Statistical Science
2004, Vol. 19, No. 3, 414–449
DOI 10.1214/088342304000000413

Statistical Issues in Studies of the Long-Term Effects of Air Pollution: The Southern California Children's Health Study

Kiros Berhane, W. James Gauderman, Daniel O. Stram and Duncan C. Thomas

Abstract. In this article we discuss statistical techniques for modeling data from cohort studies that examine long-term effects of air pollution on children's health by comparing data from multiple communities with a diverse pollution profile. Under a general multilevel modeling paradigm, we discuss models for different outcome types along with their connections to the generalized mixed effects models methodology. The model specifications include linear and flexible models for continuous lung function data, logistic and/or time-to-event models for symptoms data that account for misspecifications via hidden Markov models and Poisson models for school absence counts. The main aim of the modeling scheme is to be able to estimate effects at various levels (e.g., within subjects across time, within communities across subjects and between communities). We also discuss in detail various recurring issues such as ecologic bias, exposure measurement error, multicollinearity in multipollutant models, interrelationships between major endpoints and choice of appropriate exposure metrics. The key conceptual issues and recent methodologic advances are reviewed, with illustrative results from the Southern California Children's Health Study, a 10-year study of the effects of air pollution on children's respiratory health.

Key words and phrases: Mixed effects, time series, measurement error, ecologic regression, chronic effects, air pollution.

1. INTRODUCTION

Health effects of air pollution can be broadly classified into two types: *acute* effects associated with short-term fluctuations in pollution levels and *chronic* effects of long-term exposures to pollution. The acute health effects of ambient air pollution on daily mortality and morbidity are manifest in numerous epidemiologic (e.g., Schwartz, 1994; Dominici, Samet and Zeger, 2000) and chamber studies (Gong et al., 1998). These results have been instrumental in setting air quality standards. In contrast, health effects of longer-term exposures have not been as extensively investigated despite their importance in the regulatory process.

Some of the important cohort studies of chronic effects of air pollution in adults are the Harvard Six-Cities study (Dockery et al., 1993), the American Cancer Society (ACS) study of U.S. veterans (Pope et al., 1995), and the Seventh Day Adventist study (Abbey et al., 1999). These studies examined effects of long-term levels of pollution on mortality using multilevel analogues of the Cox proportional hazards model (Cox, 1972; Ma, Krewski and Burnett, 2000). The Harvard Six-Cities and the ACS studies have been reanalyzed by Krewski et al. (2003).

Kiros Berhane, W. James Gauderman, Daniel O. Stram and Duncan C. Thomas are members of the Department of Preventive Medicine, University of Southern California, Los Angeles, California 90089-9011, USA (e-mail: kiros@usc.edu).

The Southern California Children's Health Study, hereafter referred to as the CHS, is one of very few prospective studies of children. It was designed to examine long-term effects of air pollution on respiratory health, using comparisons between communities, between children within communities and within children over time. The CHS is the longest currently running prospective study on chronic effects of air pollution in children.

Other studies on children that have played major roles in regulatory decisions include an Austrian study (Frischer et al., 1999), the Harvard Six-Cities study (Ferris et al., 1979) and 24-Cities study (Raizenne et al., 1996). The Austrian study followed 1150 children from nine communities. Lung function tests were conducted bi-annually for three years (1994–1997) and their association with air pollution was studied. The Harvard Six-Cities study enrolled (1974–1979) 13,378 first and second grade school children from six U.S. cities. Questionnaire and lung function data were then collected until their graduation from high school. The 24-Cities study collected questionnaire and lung function data (1988–1991) from 10,251 8–12-year-old children in U.S. and Canadian communities with a diversity of pollution levels.

These studies all sought to include communities with a diverse pollution profile. They also provide the opportunity to examine within-subject effects over time. The resulting data have a rich structure, allowing comparisons (1) over time (within subjects or communities), (2) between subjects and/or (3) between communities. A comprehensive discussion of statistical methods for analyzing data from such study designs is of public health importance.

This article discusses modeling of and related methodologic issues for data from such multilevel designs. Some recurring themes that already receive methodologic attention or require further research are outlined. We discuss a broad range of issues that arise in many epidemiologic studies in environmental health, not necessarily restricted to respiratory diseases or air pollution. The CHS results will be used to highlight relevant methodologic issues. Methodologic comparisons are made to other studies when appropriate.

The design of the CHS is outlined in Section 2. Section 3 discusses statistical issues in longitudinal analysis of major endpoints and presents a general multilevel modeling approach. Section 4 discusses various recurring themes, such as ecologic bias, measurement error and choice of exposure metrics. Finally, Section 5 provides further discussion of the main methodologic issues and outlines areas that require additional research.

2. DESIGN OF THE CHILDREN'S HEALTH STUDY

The primary aims of the CHS are to assess rates of lung growth, incidence of respiratory disease and frequency of respiratory symptoms or school absences in relation to long-term air pollution levels. Secondary aims include studying the relationships between the health outcomes, the confounding or modifying effects of personal risk factors, underlying disease processes, exposure factors, and time/activity patterns, the shape of the dose–response relationships, and disentangling the effects of multiple pollutants.

Selection of communities. At the outset between-community comparisons were considered likely to be the most informative. This view motivated the selection of communities that exhibited maximum variability with respect to ambient levels of ozone (O_3), particulates (PM_{10}), nitrogen dioxide (NO_2) and acid (including nitric, acetic and formic acids). Initially, 86 Southern California communities with routine air quality monitoring were classified as "high" or "low" on each pollutant, using 1986–1990 multiyear average levels. Only 8 of the $2^4 = 16$ possible pollution "profiles" were represented in sociodemographically comparable communities that had reliable monitoring data (Table 1). Then the 12 most promising communities (Figure 1) were selected, based on cost, feasibility and statistical power (Navidi et al., 1994). On-study measurements have confirmed the original pollution patterns.

Enrollment of cohorts. Approximately 150 fourth graders and 75 seventh and tenth graders were enrolled in 1993 from each community. Whole classes were invited to participate, and the 3681 students who returned a signed consent form were enrolled. In 1994, 386 fifth and 111 eighth graders were added from the same schools. A second fourth grade cohort of 2081 children was enrolled in 1996. Thus, 6259 children have entered the study for observation. Attrition was about 8% per year, with 95% due to moving away from participating schools. A survey of children who moved away was conducted in 1998 to examine the effects of moving from less to more polluted areas or vice versa (Avol et al., 2001).

Health assessments. The primary care giver of each child completed a baseline questionnaire that covered residential history, current residential characteristics (e.g., ventilation and sources of indoor pollution),

TABLE 1

Average ambient air pollution levels for 1994 in 12 Southern California communities selected for the Children's Health Study

Design strata					1994 ambient measurements				
O_3	PM_{10}	NO_2	Acid	Communities	O_3[a]	PM_{10}[b]	$PM_{2.5}$[c]	NO_2[d]	Acid[e]
H	H	H	H	San Dimas (SD)	82.9	36.7	22.1	36.2	5.0
				Upland (UP)	73.3	49.0	24.0	42.6	4.7
H	H	H	L	Mira Loma (ML)	76.3	70.7	31.5	31.3	3.1
				Riverside (RV)	80.6	45.2	25.5	33.9	3.7
H	H	L	H	Lancaster (LN)	59.7	33.6	9.3	17.8	2.3
H	H	L	L	Lake Elsinore (LE)	76.1	34.7	13.4	21.9	3.3
H	L	H	H	Lake Gregory[f] (LA)	97.5	24.2	11.1	8.5	3.5
H	L	H	L						
H	L	L	H						
H	L	L	L	Alpine (AL)	71.3	21.3	9.2	13.2	2.6
L	H	H	H	Long Beach (LB)	41.3	38.8	16.3	36.4	3.5
L	H	H	L						
L	H	L	H						
L	H	L	L						
L	L	H	H						
L	L	H	L						
L	L	L	H						
L	L	L	L	Atascadero (AT)	50.1	20.7	7.6	14.1	1.3
				Santa Maria (SM)	35.5	29.2	6.7	4.3	1.3
				Lompoc (LM)	42.7	13.0	7.3	2.7	1.0

[a]10 AM–6 PM average (ppb).
[b]24-hour average ($\mu g/m^3$).
[c]Two-week average (ppb).
[d]24-hour average (ppb).
[e]Two-week average (HNO_3 + HCL; ppb).
[f]Lake Gregory is identified as Lake Arrowhead (LA) in subsequent years.

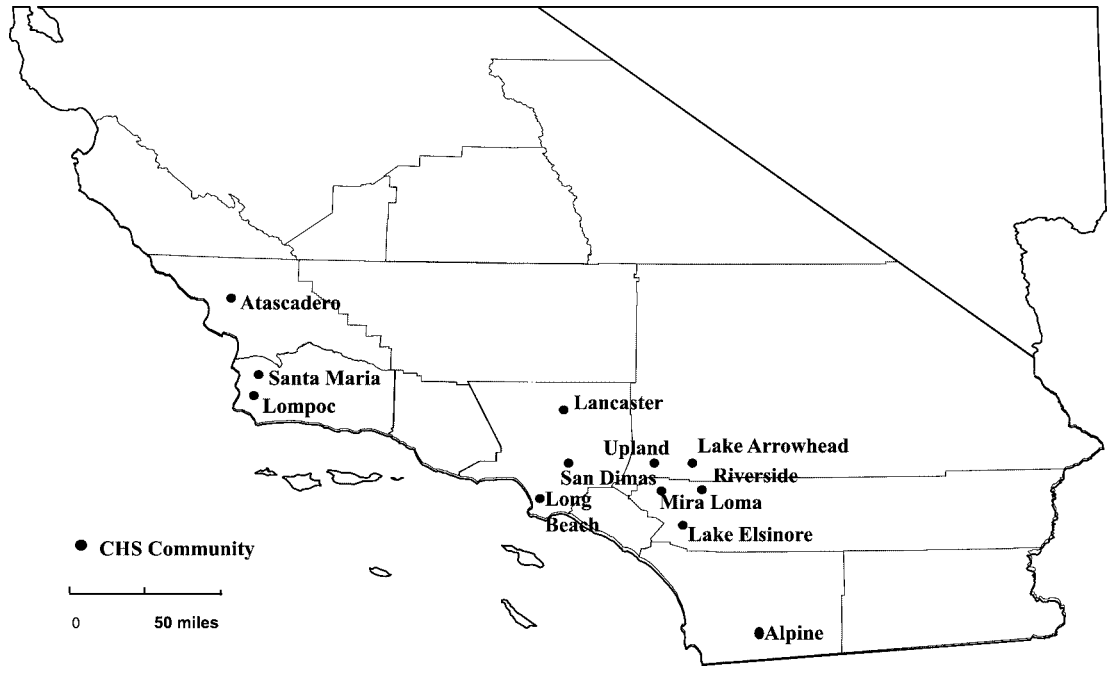

FIG. 1. *Geographical distribution of communities in the Children's Health Study.*

personal risk factors, usual respiratory symptoms and usual activities. An abbreviated yearly follow-up questionnaire collects data on chronic respiratory symptoms and diseases, and time-dependent covariates (see Peters et al., 1999a, for details).

A field team visits participating schools in the winter–spring of each year (January–June) to conduct lung function tests (Peters et al., 1999b). Our examples focus on one of the lung function measures, namely the maximum mid-expiratory flow (MMEF). In contrast to lung size measures such as the forced expiratory volume in one second (FEV_1), MMEF measures flow rates and tends to show larger deficits in susceptible subgroups (e.g., asthmatics). Daily school absences are reported weekly or bi-weekly throughout the school year. From 1996–1997, causes of absences were ascertained by phoning parents of the second fourth grade cohort within 4 weeks of each absence.

Air pollution assessments. In all 12 communities, monitoring stations provide continuous hourly ambient O_3, PM_{10} and NO_2, and 2-week measures of $PM_{2.5}$ and acid vapors. Individual exposure predictions using the "microenvironmental" approach were based on data on ambient exposure, housing characteristics and time/activity patterns. Dispersion models have been used to estimate exposures from major roadways (see Section 4.2 for details).

To date, most chronic effects analyses rely on yearly or multiyear exposure data. Table 2 summarizes the correlations of major pollutants, based on multiyear averages of 1994–1997 data from hourly (PM_{10}, O_3 and NO_2) or 2-week integrated ($PM_{2.5}$ and acid) measurements. Despite designing the study to minimize them, many correlations remain quite high. The corresponding correlations that were based on year-to-year variations in pollution levels showed a similar pattern, with the exception of those involving O_3 (which exhibited relatively higher correlations). This limits the ability to separate the effects of the various pollutants. See Section 4.3 for details on an analytic approach to deal with this problem.

TABLE 2
Correlations between multiyear averages of pollutants (1993–1997) in the 12 communities selected for the Children's Health Study

Pollutant	O_3	PM_{10}	$PM_{2.5}$	NO_2	Acid
O_3 (10 AM–6 PM)	0.73	0.28	0.29	−0.03	0.46
O_3 (24-h avg)		−0.31	−0.33	−0.54	−0.09
PM_{10}			0.96	0.67	0.73
$PM_{2.5}$				0.76	0.82
NO_2					0.85

3. STATISTICAL MODELS FOR MAJOR ENDPOINTS

Assessment of chronic effects of air pollution has relied to a large extent upon comparisons of aggregate health outcomes between communities with diverse pollution profiles. The communities included in such studies may not constitute a random sample from a larger population of communities. If so, what then is the interpretation of the p-values or confidence limits that arise from the analysis at the community level?

To address this question, we argue that methods for inference at the community level are those that would be appropriate in a grouped randomization setting. Under this scenario, air pollution levels would somehow have been assigned at random to a sample of communities, which may or may not have been selected at random. Despite the randomization, residual differences in aggregated health outcomes may remain between the communities that are left unexplained by the assigned levels of air pollution, even after adjusting for subject-specific data. The random assignment of air pollution levels to a community would permit inferences to be drawn about the effect of air pollution on health outcomes in the community-level analysis, while still allowing the residual heterogeneity to remain between communities. The way in which communities have been selected may impact the generalizability of the results of the study to children living in other places, but a nonrandom selection of communities would not invalidate the statistical tests, which derive their validity from the randomization.

Of course, in most large-scale air pollution studies, we observe, rather than manipulate, pollution levels. The use of the same statistics (for the community-level analysis) as would be appropriate to a group randomization experiment assumes that the causal forces that lead to any unexplained community differences (residual heterogeneity) in outcomes are not themselves related to air pollution exposure. This amounts to assuming that this residual heterogeneity is random relative to air pollution. Admittedly, this assumption is not directly statistically testable. However, the use of subject-specific covariates (e.g., age, height, race) to adjust for the effect of other determinants of the outcome is designed to reduce the possibility that confounding is the root cause of the effects observed.

We sought a modeling framework that can simultaneously handle all of the aspects that we have described above including (1) time-dependent adjustment variables; (2) between-community differences in aggregate air pollution measurements and outcomes; (3) within-community differences by time in aggregate air pollution and outcomes; and finally (4) individual differences in cumulative exposure or in predicted exposure based on modeling. Such considerations lead naturally to multilevel models as the analytic method of choice.

3.1 The Multilevel Modeling Paradigm

Here, except for school absences (see Section 3.3), we present the multilevel model using time-dependent pollution measures that are based on yearly averages. A discussion of the choices of exposure metrics is presented in Section 4.5. Our multilevel models are equivalent to mixed effects models for Gaussian (Laird and Ware, 1982) and non-Gaussian (Breslow and Clayton, 1993) data. These, in turn, are based on growth curve models as in, say, Harville (1977).

General formulation. Denote by y_{cij} the health endpoint for subject i in community c at time t_{cij}, where j indexes year. Predictors can be time-dependent, time-constant (depicting fixed subject-specific attributes) or community-specific. Uppercase and lowercase letters denote community-specific and subject-specific quantities, respectively. Thus, $\mathbf{z}_{cij}$ represents time-dependent covariates (e.g., height, age), $\mathbf{z}_{ci}$ represents time-constant fixed covariates (e.g., gender, ethnicity), $\mathbf{x}_{ci}$ represents subject-specific average pollution levels (e.g., from microenvironmental modeling), $\mathbf{X}_{cj}$ represents the community annual-average levels of pollution and $\mathbf{X}_c$ represents the community-specific multiyear average levels of pollution.

Consider a three-level generalized linear model of the form

$$g(\mu_{cij}) = a_{ci} + b_{ci}t_{cij} + \boldsymbol{\gamma}_1^T\mathbf{z}_{cij} + \boldsymbol{\alpha}_1^T(\mathbf{X}_{cj} - \mathbf{X}_c), \tag{1}$$

$$a_{ci} = A_c + \boldsymbol{\eta}_2^T\mathbf{z}_{ci} + \boldsymbol{\alpha}_2^T(\mathbf{x}_{ci} - \mathbf{X}_c) + e_{ci}, \tag{2}$$

$$b_{ci} = B_c + \boldsymbol{\gamma}_2^T\mathbf{z}_{ci} + \boldsymbol{\beta}_2^T(\mathbf{x}_{ci} - \mathbf{X}_c) + f_{ci}, \tag{3}$$

$$A_c = \alpha_0 + \boldsymbol{\alpha}_3^T\mathbf{X}_c + e_c, \tag{4}$$

$$B_c = \beta_0 + \boldsymbol{\beta}_3^T\mathbf{X}_c + f_c. \tag{5}$$

Here, $g(\mu_{cij})$ denotes a link function (McCullagh and Nelder, 1989), $\mu_{cij} = E(y_{cij})$ and Roman letters represent random effects. In (1), a_{ci} and b_{ci} represent subject-specific random intercept and slope of follow-up time (or age), t_{cij}, respectively. Model (1) is adjusted for time-dependent covariates and allows a test of air pollution effects via the slope $\boldsymbol{\alpha}_1$ on deviations of community annual-average ambient levels from their long-term averages.

