Estimating Short-term PM Effects Accounting for Surrogate Exposure Measurements from Ambient Monitors

Lianne Sheppard Doris Damian

NRCSE

Technical Report Series

NRCSE-TRS No. 035

The **NRCSE** was established in 1996 through a cooperative agreement with the United States Environmental Protection Agency which provides the Center's primary funding.

EPA

Estimating Short-term PM Effects Accounting for Surrogate Exposure Measurements from Ambient Monitors

Lianne Sheppard and Doris Damian

August 17, 1999

ABSTRACT

We outline a model for the health effects of air pollutants on individuals given true pollutant exposure. This is aggregated to form the typical group-level analysis of pollutant effects. However only surrogate exposure data are available from ambient monitors. We therefore also consider the exposure distribution and measurement characteristics of the pollutant data. We combine these with the disease model of interest and discuss estimation of the exposure effect in the presence of the additional modeling. We apply these to asthma hospital admissions data from Seattle. We note that these analyses are limited by the restriction to ambient monitor data. While we set up a framework for incorporating personal exposures, estimation cannot proceed without data beyond those available to us.

Key words: air pollution epidemiology, exposure measurement error, aggregate data, particulate matter, health effects

Acknowlegements: This work was partially supported by grants CR825173-01-0 from the US Environmental Protection Agency and ES08062-03 from the National Institute of Environmental Health Sciences, NIH. The contents are solely the responsibility of the authors and do not necessarily represent the official views of EPA or NIEHS.

1 Introduction

The U.S. Clean Air Act mandates regulation of ambient air pollutants to protect public health and welfare from any actual or potential adverse effects. The ubiquity of air pollutant exposures in the U.S. population and this legislative mandate have spawned numerous population-based health effect studies. Most of these are opportunistic studies using an ecologic time series design. Typically they assess the association between ambient air pollutant measurements and health event counts over time within a single geographic area. Little attention has been given to the ability of these studies to accurately assess the putative individual health effects of interest.

Epidemiologists are nearly always interested in estimating the individual-level (as opposed to group-level) effects of exposure on disease. This argues for using individual-level study designs with individual-level data. However, most often in the air pollution field this approach is not feasible since the costs of obtaining personal pollutant exposure data in a population are prohibitive and many of the outcomes of interest are rare. One alternative is to use an aggregate study design that preserves the individual-level parameter interpretation. One initially specifies a plausible *disease model* at the individual level and then sums over individuals to produce an aggregate model. Most naturally the model will be conditional on an individual's true personal pollutant exposure. However, true personal exposures cannot be observed and personal exposure measurements will not distinguish between ambient source and non-ambient source exposures, a distinction that is clearly relevant to the Clean Air Act mandate. Further, often the best available exposure data are surrogate exposures taken from fixed-site ambient monitors. In order to use the available data and still account for the true exposure data we wish we had, one can expand the modeling to incorporate two additional models, an *exposure distribution* and a *measurement error* model. We assume that only the disease model incorporates the exposure effect parameters of interest. Since the true exposure cannot be observed, it is necessary to model the exposure distribution in order to obtain the more general model of the disease conditional on the measured exposure.

In this paper we will specify plausible disease, measurement, and exposure distribution models with the goal of estimating the health effects of air pollutants. We will propose a quasi-likelihood approach for estimation. We will apply our models and approach to a dataset of asthma hospital admissions in Seattle. By making comparisons with previous work, we will be able to compare estimates conditioned on measured ambient exposure data with those that focus directly on the personal exposure of interest and suggest areas for further research.

2 Approach

We wish to estimate how much an increase in pollution affects the risk of disease morbidity or mortality in the population. We begin by specifying an individual disease model for the probability of a particular outcome and an overdispersed variance model as

$$\begin{aligned} \mathrm{E}(Y_{it}|X_{it}, Z_{it}, b_{it}; \theta) &= \lambda_{it} \\ &= \lambda_{0t} b_{it} \exp(X_{it}\beta + Z_{it}\gamma) \\ \mathrm{var}(Y_{it}|X_{it}, Z_{it}, b_{it}; \theta) &= \sigma_\lambda^2 \lambda_{it}(1 - \lambda_{it}). \end{aligned}$$

Here Y_{it} is an outcome indicator for individual i on day t, λ_{0t} is the baseline disease probability, b_{it} is an individual frailty or susceptiblity to disease, $\exp(\cdot)$ the relative risk function with exposures X_{it} and confounders Z_{it}, and $\theta = \{\beta, \gamma\}$ are the parameters to be estimated. For simplicity here we will assume that the confounders vary only as a function of time, i.e. $Z_{it} \equiv Z_t$. We may allow the confounders to enter as arbitrary smooth functions, while typically the pollutants will enter as linear effects. Our main focus is on the exposure effect parameters β for the true exposure to ambient pollutants, X_{it}.

