show similarities from visual inspection and the two papers share many similar conclusions regarding error in the independent variable in spite of/notwithstanding the LOESS technique. From the graphs, it is also clear that even if a threshold is found, it is likely an underestimate of the true threshold. Though (Cakmak et al., 1999) often reported estimated thresholds to be larger than true thresholds, this could be an artifact of the LOESS nonparametric smoothing technique.

## 7. Conclusions

In nonlinear regression, independent variable error results in biased parameter estimates and the masking of true features, as well as the loss of statistical power. It also tends to smooth out and flatten, essentially linearize, all of these curves, including threshold functions. Even when a threshold is detected, it is likely to be biased based on the independent variable error. Importantly, it has been shown that the degree of bias known to apply to actual studies is sufficient to produce a false linear result, and that although nonparametric smoothing and other error-mitigating techniques may assist in identifying a threshold, they do not guarantee detection of a threshold. Thus, exposure measurement errors as practically encountered in real environmental epidemiology data can result in biases that can affect the interpretation and use of the apparent exposure-response shapes in risk assessment applications. These errors result in an overestimation of risk at low exposures and an underestimation of risks at high exposures. The consequences of this could be great, as it could lead to a misallocation of resources towards regulations that do not offer any benefit to public health,

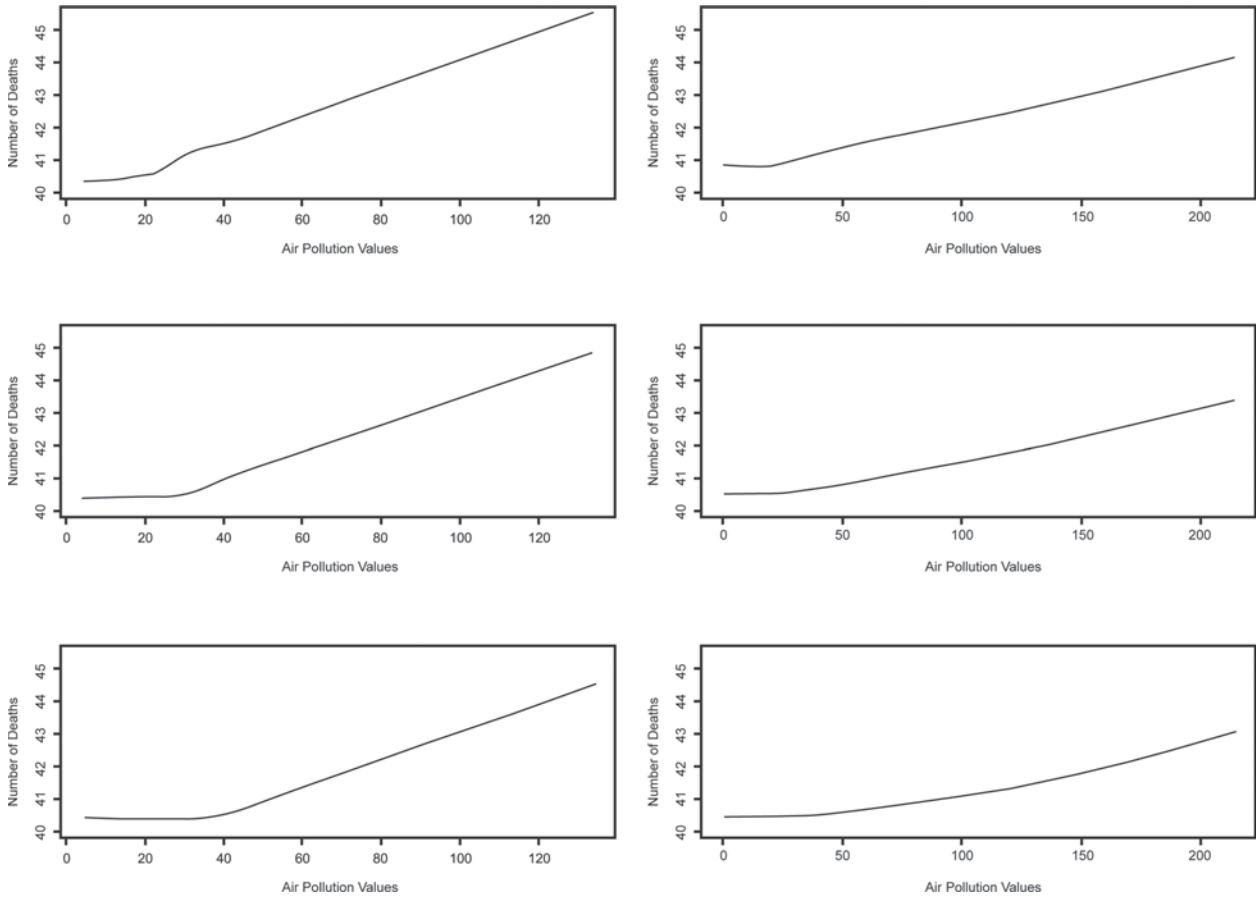


Figure 8. Simulation results reproduced from Cakmak et al. (1999). Pollution levels in μg/m³ are on each *x*-axis and number of deaths is indicated on each *y*-axis. All data in the graphs were processed by the LOESS smoothing technique. Graphs on the left present data simulated without measurement error and graphs on the right simulate data with measurement error. The uppermost graphs have an underlying threshold of 12.8 μg/m³; the middle graphs, 24.6 μg/m³; the bottom graphs, 34.4 μg/m³.

and may in fact cause harm owing to the underestimation of risks at higher exposures.