Models (2) and (3) include random effects for community (A_c and B_c), and subject-specific covariates. They allow for a second test of air pollution effects via the regression of subject-specific slopes or intercepts from (1) on deviations of personal exposures from the community means. The residuals e_{ci} and f_{ci} are assumed to be uncorrelated across subjects.

Models (4) and (5) relate the community mean adjusted intercepts and slopes to the long-term average ambient pollution levels, allowing independent random error terms. Community-specific covariates could also be included (see Section 4.1).

Although the model can be fitted using a sequence of regressions, the generalized linear mixed effect model (Diggle, Liang and Zeger, 1994), where all levels are combined with multiple error terms, provides a more unified approach as given by

$$\begin{aligned} g(\mu_{cij}) = {} & \alpha_0 + \boldsymbol{\alpha}_3^T\mathbf{X}_c + \boldsymbol{\eta}_2^T\mathbf{z}_{ci} \\ & + \boldsymbol{\alpha}_2^T(\mathbf{x}_{ci} - \mathbf{X}_c) + e_c + e_{ci} \\ & + [\beta_0 + \boldsymbol{\beta}_3^T\mathbf{X}_c + \boldsymbol{\gamma}_2^T\mathbf{z}_{ci} \\ & \quad + \boldsymbol{\beta}_2^T(\mathbf{x}_{ci} - \mathbf{X}_c) + f_c + f_{ci}]t_{cij} \\ & + \boldsymbol{\gamma}_1^T\mathbf{z}_{cij} + \boldsymbol{\alpha}_1^T(\mathbf{X}_{cj} - \mathbf{X}_c), \end{aligned} \tag{6}$$

where $e_{ci} \sim N(0, \sigma^2_{e,ci})$ and $e_c \sim N(0, \sigma^2_{e,c})$ are random subject- and community-specific intercepts. Similarly, $f_{ci} \sim N(0, \sigma^2_{f,ci})$ and $f_c \sim N(0, \sigma^2_{f,c})$ are the corresponding random slopes. For continuous outcomes (e.g., lung function), an overall error term $e_{cij} \sim N(0, \sigma^2_{cij})$ is given. Model (6) assumes independence between the random effects. This assumption could be relaxed to allow for more complex temporal and/or spatial correlation structures.

When model (6) is applied to non-Gaussian data (e.g., disease symptoms), fixed effects are interpreted conditional on the random effects. If distributional assumptions for the random effects are violated, the estimates for ecologic comparisons (e.g., $\boldsymbol{\beta}_3$) could be biased. An alternative is to use marginalized multilevel models that lead to fixed effects with marginal interpretations. See Heagerty and Zeger (2000) and references therein for details.

3.2 Longitudinal Analysis

Linear models. To assess the chronic effects of air pollution on lung function, the mixed effects model (6) is used with the identity link, perhaps with a transformation (e.g., log transformation) to improve normality.

For the CHS, Peters et al. (1999b) applied a two-level model to the baseline lung function measurements, with subject- and community-level random effects, in essence modeling the intercepts. We refer to this model as a "cross-sectional" model. Gauderman et al. (2000) applied model (6) to the first five years of lung function data, but treated the intercept terms a_{ci} as fixed effects rather than as random effects using (2) and (4). Treating the intercepts as fixed effects is equivalent to forcing the variance of the intercept terms in a multilevel model to infinity. This ensures robustness to possible misspecification of the models for the intercept terms (Dempster, Rubin and Tsutakawa, 1981).

The above two models are compared in Table 3 with a full model that includes all five levels as in (1)–(5), with ambient pollution levels appearing only in the community-level model (multiyear averages in the longitudinal models and baseline levels in the cross-sectional model). All models included adjustments for age, height, sex and race/ethnicity. Figure 2 depicts the third-level "ecologic" regression of adjusted MMEF growth rates on NO_2, showing a significant negative relationship. The standard error bars (Figure 2) illustrate the degree of within-community variance in the estimated growth rates and the relative homogeneity in these variances across communities. Similar results were observed for PM_{10} and acid, but not for O_3 (Table 3). The slope estimates and their standard errors

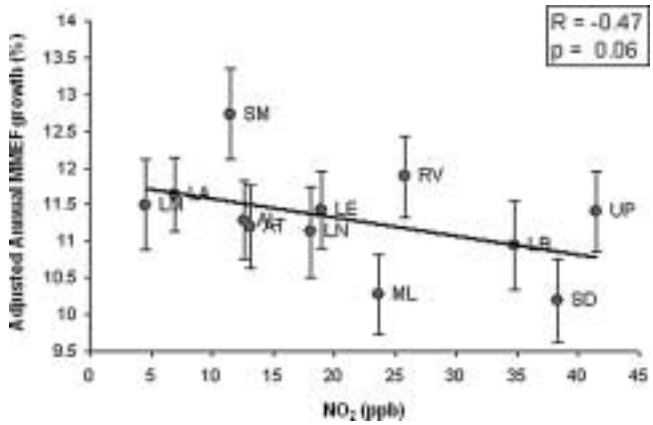

FIG. 2. *Adjusted annual MMEF growth rates by community-specific multiyear average pollution levels (1993–1997) of 24-h mean NO_2. The community labels are as defined in Table 1. The solid line depicts the ecologic linear fit of the adjusted MMEF growth rates on multiyear (1993–1997) average levels of 24-h mean NO_2.*

were insensitive to whether the subject-specific intercepts were modeled as random or fixed effects.

As in any modeling process, diagnostic analysis should be conducted to make sure that modeling assumptions are not grossly violated. This could be done via visual plots or by trial fitting of more complex models that test modeling assumptions. For the CHS data, assumptions of linearity, normality and homoscedasticity appear to be well supported. More research is needed in the development of diagnostic techniques for multilevel models.

Flexible models. As duration of follow-up is extended, the constant growth rate assumption for lung function becomes less tenable. In the CHS this was

TABLE 3
Comparisons of ecologic regression effects of pollution from longitudinal and cross-sectional models for MMEF[a]

Pollutant	Cross-sectional intercepts (%)	Longitudinal slopes (%)	Full model	
			Intercepts (%)	Slopes (%)
O_3	0.80 (0.93)	−0.20 (0.26)	1.15 (1.14)	−0.18 (0.27)
PM_{10}	−1.54 (0.93)	−0.49 (0.20)[b]	−1.65 (0.80)[c]	−0.45 (0.21)[b]
$PM_{2.5}$	−2.62 (1.58)	−0.74 (0.34)[b]	−2.83 (1.44)[c]	−0.68 (0.37)[c]
NO_2	−1.69 (1.15)	−0.47 (0.25)[c]	−1.97 (1.07)[c]	−0.46 (0.27)[c]
Acid	−0.44 (1.15)	−0.43 (0.22)[c]	−0.78 (1.08)	−0.41 (0.24)[c]

[a]The dependent variable in these models is log(MMEF). Estimates are the predicted percent difference in level (intercepts) or growth rate (slopes) per increase of 20 ppb of O_3 and NO_2, 20 $\mu g/m^3$ of PM_{10} and $PM_{2.5}$ and 2 ppb of acid.
[b] $p < 0.05$.
[c] $p < 0.10$.

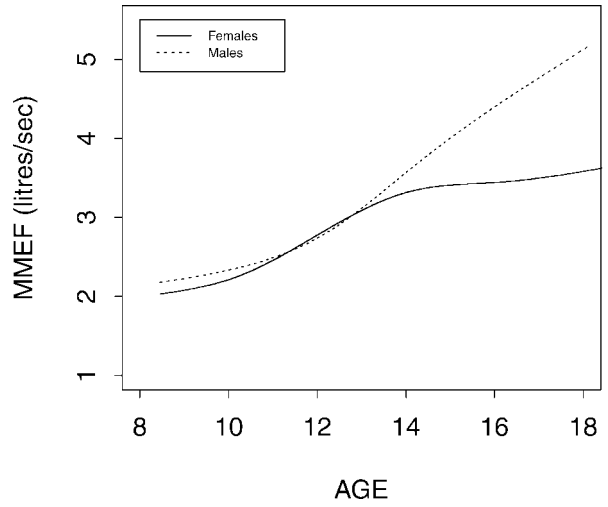

FIG. 3. *Gender-specific growth curves for MMEF in* 8–18*-year-old participants of the Children's Health Study. The curves are smooth functions estimated by natural splines from a mixed effects model with a random intercept.*

clear from the growth trajectories across the fourth, seventh and tenth grade cohorts (Figure 3). This was also recognized in the Harvard Six-Cities study (Wypij, Pugh and Ware, 1993). Exploratory analysis of CHS data revealed that height and lung function are linearly related over short age intervals, but both the intercept and the slope vary with age. This leads to a model that has a nonlinear function of age and a linear term of height, with slope that is age-dependent.

While several parametric functions have been tried for modeling lung growth in children (Wypij, Pugh and Ware, 1993), regression splines (Hastie and Tibshirani, 1990) provide a flexible way to model the growth curves. Here, the nonlinear growth trajectory is depicted via piecewise polynomials between breakpoints, known as knots. These polynomials are then smoothly joined at the knots. A set of basis functions with such properties, known as B-splines, is

$$B_q(t) = (t_{q+4} - t_q) \sum_{j=q}^{q+4} \frac{(t_j - t)_+^3}{\prod_{k=q,k\neq j}^{q+4}(t_k - t_j)}, \qquad q = 1, \ldots, m+4,$$

for a variable t at m knots. A variant set of basis functions, known as natural splines, imposes additional constraints of linearity beyond the boundary points. Natural splines are less sensitive to sparsity of information at the edges of the data due to their additional constraints. Once the basis functions are constructed, the resulting mixed effects model is then fully parametric, allowing for formal inference. For more details, see de Boor (1978). Figure 3 depicts gender-specific growth curves for MMEF using natural splines, from a mixed effects model with subject-specific random intercepts.

Berhane et al. (2000) used a flexible mixed effects model analogous to model (1) to account for nonlinear effects of age and height in modeling the effects of asthma on lung function (see Section 4.4 for details). Biologically important features of growth curves (such as peak growth rate and maximum attained value) can, in principle, be calculated for each subject using the first and second derivatives of the fitted curves. Such features of nonlinear curves, known as functionals, were studied by Ramsay and Silverman (1997). Modeling functionals directly allows for examination of air pollution effects on biologically meaningful aspects of children's growth trajectories. Generalized mixed additive models, which focus on fully nonparametric smoothing techniques, have been introduced (Lin and Zhang, 1999; Hastie and Tibshirani, 2000). Further research is needed in the development of models with random flexible terms, with inferential focus on functionals.

Logistic models. Annual reports of symptoms (e.g., bronchitis) can be modeled with a logistic mixed effects model (Breslow and Clayton, 1993). An alternative is to use conditional logistic models as outlined in Diggle, Liang and Zeger (1994, pages 175–183). This last approach is equivalent to that of stratified case-control studies, where each subject is treated as a stratum with $y_{ci\cdot} = \sum_j y_{cij}$ cases and $n_{ci} - y_{ci\cdot}$ controls.

It is sometimes helpful to distinguish between prevalence and incidence for chronic diseases such as asthma. This leads to a pair of first-level models given as

$$\text{logit}\Pr(y_{ci1} = 1) = a_{ci} + \boldsymbol{\gamma}_1^T \mathbf{z}_{ci1}, \tag{7}$$

$$\begin{aligned}&\text{logit}\Pr(y_{cij} = 1 | y_{ciq} = 0, q = 1, \ldots, j-1)\\ &\quad = b_{ci}[\Delta t_{cij}] + \boldsymbol{\gamma}_1^T \mathbf{z}_{cij} + \boldsymbol{\alpha}_1^T (\mathbf{X}_{cj} - \mathbf{X}_c).\end{aligned} \tag{8}$$

Second- and third-level models are as described in (2)–(5). Thus, person-time would be counted only up to the first appearance of a given symptom in this approach. Analyses of baseline symptoms using a two-level logistic model were reported by Dockery et al. (1989) and Peters et al. (1999a) for the Harvard Six-Cities study and the CHS, respectively.

Subjects often report being diagnosed for a chronic disease (e.g., asthma) on one or more occasions and not subsequently. In this case, was the first report correct

and the later negative reports incorrect or vice versa? Models for such scenarios are not well developed. A sensible approach to this problem would be via hidden Markov models (MacDonald and Zucchini, 1997). Let Y_{cij} represent the true disease status and let y_{cij} represent the reported status. A multilevel model would then be specified for $\text{logit}[\Pr(Y_{cij} = 1|Y_{ci,j-1} = 0)]$ and $\text{logit}[\Pr(Y_{ci1} = 1)]$ as above, together with models for misclassification rates, which take the form $\text{logit}[\Pr(y_{cij}|Y_{ci1}, \ldots, Y_{cij})]$ contribution for each subject. The likelihood could be formed by summing the product of these probabilities over all possible subject-specific outcomes.

Time-to-event models. Risk factors for disease incidence (e.g., using time to first asthma report) can be examined via proportional hazards models (Cox, 1972). For the CHS, such models revealed that the risk of asthma was elevated in those who played at least three team sports in high ozone communities (McConnell et al., 2002). Here, the effect of air pollution was investigated by stratifying the communities into low, medium or high ambient pollution levels or via a two-level proportional hazards model (Burnett et al., 2001) as in

$$\lambda(t) = \lambda_{s0}(t) \cdot \exp(A_c + B_c \tilde{z}_{ci} + \boldsymbol{\gamma}_1^T \mathbf{z}_{cij} + \boldsymbol{\gamma}_2^T \mathbf{z}_{ci}), \tag{9}$$

$$B_c = \beta_0 + \beta_3 X_c + f_c, \tag{10}$$

where $\tilde{z}_{ci}$ denotes the variable assumed to modify the effect of pollution (e.g., outdoor sports in McConnell et al., 2002), $\lambda_{s0}(t)$ denotes baseline hazards with s strata (e.g., by age groups and gender) and $f_c \sim N(0, \sigma_{f,c}^2)$ denotes a community-level random effect. A unified random effects Cox model was proposed by Ma, Krewski and Burnett (2000). This general approach was used to examine the effects of pollution on adult mortality (Krewski et al., 2003).

3.3 Analysis of School Absence Count Data

Data on school absences provide a good opportunity to study the effect of air pollution on children's health. In one notable study, Ransom and Pope (1992) studied the relationship between PM_{10} and school absenteeism in the Utah Valley between 1985 and 1990 based on weekly data from a school district and daily data from one elementary school. They showed that high PM_{10} levels were associated with significant increases in school absenteeism.

In the CHS, data on school absences are being collected from school records. In 1995–1996, a substudy known as the Air Pollution and Absences Study (APAS) was conducted to ascertain whether absences were illness related, and if so the specific health reasons. Analysis of the resulting binary time series data falls into the general multilevel framework with a logit link. However, additional refinements are warranted to handle the complex lag structure of the effects of air pollution and/or to account for the serial autocorrelation induced by unmeasured confounders (e.g., influenza epidemics). Several methods have been proposed for examining associations between daily aggregate mortality and morbidity counts. These include the filtered least squares approach (Zeger, 1988; Samet, Zeger and Berhane, 1995; Berhane and Thomas, 2002), generalized additive models (Schwartz, 1994; Kelsall, Zeger and Samet, 1999; Zanobetti et al., 2000; Dominici, Samet and Zeger, 2000) and transition models (Brumback et al., 2000).

Given this article's focus on the "chronic" effects of air pollution, we give details only for models that collapse the binary school absence data over time to yield absence counts per subject. Let y_{cij} and r_{cij} be binary indicators of an incident absence and for being "at risk" on that day j, respectively. Let $Y_{ci} = \sum_j y_{cij}$ be the total number of absences for child i. After computing an expectation under the null hypothesis of no air pollution, community or covariate effects as $E_{ci} = \sum_j \lambda_j r_{cij}$, where $\lambda_j = \sum_c Y_{cj} / \sum_c R_{cj}$, an overdispersed Poisson mixed effects model [i.e., allowing for $V(Y_{ci}|\mu_{ci}) > \mu_{ci}$ in (11)] is

$$\mu_{ci} = E(Y_{ci}) = E_{ci} \exp(A_c + \boldsymbol{\eta}_2^T \mathbf{z}_{ci}), \tag{11}$$

$$A_c = \alpha_0 + \boldsymbol{\alpha}_3^T \mathbf{X}_c + e_c. \tag{12}$$

Here, A_c are logarithms of the community mean absence rates, adjusted for personal covariates, and $\boldsymbol{\alpha}_3$ denotes a vector of parameters for long-term average pollution levels $\mathbf{X}_c$. For the CHS, high levels of body mass index ($p = 0.03$), current smoking by the mother ($p = 0.02$), wheezing ($p = 0.01$) and active asthma ($p < 0.01$) were associated with elevated number of illness related absences, but no long-term pollution effects were observed.

The models given by (11) and (12) allow adjustments for personal time-constant covariates, while they do not allow adjustments for time-dependent covariates. On the other hand, time series models that aggregate over subjects allow adjustments for temporal covariates, but not for personal characteristics. Three-level

models can be developed that incorporate both individual time-constant covariates and community-specific time-dependent exposures. Details of one such model explored by our group will be reported elsewhere.

3.4 Computational Issues

Model (6) has been implemented in several software packages. For Gaussian data, PROC MIXED in SAS (Littell, Milliken, Stroup and Wolfinger, 1996) fits mixed effects models via maximum likelihood or restricted maximum likelihoods. For non-Gaussian data, PROC NLMIXED or GLIMMIX in SAS could be used. PROC NLMIXED maximizes an approximation (e.g., a first-order Taylor series) to the likelihood integrated over the random effects. The estimation algorithm implemented in the SAS macro GLIMMIX is also based on a Laplace approximation to the integrated likelihood function (Breslow and Clayton, 1993). Inadequacies of these approximations may lead to biased estimates (Breslow and Lin, 1995; Lin and Breslow, 1996).

The R/Splus NLME library of Pinheiro and Bates (2000) provides an alternative way to fit the mixed effects model. The well developed routines for regression splines in R/Splus are particularly useful for the flexible models discussed in Section 3.2. Splus functions for fitting flexible mixed effects models can be obtained from the authors of this paper.