Conditional on the covariates and the frailty, we assume the Y_{it} are independent. We will assume the frailties can be characterized simply. Let the individual frailties be independent over individuals with moments

$$\begin{aligned} \mathrm{E}(b_{it}) &= 1 \\ \mathrm{var}(b_{it}) &= \sigma_b^2, \end{aligned}$$

independent of t and X_{it}. The independence of individual susceptibility with exposure levels will need careful consideration in specific applications. For some pollutants it is likely that some individuals, particularly those most susceptible to air pollution effects, will modify their pollution exposure on high air pollution days. The independent frailty assumption may be reasonable for particulate matter (PM) or carbon monoxide (CO) exposures, both of which have high penetration rates. Thus exposure to ambient air pollutants in outdoor air will not be very different from indoor air implying that individual behavior will not substantially impact individual exposure to ambient source levels. This same independent frailty assumption will be poorly justified for exposures such as ozone (O_3) where air pollution alerts are broadly advertized and indoor penetration rates are low so that individual behavior will have a significant impact on individual exposure.

The individual data, Y_{it}, are presence or absence of an event for individual i on day t in geographic region $\mathcal{S}$. However, the outcome data are typically only available as event counts on day t. Thus we need to model the aggregate outcomes $Y_t = \sum_{i=1}^{n_t} Y_{it}$ in a population of

size n_t. Thus

$$\begin{aligned} \mathrm{E}(Y_t|X_{it}, Z_t, b_{it}\forall i = 1, \ldots, n_t; \theta) &= \sum_{i=1}^{n_t} \lambda_{it} \\ &= n_t\overline{\lambda}_t \\ \mathrm{var}(Y_t|X_{it}, Z_t, b_{it}\forall i = 1, \ldots, n_t; \theta) &= \sigma_\lambda^2(n_t\overline{\lambda}_t - \sum_{i=1}^{n_t} \lambda_{it}^2) \end{aligned}$$

Given our frailty model assumptions we can marginalize over susceptibility:

$$\begin{aligned} \mathrm{E}(Y_t|X_{it}, Z_t\forall i = 1, \ldots, n_t; \theta) &= \lambda_{0t} \sum_{i=1}^{n_t} \exp(X_{it}\beta + Z_t\gamma) \\ &= n_t\overline{\tilde{\lambda}}_t \\ \mathrm{var}(Y_t|X_{it}, Z_t\forall i = 1, \ldots, n_t; \theta) &= \sigma_\lambda^2(n_t\overline{\tilde{\lambda}}_t + (\sigma_b^2 - \sigma_\lambda^2(1+\sigma_b^2))\sum_{i=1}^{n_t} \tilde{\lambda}_{it}^2) \end{aligned}$$

where $\tilde{\lambda}_{it} = \lambda_{0t}\exp(X_{it}\beta + Z_t\gamma)$

Analysis based on this model is still not feasible since personal exposure to ambient pollutants, X_{it}, typically is not measured for any, not to mention all individuals in a defined geographic area. Instead one can obtain ambient measurements from one or more sites within the geographic area of interest. Therefore we must link the ambient pollution measurements in spatial location s on day t, W_{st}, to the unmeasured true individual exposures X_{it}. We make three additional assumptions a this point. First we assume that for all ambient measurements $\mathbf{W} = \{W_{st} : s \in \mathcal{S}, t = 1, \ldots, T\}$, $\mathbf{W}$ is a true surrogate for X_{it} so that

$$\mathrm{E}(Y_t|X_{it}, Z_t, \mathbf{W}\ \forall i = 1, \ldots, n_t; \theta) = \mathrm{E}(Y_t|X_{it}, Z_t; \theta).$$

Second, we assume the pollutant effect and thus $X_{it}\beta$ will be small so we can use a polynomial approximation for $\exp\{X_{it}\beta\}$ yielding

$$E(\exp\{X_{it}\beta\}|Z_t, \mathbf{W}; \tau) \doteq 1 + \sum_{j=1}^{J} \mathrm{E}(X_{it}^j|Z_t, \mathbf{W}; \tau)\beta^j/j!. \qquad (2{\cdot}1)$$

This means we need only characterize the moments of $X_{it}|Z_t, \mathbf{W}$ in order to substitute ambient pollution measurements for true personal pollution exposures in the disease model. We will use this approximation for lognormally distributed predictors where the moment generating function does not converge. It will be important to assess the contribution of higher moments in any particular application. Finally, to greatly simplify our variance model, we assume the disease probability is sufficiently small so that any terms involving λ_{it}^2 will be negligible.