Our fundamental overall conclusion is that appropriate skepticism should be maintained about the true low-dose shapes of empirically based dose-response relationships in environmental epidemiology. The amount and nature of potential exposure measurement error, and its ability to mask nonlinearities and thresholds, need to be borne in mind. One can also refer to mode of action and biological considerations about the nature of disease processes to inform the likely properties of environmental agents at low exposure levels.

## Acknowledgment

The authors wish to thank Eric Wannamaker for his contributions to the manuscript.

## Declaration of interest

This paper was prepared with financial support to Gradient, a private environmental consulting firm, from ExxonMobil Biomedical Sciences, Inc. We have done similar work in the past for trade organizations and other clients, but these clients did not contribute funding for this work. The work reported in the paper was conducted during the normal course of employment by Gradient. The authors have the sole responsibility for the writing and contents of this paper. The views and opinions expressed are not necessarily those of ExxonMobil Biomedical Sciences, Inc.

## References

Adar SD, Gold DR, Coull BA, Schwartz J, Stone PH, Suh H. (2007). Focused exposures to airborne traffic particles and heart rate variability in the elderly. Epidemiology 18(1):95–103.

Allen R, Larson T, Sheppard L, Wallace L, Sally Liu LJ. (2003). Use of real-time light scattering data to estimate the contribution of infiltrated and indoor generated particles to indoor air. Environ Sci Technol 37:3484–3492.

Arbuckle TE, Hrudey SE, Krasner SW, Nuckols JR, Richardson SD, Singer P, Mendola P, Dodds L, Weisel C, Ashley DL, Froese KL, Pegram RA, Schultz IR, Reif J, Bachand AM, Benoit FM, Lynberg M, Poole C, Waller K. (2002). Assessing exposure in epidemiologic studies to disinfection by-products in drinking water: Report from an international workshop. Environ Health Perspect 110(Suppl 1):53–60.

Avery CL, Mills KT, Williams R, McGraw KA, Poole C, Smith RL, Whitsel EA. (2010a). Estimating error in using ambient PM2.5 concentrations as proxies for personal exposures. Epidemiology 21(2):1–9.

Avery CL, Mills KT, Williams R, McGraw KA, Poole C, Smith RL, Whitsel EA. (2010b). Estimating error in using residential outdoor

PM2.5 concentrations as proxies for personal exposures: A meta-analysis. Environ Health Perspect 118:673–678.