Other computationally efficient software [e.g., MLn (Rasbash and Woodhouse, 1995) and HLM (Bryk, Raudenbash and Congdon, 1996)] also is available. Model (6) could be extended to allow random components of $\boldsymbol{\eta}_2$ and $\boldsymbol{\alpha}_2$, which allow subject-specific effects of time-dependent variables such as height. Because age and height are correlated and due to the general monotonic pollution trends, it would be difficult to distinguish between subject-specific variation in these terms and the variation in subject-specific intercepts and slopes on age.

We now briefly discuss the conditions under which the combined mixed effects model (6) and the sequence of regressions (1)–(5) could be equivalent. This is important because some modeling situations are not suitable for fitting the unified mixed effects model as in (6).

In a pure growth curve model for lung function, the fixed effects may all be regarded as modifying the means of the subject-specific random effects. More formally, the columns of the fixed effects design matrices in a growth curve model are all linear combinations of the columns of the random effects design matrices (Laird and Ware, 1982). The main interest here is in such main effects; for example, we are interested in how air pollution affects lung function level at baseline (the subject-specific intercept parameters) or growth in lung function over time (the slopes). The inclusion of other time-dependent variables that are not modifiers of subject-specific intercepts or slopes is required to incorporate adjustments for factors such as temperature at the time of the examination. A pure growth curve model would correspond to eliminating the time-dependent covariates and air pollution variables in (1) and the subject-specific adjustment and air pollution variables in level 2. Then, focusing on the subject-specific slopes, models (3) and (5) could be modified to allow for uncertainties in the estimates of the subject-specific slopes, that is,

$$\begin{aligned} b_{ci} &= B_c + \boldsymbol{\gamma}_2^T \mathbf{z}_{ci} + \boldsymbol{\beta}_2^T (\mathbf{x}_{ci} - \mathbf{X}_c) + f_{ci} + \psi_{ci}, \\ B_c &= \beta_0 + \boldsymbol{\beta}_3^T \mathbf{X}_c + f_c + \psi_c, \end{aligned} \tag{13}$$

where $\psi_{ci} \sim N(0, V_{ci})$ and $\psi_c \sim N(0, V_c)$ are additional random effects with V_{ci} and V_c given by the sampling variances of b_{ci} and B_c, respectively. Here, f_{ci} and f_c are residual error terms as in (3) and (5), respectively. In either level, estimation proceeds iteratively between estimations for regression parameters and the residual variances.

A multilevel model that uses this meta-analytic scheme gives results that are nearly equivalent to those from (6) for a growth curve model (Ware and Stram, 1988; Stram, 1996). Complications arise when time-dependent covariates are included in (1), because correlations are induced between estimates of B_c for various communities, due to shared sampling errors in the estimation of $\boldsymbol{\eta}_2$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ and $\boldsymbol{\beta}_2$ in (6). Incorporating these covariances into V_{ci} and V_c in the multilevel approach would lead to results which are nearly equivalent to those obtained using (6). This near equivalence of estimates follows from the fact that, for normally distributed data, the full mixed effects model can be fitted using a set of sufficient statistics for the random effects rather than using the full data set of all measurements. In fact, this is how the multilevel modeling software HLM fits the full model (Bryk, Raudenbash and Congdon, 1996).

4. RECURRING THEMES

4.1 Ecologic Bias

In the classic "ecologic correlation" study, the rate of disease or some average health effect Y_c in a set of populations (typically geographically defined) is

related to some measure of average exposure X_c, possibly adjusted for group-level covariates Z_c. The so-called ecologic fallacy (Selvin, 1958) or cross-level bias (Firebaugh, 1978) concerns the difference between the regression estimates from such an analysis and those estimated from individual data, that is, a regression of y_{ci} on x_{ci} and z_{ci}. Greenland and Morgenstern (1989) described three ways such bias can come about: (1) within-group confounding that acts differentially across groups, (2) confounding by group effects and (3) effect modification by group. Omitting covariates and focusing only on baseline data, the multilevel model (6) with identity link is $y_{ci} = \alpha_0 + \alpha_2(x_{ci} - X_c) + \alpha_3 X_c + e_{ci} + e_c$. Then the absence of cross-level bias can be written as $\alpha_2 = \alpha_3$. Equivalently, rewriting the model as $y_{ci} = \alpha_0 + \alpha_2 x_{ci} + (\alpha_2 - \alpha_3)X_c + e_{ci} + e_c$, we see that the absence of ecologic bias corresponds to no effect of X_c on y_{ci} beyond its effect through x_{ci}. Such a group-level effect could arise, however, not as a causal effect, but by confounding by some omitted group-level covariate Z_c. This understanding of ecologic bias appears to have been expressed first by Robinson (1950) and has been treated in numerous reviews (Greenland, 2002; Morgenstern, 1982, 1995; Wakefield and Salway, 2001). A lengthy series of articles (with numerous letters to the editor and rejoinders) on the subject of the ecologic regression of lung cancer rates on domestic radon levels is particularly revealing about these issues (Cohen, 1990; Darby et al., 2001; Greenland and Robins, 1994; Lubin, 1998; Stidley and Samet, 1994).

The CHS differs from the classic ecologic study in that outcome and covariate data are available on individuals, but the exposure variable of primary interest—ambient air pollution—is measured only at a central site and varies much less within communities than between them. Several authors (Künzli and Tager, 1997; Sheppard, 2003) distinguished four types of design: the truly individual design where all variables are measured at the individual level; the "semi-individual" design, in which y and z are measured at the individual level, but X_c is measured only at the aggregate level; the "aggregate" design, in which x and z are measured at the individual level (say via sample surveys in each group), but Y_c is measured only at the aggregate level (Prentice and Sheppard, 1995); and "ecologic" designs, where all variables are measured only at the individual level. The CHS corresponds to the semi-individual design for which multilevel models are particularly relevant (Greenland, 2002).

In Section 4.2, we describe some approaches, based on microenvironmental and spatial modeling, to assess interindividual variation in personal pollution exposures x_{ci}. To date, such variation appears to be small compared to the between-community variation in ambient pollution. Community-level confounding variables, such as altitude or weather, may interact with individual-level exposure (or confounding) variables. For example, the effect of personal variation in exposure (due to time–activity patterns, indoor sources or within-community spatial variation in pollution) may have a relatively larger effect in low ambient pollution communities. It is also possible that exposure measurement error could act differently at the different levels. Suppose that temperature is measured with less error than air pollution levels. An analysis of health endpoints may then provide stronger statistical evidence of a temperature effect than an air pollution effect, even if temperature is only a determinant of pollution levels and has no direct impact on health (Zidek, Wong, Le and Burnett, 1996; Zeger et al., 2000). See Brenner et al. (1992) and Wakefield and Elliott (1999) for the effect of measurement error on ecologic regressions and see Greenland and Brenner (1993) for methods for correction.

With only a few communities, the prospects for including many ecologic covariates are limited and there is some danger of "overadjustment"—controlling for variables which do not have a causal effect on health outcomes, but are simply determinants of pollution variables that are the real causal factors. Weather patterns, for example, are major determinants of pollution levels and thus one must think carefully about whether they are plausibly direct risk factors for health outcomes. There is evidence that temperature and humidity are associated with mortality and hospitalization rates, independent of air pollution (Schwartz, 1994), so inclusion of such variables in the third-level model might be justified. However, there is less evidence that wind is associated with health outcomes. Inclusion of wind in the model might constitute overadjustment, since it is probably an even stronger determinant of pollution level than is temperature.

We analyzed personal income as a potential confounder of the association between air pollution and lung growth. The estimated difference in MMEF growth rate per increase of 20 ppb of NO_2 was -0.46%, with standard error 0.27% (Figure 2). Personal income at the individual level was a significant (inversely related) predictor, but did not appear to be a confounder of the third-level NO_2 effect. It has been

argued, however (cf. Pearce, 2000), that the within-community income disparities may have a stronger impact on health than does one's own personal income. Thus we may consider as predictors, in addition to personal income, the average level within the community or some other statistic (e.g., the variance of income within each community). Using the community-average personal income as an adjustment variable in (4) and (5) produced little change in the estimated NO_2 effect (-0.45%), although the standard error got larger (0.31%). Using an average income variable computed at the neighborhood school level (between levels 2 and 3) also had a somewhat more important effect. From this model the estimated NO_2 effect was reduced to -0.35% with standard error 0.39%. Although school-zone average income was not significantly related to MMEF growth rate ($p = 0.49$), some people would argue that it should be treated as a confounder since it caused a moderate change (24% reduction) in the NO_2 effect estimate (of course the difference seen here in NO_2 effects between these analyses was not itself statistically significant). Additional models did not reveal any interaction of personal income with community-level income. Nevertheless, it as important to consider the potential "contextual" effect (Greenland, 2001) of possible confounders, even when they are measured at the individual level.

4.2 Personal Exposure Models and Measurement Error

Most studies tend to focus on between-communities comparisons of pollution effects. Some additional comparisons may also be made at the temporal level (by year-to-year or daily variation in ambient exposures as in Sections 3.2 and 3.3). While such analyses address questions of immediate public policy concern, the evidence for causality would be enhanced if it were possible to assess exposure–response relationships at the individual level. There are three principal approaches to quantifying individual variation in exposure: (1) using time–activity patterns and housing characteristics to model personal exposure; (2) exploiting spatial variation in measured pollutant concentrations and traffic density within communities; and (3) comparing lifetime exposures of permanent residents and those who moved from areas of higher or lower pollution (or subsequently moved away). Any of these comparisons could entail substantial measurement error. In this section we describe our approaches to estimating these three sources of interpersonal variation and their uncertainties, and then discuss their use in the hierarchical model for health effects in the context of measurement error models.

Microenvironmental models. A standard approach to estimation of personal exposure in the occupational and environmental hygiene literature, the "microenvironmental model," was first introduced by Duan (1982, 1991) and was further developed by Lioy et al. (1992), McCurdy (1995), Johnson, Long and Ollison (2000) and Burke, Zufall and Özkaynak (2001). This approach has been implemented for the CHS as in Navidi and Lurmann (1995). For each subject, we obtained annual questionnaire data on usual time–activity patterns, namely the proportion of time p_{cijm} spent in microenvironment m (home, school, outdoors, car, etc.) and data used to estimate ventilation rates (sports, etc.). Using measurements made in a sample of homes and schools, supplemented with data from the literature, we have constructed models $f_m(\mathbf{w}_{cijm}, X_{cj})$ for the mean exposure level in each microenvironment as a function of ambient exposure X_{cj} and various characteristics of the microenvironment $\mathbf{w}_{cijm}$ (indoor sources, air conditioning, etc.). Combining the two components, we estimate personal exposure as

$$x_{cij} = \sum_m p_{cijm} f_m(\mathbf{w}_{cijm}, X_{cj}). \tag{14}$$

The time–activity patterns and the microenvironmental models may have uncertainties characterized by probability distributions. These distributions are informed in part by data from a sample of children with short-term personal measurements of O_3 and concurrent daily diaries of time–activity patterns (Avol, Navidi and Colome, 1998). To allow for these uncertainties, we repeatedly evaluate (14) with random samples from the various inputs (times, exposures, model parameters, etc.), and summarize these by the mean $\bar{x}_{cij}$ and its variance.

In general, between-subjects variation in assigned exposures was small compared with between-community—9% of the total variance for O_3, 33% for PM_{10} and 3% for NO_2—and the within-person uncertainties were even smaller—1, 3 and 0.3% of the total, respectively. While the use of community mean personal exposures as $\mathbf{X}_c$ instead of the central site measurement led to modestly improved significance in some cases, the small within-community variance has so far precluded finding any significant associations at the individual level. We view these "assigned" exposures as having a Berkson error structure (Fuller, 1987), that is, the subjects' true exposures are randomly distributed

around these model predictions, as discussed further below. Below we also discuss a Monte Carlo approach to the problem that some components of these uncertainties are correlated across children.

One approach to examining the effect of personal modifying factors such as time–activity patterns views them as determinants of personal *exposure* x_{cij}, which is the covariate of primary interest. This is the approach taken to allow for time spent in different microenvironments in the models discussed in this section and could, in principle, be extended to allow for breathing rate differences in different activities to compute personal *dose* to the lung.

Another approach views time–activity patterns as *modifiers* of the effect of ambient exposure X_{cj}, now viewed as the covariate of primary interest, together with its interaction with such factors. For example, Gauderman et al. (2000) showed that the effect of ambient exposure was stronger in children who spent more time outdoors. The difference in annual MMEF growth rate per 20 ppb increase in NO_2 levels was -1.04% ($p = 0.01$) in more-outdoors children, but only -0.62% ($p = 0.17$) in less-outdoors children. This difference may be thought of as (1) due to better correlation between ambient and personal exposures or (2) due to modification of dose or risk by increased ventilation rates for the more-outdoors children.

Modeling local dispersion patterns. Extensive efforts are being made in the CHS to assess the heterogeneity of ambient exposure levels *within* study communities. Due to cost constraints, we can only obtain sample data at selected locations at particular times. Low-cost integrating monitors (Palmes tubes) were deployed for NO_2 at all 34 elementary schools and at a sample of 287 of the subjects' homes for 2-week periods in winter and summer. A means of using these data to predict pollution levels at any location is desirable.

Spatial interpolation techniques like kriging (Cressie, 1993) could be used descriptively. However, the sparseness of available measurements in space and time could render this approach unreliable as an indicator of the extent of the true variability in concentrations within communities. A promising alternative is to build models for dispersion from known sources. The strong observed associations with PM_{10} and NO_2 suggest that the most relevant source of pollution may be fresh motor vehicle exhaust. Fortunately, there are abundant data available in geographic information systems on traffic density patterns. Proximity to roadways may thus be a suitable surrogate for exposure to air pollution from motor vehicles (generally the dominant source of particulate and NO_2 pollution in Southern California). Furthermore, since traffic patterns are fairly stable, it may be a more accurate predictor of long-term average exposure than actual measurements, particularly where spatial interpolation is required.

Several models for predicting pollution have been proposed. These range from empirical models based on summing the traffic density on each nearby roadway with weights depending on the shortest distance (Pearson, Wachtel and Ebi, 2000; Rijnders et al., 2001) to complex models that account for traffic speeds, vehicle types and densities in each segment, along with meteorological information [e.g., U.S. EPA's MOBILE6 model (U.S. EPA, 2002)]. The former do not account for prevailing wind patterns, whereas the latter require more detailed information and more computation than would be feasible for large studies. We used the CALINE4 model (Benson, 1989), which incorporates traffic counts on major roadways along with local wind-rose data (the distribution of daily wind direction and speed) to estimate traffic-derived pollutant levels (e.g., NO_2, CO) at each child's residence and at study schools.

For the CHS, both traffic density and measured NO_2 concentrations vary significantly from house to house within all study communities. Spatial mixed effects models are being used (e.g., via PROC MIXED in SAS) to describe the dependence of measured NO_2 concentrations on traffic density and other factors. The basic model is of the form

$$\mathbf{X}_c \sim \mathrm{MVN}\big(\mathbf{Z}_c'\boldsymbol{\beta}, \sigma^2\mathbf{I} + \omega^2\mathbf{B}(\phi) + \tau^2\mathbf{A}_c(\theta)\big),$$

where $\mathbf{X}_c$ is the vector of all the available home measurements for community c, $\mathbf{Z}_c$ is a design matrix of covariates such as the ambient levels at the central site and the schools, and traffic density estimates of homes. The covariance structure has three components: an uncorrelated residual variance σ^2, a spatially correlated community random effect with variance ω^2 and distance parameter ϕ, and a spatially correlated home random effect with variance τ^2 and distance parameter θ. Both the community mean and several measures of traffic density were significant predictors. As a proportion of the total variance, 21% was independent residual, 68% was spatially correlated community effect and 10% was spatially correlated within the community. The predictions from this model can then be used to assign exposures to all study homes and estimate their uncertainties for inclusion in the exposure–

response analyses. Preliminary analyses suggest a modest influence of the within-community deviations on the subject-specific intercepts a_{ci} of MMEF, but not on their slopes. The asthmatics (at baseline) among the measured subset also tended to have the highest NO_2 levels within most communities.

Lifetime exposure. Cumulative exposures could vary substantially due to subjects' differing residential histories. We constructed exposure histories using data from the U.S. EPA's AIRS data base for the closest monitoring stations for each child who moved from outside of the study area. Here we required that at least 90% of the residential history be complete and located within the United States. A priori, one would expect between-community comparisons of ambient exposure to show a stronger effect on initial lung function in nonmigrants than in migrants, since the latter would also be influenced by their prior exposures. On the other hand, within-community comparisons of lifetime exposure up to entry into the study would be more informative only among migrants, since there would be no variation in nonmigrants at the same age. Differences in outcomes between migrants and nonmigrants could be due to differences in their mean exposures, to selection bias, to uncontrolled confounding or to differential effects of exposure measurement error. Exposures at certain critical periods of lung development could also be important. Hence, we plan to examine the influence of personal exposures accumulated over specific time windows to both baseline lung function and subsequent rates of change. To explore these issues, a two-level version of model (6) could be set up for baseline lung function, including personal covariates, separate effects of ambient exposure in migrants and nonmigrants, and the deviation of individual from community exposure (zero for nonmigrants), together with individual and community random effects:

$$Y_{ci} = \begin{cases} \alpha_{0,N} + \alpha_{3,N} X_{c,N} + \boldsymbol{\eta}_2^T \mathbf{z}_{ci} + e_c + e_{ci} \\ \qquad\qquad\qquad \text{for nonmigrants,} \\ \alpha_{0,M} + \alpha_{3,M} X_{c,M} + \alpha_{2,M}(x_{ci} - X_{c,M}) \\ \quad + \boldsymbol{\eta}_2^T \mathbf{z}_{ci} + e_c + e_{ci} \quad \text{for migrants.} \end{cases}$$

The above setup allows tests on whether $\alpha_{3,N} = \alpha_{3,M}$, $\alpha_{2,M} = 0$ or $\alpha_{2,M} = \alpha_{3,M}$. Preliminary results indicate a significant effect of individual lifetime exposure to O_3 on MMEF in migrants. In between-community comparisons for nonmigrants, associations tended to be negative (but nonsignificant) with all three pollutants. For migrants there was a significant protective effect of NO_2 that could be a reflection of its strong inverse correlation with O_3. There were no significant differences between $\alpha_{0,N}$ and $\alpha_{0,M}$. Further analysis with longitudinal data is planned that will allow for within-subject comparisons over time, in the hope of learning whether differential migration rates explain these observations. This also allows tests of whether previous exposure affects only the lung function levels at entry to the study or modifies the effect of current exposures on subsequent rates of change.