Then we may write the aggregate disease model conditional on ambient pollution measurements as

$$\begin{aligned} \mathrm{E}(Y_t|\mathbf{W}, Z_t) &= \mu_t \\ &= \lambda_{0t} \exp\{Z_t\gamma\} \sum_{i=1}^{n_t} [1 + \sum_{j=1}^{J} \mathrm{E}(X_{it}^j|Z_t, \mathbf{W}; \tau)\beta^j/j!] \\ \mathrm{var}(Y_t|\mathbf{W}, Z_t) &\doteq \sigma_\lambda^2 \mu_t. \end{aligned}$$

The term $\lambda_{0t} \exp\{Z_t\gamma\}$ includes the baseline risk, known offsets, and confounding factors.

Note that these define the mean and variance of an overdispersed Poisson model. They differ from the typical model used in ecologic time series studies of air pollution effects in how the ambient pollutant data are incorporated. This model indicates that one needs the expected individual exposure given ambient measurements as well as higher order moments. These will vary as a function of time and thus cannot be neglected if one wants an unbiased estimate of the exposure effect parameter β. Our goal in this paper is to evaluate the impact of ignoring the exposure modeling in the analysis of air pollution health effect studies. We will formulate a model for the air pollution exposure, use these quantities in the above regression model, and compare our results with those obtained when the exposure distribution is ignored.

3 Exposure Model

Exposure modeling poses significant challenges because the true exposure and its distribution are not observed. Thus we will need to rely on assumptions, and we expect only some of these may be verified independently. We emphasize that our interest is in the ambient source personal exposure. This differs from total personal exposure, the type of personal exposure that can be feasibly measured with current technology. We hypothesize the following model for true personal exposure to person i on day t:

$$X_{it} + I_{it} = \delta_{it} A_{s_i t} + I_{it}$$

where X_{it} is the true personal exposure to ambient source pollutants, I_{it} is the air pollutant exposure to person i on day t from non-ambient sources (e.g. the personal cloud and indoor sources), $A_{s_i t}$ is the true ambient level at location s_i and time t, and δ_{it} is the fraction of ambient air person i is exposed to on day t for $0 \leq \delta_{it} \leq 1$. We further assume that each person i is confined to a single location s. This implies notationally that s_i is equivalent to i.

We note now that even if we specify the disease model conditional on the true personal exposure $X_{it} + I_{it}$, we can dispense with the non-ambient source component of personal exposure in this design. Since the only available exposure measurements are from ambient monitors, there is no information in the data about this source. More important, since we

are aggregating over individuals, under certain conditions the non-ambient source exposure will not bias the effect estimate in an aggregate model. Key properties to verify are:

$$\begin{aligned} \mathrm{E}(I_{it}|\mathbf{W}) &= \mu_I \\ \mathrm{var}(I_{it}|\mathbf{W}) &= \sigma_I^2. \end{aligned}$$

When neither of these vary over time, they do not impact modeling in an aggregate time series design.

The other two components of the true exposure to ambient air pollutants are the fraction of ambient air exposure, δ_{it}, and the true ambient level $A_{s_i t}$. Most appropriately the ambient level will vary over space. While we will define exposure variables spatially, for the purposes of this paper we will make the further simplifying assumption that there is no spatial variation, i.e.