Bateson TF, Coull BA, Hubbell B, Ito K, Jerrett M, Lumley T, Thomas D, Vedal S, Ross M. (2007). Panel discussion review: Session three—Issues involved in interpretation of epidemiologic analysesmp_semiStatistical modeling. J Expo Sci Environ Epidemiol 17(Suppl 2):S90–S96.

Baxter LK, Wright RJ, Paciorek CJ, Laden F, Suh HH, Levy JI. (2010). Effects of exposure measurement error in the analysis of health effects from traffic-related air pollution. J Expo Sci Environ Epidemiol 20:101–111.

Brauer M, Brumm J, Vedal S, Petkau AJ. (2002). Exposure misclassification and threshold concentrations in time series analyses of air pollution health effects. Risk Anal 22:1183–1193.

Burke TA, Dominici F, Fox M, White RH. (2009). Linear low-dose extrapolation for non-cancer responses: Burke *et al.* respond. Environ Health Perspect 117:A142.

Cakmak S, Burnett RT, Krewski D. (1999). Methods for detecting and estimating population threshold concentrations for air pollution—Related mortality with exposure measurement error. Risk Anal 19:487–496.

Carroll RJ, Galindo CD. (1998). Measurement error, biases, and the validation of complex models for blood lead levels in children. Environ Health Perspect 106(Suppl 6):1535–1539.

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. (2006). Measurement Error in Nonlinear Models: A Modern Perspective. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC, Taylor & Francis Group.

Clark RR. (1982). The errors-in-variables problem in the logistic regression model. Unpublished PhD thesis, University of North Carolina, Chapel Hill.

Crump KS. (2005). The effect of random error in exposure measurement upon the shape of the exposure response. Dose Response 3:456–464.

Heid IM, Küchenhoff H, Miles J, Kreienbrock L, Wichmann HE. (2004). Two dimensions of measurement error: Classical and Berkson error in residential radon exposure assessment. J Expo Anal Environ Epidemiol 14:365–377.

Janssen NAH, Schwartz J, Zanobetti A, Suh HH. (2002). Air conditioning and source-specific particles as modifiers of the effect of PM10 on hospital admissions for heart and lung disease. Environ Health Perspect 110:43–49.

Küchenhoff H, Carroll RJ. (1997). Segmented regression with errors in predictors: Semi-parametric and parametric methods. Stat Med 16:169–188.

Kuha J, Temple J. (2003). Covariate measurement error in quadratic regression. Int Stat Rev 71:131–150.

Lioy PJ, Waldman JM, Buckley T, Butler J, Pietarinen C. (1990). The personal, indoor and outdoor concentrations of PM-10 measured in an industrial community during the winter. Atmos Environ 24:57–66.

Lipfert FW, Wyzga RE. (1996). The effects of exposure error on environmental epidemiology. In: Phalen RF, Mannix RC, Tonini MC, eds. Proceedings of the Second Colloquium on Particulate Air Pollution and Human Mortality and Morbidity. Irvine, CA: University of California, College of Medicine, 295–302. Available at: http://www.arb.ca.gov/research/abstracts/95-323.htm. Accessed 23 February 2010.

Meng QY, Turpin BJ, Polidori A, Lee JH, Weisel C, Morandi M, Colome S, Stock T, Winer A, Zhang J. (2005). PM2.5 of ambient origin: Estimates and exposure errors relevant to PM epidemiology. Environ Sci Technol 39:5105–5112. Available at: http://pubs.acs.org/doi/pdf/10.1021/es048226f. Accessed 17 February 2010.

Michalik JE, Tripathi RC. (1980). The effect of errors in diagnosis and measurement on the estimation of the probability of an event. J Am Stat Assoc 75:713–721.

National Research Council (NRC). (1998). Research Priorities for airborne Particulate Matter: I. Immediate Priorities and a Long-Range Research Portfolio (Executive Summary). Washington, DC: National Academy Press.