There are two components of uncertainty in lifetime exposure estimates: errors or gaps in residence histories and uncertainties in community- and year-specific ambient concentrations. We plan to address these areas by using Monte Carlo methods similar to those described above to repeatedly sample exposure levels from their uncertainty distribution for each place of residence (if location is known precisely) or for the corresponding region (if known only approximately) or from the distribution of all levels in that year (if completely unknown).

Measurement error models. The approaches discussed above to estimating personal exposure could be combined by using the spatial estimates of local outdoor exposures as input to the microenvironmental model and integrating over time. The goal here is to propagate the uncertainty in these personal exposure assignments through to the exposure–response analysis, so as to adjust for the bias due to measurement error and to conduct inference that reflects this uncertainty. There is a large literature on measurement error correction models (see, e.g., Fuller, 1987; Carroll, Ruppert and Stefanski, 1995; Thomas, Stram and Dwyer, 1993; Zeger et al., 2000).

There are two main conceptual models for measurement error: the "classical model," in which measured exposures are seen as distributed around the true (unknown) exposures for each individual, and the "Berkson model," in which the true exposures of individuals are seen as distributed around some assignment for a group with otherwise indistinguishable characteristics. The classical model would be appropriate for analysis of data derived from samples where personal exposures have been measured, say using a passive O_3 badge or household NO_2 exposures using Palmes tubes. The Berkson model is more appropriate for predictions from microenvironmental or traffic density models. Both models conventionally assume that measurement errors are independent across subjects. This assumption could be violated when subjects share exposures (e.g., through attending the same school) or

when the inputs to a prediction model are uncertain; see Stram (2002).

To incorporate the uncertainties in exposure–response modeling, the full likelihood that incorporates the measurement error process is given by

$$L(\beta) = \int L(\mathbf{y}|\mathbf{x};\beta)\Pr(\mathbf{x}|\mathbf{z})\,d\mathbf{z},$$

where $\mathbf{z}$ denotes the available measurements or inputs to the model predictions and $\mathbf{x}$ denotes the unknown true exposures. We used a Monte Carlo approximation to this likelihood described by Thomas, Stram and Dwyer (1993) via repeated realizations of predicted exposures $\hat{x}_{cijr}$ as

$$L(\beta) = R^{-1}\sum_{r=1}^{R} L(\mathbf{y}|\mathbf{x}_r;\beta).$$

This likelihood is evaluated via hierarchical sampling of realizations that reflect those components of uncertainty that are shared between subjects or across time. These could be shared uncertainties in true ambient exposures by all members of the same community. For example, for a given sample of model parameters and ambient exposures, we drew multiple samples of time–activity profiles and personal–microenvironmental covariates for each subject to generate distributions of $\mathbf{x}_r$ vectors with appropriate correlational structure.

4.3 Multipollutant Models

Due to high correlations between pollutants (Table 2), it is difficult to separate the effects of the different pollutants. In multipollutant models, rarely do two pollutants both contribute significantly to the same model. Rather than try to resolve the question of which is the "best" model, we wish to draw inferences on each pollutant's effect that take into account our uncertainty as to which other pollutants should be adjusted for. This is the problem that "Bayes model averaging" (Raftery, Madigan and Hoeting, 1997) attempts to address. The basic idea is to fit all possible models and then draw inference on the marginal distribution of model coefficients. These calculations are difficult, but have been facilitated by recent developments in Markov chain Monte Carlo (MCMC) methods (Gilks, Richardson and Spiegelhalter, 1996). We implemented a MCMC version of an approach described by George and Foster (2000) for application to the regression of the community-level effects B_c on a vector of $p = 1,\ldots,P$ predictor variables $\mathbf{X}_c$ (e.g., multiple pollutants) in the third-level model (5). The approach is general and could be applied to other parts of the mixed effects model described in Section 3.1.

Let $\mathcal{D} = \{(B_c, \mathbf{X}_c),\ c = 1,\ldots,C\}$ denote the data, let $m = 1,\ldots,2^P$ index the set of all possible regression models with subsets of the P variables, let q_m denote the number of variables in model m and let SS_m denote the corresponding regression sum of squares. George and Foster (2000) suggested priors for $\boldsymbol{\beta}_m$ and m of the form

$$\Pr(\boldsymbol{\beta}_m|m,\gamma) = N_{q_m}\{0, \gamma\sigma^2(X_m'X_m)^{-1}\} \quad \text{for } \gamma > 0, \tag{15}$$

$$\Pr(m|w) = w^{q_m}(1-w)^{P-q_m} \quad \text{for } w \in (0,1), \tag{16}$$

where γ and w are hyperparameters that control the variance of the coefficients and the parsimony of the models. They derived expressions for the full conditional distributions of m and β_m given $\mathcal{D}$ and the current estimates of γ and w, together with a marginal likelihood for γ and w given only $\mathcal{D}$. Finally, they suggested a form of Bayes model averaging in which the posterior density of β is obtained by averaging over the set of all possible models,

$$\Pr(\beta|\mathcal{D},\hat{\gamma},\hat{w}) = \sum_m \Pr(\beta_m|\mathcal{D},m,\hat{\gamma},\hat{w})\Pr(m|\mathcal{D},\hat{\gamma},\hat{w}). \tag{17}$$

Rather than evaluating these probabilities at the maximum likelihood estimators (MLEs) of γ and w, we use a fully Bayesian approach to integrate over the posterior distributions of these parameters using MCMC methods. This entails five types of sampling: $[m|\mathcal{D},\gamma,w]$, $[\boldsymbol{\beta}|\mathcal{D},\gamma,w]$, $[\gamma|\mathcal{D},m,w]$, $[w|\mathcal{D},m,\gamma]$ and $[\sigma^2|\mathrm{SS}_m,q_m]$. Details will be reported elsewhere. Our results below are based on 1,000,000 iterations of this process, after discarding 100,000 iterations to allow for convergence.

Inference can be based on the posterior distributions for m and $\boldsymbol{\beta}$ using informative priors for γ and w. This takes into account prior knowledge about the anticipated degree of parsimony, if available. Lacking such knowledge, we prefer to base our inferences on Bayes factors (BF), which minimize the influence of prior specifications by taking the ratio of the posterior to the prior and can be thought of as a form of marginal likelihood ratio (Kass and Raftery, 1995). For m, the Bayes factor is defined as $\mathrm{BF}(m) = \Pr(m|\mathcal{D})/\Pr(m)$, where the numerator is simply the frequency distribution of m from the MCMC iterations, and the denominator is $\int \Pr(m|w)\Pr(w)\,dw$, computed by sampling a large

number of values of w from the prior and averaging the resulting probabilities of m. From $\Pr(m|\mathcal{D})$ other marginals such as $\Pr(\beta_p \neq 0)$ and $\Pr(q_m)$ can be computed readily for comparison with their corresponding posterior distributions. For β_p we tabulated $\Pr(\beta_p \neq 0|\mathcal{D})$, $\Pr(\beta_p < 0|\beta_p \neq 0, \mathcal{D})$, $\Pr(\beta_p < 0|\mathcal{D})$, $E(\beta_p|\beta_p \neq 0, \mathcal{D})$ and $\text{var}(\beta_p|\beta_p \neq 0, \mathcal{D})$, and compared them with their corresponding priors to compute Bayes factors. We also computed Bayes factors to adjust for other pollutants [e.g., based on $\Pr(\beta_p < 0|\beta_q \neq 0, \mathcal{D})$ and its corresponding prior probability].

We applied this approach to estimates of the community-specific adjusted 4-year rates of change in MMEF separately for the two fourth-grade cohorts (enrolled in 1993 and in 1996), as defined earlier; thus $C = 24$ observations in total. They were regressed on the community mean ambient levels over 1992–1995 and 1996–1999 for seven pollutants, respectively: O_3, NO_2, PM_{10}, $PM_{2.5}$, elemental carbon (EC), organic carbon (OC) and acid (EC and OC are constituents of $PM_{2.5}$). Linear regression (adjusted for cohort) showed significant associations with all pollutants except O_3, with the strongest association for acid ($p = 0.0012$). No two-pollutant model had significant contributions from both pollutants, although O_3 was marginally significant ($p = 0.056$) in a model which included NO_2 ($p = 0.006$).

The posterior probability that no variables contribute to the model was 0.112, much less than its prior probability 0.419, for a $\text{BF}(m)$ of 0.268 relative to all models combined. Models with one or two variables had BFs greater than 1 relative to the set of all possible models, or 7.58 and 5.29, respectively, relative to the null model. According to guidelines suggested by Kass and Raftery (1995), a BF of 1–3 is interpreted as "very mild evidence," 3–20 as "positive," 20–150 as "strong" and greater than 150 as "very strong." By these criteria, the evidence for the single best-fitting model that contains only acid is "strong" ($\text{BF} = 30$ relative to the null model). Marginally, the evidence that implicates acid is only "positive." Eight two-pollutant models had BFs greater than 2, including all models with acid. No three-pollutant or higher models had $\text{BF}(m) > 2$, although 22 of the 35 possible three-pollutant models had BFs greater than 1, again predominantly those which included acid. Single-pollutant models with NO_2, PM_{10} and EC also had BFs greater than 1, relative to all models.

Table 4 summarizes the marginal distribution of the β_p for each pollutant, averaging over the entire model space. The evidence for acid is the strongest, with a $\text{BF}(\beta_p \neq 0) = 5.27$, $\text{BF}(\beta_p < 0|\beta_p \neq 0) = 3.79$ and $\text{BF}(\beta_p < 0) = 7.53$. However, the marginal distribution has a much larger variance than the conditional distribution for the model with only acid: the conditional MLE and its standard error (SE) is -0.124 ± 0.033, whereas the marginal mean and standard deviation (SD) is only -0.024 ± 0.054. This difference reflects the larger conditional SEs for the multipollutant models which include acid and the between-model variance in the $\hat{\beta}_p$'s.

Ultimately, we suspect that even the use of sophisticated techniques may not resolve the multipollutant issue without exploiting comparisons at other spatial and/or temporal levels.

Summarization across subgroups. Bayesian model averaging techniques also could be used to explore ef-

TABLE 4
Marginal distribution of β_p for the seven pollutants under consideration in the Children's Health Study

Posterior summary	O_3	NO_2	PM_{10}	$PM_{2.5}$	EC[a]	OC[b]	Acid[c]
$\Pr(\beta_p \neq 0|\mathbf{B})$	0.1312	0.1653	0.1116	0.1339	0.1051	0.2039	0.5061
$\text{BF}(\beta_p \neq 0)$	0.78	1.02	0.65	0.79	0.60	1.32	5.27
$\Pr(\beta_p < 0|\beta_p \neq 0, \mathbf{B})$	0.6280	0.6672	0.5210	0.5926	0.5525	0.7118	0.7911
$\text{BF}(\beta_p < 0|\beta_p \neq 0)$	1.69	2.00	1.09	1.45	1.23	2.47	3.79
$\Pr(\beta_p < 0|\mathbf{B})$	0.0824	0.1103	0.0582	0.0793	0.0580	0.1452	0.4004
$\text{BF}(\beta_p < 0)$	1.01	1.40	0.70	0.97	0.70	1.92	7.53
$E(\beta_p|\beta_p \neq 0, \mathbf{B})$	−0.0034	−0.0051	0.0006	−0.0039	−0.0061	−0.2093	−0.0338
$\text{SD}(\beta_p|\beta_p \neq 0, \mathbf{B})$	0.0121	0.0196	0.0218	0.0337	0.0706	0.5344	0.0535

NOTE: $\Pr(\beta_p \neq 0) = 0.1629$; $\Pr(\beta_p < 0|\beta_p \neq 0) = 0.5000$; $\Pr(\beta_p < 0|\beta_p \neq 0) = 0.5000$.
[a]Elemental carbon (in ppb).
[b]Organic carbon (in ppb).
[c]Composed of HNO_3 and HCl (in ppb).

fect modification and to assess the consistency of findings over several stratification factors. Some factors are expected to modify the slope of air pollution effects and some of these modifying factors may interact with each other, for example, girls reach their growth spurts earlier than boys, leading to the need for joint stratification by age and sex due to the possible three-way age × sex × pollution interaction effect. As the number of such stratification factors increases, we expect to see some spurious interaction effects simply by chance, due to the increase in the number of strata and the corresponding reduction in sample size per stratum. Hence, lacking a strong prior belief that there ought to be many higher-order interactions, one ought to try to average stratum-specific slope estimates over similar strata and thus improve the power for detecting real heterogeneity. Bayes model averaging techniques offer a way to average over the set of all possible interaction models. See Chipman (1996) for more discussion on models with main effects and interactions.

4.4 Integrated Analysis of Several Longitudinal Outcomes

The methods discussed so far deal with each type of outcome (lung function, symptoms, school absences) univariately. It is of interest to perform integrated analysis of several outcomes to be able to examine patterns in lung function tests and/or to identify biologically meaningful trends in the effects of air pollution on children's health. For example, we may wish to test whether respiratory illnesses lead to long-term decrements in lung function or, conversely, whether poor lung function is a risk factor for respiratory illness. A major strength of longitudinal studies is that they offer a way to disentangle such complex interrelationships. Nevertheless, the autocorrelation of the different endpoints may make it difficult to infer the direction of causation even in longitudinal data.

An overall strategy for addressing this question could be set up as follows. Let $\mathbf{y}_{cij} = (y_{cij}^{(1)}, y_{cij}^{(2)})$ denote a vector of multivariate responses over time, where, for example, $y_{cij}^{(1)}$ might be lung function measurements at time t_{cij} and $y_{cij}^{(2)}$ might be the number of school absences for respiratory illness between $t_{ci,j-1}$ and t_{cij}. For simplicity, we treat both endpoints as continuous, although other types of outcome as in the generalized linear models framework could be considered. We consider a coupled set of models of the form

$$E(y_{cij}^{(1)} - y_{ci,j-1}^{(1)}) = a_{ci}^{(1)} + b_{ci}^{(1)}(t_{cij} - t_{ci,j-1}) + \boldsymbol{\gamma}_1^{(1)T}\mathbf{z}_{cij} + \phi^{(1)} y_{ci,j-1}^{(2)},$$

$$E(y_{cij}^{(2)}) = a_{ci}^{(2)} + b_{ci}^{(2)}(t_{cij} - t_{ci,j-1}) + \boldsymbol{\gamma}_1^{(2)T}\mathbf{z}_{cij} + \phi^{(2)} y_{ci,j-1}^{(1)}.$$

Here, $\phi^{(1)}$ measures the dependence of lung function changes on previous school absences and $\phi^{(2)}$ measures the dependence of school absences on previous lung function levels. The models account for intrasubject serial correlation.

In the CHS, all lung functions [except for forced vital capacity (FVC)] showed qualitatively similar associations with all pollutants. Associations were somewhat stronger with measures of flow rates (e.g., MMEF) than with measures of lung volume (e.g., FEV_1). This suggests that air pollution may be a stronger determinant of small airways obstructive disease. Univariate analysis provides a means to describe such patterns, but it may fail in formally testing whether the associations for several measures are significantly different. Moreover, it may lead to a proliferation of significance tests with elevated risk of false positive inferences. However, because of their relatively high correlation, simple Bonferroni adjustment of p-values may not suffice. Some form of multivariate analysis would therefore be desirable.

A two stage least squares approach for modeling two continuous outcomes by modeling one outcome, followed by a model for the second outcome with predicted values from the first model as covariates, was proposed by Amemiya (1985). This was later extended to allow for joint analysis of a discrete and a continuous outcome via a latent variable approach (Bartholomew, 1987; Catalano and Ryan, 1992). Fitzmaurice and Laird (1995) proposed an alternative model that instead focused on the discrete outcome. Direct extensions of the work of Laird and Ware (1982) for inference on multiple outcomes also have been proposed for balanced data (Reinsel, 1984) and unbalanced designs with possibly missing data (Shah, Laird and Schoenfeld, 1997).

A descriptive analysis of the relationship between lung function tracking and asthma status by age at diagnosis was reported by Berhane et al. (2000) using the model

$$\mathrm{E}[\log(y_{cij}^{(1)})] = a_{ci} + f_1(t_{cij}) + f_2(t_{cij}) \cdot \log(\text{height}) + \alpha_{y_{cij}^{(2)}} + \beta_{y_{cij}^{(2)}} t_{cij} + \boldsymbol{\gamma}_1^{(1)T}\mathbf{z}_{cij},$$

where $a_{ci} \sim N(0, \sigma_a^2)$ is a subject-specific random intercept, $f_1(t)$ is a smooth function of age (t) and $f_2(t)$ is a smooth function of age that depicts the age-dependent

TABLE 5
Gender-specific effects of asthma on MMEF by age at diagnosis based on 4-year follow-up data on participants of the Children's Health Study

Gender	Asthma by age at diagnosis (years)	Main effect of asthma		Trend (*asthma* × *age*)	
		% difference	95 C.I.	% difference	95 C.I.
Female	0–2	−15.7[a]	(−25.6, −4.4)	−0.8	(−3.4, 1.8)
	3–5	−2.8	(−15.8, 12.2)	−0.4	(−3.5, 2.8)
	6–9	−2.2	(−9.4, 5.7)	0.1	(−1.6, 1.8)
	≥ 10	−5.7[b]	(−9.7, −1.5)	−0.01	(−0.9, 0.9)
Male	0–2	−18.8[b]	(−26.0, −11.4)	0.6	(−1.0, 2.2)
	3–5	4.3	(−5.9, 15.6)	−0.1	(−2.0, 1.8)
	6–9	−7.3[a]	(−13.4, −0.8)	−1.0	(−2.3, 0.4)
	≥ 10	−6.5[b]	(−11.0, −1.7)	−0.2	(−1.1, 0.8)

[a] $p < 0.01$.
[b] $p < 0.05$.

slopes of log(height), both fitted using regression splines (see Section 3.2 for details). Here, α_y measures parallel deviations between the four asthma groups $y^{(2)}$, and β_y depicts the trend in the deviations over time between the four groups. Table 5 summarizes the estimates of the deviation and trend parameters. Early diagnosis of asthma is associated with a significant deficit in MMEF in both females and males, but the percent differences do not seem to change with increasing age. Figure 4 depicts this relationship for females in various "age at diagnosis" categories.