$$A_t = A_{s_i t}.$$

We also expect δ_{it} to vary over individuals and time. Work by Mage et al. (1999) suggests a reasonable model will be a function of time spent outdoors, indoor air exchange rates, and pollutant-specific penetration and deposition parameters:

$$\delta_{it} \doteq r_{it} + (1 - r_{it}) P a_{it}/(k + a_{it}),$$

where r_{it} is the fraction of time an individual spends outdoors, P is the penetration rate for a specific pollutant ($0 < P < 1$), a_{it} is the number of air exchanges per hour, and k is the pollutant-specific deposition rate. We assume that $\delta_{it} = r_{it} = 1$ for any individual who spends their entire 24-hour day outdoors and $\delta_{it} = P a_{it}/(k + a_{it}) < 1$ for any individual who spends a whole day indoors. We reasonably expect δ_{it} to vary seasonally since people spend more time outdoors in the warm months than in the cold season. Air exchange rates vary by individual and seasonally depending upon use of open windows and building-specific ventilation system characterisitcs. Penetration rates are a function of individual pollutants with fine particulates and carbon monoxide having high penetration rates while ozone and sulfur dioxide bind readily to surfaces and thus don't penetrate well indoors. This model suggests analyses that rely exclusively on ambient monitoring measurements will produce biased health effect estimates for pollutants, subpopulations, and/or seasons when $\mathrm{E}(\delta_{it}) < 1$.

We currently have no data on δ_{it}. This is a potentially fruitful area of research that will improve health effect estimates from air pollution studies. We will make the strong (and poorly justified) assumption here that $\delta_{it} \equiv \delta = 1$.

We will also need a measurement model for the ambient pollutant measurements indexed by space and time. Define W_{st} to be the measured ambient air pollutant level at time t and spatial location s, A_{st} to be the "true" global ambient air pollutant level on day t at location s, L_{st} to be local source contributions to the ambient monitor, and U_{st} the measurement

error. In this paper we will average over space and use $\overline{W}_t$ for W_t. (We will drop the overbar from further notation.) Assume

$$\begin{aligned} W_t &= A_t\ L_t\ U_t \\ \mathrm{E}(L_t\ U_t) &= 1 \\ \mathrm{var}(L_t\ U_t) &= \exp(\sigma_W^2\sigma_e^2) - 1 \\ \mathrm{cov}(L_t\ U_t, L_{t'}\ U_{t'}) &= 0 \quad \forall t \neq t' \end{aligned}$$

Temporal dependence in the local effects will not be separable from the global effects for a single pollution time series. We expect better resolution from spatio-temporal data; the potential for better spatio-temporal modeling to improve the health effect estimate remains to be explored.

Our experience suggests (see Section 6) that a reasonable model for W_t depends upon daily covariates, Z_t, a smooth function of time with n degrees of freedom, $s(t,n)$, has a zero mean with constant variance measurement error component, $\sigma_W e_t$, and has a residual correlation of up to 4 days for the stationary residuals $\sigma_W\epsilon(t)$. We will assume a model with nonzero autocorrelation for three lags. Thus our model is

$$\log(W_t)|Z_t = \mu(t,Z_t) + \sigma_W(\epsilon(t) + e_t) \tag{3·1}$$

$$\mu(t,Z_t) = Z_t\eta + s(t,n) \tag{3·2}$$

where $\epsilon(t)$ and e_t have independent multivariate normal distributions with zero mean, and

$$\begin{aligned} e_t &\sim i.i.d.\ N(0,\sigma_e^2)\ \forall\ t \\ \mathrm{var}(\epsilon(t)) &= 1 - \sigma_e^2 \\ \mathrm{corr}(\epsilon(t),\epsilon(t-1)) &= \rho_1 \\ \mathrm{corr}(\epsilon(t),\epsilon(t-2)) &= \rho_2 \\ \mathrm{corr}(\epsilon(t),\epsilon(t-3)) &= \rho_3 \\ \mathrm{corr}(\epsilon(t),\epsilon(t-t')) &= 0\ \forall\ t' > 3. \end{aligned}$$

This suggests we define the components of W_t as

$$\log(A_t)|Z_t = \mu(t,Z_t) + \sigma_W(\epsilon(t) + \sigma_W\sigma_e^2/2), \tag{3·3}$$

and

$$\log(L_t\ U_t) = \sigma_W(e_t - \sigma_W\sigma_e^2/2). \tag{3·4}$$

This definition constrains $\mathrm{E}(L_t\ U_t) = 1$. Thus the exposure and measurement model parameters τ include the mean parameters η and parameters making up $s(\cdot)$ as well as the variance parameters $\rho_1, \rho_2, \rho_3, \sigma_W^2, \sigma_e^2$.