Sarnat JA, Schwartz J, Catalano PJ, Suh HH. (2001). Gaseous pollutants in particulate matter epidemiology: Confounders or surrogates? Environ Health Perspect 109:1053–1061.

Sarnat JA, Brown KW, Schwartz J, Coull BA, Koutrakis P. (2005). Ambient gas concentrations and personal particulate matter exposures: Implications for studying the health effects of particles. Epidemiology 16:385–395.

Sarnat JA, Wilson WE, Strand M, Brook J, Wyzga R, Lumley T. (2007). Panel discussion review: Session one—Exposure assessment and related errors in air pollution epidemiologic studies. J Expo Sci Environ Epidemiol 17:S75–S82.

Sarnat JA, Brown KW, Bartell SM, Sarnat SE, Wheeler AJ, Suh HH, Koutrakis P. (2009a). The relationship between averaged sulfate exposures and concentrations: Results from exposure assessment panel studies in four U.S. cities. Environ. Sci Technol 43:5028–5034. Available at: http://pubs.acs.org/doi/pdf/10.1021/es900419n. Accessed 18 September 2009.

Sarnat SE, Klein M, Sarnat JA, Flanders WD, Waller LA, Mulholland JA, Russell AG, Tolbert PE. (2009b). An examination of exposure measurement error from air pollutant spatial variability in time-series studies. J Expo Sci Environ Epidemiol 20:135–146.

Schwartz J, Zanobetti A. (2000). Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. Epidemiology 11:666–672.

Schwartz J, Sarnat JA, Coull BA, Wilson WE. (2007). Effects of exposure measurement error on particle matter epidemiology: A simulation using data from a panel study in Baltimore, MD. J Expo Sci Environ Epidemiol 17(Suppl 2):S2–S10.

Sheppard L, Slaughter JC, Schildcrout J, Liu LJ, Lumley T. (2005). Exposure and measurement contributions to estimates of acute air pollution effects. J Expo Anal Environ Epidemiol 15:366–376.

Stefanski LA, Carroll RJ. (1985). Covariate measurement error in logistic regression. Ann Stat 13:1335–1351.

US EPA. (2006a). Air Quality Criteria for Lead. Volumes I and II (Final). Washington, DC: US EPA. Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid = 158823. Accessed October 2006.

US EPA. (2006b). Air Quality Criteria for Ozone and Related Photochemical Oxidants. Volume I. Washington, DC: US EPA. EPA 600/R-05/004aF, February.

US EPA. (2008). Integrated Science Assessment for Oxides of Nitrogen. Washington, DC: US EPA. EPA/600/R-08/071, July.

US EPA. (2009). Integrated Science Assessment for Particulate Matter (Final). Washington, DC: US EPA. EPA/600/R-08/139F. December. Available at: http://www.epa.gov/ncea/pdfs/partmatt/Dec2009/PM_ISA_full.pdf. Accessed 3 February 2010. Includes errata sheet created on February 10, 2010.

Wallace L, Williams R. (2005). Use of personal-indoor-outdoor sulfur concentrations to estimate the infiltration factor and outdoor exposure factor for individual homes and persons. Environ Sci Technol 39:1707–1714. Available at: http://pubs.acs.org/doi/pdf/10.1021/es049547u. Accessed 20 January 2009.

Watt M, Godden D, Cherrie J, Seaton A. (1995). Individual exposure to particulate air pollution and its relevance to thresholds for health effects: A study of traffic wardens. Occup Environ Med 52:790–792.

White RH, Cote I, Zeise L, Fox M, Dominici F, Burke TA, White PD, Hattis DB, Samet JM. (2009). State-of-the-science workshop report: Issues and approaches in low-dose-response extrapolation for environmental health risk assessment. Environ. Health Perspect 117:283–287.

Wright JM, Bateson TF. (2005). A sensitivity analysis of bias in relative risk estimates due to disinfection by-product exposure misclassification. J Expo Anal Environ Epidemiol 15:212–216.

Yoshimura I. (1990). The effect of measurement error on the dose-response curve. Environ Health Perspect 87:173–178.

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.