4.5 Exposure Metrics

For annual outcomes such as lung function, annual averages X_c of daily summary exposure measures (e.g., 24-hour mean) could be used. However, the choice of an appropriate "metric" is crucial, since pollution levels vary across the year and even within days, and the biological effect of this variation is not well understood.

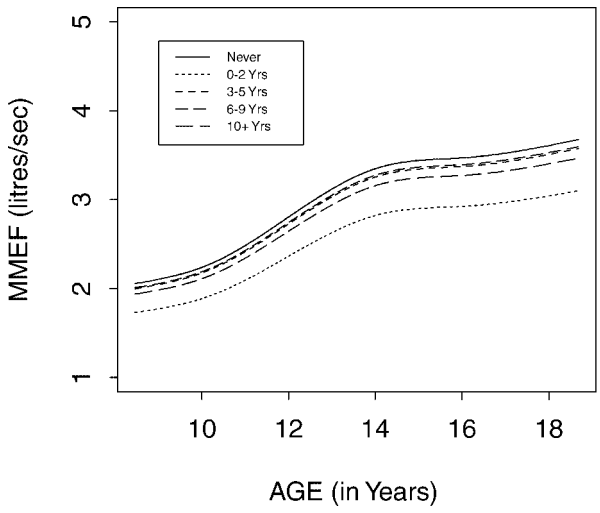

FIG. 4. *Growth curves for MMEF by age at diagnosis of asthma for 8–18-year-old participants of the Children's Health Study.*

For the CHS, our choice of the time-weighted average is based on a hypothesis that chronic effects represent an accumulation of small insults received continuously over time and these insults depend linearly on pollution levels. This approach can be extended to allow for nonlinearities in the instantaneous dose–response relationship or interactions between pollutants, possibly lagged over time. Our approach is not to try to estimate the parameters of this unobserved continuous process. Instead, we use the predictions of such a process to motivate the choice of exposure metrics for the chronic effects analysis that could test hypotheses about nonlinearities and interactions.

We are currently evaluating a family of additional exposure metrics which are motivated by a Taylor series approximation to a general dose–response relationship of the form $Y = \int g[X(t)]\,dt$, where $g(X)$ denotes the effect of ambient exposure on the instantaneous rate of change of Y. Following this line of reasoning, a general test of nonlinearity in $g(X)$ is obtained by adding as a covariate in a community-level model the variance V_c of exposure over time in each community (Thomas, 1988). These variances can be decomposed into hourly, daily, weekly, seasonal and annual components.

To distinguish nonlinearities in the instantaneous dose–response from nonlinearity in the relationship between Y and X, variance-based metrics could be

added to a model that includes not just X, but also some transformation(s) thereof, such as X^2 or $X \ln X$. Transformations of the X axis leave the rankings of the communities intact, whereas addition of V_c to the model can change their rankings and, hence, contribute additional information.

This approach could be extended to allow for (1) effects of rate of change of exposure $X'(t) = dX/dt$ via metrics of the form $\int X(t)X'(t)\,dt$; (2) interactions between pollutants via metrics of the form $\int X_1(t)X_2(t)\,dt$ to be added to a model that includes the main effects of X_p $(p = 1, 2)$, their product and their respective variances V_p; and (3) lag effects via metrics of the form $\int X_1(t)X_2'(t)\,dt$. Various indices of interest to the regulatory community may be considered, such as threshold models of the form $\int \max\{X(t) - \tau, 0\}\,dt$ for values of thresholds, τ, such as the current standards and proposed smaller values.

5. CONCLUSIONS

The Southern California Children's Health Study is an ongoing longitudinal study that is already yielding results that are likely to help in understanding the chronic effects of air pollution on children's health. Because of the regulatory significance of such results, the development and use of appropriate state-of-the-art statistical methods are of paramount importance. The unique study design and the complexity of the substantive research questions provide an exciting opportunity for development of new statistical methodology. By publishing this general approach and review of methodologic issues and developments while the study is in progress, we invite input from the broader scientific and statistical community about ways to refine the approach or alternative approaches we should consider.

This study is presently the longest running cohort study in the United States of the health effects of air pollution specifically targeted at children. Although the original funding will terminate upon graduation of the last fourth-grade cohort from high school, a new study will investigate the potential for long-term effects to continue into adulthood and will enroll a new cohort of kindergarten children that will have a special emphasis on early life influences on asthma incidence.

Although the methods described here are motivated by cohort studies on the effects of air pollution, such as the CHS, they do have potential applicability to a broader range of studies in environmental health, not necessarily restricted to air pollution, to respiratory disease or even necessarily to designs that entail longitudinal observation and comparisons at both the individual and ecologic levels. Many environmental agents are geographically determined, necessitating some form of ecologic comparison (Thomas, 2000).

The development of optimal study designs that involve careful selection of geographic areas and appropriate balancing of resources between ecologic, individual and temporal observations should be of high priority. Numerous analytical challenges remain, many of which have been touched on herein. In particular, the problem of exposure measurement error is ubiquitous and serious in environmental epidemiology, and has been an active area of statistical research; epidemiologic applications, however, remain in their infancy. It is hoped that this article will stimulate further research along these lines.

ACKNOWLEDGMENTS

We thank John Peters, Frank Gilliland, Rob McConnell, Ed Avol and Fred Lurmann for valuable discussions, and Ed Rappaport, Hita Vora, Talat Islam and Jassy Molitor for data analytic assistance. Comments from the external advisory committee of the Children's Health Study are greatly appreciated. We thank the Editor and two anonymous referees for their extensive comments that strengthened the article. This research was supported by the California Air Resources Board (Contract A033-186), the NIEHS (Grants 5P30ES07048-04 and 1PO1ESO939581-01) and the U.S. EPA (Contract CR824034-01-3 and Grant R826708-01-0).

REFERENCES

ABBEY, D. E., NISHINO, N., MCDONNELL, W. F., BURCHETTE, R. J., KNUTSEN, S. F., BEESON, W. L. and YANG, J. X. (1999). Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *American J. Respiratory and Critical Care Medicine* **159** 373–382.

AMEMIYA, T. (1985). *Advanced Econometrics*. Blackwell, Oxford.

AVOL, E. L., GAUDERMAN, W. J., TAN, S. M., LONDON, S. J. and PETERS, J. M. (2001). Respiratory effects of relocating to areas of differing air pollution levels. *American J. Respiratory and Critical Care Medicine* **164** 2067–2072.

AVOL, E. L., NAVIDI, W. C. and COLOME, S. D. (1998). Modeling ozone levels in and around Southern California homes. *Environmental Science and Technology* **32** 463–468.

BARTHOLOMEW, D. (1987). *Latent Variable Models and Factor Analysis*. Oxford Univ. Press.

BENSON, P. (1989). CALINE4—A dispersion model for predicting air pollution concentration near roadways. Report, Office of Transportation Laboratory, California Department of Transportation, Sacramento, CA.

BERHANE, K., MCCONNELL, R., GILLILAND, F., ISLAM, T., GAUDERMAN, W. J., AVOL, E., LONDON, S. J., RAPPAPORT, E., MARGOLIS, H. G. and PETERS, J. M. (2000). Sex-specific effects of asthma on pulmonary function in children. *American J. Respiratory and Critical Care Medicine* **162** 1723–1730.

BERHANE, K. and THOMAS, D. (2002). A two-stage model for multiple time series data of counts. *Biostatistics* **3** 21–32.

BRENNER, H., SAVITZ, D. A., JÖCKEL, K.-H. and GREENLAND, S. (1992). Effects of nondifferential exposure misclassification in ecologic studies. *American J. Epidemiology* **135** 85–95.

BRESLOW, N. E. and CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88** 9–25.

BRESLOW, N. E. and LIN, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82** 81–91.

BRUMBACK, B., RYAN, L., SCHWARTZ, J., NEAS, L., STARK, P. and BURGE, H. (2000). Transitional regression models, with application to environmental time series. *J. Amer. Statist. Assoc.* **95** 16–27.

BRYK, A., RAUDENBASH, S. and CONGDON, R. (1996). *HLM*: *Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Scientific Software International, Chicago.

BURKE, J. M., ZUFALL, M. J. and ÖZKAYNAK, H. (2001). A population exposure model for particulate matter: Case study results for $PM_{2.5}$ in Philadelphia, PA. *J. Exposure Analysis and Environmental Epidemiology* **11** 470–489.

BURNETT, R., MA, R., JERRETT, M., GOLDBERG, M. S., CAKMAK, S., POPE, C. A. and KREWSKI, D. (2001). The spatial association between community air pollution and mortality: A new method of analyzing correlated geographic cohort data. *Environmental Health Perspectives* **109**(suppl.) 375–380.

CARROLL, R. J., RUPPERT, D. and STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall, London.

CATALANO, P. J. and RYAN, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *J. Amer. Statist. Assoc.* **87** 651–658.

CHIPMAN, H. (1996). Bayesian variable selection with related predictors. *Canad. J. Statist.* **24** 17–36.

COHEN, B. L. (1990). Ecological versus case-control studies for testing a linear-no threshold dose–response relationship. *International J. Epidemiology* **19** 680–684.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*, revised ed. Wiley, New York.

DARBY, S., DEO, H., DOLL, R. and WHITLEY, E. (2001). A parallel analysis of individual and ecological data on residential radon and lung cancer in south-west England. *J. Roy. Statist. Soc. Ser. A* **164** 193–203.

DE BOOR, C. (1978). *A Practical Guide to Splines*. Springer, New York.

DEMPSTER, A. P., RUBIN, D. B. and TSUTAKAWA, R. K. (1981). Estimation in covariance components models. *J. Amer. Statist. Assoc.* **76** 341–353.

DIGGLE, P. J., LIANG, K.-Y. and ZEGER, S. L. (1994). *Analysis of Longitudinal Data*. Oxford Univ. Press.

DOCKERY, D. W., POPE, C. A., XU, X., SPENGLER, J. D., WARE, J. H., FAY, M. E., FERRIS, B. G. and SPEIZER, F. E. (1993). An association between air pollution and mortality in six U.S. cities. *New England J. Medicine* **329** 1753–1759.

DOCKERY, D. W., SPEIZER, F. E., STRAM, D. O., WARE, J. H., SPENGLER, J. D. and FERRIS, B. G. (1989). Effects of inhalable particles on respiratory health of children. *American Review of Respiratory Disease* **139** 587–594.

DOMINICI, F., SAMET, J. and ZEGER, S. L. (2000). Combining evidence on air pollution and daily mortality from the largest 20 US cities: A hierarchical modeling strategy (with discussion). *J. Roy. Statist. Soc. Ser. A* **163** 263–302.

DUAN, N. (1982). Models for human exposure to air pollution. *Environment International* **8** 305–309.

DUAN, N. (1991). Stochastic microenvironment models for air pollution exposure. *J. Exposure Analysis and Environmental Epidemiology* **1** 235–257.

FERRIS, B. G., SPEIZER, F. E., SPENGLER, J. D., DOCKERY, D., BISHOP, Y. M., WOLFSON, M. and HUMBLE, C. (1979). Effects of sulfur oxides and respirable particles on human health: Methodology and demography of populations in study. *American Review of Respiratory Disease* **120** 767–779.

FIREBAUGH, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review* **43** 557–572.

FITZMAURICE, G. M. and LAIRD, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *J. Amer. Statist. Assoc.* **90** 845–852.

FRISCHER, T., STUDNICKA, M., GARTNER, C., TAUBER, E., HORAK, F., VEITER, A., SPENGLER, J., KÜHR, J. and URBANEK, R. (1999). Lung function growth and ambient ozone: A three-year population study in school children. *American J. Respiratory and Critical Care Medicine* **160** 390–396.

FULLER, W. (1987). *Measurement Error Models*. Wiley, New York.

GAUDERMAN, W. J., MCCONNELL, R., GILLILAND, F., LONDON, S., THOMAS, D., AVOL, E., VORA, H., BERHANE, K., RAPPAPORT, E. B., LURMANN, F., MARGOLIS, H. G. and PETERS, J. M. (2000). Association between air pollution and lung function growth in Southern California children. *American J. Respiratory and Critical Care Medicine* **162** 1383–1390.

GEORGE, E. I. and FOSTER, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87** 731–747.

GILKS, W., RICHARDSON, S. and SPIEGELHALTER, D., eds. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

GONG, H., JR., WONG, R., SARMA, R., LINN, W. S., SULLIVAN, E. D., SHAMOO, D. A., ANDERSON, K. R. and PRASAD, S. B. (1998). Cardiovascular effects of ozone exposure in human volunteers. *American J. Respiratory and Critical Care Medicine* **158** 538–546.

GREENLAND, S. (2001). Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International J. Epidemiology* **30** 1343–1350.

GREENLAND, S. (2002). A review of multilevel theory for ecologic analyses. *Statistics in Medicine* **21** 389–395.

GREENLAND, S. and BRENNER, H. (1993). Correcting for non-differential misclassification in ecological analyses. *Appl. Statist.* **42** 117–126.

GREENLAND, S. and MORGENSTERN, H. (1989). Ecological bias, confounding, and effect modification. *International J. Epidemiology* **18** 269–274.

GREENLAND, S. and ROBINS, J. (1994). Ecologic studies—biases, misconceptions, and counterexamples (with discussion). *American J. Epidemiology* **139** 747–771.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems (with discussion). *J. Amer. Statist. Assoc.* **72** 320–340.

HASTIE, T. and TIBSHIRANI, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

HASTIE, T. and TIBSHIRANI, R. (2000). Bayesian backfitting (with discussion). *Statist. Sci*. **15** 196–223.

HEAGERTY, P. and ZEGER, S. L. (2000). Marginalized multilevel models and likelihood inference (with discussion). *Statist. Sci*. **15** 1–26.

JOHNSON, T., LONG, T. and OLLISON, W. (2000). Prediction of hourly microenvironmental concentrations of fine particles based on measurements obtained from the Baltimore scripted activity study. *J. Exposure Analysis and Environmental Epidemiology* **10** 403–411.

KASS, R. and RAFTERY, A. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90** 773–795.

KELSALL, J. E., ZEGER, S. L. and SAMET, J. M. (1999). Frequency domain log-linear models; air pollution and mortality. *Appl. Statist*. **48** 331–344.

KREWSKI, D., BURNETT, R., GOLDBERG, M., HOOVER, K., SIEMIATYCKI, J., JERRETT, M., ABRAHAMOWICZ, M. and WHITE, W. (2003). Overview of the reanalysis of the Harvard six-cities study and the American Cancer Society study of particulate air pollution and mortality. *J. Toxicology and Environmental Health* **66** 1507–1551.

KÜNZLI, N. and TAGER, I. B. (1997). The semi-individual study in air pollution epidemiology: A valid design as compared to ecologic studies. *Environmental Health Perspectives* **105** 1078–1083.

LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963–974.

LIN, X. and BRESLOW, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Amer. Statist. Assoc.* **91** 1007–1016.

LIN, X. and ZHANG, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 381–400.

LIOY, P. J., FREEMAN, N. C., WAINMAN, T., STERN, A. H., BOESCH, R., HOWELL, T. and SHUPACK, S. I. (1992). Microenvironmental analysis of residential exposure to chromium-laden wastes in and around New Jersey homes. *Risk Analysis* **12** 287–299.

LITTELL, R. C., MILLIKEN, G. A., STROUP, W. W. and WOLFINGER, R. D. (1996). *SAS System for Mixed Models*. SAS Institute, Inc., Cary, NC.

LUBIN, J. H. (1998). On the discrepancy between epidemiologic studies in individuals of lung cancer and residential radon and Cohen's ecologic regression. *Health Physics* **75** 4–10.

MA, R., KREWSKI, D. and BURNETT, R. (2000). Random effects Cox models: A Poisson modeling approach. Report, Laboratory for Research in Statistics and Probability, Univ. Ottawa.

MACDONALD, I. L. and ZUCCHINI, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall, London.

MCCONNELL, R., BERHANE, K., GILLILAND, F., LONDON, S. J., ISLAM, T., GAUDERMAN, W. J., AVOL, E., MARGOLIS, H. G. and PETERS, J. M. (2002). Asthma in exercising children exposed to ozone: A cohort study. *Lancet* **359** 386–391.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London.

MCCURDY, T. (1995). Estimating human exposure to selected motor vehicle pollutants using the NEM series of models: Lessons to be learned. *J. Exposure Analysis and Environmental Epidemiology* **5** 533–550.

MORGENSTERN, H. (1982). Uses of ecologic analysis in epidemiologic research. *American J. Public Health* **72** 1336–1344.

MORGENSTERN, H. (1995). Ecologic studies in epidemiology: Concepts, principles, and methods. *Annual Review Public Health* **16** 61–81.

NAVIDI, W. and LURMANN, F. (1995). Measurement error in air pollution exposure assessment. *J. Exposure Analysis and Environmental Epidemiology* **5** 111–124.

NAVIDI, W., THOMAS, D., STRAM, D. and PETERS, J. M. (1994). Design and analysis of multilevel analytic studies with applications to a study of air pollution. *Environmental Health Perspectives* **102** 25–32.