Given $X|\mathbf{W}$ is assumed to be lognormal, this implies

$$\mathrm{E}(X_{it}^j|Z_t,\mathbf{W};\tau) = \exp\{j\mu_{X|W} + j^2\sigma_{X|W}^2/2\}$$

where $\mu_{X|W}$ and $\sigma_{X|W}^2$ are the conditional mean and variance of $X_t|\mathbf{W}$.

4 Estimaton

Quasi-likelihood estimating equations can be used to estimate the disease model parameters θ while treating the exposure and measurement model paremeters τ as nuisance parameters. In general terms, we write the original likelihood for the data as

$$\begin{aligned} L_O(\theta,\tau) &= f(Y,W|Z;\theta,\tau) \\ &= \int_X f_1(Y|X,Z;\theta)\ f_2(W|X,Z;\tau)\ f_3(X|Z;\tau)\ dX. \end{aligned}$$

where $f_1(Y|X,Z;\theta)$ is the disease model, $f_2(W|X,Z;\tau)$ is the measurement model, and $f_3(X|Z;\tau)$ is the exposure model. Liang & Liu (1991) propose basing the estimation on the modified likelihood

$$L_M(\theta,\tau) = f(Y|W,Z;\theta,\tau) = \int_X f_1(Y|X,Z;\theta) f(X|W,Z;\tau) dX.$$

since $L_O(\theta,\tau) = L_M(\theta,\tau) f(W|Z;\tau)$ and it is reasonable to assume that $L_M(\theta,\tau)$ carries little information about τ. Thus we can replace τ by an estimate and maximize the pseudo-likelihood $L_M(\theta,\hat{\tau})$. When $E(Y|W,Z;\tau)$ and $\text{var}(Y|W,Z;\tau)$ are easy to calculate, we may avoid integrating over X and instead apply estimating equations. The score function based on $L_M(\theta,\hat{\tau})$ is

$$U(\theta,\hat{\tau}) = \sum_{t=1}^{T} U_t(\theta,\hat{\tau}) = \sum_{t=1}^{T} \{D_\theta\}_t V_t^{-1}(Y_t - \mu_t),$$

for $\mu_t = E(Y_t|\mathbf{W}_t,\mathbf{Z}_t;\theta,\hat{\tau})$, $\{D_\theta\}_t = \partial\mu_t/\partial\theta$, and $V_t = \text{var}(Y_t|\mathbf{W}_t,\mathbf{Z}_t;\theta,\hat{\tau})$. In order to appropriately estimate the variance of $\hat{\theta}$, we must account for estimation of τ and any residual time dependence induced by modeling the exposure. Liang & Liu (1991) showed that for independent data, $\sqrt{T}(\hat{\theta}-\theta)$ is asymptotically normal with mean 0 and variance V_Q derived from partitioning the joint distribution of $U(\theta,\hat{\tau})/T$ and $\hat{\tau}-\tau$. By substituting empirical estimates we obtain

$$\hat{V_Q} = B^{-1}\ [C + 2A\ \text{cov}(U_t(\theta,\tau),\hat{\tau}) + A\ \text{cov}(\hat{\tau})\ A^T]\ B^{-1}$$

where

$$\begin{aligned} B &= \frac{1}{T}\sum_{t=1}^{T} \{D_\theta\}_t^T\ V_t^{-1}\ \{D_\theta\}_t \\ C &= \frac{1}{T}\sum_{i,j=1}^{T} U_i(\theta,\tau)^T\ U_j(\theta,\tau) \\ A &= \frac{1}{T}\sum_{t=1}^{T} \{D_\theta\}_t^T\ V_t^{-1}\ \{D_\tau\}_t \end{aligned}$$

for $\{D_\tau\}_t = \partial\mu_t/\partial\tau$ and estimates of $\hat{\theta}, \hat{\tau}$ are substituted for θ, τ. When the secondary parameter τ is assumed to be known, the second two terms in the variance drop out leaving only the sandwich variance estimate, $B^{-1}CB^{-1}$. Even when τ is estimated, typically we will assume that $\text{cov}(U_t, \hat{\tau}) = 0$ since $\hat{\tau}$ does not depend on Y_t, leaving only the first and last terms in the variance. Lumley & Heagerty (1999) note that for correlated data, the plug-in estimate C will tend to 0. They propose weighted estimates in order to correct for this inconsistency:

$$\hat{C} = \frac{1}{T}\, w^* \sum_{i,j=1}^{T} w_{ij} U_i(\theta, \hat{\tau})^T U_j(\theta, \hat{\tau})$$

where $w^* \geq 1$, $w_{ii} = 1$ and $w_{ij} \leq 1$ for $i \neq j$. We apply their weighted empirical adaptive variance estimators (WEAVEs) with truncated weights.

All that remains is to estimate the nuisance parameters τ and their covariances. We estimate the secondary mean parameters, η and the parameters of the regression spline, $s(.)$, using least squares. In estimates of their covariances, we incorporate the temporal dependence (Seber (1977)). We use the sample variance of the residuals as an estimate of σ^2_W. The variance of this estimate is obtained through direct implementation of theorem 1.8 in Seber (1977). After dividing the residuals by the estimated σ_W, we obtain sample estimates for (ρ_1, ρ_2, ρ_3), and then apply Bartell's formula (Brockwell & Davis (1991)) to estimate their covariance matrix. We assume that the mean and the variance estimates are uncorrelated, as well as the estimate of σ^2_W with the estimates of (ρ_1, ρ_2, ρ_3). We estimate σ^2_λ from the usual Pearson residuals.

5 Simulation Studies

We studied the effect of monitor measurement error in a series of simulation studies. We simulated 3-year ($T = 1096$) exposure, measurement and disease outcome series using parameters derived from an 8-year Seattle time series (Sheppard et al. (1999)). The measured PM, W_t, was modeled as given in (3·1). We assumed the PM exposure mean was known with effects for stagnation, season, and day of week (3·2). We defined stagnation to be the number of hours per day that the wind speed was below the 25th percentile for all hourly measurements from that monitor. We used regression splines to model season with 3 degrees of freedom per year in the 3-year series. Day of week was a factor variable. The stochastic component of the exposure, $\sigma_W \epsilon(t)$, had parameters $\sigma^2_W = 0.11$, $(\rho_1, \rho_2, \rho_3) = (0.45, 0.24, 0.15)$ or $(0.60, 0.40, 0.05)$. We studied measurement error variances σ^2_e of 0.1 and 0.2. We again assumed the fraction of ambient exposure parameter, δ_{it}, was fixed at 1.

We assumed a simple disease model with no confounding factors. Asthma hospitalizations had a Poisson distribution with mean $\exp(\alpha + X_t\beta)$ for $\alpha = .888$ and $\beta = .003$. In the estimation we used the first six conditional moments in the approximation (2·1).

Table 1 gives estimated parameter estimates, standard error estimates and coverage probabilities for models using the known and measured PM exposures as well as from the model using expected personal exposure. The first standard error estimate is estimated from the distribution of $\hat{\beta}$. The second is the square root of the average of the asymptotic variance estimates. All coefficient and standard error estimates are scaled by 10^3. Under the four conditions studied, there is a small amount of attenuation bias in the measured predictor while the expected exposure conditional on the measurements has some positive bias, particularly with the larger measurement error variance of 0.2. Overall however, all approaches give estimates close to the true value with good coverage.

6 Analysis

In this section we did an analysis on data collected in Seattle for eight years (1987-1994) similar to that reported by Sheppard et al. (1999). The outcome of interest was the count of asthma hospital admissions for all residents in a subset of King County below age 65. We used the ambient exposure data directly as a predictor, as well as expected exposure conditional on ambient measurements. Here we restricted the exposure measurements to PM_{10} and included temperature as a confounder, time and day of the week as both confounders and as exposure predictors, and stagnation as an additional exposure predictor. We defined stagnation as given in the previous section by site and then averaged over all sites. This was a strong predictor in the exposure model (see Figure 1) and its inclusion in the model reduced the residual autocorrelation and removed much of the residual structure in the autocorrelation (Figure 2). We used the same exposure model described in the previous section. Using replicate measurements from two monitoring sites, we estimated the pure measurement error variance to be only 0.02. We were concerned that this estimate does not account for any local area variation in the measurement error, so we examined σ_e^2 at 0.02, 0.1, and 0.2 in order to allow up to 20% of the variance due to local spatial variation.

In the disease model we adjusted for seasonal variation using a 32 degree of freedom regression spline and allowed temperature to vary as a 4 degree of freedom regression spline. Day of the week entered as a factor variable, yielding 44 parameters in the disease model.