PEARCE, N. (2000). The ecological fallacy strikes back. *J. Epidemiology and Community Health* **54** 326–327.

PEARSON, R. L., WACHTEL, H. and EBI, K. L. (2000). Distance-weighted traffic density in proximity to a home is a risk factor for leukemia and other childhood cancers. *J. Air and Waste Management Association* **50** 175–180.

PETERS, J. M., AVOL, E., NAVIDI, W., LONDON, S. J., GAUDERMAN, W. J., LURMANN, F., LINN, W. S., MARGOLIS, H., RAPPAPORT, E., GONG, H., JR. and THOMAS, D. (1999a). A study of twelve southern California communities with differing levels and types of air pollution. I. Prevalence of respiratory morbidity. *American J. Respiratory and Critical Care Medicine* **159** 760–767.

PETERS, J. M., AVOL, E., GAUDERMAN, W. J., LINN, W. S., NAVIDI, W., LONDON, S. J., MARGOLIS, H., RAPPAPORT, E., VORA, H., GONG, H., JR. and THOMAS, D. (1999b). A study of twelve southern California communities with differing levels and types of air pollution. II. Effects on pulmonary function. *American J. Respiratory and Critical Care Medicine* **159** 768–775.

PINHEIRO, J. C. and BATES, D. M. (2000). *Mixed-Effects Models in S and S-Plus*. Springer, New York.

POPE, C. A., THUN, M. J., NAMBOODIRI, M. M., DOCKERY, D., EVANS, J., SPEIZER, F. and HEATH, C. M., JR. (1995). Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American J. Respiratory and Critical Care Medicine* **151** 669–674.

PRENTICE, R. L. and SHEPPARD, L. (1995). Aggregate data studies of disease risk factors. *Biometrika* **82** 113–125.

RAFTERY, A. E., MADIGAN, D. and HOETING, J. A. (1997). Bayesian model averaging for linear regression models. *J. Amer. Statist. Assoc.* **92** 179–191.

RAIZENNE, M., NEAS, L. M., DAMOKOSH, A. I., DOCKERY, D. W., SPENGLER, J. D., KOUTRAKIS, P., WARE, J. H. and SPEIZER, F. E. (1996). Health effects of acid aerosols on North American children: Pulmonary function. *Environmental Health Perspectives* **104** 506–514.

RAMSAY, J. O. and SILVERMAN, B. W. (1997). *Functional Data Analysis*. Springer, New York.

RANSOM, M. R. and POPE, C. A. (1992). Elementary school absences and PM_{10} pollution in Utah Valley. *Environmental Research* **58** 204–219.

RASBASH, J. and WOODHOUSE, G. (1995). MLn command reference, Version 1.0. Multilevel Models Project, Institute of Education, Univ. London.

REINSEL, G. (1984). Estimation and prediction in a multivariate random effects generalized linear model. *J. Amer. Statist. Assoc.* **79** 406–414.

RIJNDERS, E., JANSSEN, N. A. H., VAN VLIET, P. H. N. and BRUNEKREEF, B. (2001). Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density. *Environmental Health Perspectives* **109**(suppl.) 411–417.

ROBINSON, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* **15** 351–357.

SAMET, J., ZEGER, S. and BERHANE, K. (1995). The association of mortality and particulate air pollution. Phase I Report, Particulate Epidemiology Evaluation Project. Health Effects Institute, Boston, 1–104.

SCHWARTZ, J. (1994). Air pollution and daily mortality: A review and meta-analysis. *Environmental Research* **64** 36–52.

SELVIN, H. C. (1958). Durkheim's suicide and problems of empirical research. *American J. Sociology* **63** 607–619.

SHAH, A., LAIRD, N. and SCHOENFELD, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *J. Amer. Statist. Assoc.* **92** 775–779.

SHEPPARD, L. (2003). Insights on bias and information in group-level studies. *Biostatistics* **4** 265–278.

STIDLEY, C. A. and SAMET, J. M. (1994). Assessment of ecologic regression in the study of lung cancer and indoor radon. *American J. Epidemiology* **139** 312–322.

STRAM, D. O. (1996). Meta-analysis of published data using a linear mixed-effects model. *Biometrics* **52** 536–544.

STRAM, D. O. (2002). Shared uncertainty. In *Letter Report*: *Review of a Research Protocol Prepared by the University of Utah.* Board on Radiation Effects Research, National Academy of Sciences, Washington.

THOMAS, D. C. (1988). Models for exposure-time–response relationships with applications to cancer epidemiology. *Annual Review of Public Health* **9** 451–482.

THOMAS, D. C. (2000). Some contributions of statistics in environmental epidemiology. *J. Amer. Statist. Assoc.* **95** 315–319.

THOMAS, D. C., STRAM, D. and DWYER, J. (1993). Exposure measurement error: Influence on exposure–disease relationships and methods of correction. *Annual Review of Public Health* **14** 69–93.

U.S. EPA (2002). *Users Guide to MOBILE6 Emission Factor Model*. Environmental Protection Agency, U.S. GPO, Washington.

WAKEFIELD, J. and ELLIOTT, P. (1999). Issues in the statistical analysis of small area health data. *Statistics in Medicine* **18** 2377–2399.

WAKEFIELD, J. and SALWAY, R. (2001). A statistical framework for ecological and aggregate studies. *J. Roy. Statist. Soc. Ser. A* **164** 119–137.

WARE, J. H. and STRAM, D. O. (1988). Statistical issues in epidemiologic studies of the health effects of ambient acid aerosols. *Canad. J. Statist.* **16** 5–13.

WYPIJ, D., PUGH, M. and WARE, J. H. (1993). Modeling pulmonary function growth with regression splines. *Statist. Sinica* **3** 329–350.

ZANOBETTI, A., WAND, M., SCHWARTZ, J. and RYAN, L. (2000). Generalized additive distributed lag models: Quantifying mortality displacement. *Biostatistics* **1** 279–292.

ZEGER, S. L. (1988). A regression model for time series of counts. *Biometrika* **75** 621–629.

ZEGER, S. L., THOMAS, D., DOMINICI, F., SAMET, J. M., SCHWARTZ, J., DOCKERY, D. W. and COHEN, A. (2000). Exposure measurement error in time-series studies of air pollution: Concepts and consequences. *Environmental Health Perspectives* **108** 419–426.

ZIDEK, J. V., WONG, H., LE, N. D. and BURNETT, R. (1996). Causality, measurement error and multicollinearity in epidemiology. *Environmetrics* **7** 441–451.

Comment

Wendy Meiring

1. INTRODUCTION

I thank Berhane, Gauderman, Stram and Thomas (hereafter referred to as BGST) for providing a valuable review of the analyses used in the Southern California Children's Health Study (CHS). This cohort study of chronic air pollution effects follows children over several years. The task of designing, implementing and analyzing such a study presents enormous challenges, especially since pollution exposure monitoring is limited by cost. The authors and the CHS researchers are to be congratulated on every-

Wendy Meiring is Associate Professor, Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, USA (e-mail: meiring@pstat.ucsb.edu).

thing they have achieved with limited resources. There remain numerous challenging opportunities for continued research, combining statistics and science to assess air pollution health impacts on children. I refer to Dominici, Sheppard and Clyde (2003) for a complementary review of methods and designs for both acute and chronic health effect studies on a variety of time scales and age groups, and to Piegorsch, Smith, Edwards and Smith (1998) for an environmental statistics review that also discusses health impacts.

2. GENERALIZED LINEAR MIXED MODEL, DIAGNOSTICS AND INFERENCE

As BGST indicate, their model (6) is a generalized linear mixed model (GLMM). Many recent GLMM research advances were reviewed by McCulloch (2003), together with further research needs. For example, additional diagnostics are needed to evaluate sensitivity to the random effects distribution, and likelihood inference for GLMM variance components is not well understood for small samples. BGST also mention the need for additional research on both diagnostics and inference. In this section I elaborate further on this need. I later pose several scientific questions that relate to model specification in the CHS study and discuss spline-type models to study interactive air pollution health impacts.

By grouping together subject- and community-specific terms, Equation (6) in BGST may be written in the general GLMM form (see, e.g., McCulloch, 2003, Equation 4.5)

$$Y_i|\mathbf{u} \sim f_{Y_i|\mathbf{u}}(y_i|\mathbf{u}), \tag{1}$$

$$E[Y_i|\mathbf{u}] = \mu_i,$$

$$g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{u},$$

$$\mathbf{u} \sim f_{\mathbf{U}}(\mathbf{u}), \tag{2}$$

where $f_{Y_i|\mathbf{u}}(y_i|\mathbf{u})$ is an exponential family probability density/mass function, $\mathbf{u}$ is a vector of random effects, $\mathbf{w}_i$ indicates the multipliers of random effects for response i, $\mathbf{x}_i$ are the fixed effect explanatory variables for response i, $\boldsymbol{\beta}$ are parameters that correspond to the fixed effects and g is a link function. Correlation is incorporated through correlated random effects $\mathbf{u}$ in (2) or by random effects that are common to sets of observations. The conditional distributions in (2) often are assumed independent, but this assumption may be relaxed. Generalized linear mixed models extend several commonly used families of models, including generalized linear models and linear mixed models (for Gaussian data), as reviewed by McCulloch and Searle (2001) and McCulloch (2003).

Several components in the GLMM [BGST, (6)] were chosen, including the random effects distribution $f_{\mathbf{U}}(\mathbf{u})$, and the fixed and random effect design matrices. Similar components were chosen in their hierarchical models (1)–(5). Please would the authors comment further on their choice of design matrices, including whether interactions were included. In BGST model (6), the authors assume normally distributed random effects on the link scale and mention potential bias of ecological comparisons if the distributional assumptions are violated. I would value further discussion on available diagnostics for nonnormality of these effects, especially for non-Gaussian data, and sensitivity to these assumptions.

3. FLEXIBLE MODELS

In BGST models (1)–(5) and (6), fixed effect parameters and random effects enter linearly in the intercept and slope components. BGST later describe flexible B-spline models for age-dependent growth rate and the need for further inference methodology for functionals. The methodology for flexible models for functional/longitudinal/spatial data is expanding rapidly, including approaches with spline and mixed model connections (in addition to references provided by BGST, see, e.g., Diggle, 1997; Brumback and Rice, 1998; Ke and Wang, 2001; Ramsay and Silverman, 2002; Wand, 2003; Liu, Meiring and Wang, 2005; Zhang and Lin, 2003). Flexible spline-based models may prove valuable in both modeling and model diagnostics, in the spirit of BGST's comment on fitting "more complex models that test modeling assumptions."

Certain families of flexible models have earned popularity in epidemiological health effect studies, including generalized additive models (GAM; Hastie and Tibshirani, 1990; Dominici, McDermott, Zeger and Samet, 2002; Dominici, McDermott and Hastie, 2003, and references therein). The GAMs also may be viewed as special cases of functional analysis of variance (ANOVA) models. Functional ANOVA models extend analysis of variance concepts to include functional spaces, enabling the study of main effects and interactions between combinations of functional or discrete covariates. Suppose that a vector of explanatory variables $\mathbf{v} = (v_1, \ldots, v_M)$ has M components, each of which may be multidimensional. If n observations were made, the ith observation would correspond to

a vector of explanatory variables $\mathbf{v}_i = (v_{i1}, \ldots, v_{iM})$, for each $i \in \{1, \ldots, n\}$. A functional ANOVA model decomposes a modeled function

$$f(\mathbf{v}_i) = \nu + \sum_{k=1}^{M} f_k(v_{ik}) + \sum_{k<l} f_{kl}(v_{ik}, v_{il}) + \cdots + f_{1\cdots M}(v_{i1}, \ldots, v_{iM}),$$

where the first term ν is an overall intercept and the second term on the right-hand side is the sum of M flexible main effects, followed by additive functional interaction terms of different orders. Some components may be parametric (such as linear terms) and other components may be spline-based. Functional ANOVA models include smoothing spline (SS) ANOVA (e.g., Wahba et al., 1995; Gu, 2002; Wang, 1998; with smoothness controlled by multiple smoothing parameters) and the reduced basis approach of Stone, Hansen, Kooperberg and Truong (1997). The GAM corresponds to a model with only the intercept and main effects. Each estimation approach has its own challenges. I do not attempt a unified review here, but I note that GAM standard errors and the choice of smoothness recently have been improved for certain time series health impact studies (see Dominici, McDermott and Hastie, 2003, and references therein). Concerns have been raised about biased GAM standard error estimates for spatial air pollution data due to concurvity (Ramsay, Burnett and Krewski, 2003). The GAM constraints for model identifiability may need investigation for certain sampling schemes and covariate models.

These models also have been extended. Of particular relevance to BGST, Karcher and Wang (2001) developed generalized nonparametric mixed effect model extensions of GLMM that include SSANOVA-type components in the explanatory fixed and random effects. Such approaches hold promise for investigating model form and interactions in health impact studies, motivating continued research. Further understanding is needed on sensitivity to multiple smoothing parameter selection and the *curse of dimensionality* for multiple explanatory variables, together with computational improvements for large data sets. Rapid advances in the areas of GLMM and flexible models testify to their value to address particular questions, with further need of diagnostic and inference methods, including inference about interactions.

4. INTERACTIVE EXPOSURE PROCESSES

Many of the pollutants under study are produced or depleted in complex chemical reactions that rely on other pollutants and meteorological factors (see, e.g., Meng, Dabdub and Seinfeld, 1997; Seigneur, 2001). They also are measured subject to error at both the community and individual levels. As BGST notes, it is extremely difficult in observational studies to identify the effects of individual pollutants due to these chemical associations. The difficulties of studying health impacts of individual pollutants potentially also rise from interactive and compounded health impacts within the human body. For example, Brunekreef and Holgate (2002, referred to as BH in the remainder of this comment) reviewed many recent air quality health impact studies at a variety of time scales, together with state of knowledge physical mechanisms of health impact. BH discussed mechanisms by which ozone acts as a strong oxidant in the human body and antioxidants in the lung reduce the health impacts of inhaled ozone. Less is known about the damage mechanisms and long-term impact of nitrogen dioxide and particulate matter. However, BH also indicated that both particulate matter and nitrogen dioxide have the potential to *activate oxidant pathways* that impact on the respiratory and cardiovascular systems. While the mechanisms may be different and while pollutants may travel to different parts of the lung (with ultra-fine particles penetrating further into the lung than larger particles), I would value the authors' comments on the potential presence and form of interactive pollution health impacts. For example, if different pollutants damage different aspects of the respiratory system simultaneously, even at individual low to moderate levels, is there potential that the health impacts will be more severe and reported more frequently? Is there a combined point where the effects increase in severity nonlinearly? Might flexible spline-based models aid the investigation of possible combined health impacts, including possible nonlinearities in the severity of health impact and frequency of chronic disease diagnosis?

5. SPACE–TIME SCALES OF POLLUTION VARIATION

Ambient pollution levels and personal exposure activities vary on a variety of time scales, with long-term trends as well as seasonal and daily cycles. To study chronic effects, BGST usually use yearly averaged pollutant levels as the shortest time scale (although there is discussion of the need to examine other time scales in

the paper, and it is not clear to me currently if models at other times scales were implemented). Yearly pollutant averages eliminate information on smaller time scales that may be important in chronic air pollution studies. It will be important to investigate whether there may be increased cumulative risk for repeated exposure at high levels compared with constant exposure at medium levels. The amplitude and shape of the daily cycle may be a very important factor for certain pollutants. For example, surface ozone is a secondary pollutant produced in photochemical reactions. Daily patterns in primary pollutant emissions (anthropogenic and biogenic), photochemical activity and transport lead to a daily cycle in ozone levels. The shape of this daily cycle varies in space and time (see, e.g., Guttorp, Meiring and Sampson, 1994). In 1997, the United States Environmental Protection Agency changed the ozone regulation from a standard based on daily maximum hourly ozone values to one based on daily 8-hour average ozone levels. Have the authors tried to include counts of days exceeding some standard, and possibly also the duration of individual episodes (perhaps due to persistent meteorological conditions) in their statistical models, related to the exposure metrics discussed in Section 4.5? This may be a first step toward detailed consideration of multiple time scales to investigate whether the severity of the chronic impact increases nonlinearly with exposure on a variety of time scales. I would value the authors' additional comments on this.

Due to large variation in NO_2 levels within each community, BGST use a spatial mixed model combined with traffic estimates to estimate NO_2 levels at unmonitored locations. Spatial mixed models have also been extended to spatial GLMM (see, e.g., Diggle, Tawn and Moyeed, 1998; Christensen and Waagepetersen, 2002; Kammann and Wand, 2003), which may be of value for predicting the number of days exceeding pollution standards at an unmonitored location/microenvironment, given the corresponding number of days at monitoring sites. Space–time estimation methodology is constantly advancing and detailed reviews appear elsewhere. In particular, hierarchical space–time dynamic models are enabling the inclusion of science in the statistical model through a conditionally specified hierarchy (as reviewed by Wikle, 2003). Many of the pollutant fields are nonseparable and nonstationary in space and time (see, e.g., Guttorp, Meiring and Sampson, 1994; Sampson, Damian and Guttorp, 2001; Zidek, Sun, Le and Özkaynak, 2002). Stochastic simulation of potential air pollution fields, conditional on observations, may offer promise for assessing exposure uncertainty. This ideally would be combined with scientific knowledge about atmospheric chemistry, microenvironment variation and measurement error.

The error in using ambient measurements or microenvironment models versus true exposures varies substantially with pollutant. For example, exposures to some $PM_{2.5}$ constituents are poorly represented by ambient measurements, whereas some others correlate well with ambient outdoor measurements (e.g., Ebelt et al., 2000). Also, there is a *personal cloud* phenomenon related to activities that increase particle suspension levels, leading to higher personal exposures than microenvironment measurements. Health impacts of particulate matter currently receive high research priority (see, e.g., National Research Council, 1998; Lippmann et al., 2003) and new regulations are introduced. However, little is known about the impact of individual particulate matter constituents, which vary substantially in relative proportion within the United States and globally. Many analyses rely only on size distribution (for $d \in \{2.5, 10\}$, PM_d consists of particles less than d μm in aerodynamic diameter). Both elemental and organic carbon (EC and OC) constituents of particulate matter were used by BGST in Section 4.3. Please would the authors comment on the space–time variation in these (and other) particulate matter constituents and their correlations on a variety of space–time scales. Is there reason to believe that the $PM_{2.5}$ relative composition is similar across all the sites in the CHS on an annual basis and throughout the year? Do the authors have additional comments about individual particulate matter composition and impact?