Table 2 gives estimates of the PM_{10} health effect parameter and standard error using the ambient data and the expected personal exposure conditional on the ambient data given the three measurement error variance assumptions. The exposure effect parameter estimate and its standard error increase with the size of the assumed measurement error variance. The change is small and does not influence the relative risk estimate.

7 Discussion

We have proposed an approach to modeling short-term pollutant exposure effects of ambient source personal exposure using ambient monitoring data. We began development from an

individual-level disease model conditional on true personal exposure. Lacking appropriate data, aggregation of the individual-level disease model and additional modeling assumptions were necessary. Interpretation of the exposure effect parameter is preserved under this development as long as the hypothesized models and simplifying assumptions are valid. By clearly defining a set of models and underlying assumptions we create a framework for discussion among PM researchers, expose assumptions needing verification, and identify gaps in current data. We currently lack data to assess the impact of many of our simplifying assumptions.

In our simulation study and Seattle analysis we observed that the correction for pure measurement error without any local spatial or personal exposure variation yielded a trivial change in the exposure effect parameter estimate. If this captured all the sources of variation and error, then ecologic time series health effects studies as currently practiced do a reasonable job of estimation air pollution health effects. However, we do not believe that the replicate ambient measurements within sites we used to estimate the measurement error adequately captured the unexplained variation in true personal exposure. In particular, in our analyses there was no exposure variation across individuals. Explicitly including the aggregation became unnecessary under this simplification. Incorporation of spatial variation in the ambient PM levels along with individual-specific ambient air exposure fractions will increase the within-day personal exposure variation of expected personal exposures. An aggregate disease model with time- and individual-specific personal exposures can estimate the relative risk parameter specified in an individual-level exponential relative risk model. Further research is necessary to understand the impact of incorporating a spatio-temporal personal exposure model and the ability of the simpler ecologic time series studies to adequately estimate this parameter.

This work suggests there is considerable room for improved personal exposure studies. These studies should focus on estimation of the parameters in the personal exposure model and verification of key assumptions regarding source apportionment, fraction of outdoor air exposure, and independence of fraction of outdoor air exposure with other factors, namely individual frailties and spatial variation in ambient levels. The emphasis of such studies should be on capturing the major sources of exposure variation in order to improve health effect parameter estimates rather than focusing on optimization of exposure predictions per se.

For estimation we used a quasi-likelihood estimating equations approach. This approach relies on specification of only mean and variance models for the disease model of interest. However when incorporating a measurement error model, additional modeling assumptions for the exposure and measurement error distribution are unavoidable, diminishing the natural appeal of quasi-likelihood. An alternative would be to use a Bayesian approach.

References

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, New York, second edition.

Liang, K.-Y. & Liu, X.-H. (1991). *Estimating Functions*, chapter 4: Estimating Equations in Generalized Linear Models with Measurement Error, pages 47–63. Oxford University Press, Oxford. V. P. Godambe, editor.

Lumley, T. & Heagerty, P. (1999). Weighted empirical adaptive variance estimators for correlated data regression. *Journal of the Royal Statistical Society, Series B* **61**, 459–477.

Mage, D. T., Wilson, W., Hasselblad, V., & Grant, L. (1999). Assessment of human exposure to ambient particulate matter. *Journal of The Air Waste Managment Association* in press.

Seber, G. A. F. (1977). *Linear Regression Analysis*. John Wiley & Sons.

Sheppard, L., Levy, D., Norris, G., Larson, T. V., & Koenig, J. Q. (1999). Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994. *Epidemiology* **10**, 23–30.

Table 1
Simulation Studies (150 sims): Properites of the Exposure Effect Parameter ($\beta = 3 \times 10^{-3}$) under Different Exposure and Measurement Distribution Assumptions and Using Different Exposure Metrics

		$\sigma_e^2 = 0.1$			$\sigma_e^2 = 0.2$		
$(\sigma_W^2, \rho_1, \rho_2, \rho_3)$		$\hat{\beta}$	$(\sqrt{\text{var}}\hat{\beta}\ ;\ \sqrt{\widehat{\text{var}}}\hat{\beta})$	95% CI Coverage	$\hat{\beta}$	$(\sqrt{\text{var}}\hat{\beta}\ ;\ \sqrt{\widehat{\text{var}}}\hat{\beta})$	95% CI Coverage
(.11,.45,.24,.15)	W	2.90	(1.03 ;1.06)	97	2.82	(0.92 ;1.00)	97
	$X\|W$	3.10	(1.10 ;1.12)	97	3.27	(1.08 ;1.17)	93
	X	3.02	(1.04 ;1.08)	97	3.05	(0.95 ;1.05)	98
(.11,.60,.40,.05)	W	2.83	(1.13 ;1.07)	95	2.77	(1.01 ;1.00)	96
	$X\|W$	3.01	(1.20 ;1.14)	95	3.19	(1.15 ;1.15)	95
	X	2.94	(1.13 ;1.10)	96	3.02	(1.06 ;1.04)	95

Table 2

Health Effect Parameter Estimates ($\times 10^3$) Using Different Exposure Measures

Exposure Estimate	$\hat{\beta}$	SE ($\hat{\beta}$)	95% CI
W	1.88	0.70	(0.51,3.25)
$X\|W, \sigma_e^2 = 0.02$	1.90	0.77	(0.39,3.41)
$X\|W, \sigma_e^2 = 0.10$	1.99	0.81	(0.40,3.58)
$X\|W, \sigma_e^2 = 0.20$	2.09	0.86	(0.40,3.78)

Figure 1: Stagnation *vs.* log(PM)

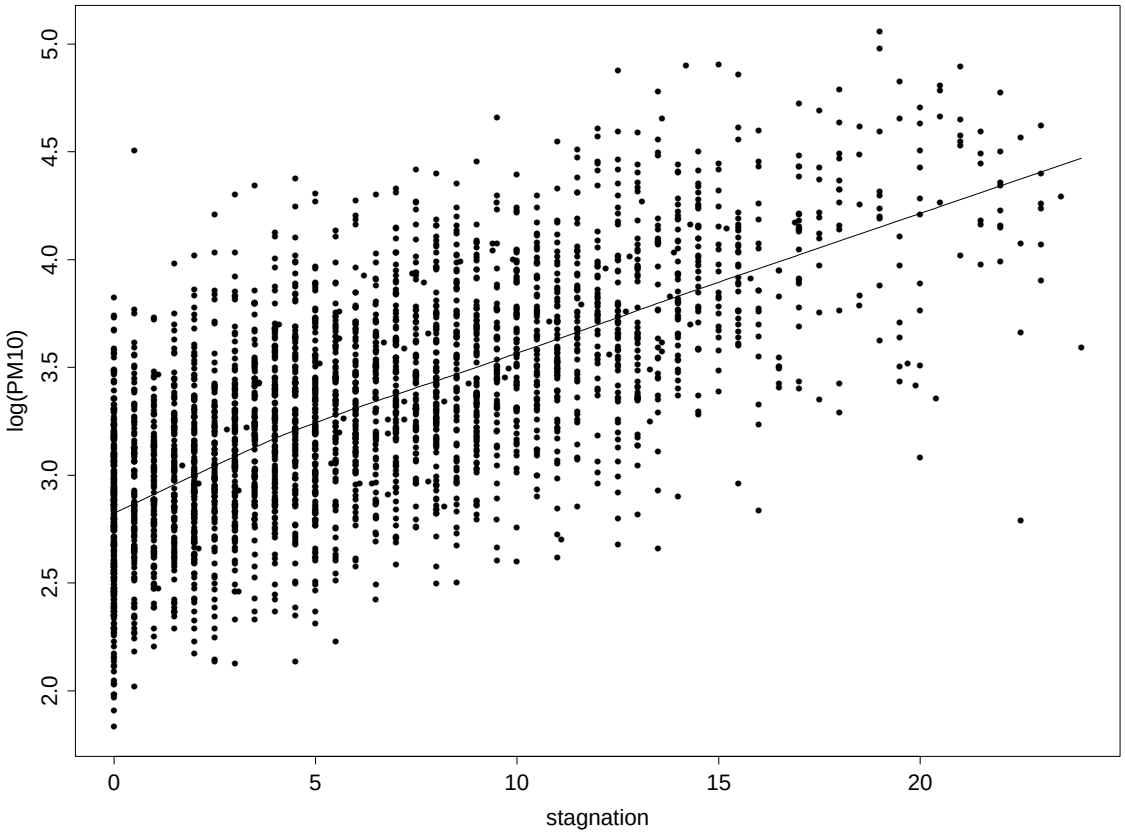

Figure 2: Autocorrelations for ambient log(PM) residuals adjusted for season and day of week, without stagnation in the model (upper plot) or after adjusting for stagnation (lower plot)