6. REGULATION POLICY AND MODEL UNCERTAINTY

The development of regulations depends on the perceived health impact. Sensitivity of analysis conclusions to model choice has been found in mortality studies on shorter time scales (e.g., Smith, Davis and Speckman, 1999). Bayesian model averaging provides valuable information on statistical model uncertainty by combining inference from a family of statistical models (BGST; Clyde, 2000; Dominici, Sheppard and Clyde, 2003). This requires specification of the general family of candidate models, including the nature of any interactions. Model specification diagnostics remain crucial.

The valuable work by BGST and other health impact researchers leads toward the challenge of incorporating uncertainty about individual and combined pollution health impacts into emissions control deci-

sions. Improved statistical approaches are needed to contribute to decision making (Barnett and O'Hagan, 1997; Novartis Foundation, 1999). Atmospheric chemistry models are developed to investigate the changes in pollution levels under different emissions scenarios. These models must be evaluated under the current scenario by comparison with observations (Sampson and Guttorp, 1999) and ideally must be combined with uncertainty measures. Statisticians must continue to improve cost-effective sampling designs and probabilistic models for exposure assessment at the individual and community levels (see, e.g., Zidek and Le, 1999; Zidek et al., 2003). Continued study of the health impacts of multiple air pollutants on children will be a vital component of policy development to protect people of all ages.

Comment

Lianne Sheppard and Jonathan C. Wakefield

The article by Berhane, Gauderman, Stram and Thomas (BGST) addresses an array of statistical aspects that relates to estimation of the long-term effects of air pollution. The Southern California Children's Health Study (CHS) is an important resource for the community of scientists and policy makers who are trying to understand the long-term effects of air pollution on health. Many complex topics are discussed by BGST, but we focus our discussion here on just a small number of the issues considered. In particular, we discuss the role of complex hierarchical models in environmental health research, the role of exposure variation and measurement, and ecological inference.

1. HIERARCHICAL MODELS

Most air pollution studies attempt to address the broad questions of whether and how air pollution is associated with health outcomes. In the CHS, the ultimate goal is to understand the long-term effects of air pollution on children's health. Since this goal is broad, it must be refined and translated into contrasts (parameters or functions of parameters) that can be estimated from the available data. The approach described in BGST is the specification of a complex hierarchical model that is sufficiently general to allow for multiple levels of variation and types of confounding, and includes an array of parameters for the exposure effect $(\alpha_1, \alpha_2, \alpha_3, \beta_2, \beta_3)$ that allow the consideration of many possible scientific questions.

The hierarchical model described in (1)–(5) of the paper includes exposure effects of yearly exposure (α_1), individual exposure (α_2), community exposure (including contextual contributions) (α_3), individual exposure modified by time (β_2), and community exposure (including contextual contributions) modified by time (β_3). The exposure parameterization therefore allows for different effects for exposures that vary between cities (X_c), over time within cities $(X_{cj} - X_c)$ and across individuals within cities $(x_{ci} - X_c)$. This flexibility is a strength in the sense that one source of exposure variation does not influence estimation of the effect of another. For instance, the purely ecological comparisons can be separated from the effects of exposure variation across individuals or over time within cities. However, scientifically, it becomes imperative to question whether we expect these parameters to be the same. For instance, cross-sectional (α_2, α_3) versus longitudinal (α_1) effects are often believed to be different. A *contextual effect* of exposure is the additional modifying effect on the outcome of exposure through belonging to a group. Cross-sectional contextual effects are present in BGST's model when $\alpha_2 \neq \alpha_3$ (*and* when the associated predictors have the *same* definition; see Section 2). However, it is often reasonable to assume that contextual effects are absent for environmental exposures (Sheppard, 2002). Furthermore, even when the parameters are the same, their estimates can be different, because of uncontrolled confounding or covariate measurement error. Sheppard (2003) argued that biases operating at one level (or differently at several levels) can drive the differences in the es-

Lianne Sheppard and Jonathan C. Wakefield are Research Associate Professor and Professor, Department of Biostatistics, Box 357232, University of Washington, Seattle, WA 98195-7232, USA (e-mail: sheppard@u.washington.edu).

timates, even when the underlying parameters are the same (an example of this follows). This suggests that side-by-side comparisons of parameter estimates, such as the cross-sectional versus longitudinal effects given in Table 3, are a crucial step to identifying the potential biases in these studies. Note that the cross-sectional estimates in Table 3 are 2–3 times larger than the longitudinal estimates (although nearly all the estimates are within 2 standard errors of 0). It remains to be determined whether the differing estimates reflect real differences in the parameters or are merely different biases. Separation of the parameters is useful for comparing estimates and identifying different sources of bias (and this step should not be neglected). However, because the exposure variation is partitioned into several variables, reporting separate parameter estimates may not be a desirable approach to answering the scientific questions of whether there are, and how big are, the exposure effects.

We demonstrate how contextual effects may be induced by unmeasured confounding in a simple situation. Suppose the true model is given by

$$\mathrm{E}[Y_{ci}|X_{ci}, Z_{ci}] = \exp(\alpha + \beta X_{ci} + \gamma Z_{ci}),$$

where within areas

$$\begin{bmatrix} X_{ci} \\ Z_{ci} \end{bmatrix} \sim N\left(\begin{bmatrix} X_c \\ Z_c \end{bmatrix}, \quad \begin{bmatrix} W_x^2 & \rho_W W_x W_z \\ \rho_W W_x W_z & W_z^2 \end{bmatrix} \right),$$

and between areas

$$\begin{bmatrix} X_c \\ Z_c \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_z \end{bmatrix}, \quad \begin{bmatrix} B_x^2 & \rho_B B_x B_z \\ \rho_B B_x B_z & B_z^2 \end{bmatrix} \right).$$

If the confounders Z_{ci} and Z_c are unmeasured we obtain

$$\mathrm{E}[Y_{ci}|X_{ci}, X_c] = \exp\left(\alpha^\star + X_{ci}\left\{ \beta + \frac{\gamma \rho_W W_z}{W_x} \right\} + X_c \left\{ \frac{\rho_B B_z}{B_x} - \frac{\rho_W W_z}{W_x} \right\} \gamma \right),$$

where $\alpha^\star$ does not depend on X_{ci} or X_c. Hence a contextual effect has been *induced* by unmeasured confounding. The effect of individual exposure X_{ci} is only confounded in the presence of within-community dependence ($\rho_W \neq 0$), and confounding does not depend on the between-community dependence between the average exposure and confounder, ρ_B, because a contextual effect of exposure (X_c) has been included in the model. If this term is excluded from the model, then there will be confounding due to ρ_B also. Hence to prevent confounding from this source, a contextual effect should be included in the above model. A contextual effect is induced when there is either within- or between-community dependence ($\rho_W \neq 0$ or $\rho_B \neq 0$), unless they cancel out, which is extremely unlikely. The size of the induced contextual effect is reduced when either of the ratios of between- or within-community variability in confounder to exposure is small. This indicates that the interpretation of contextual effects requires great care. The above development is analogous to the role of unmeasured confounding in longitudinal studies (see Palta and Yao, 1991).

The underlying scientific questions must be related to the proposed contrasts. It is a dilemma for statisticians to decide whether to present a simple model that highlights a specific contrast of interest or to use a very complex model within which the same contrast is embedded. Presentation of the simple model is often much easier, but risks hidden bias. Complex hierarchical models are difficult to convey to a broad audience and are less straightforward to interpret since estimates are only rarely direct functions of the data. However, specification of the complex hierarchical model can inform the simple contrasts by giving a structure for deciding which sensitivity studies should be done and how to approach them. Simpler models can be derived from the complex multilevel model by collapsing over one or more levels or by dropping one or more sources of variation. In BGST, very interesting unanswered questions remain regarding whether either of the simpler cross-sectional or longitudinal change models shown in Table 3 is adequate relative to the full multilevel model. The simpler models, while not explicitly shown in this paper, are easier to understand. The coefficients suggest the simpler models perform similarly for all pollutants except O_3. In general, the combination of data and comparison of results from different analysis approaches and study designs, including cohort, case-control, ecological and longitudinal studies, is a vital area of research since the strengths of each can be exploited and the potential for hidden biases can be decreased.

We now discuss parameter interpretation with reference to a simplified form of the model given by (1)–(5) of the paper, in which we ignore the time effects. (We would be interested in the authors' interpretation of β_2 and β_3.) For clarity of exposition we take a log link, a common choice for rare events. Hence we have the model

$$\mu_{ci} = \exp\{\alpha_0 + \alpha_3 X_c + e_c + \eta_2 z_{ci} + \alpha_2(x_{ci} - X_c) + e_{ci}\}.$$

Suppose all of the exposures in area c are increased by 1 unit. Then the relative risk associated with this change for each of the individuals in community c, assuming that individual confounders and random effects remain constant, is given by $\exp(\alpha_3)$. The relative risk between two individuals, i and i' say, in the same community whose exposures differ by 1 unit but have the same confounders, is $\exp(\alpha_2 + e_{ci} - e_{ci'})$. Finally consider individuals i and i' in two communities c and c', and suppose they have the same individual exposures and confounders, but the average exposure in community c is 1 unit higher than in community c'. Then the relative risk is $\exp(\alpha_3 - \alpha_2 + e_{ci} - e_{ci'} + e_c - e_{c'})$. The presence of random effects makes interpretation of these relative risks more difficult, as discussed by Heagerty and Zeger (2000).

The form of the random effects model may be secondary to other considerations when effect estimation is the goal of the analysis. Wakefield (2003) argued that in the ecological setting, spatial effects will frequently be of secondary importance compared to issues of confounding and pure specification bias (see Section 3). Guthrie, Sheppard and Wakefield (2002) showed that the efficiency of an aggregate data model was unaffected by ignoring the spatial dependence, unless the exposures varied on the same scale as the random effects. When the exposures varied quickly in comparison to the residual variation over space, the standard error of the exposure effect estimates was similar for models that incorporated spatial dependence as with models that ignored its presence.

To the discussion of software for hierarchical models in Section 3.4, we mention the WinBUGS software (Spiegelhalter, Thomas and Best, 1998). It is straightforward to fit complex hierarchical models within a Bayesian framework using WinBUGS.

2. EXPOSURE MEASUREMENT

The relative variation of exposure at each level of the hierarchy has implications for the relative information about and interpretability of each of the exposure effect parameters in the model. When there is limited exposure variation, the parameters are poorly estimated. The standard error estimates given in Table 3 suggest the between-community variability is lower than the average temporal variability within community, even though the design of CHS would suggest the opposite. Reporting numerical summaries of the key predictors, particularly exposure, at each level of analysis should be a required part of a hierarchical analysis that focuses on interpretation of exposure effects.

The exposure definition can be potentially problematic too. BGST mention that two of the exposure predictors, X_c and X_{cj}, are community levels of pollution, implying that these are ambient *concentration* measures. However, distinguished by its notation, x_{ci} is described as a subject-specific average pollution level from microenvironmental models, suggesting that this term is a measure of individual *exposure*. Since such ambient exposures are often attenuated relative to ambient concentrations due to the amount of time people spend indoors (Ott, Wallace and Mage, 2000; Wilson, Mage and Grant, 2000; Sheppard and Damian, 2000), this difference makes the parameter α_2 inherently different from α_1 or α_3. Thus the comparability between the α's and their interpretation will be lost when the comparability between exposure predictors is not maintained.

There are additional questions regarding exposure that need to be incorporated into the modeling framework. How do we know which exposure metric to choose in these models? What implications does this choice have on the parameters in the model? Furthermore, measurement error distributions can vary by level of analysis and exposure measurement properties, with a consequent impact on the parameter estimates. For instance, while temporal variation within a city may be reasonably assessed by a single fixed-site monitor, the city-specific mean from that same fixed-site monitor may still be subject to measurement error.

3. ECOLOGICAL INFERENCE

Although the CHS provides individual-level data, since BGST devote a section to ecological bias we feel it is useful to examine the effects of aggregation in a purely ecological setting. In our own research in this setting we have found it beneficial to begin with an individual-level model and then average to determine the effects of aggregation. For simplicity, consider a single exposure x and a single confounder z, and suppose that for individual i in community c the risk is given by the log-linear form

$$p_{ci} = \exp(\alpha + \beta x_{ci} + \gamma z_{ci}),$$

which, upon aggregation, yields average risk

$$p_c = \exp(\alpha) \int \exp(\beta x + \gamma z) f_c(x, z)\, dx\, dz,$$

where $f_c(x, z)$ represent the joint distribution of x and z within area c. Unfortunately only marginal in-

formation is typically available and the most difficult aspect of ecological analysis is to control for within-community confounding (between-community confounding is analogous to individual-level confounding in a study at the level of the individual). Wakefield (2004) highlighted two sources of ecological bias: that due to the noncollapsibility of nonlinear association measures, and that due to confounding. The presence of the former illustrates that even in the absence of confounding, bias may occur when a nonlinear individual risk model is distorted upon aggregation. Greenland (1992) referred to this as *pure specification bias*. Note that if there is no variability in exposure within an area, then this bias will be absent, which provides one motivation for utilizing small area data. As an example, Wakefield (2003) considered the case in which the within-community variability is given by $N(X_c, \sigma_c^2)$, where $\sigma_c^2 = a + bX_c$. In this scenario we obtain the aggregate model

$$p_c = \exp(\alpha + \beta^2 a/2 + X_c[\beta + \beta^2 b/2]),$$

showing that the variance is acting like an unmeasured confounder (so that if the variance is independent of the mean, no bias will result). For environmental exposures the variance typically increases with the mean ($b > 0$), so for a harmful exposure ($\beta > 0$) the effect will be overestimated (in the absence of other biases). Contextual effects are an example of confounding in which Z_c corresponds to the average exposure.

There is a long history of ecological analysis in the social sciences; this literature was reviewed by Wakefield (2004), while Salway and Wakefield (2004) compared and contrasted the aims and models of ecological inference in epidemiology and the social sciences. In a highly influential paper, Robinson (1950) highlighted the inconsistency of summary association measures across different levels of aggregation (noncollapsibility) and, by example, illustrated that the correlation between literacy and race ranged between 0.95 and 0.20 across different geographical units. Selvin (1958) later coined the term *ecological fallacy* for the situation in which, "relationships between characteristics of individuals are wrongly inferred from data about groups." Much of the discussion in the social sciences literature concerns the difficulties associated with simultaneous estimation of individual and contextual exposure effects. This difficulty is simply illustrated by consideration of what is known in the social sciences literature as extended ecological regression (Goodman, 1959; Achen and Shively, 1995). In an epidemiological context, consider a single binary exposure and the individual risk model

$$p_{0c} = a_0 + b_0 x_c$$

for unexposed individuals in a community with exposed proportion x_c and

$$p_{1c} = a_1 + b_1 x_c$$

for exposed individuals, which leads to the aggregate form

$$p_c = p_{0c}(1 - x_i) + p_{1c} = \alpha + \beta x_c + \gamma x_c^2,$$

where $\alpha = a_0$, $\beta = a_1 + b_0 - a_0$ and $\gamma = b_1 - b_0$, so that the effects of interest are nonidentifiable, even under the simplified model in which a common contextual effect for unexposed and exposed individuals, $b_0 = b_1$, is assumed.

In an epidemiological context, a linear risk model is less plausible and a log-linear form is more typically used. Contextual effects are of great interest in social epidemiology. For example, suppose we begin with the individual-level model

$$Y_{ci} | x_{ci}, x_c \sim \text{Bern}\{\exp(\alpha + \beta x_{ci} + \delta x_c)\},$$

so that the contextual effect, δ, is the same for exposed and unexposed individuals. We then obtain the aggregate-level model

$$p_c = \exp(\alpha + \delta x_c)\{(1 - x_c) + x_c \exp(\beta)\}.$$

Simultaneous estimation of the individual and contextual effects is, therefore, theoretically possible, but hinges on a nonlinearity, which is uncheckable from the aggregate data and so would not be recommended. Similar arguments (e.g., Little, 1985; Copas and Li, 1997) have been made against a class of methods for avoiding selection bias, for example, those proposed by Heckman (1979). Similarly, estimation of both effects is possible with the aggregate data approach of Prentice and Sheppard (1995), but the amount of information available on the contextual effect (in the presence of individual effects) is small (Sheppard, 2002).

As we hope is obvious from this discussion, we found the paper very stimulating and we would like to encourage the authors in their pursuit of the difficult yet vital endeavor of investigating the complex relationship between air pollution and health.

Comment

Nhu D. Le and James V. Zidek

We congratulate the authors on a stimulating and comprehensive article on spatial epidemiology and environmental health risk. They provide a broad survey of the methods that have been developed in this area as well as the issues that arise in the analysis of that risk. By couching their survey in terms of an ongoing investigation on the long-term effects of air pollution, they endow their presentation with a sense of timeliness and importance. Furthermore, their survey possesses a liveliness that a mere review might well lack.

The importance of the study on which this paper focuses cannot be overestimated, for the potential impact of chronic diseases greatly outweighs that of its acute cousins. Indeed, cynics have suggested that the treatment of acute morbidity would be a cheaper and preferable option to the costly abatement programs that would be needed to reduce pollution to the levels needed to eliminate it! We find it hard to imagine that the same could be said of chronic morbidity if, in fact, that is a product of excessively high levels of pollution. However, that brings us to the second important aspect of studies like the one described in this paper.

Such studies are extremely difficult and expensive to carry out, both in measuring health outcomes and exposure to $PM_{2.5}$. Thus, both responses and predictors are susceptible to high levels of error, making detection of association with chronic disease effects difficult. The study in this paper is all the more remarkable in that it is prospective, rather than retrospective, meaning that subjects are subject to long-term follow-up and a large response burden.

For comparison, we briefly describe another such study, centered in the Northwest Center for Particulate Matter and Health Effects at the University of Washington (UW), that seeks to make a comprehensive exposure and health effect assessment in susceptible subpopulations. Like the study addressed in this paper, that at the UW integrates personal exposure assessment, exposure characterization and the study of health effects. Three subprojects are underway:

1. The health of three susceptible and one healthy subpopulation in Seattle, Spokane and other cities is being monitored. The susceptible individuals are 65 or older and have chronic obstructive pulmonary disease, cardiovascular disease or asthmatic children. The health endpoints include such things as pulse rate and blood pressure in the case of the adults and symptoms of asthma in the case of the children.
2. In this subproject personal exposure measurements are being collected for the subjects to determine the contribution of ambient sources to personal PM exposures. Models are being developed to predict such exposures for nonmonitored subjects. Finally, the association between these measurements/predictions and acute health outcomes will be determined.
3. The third subproject characterizes chemical and physical parameters of different sizes of ambient and indoor aerosols in simulated airway conditions.

The UW study, in particular, is seeking to identify specific components of PM such as chemical constituents that cause ill health. Of particular interest are the high exposures to PM and associated products of combustion in the Northwestern United States and whether these are associated with acute cardiorespiratory physiologic health measures.

These two studies can be contrasted with that of Le, Mao, Sun and Zidek (2004), which is a retrospective analysis of the cancer effects of air pollution. In this case-control study, cancer patients and "healthy" (noncancer) individuals were identified through population-based provincial registries. The participants' residential histories, along with information on important confounding factors such as smoking, diet and occupational histories, were collected using the self-administered questionnaire approach. The key feature of this study is the estimation of lifetime exposure to air pollutants which is obtained through the use of residential histories in conjunction with historical air pollution measurements from fixed monitoring stations; some have been in operation for over 20 years. Concentration levels for pollutants at residential locations are obtained through a Bayesian spatial interpolation method which does not assume

Nhu D. Le is Senior Scientist, British Columbia Cancer Research Centre, Vancouver, BC, Canada V5Z 4E6. James V. Zidek is Professor, Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2 (e-mail: jim@stat.ubc.ca).

stationarity of the environmental fields. Cumulative exposure levels for individual pariticipants are then obtained by aggregating over the predictions at their residential locations. Unlike cohort studies, this type of study can only examine impact for one specific adverse health outcome. However, for chronic and rare diseases with long latency such as cancer, it can be very cost-effective.

It should be added that in studies where the cumulative exposure levels for individual participants are available (e.g., cohort studies considered in this paper and the type of case-control studies described above), investigators have the flexibility to adjust for potential latency or for the induction period of the chronic disease under examination. This flexibility may not be available for ecological studies due to the mobility of the population.

The authors focus on the methodological problems and the substantive issues that arise in such studies, and they provide a remarkably comprehensive list. However, although they describe their study design, they do not explicitly discuss its merits in comparison to others. In contrast, "design" does rate a separate section in Dominici, Sheppard and Clyde (2003), who considered strengths and weaknesses of various types of studies: (1) ecological time series; (2) case crossover; (3) panel (repeated measures); (4) cohort studies (time to event), which in this paper would be called a panel study. In particular, they note that chronic effects and acute effects may not be separable in a cohort study. They conclude that "[u]ltimately, the choice of an optimal design depends upon the research question and the availability of data."

Dominici, Sheppard and Clyde (2003) also considered methodological problems, although not with respect to a specific study like this paper. Their paper, unlike this one, takes a hierarchical Bayesian modeling approach that comes equipped with an inferential base. These authors face some conceptual difficulties because of their seeming reluctance to adopt a more subjectivist position. They note that their communities are not a random sample. So what is to be made of their results? More precisely, "... what is then the interpretation of the p-values or confidence limits arising from the analysis at the community level." Their answer is that "... air pollution levels would have somehow been assigned at random to a sample of communities, which may or may not have been selected at random." Hmmm .... No doubt the reader will be left with some uncertainty about how to assess the validity of that claim.

The (by now fairly standard) multilevel modeling paradigm is imaginatively invoked here and yields an analysis that allows time-dependent covariates, on the one hand, while enabling a wide range of comparisons at both individual and aggregate levels.

Rejoinder

Kiros Berhane, W. James Gauderman, Daniel O. Stram and Duncan C. Thomas

We thank all the discussants for their kind words and their many insightful comments on the issues raised by the paper. We truly believe that their discussions, without exception, have immensely enriched our paper.

First, we feel obliged to explain the reasons for our exclusive focus on cohort studies that examine chronic effects of air pollution, as opposed to the many other study designs that could potentially examine acute and/or subacute effects of air pollution. The reasons were twofold: (1) the CHS (and hence most of our related methodologic work) was mainly intended to handle chronic effects of air pollution and (2) the discussion easily could have gotten out of hand if we tried to accommodate both acute and chronic effects. That said, methods for acute effects (as ably reviewed by Dominici, Sheppard and Clyde, 2003) are of interest to us. In fact, one of our major substudies, the Air Pollution and Absenteeism Study (APAS), deals with the effects of day-to-day variation in air pollution on school absenteeism. Upon analysis of daily time series data from APAS, we showed that a 20-ppb increase in daily 10 AM–6 PM levels of ozone is associated with an 83% increase in illness related absenteeism (Gilliland et al., 2001). Technical details on a two-stage model for daily time series of counts that is based on the polynomial distributed-lag approach were reported by Berhane and Thomas (2002) and a three-level model for binary time series was reported by Rondeau, Berhane and Thomas (2004). We also believe that a systematic review of methods that examine the interrelationships between acute and chronic effects of air pollution is timely. Such a review could tie

together the issues that have been raised by BGST and Dominici, Sheppard and Clyde (2003).

Meiring highlighted our point on the need for more research on diagnostics and inference in the longitudinal setting and we thank her for her elaboration on the current state of methodologic work in this area. Our approach in the CHS for examining the adequacy of the normality assumption about the various random effects, for both the Gaussian and non-Gaussian outcome data, has been to fit preliminary models separately by levels and then examine the residuals from each of the levels. In our experience so far, this assumption was not found to be violated. For the lung function models, a log transformation was required to ensure the normality of the overall residual error term. We also note that in situations where the random effects themselves are of interest, one needs to check the adequacy of the normality assumption. Currently available methods for checking the adequacy of the normality assumption include the graphical approach to testing the adequacy of normality (Lange and Ryan, 1989) and nonparametric approaches to estimating the distribution of the random effect (Davidian and Gallant, 1993). The former is limited to continuous outcomes and the latter could be too complicated and computationally intensive for our applications. We are currently exploring a class of simple, yet intuitive and quite general, approaches [based on the Box–Cox transformation technique (Box and Cox, 1964)] to assessing the adequacy of the normality assumption of the random effects and potential solutions for any detected nonnormality. Development of methods to handle diagnostics for mixed effects models is an area that could benefit from more research. Examining the effects of outliers and influential observations in the longitudinal setting becomes challenging because outliers and/or influential observations could appear at the observation and/or the subject level.

We agree with Meiring that flexible models (e.g., generalized additive models) could also be used to assess the adequacy of functional forms of covariate effects. This was the approach we used to flexibly depict the nonlinear growth trajectory of lung function measures in children. Based on this exploratory finding, our recent focus has been on developing functional based multilevel models for our data on lung function growth patterns. This has allowed us to examine the effects of air pollution on biologically important aspects (e.g., maximum rate of growth) of the lung function trajectories. We welcome the comprehensive review of related methodologic work in this area that Meiring provided, including the more general functional ANOVA models. To this, we add the work by Guo (2002) that developed a general GLMM model that allows for nonparametric modeling at both the fixed and the random parts of the GLMM. We share Meiring's enthusiasm regarding the usefulness of GLMM and flexible models, and the continued need for more research in dealing with problematic areas, such as the well documented *curse of dimensionality* in models with multiple flexible terms (Hastie and Tibshirani, 1990).

Meiring asks for some elaboration on our choice of design matrices and whether interaction effects were included. This question can be addressed separately at each level of the model. At the temporal level, time-specific covariates z_{cik} included such factors as the presence of an acute illness at the time of testing and indicator variables for technician and spirometer. At the subject level, z_{ci} included time-fixed covariates such as race and baseline asthma status. Some interaction effects were included, such as sex by race and sex by asthma at the community level. Most analyses included only a single pollutant at a time in X_c, with no adjustment for ecologic confounders except in an exploratory mode, as discussed in our paper. With only 12 communities at this level of comparison, we generally had inadequate degrees of freedom to fit multipollutant models. While we can fit the models including two-way interactions, we just do not have enough information to be able to parse out effects, let alone test for interactions between pollutants. However, it is also possible to test pollution effects at the temporal and individual levels, which we did by including year-to-year deviations $(X_{ck} - X_{c\cdot})$ of the ambient concentrations from the long-term average in the first level, and deviations $(x_{ci} - X_{c\cdot})$ of person-specific exposures (based on spatial or microenvironmental modeling) from the ambient level in the subject-level model. This can be helpful for unscrambling multipollutant effects because the correlations between pollutants can be quite different within and between communities, and across time as shown in our study of air pollution and asthma exacerbation (McConnell et al., 2003). In some cases we also tested for interactions between exposures at different levels. For example, in our analyses of school absences, we found that daily O_3 levels had a bigger effect in communities with low than high PM exposure (Gilliland et al., 2001; Berhane and Thomas, 2002).

This question is related to Meiring's later discussion of biological processes, multipollutant models, space–time correlations and time scales. The oxidative stress hypothesis she discusses indeed underlies our

thinking about mechanistic hypotheses, as discussed by Gilliland et al. (1999). This indeed could guide the selection of models for investigating multipollutant effects, such as providing a rationale for the observation noted above about the apparent interaction between O_3 and PM on school absences. However, we feel that the most promising line of research involves characterization of the specific genes involved in modulating these processes (Gilliland et al., 2002b, c, 2003), as well as dietary factors such as antioxidant intake, which could modify host susceptibility (Gilliland, Li and Peters, 2001; Gilliland et al., 2002a). Ultimately, we hope to build comprehensive models for the entire oxidative stress pathway, incorporating environmental, genetic and host factors, perhaps using some of the techniques discussed by Conti et al. (2003). Other promising approaches to the multipollutant problem include using source apportionment methods and exploring alternative temporal metrics.

In the source apportionment approach (Schauer et al., 1996), the chemical species in PM are used to estimate the proportions of pollutants that are derived from such sources as automobile and diesel emissions, wood burning and tire wear. We plan to extend the Bayes model averaging approach described in our paper to include such data as "prior covariates" in a hierarchical model, thereby allowing improved estimates of specific pollutant effects by borrowing strength from other pollutants derived from similar sources and by providing estimates of the overall health effects of the source contributions themselves, information that would be particularly useful for regulation.

We are exploring alternative exposure metrics based on different temporal patterns of exposure. Most of the results given in BGST are based on the long-term average concentration, under the hypothesis that chronic effects represent the cumulative burden of incremental exposure effects that are linear and additive. However, it is certainly plausible (as suggested by Meiring) that there are threshold, saturation or interactive effects that act at different time scales and could violate this assumption, implying that exposure metrics that allow for these nonlinearities might predict chronic effects better. Preliminary exploration of such effects for 4-year changes in MMEF in the two fourth-grade cohorts showed no evidence of nonlinear effects for the cluster of highly correlated NO_2/PM/acid pollutants, but suggested a possible nonlinear effect for O_3. While the long-term average exposure to O_3 was not significantly associated with any lung function measurement, the variance in O_3 levels between hours (within days) and between days (within weeks), and to a lesser extent between weeks (within seasons), was significantly associated with slower lung function growth (unpublished data), suggesting a possible threshold effect. We are continuing to explore this phenomenon by studying a flexible spline-based class of exposure indices that involve the percentage of time above thresholds, which may be useful for regulatory purposes.

We strongly concur with Meiring's comments on the regulatory policy implications of our work and, in particular, the importance of accounting for exposure and model uncertainties. Künzli et al. (2003) elaborated some of these implications, in particular, the trade-offs between primary (emission) and secondary (personal exposure) interventions. Further research on propagation of uncertainties through health effects analyses and risk assessment is needed, as well as vigorous public debate about the appropriate interpretation of uncertainty in risk estimates to establish regulatory policy without leading to paralysis.

On the issue of parameter interpretation, Sheppard and Wakefield accurately point out that understanding pollutant effects from a multilevel model can be difficult in the context of a log link (or other nonidentity link functions). They demonstrate this issue in the context of a model for the main effect of pollution on a binary outcome at the individual and community levels [as parameterized by α_2 and α_3, respectively, in (1)–(5) of BGST]. Although these equations also parameterize an effect of time on the outcome (through the β parameters), we do not envision these terms being used in a model for a binary outcome. Rather, we have applied the full model described in (1)–(5) to analyze continuous outcomes in the CHS, most notably lung function. In this case, we have adopted the identity link. The parameters β_2 and β_3 then quantify the effect of pollution at the individual and community levels, respectively, on average lung function. The parameters α_2 and α_3 quantify the corresponding effects of pollution on change in lung function over some time interval. Under the identity link, all random effects in (1)–(5) drop out (as they have expectation zero) in any comparison of exposure effects on expected outcome. All other things being equal, the expected effect of one unit change in pollution will have similar interpretations at either the individual or community level.

We should point out one subtlety in the interpretation of the α's from the model in (1)–(5) of BGST. As in any model, the intercept quantifies some expectation of the outcome at "baseline," that is, when all other terms in the model drop out. For the model in (1), a_{ci} denotes

the individual specific intercept when all covariates (z_{cij} and $X_{cj} - X_c$) and the time variable (t_{cij}) equal zero. Whereas the a_{ci} are then treated as random effects to be modeled in the second and third level models, understanding what is being quantified by a_{ci} in the first model is important for interpreting pollutant effect estimates. As an example, consider a simplification of the model in (1) of the form $\mu_{cij} = a_{ci} + b_{ci}(t_{cij} - t^*)$. The value t^* can be chosen to estimate pollutant effects on lung function level at any point along the observed age range. For example, if t^* is chosen to be the average age at study entry, the α parameters will approximate the effects that would have been estimated from a cross-sectional analysis of baseline data only. This choice of t^* was used for the "full model" (Table 3) to facilitate comparisons of the pollutant effects on intercepts from this model with those from the intercepts from the pure cross-sectional analysis. However, one may alternatively choose t^* to be the average age over some study period to quantify the overall average effect of pollution on level or the average age at the end of the study period to quantify the effect of air pollution on attained level.

We welcome the comments of Sheppard and Wakefield about ecologic inference and exposure measurement. In particular, we agree that associations with the community ambient concentrations X_c and with personal microenvironmental exposures x_{ci} estimate conceptually different parameters. Our use of the same symbol for both quantities was perhaps confusing and was adopted only to avoid proliferation of notation. Conceptually, it might be simpler to think of X_c as the community mean of personal exposures, although in practice we do not have adequate measures of the latter for most pollutants. Setting this issue aside, we note that some of the complexities of contextual effects described arise from Sheppard and Wakefield's use of a log-linear model, whereas our discussion of ecologic bias was focused on linear models. While certainly convenient for analysis of event data, many of our analyses concern continuous normally distributed traits such as lung function, for which we use an identity link. In this case, the dependence of the community mean outcome on the variance of exposure induced by a log-linear model at the individual level disappears, as does the "contextual" effect of X_c in a model that includes X_{ci} (but not Z_{ci}) whenever the between- and within-community correlations in X and Z differ. This is also true for the comparison between individuals in communities with different ambient exposures but the same personal exposure, where the expected values of their random effects cancel out in a linear but not in a log-linear model. Nevertheless, we agree that it is generally helpful to test for a contextual effect. We believe that our multilevel modeling approach, in which the effects of individual exposures and confounders are assessed by deviations from the community means, and community mean exposures are also included, accomplishes this.

Le and Zidek provide some interesting comparisons with their own Bayesian spatial modeling of pollution levels. We are exploring similar approaches using our data on NO_2 and O_3 levels at sampled residences and traffic density measures available on all homes. Our approach relies on Bayesian spatial modeling, in which the (log) true pollution levels are assumed to be normally distributed with means given by a regression on traffic density and a spatial covariance within communities (note that, like Le and Zidek, this does not assume stationarity of the environmental field). Measured values are assumed to be distributed independently around these true values and the health effects are regressed on the true exposures. We have implemented this model using the WinBUGS software, thereby providing an estimate of the relationship between health outcomes and exposure at the individual level in the entire cohort, combining the actual measurements on the subset and the predictions on everybody. For MMEF and NO_2, we find a marginally significant negative association using this approach, whereas the small sample of actual measurements alone is inadequate to demonstrate such an effect (Molitor et al., private communication). We are currently attempting to extend this approach to joint modeling of NO_2 and O_3, allowing for the predicted negative correlation between the two due to scavenging.

The design of the CHS was aimed purely at estimating chronic effects of air pollution, unlike those reviewed by Dominici, Sheppard and Clyde (2003). A major outstanding question is whether acute and chronic effects studies estimate the same quantity (Künzli et al., 2001; Rabl, 2003) and, if not, whether both can be derived from an appropriate multilevel study design that incorporates temporal and spatial comparisons.

Le and Zidek raise interesting points regarding our choice of communities and its implications on interpreting results from the study. The evaluation of long-term effects of air pollution exposure has relied to a very large extent (and not just in our study) on the comparisons of health outcomes in communities with different levels of air pollution. It is often asked whether

the communities studied constitute a random sample of communities from a larger population and, if not, then how are p-values to be interpreted? In the main paper we rationalize our statistical inferences not by assuming that the communities are a random sample of a population of communities, but rather by assuming that air pollution levels are independent of other unmeasured confounders that influence disease risk so that air pollution may be regarded as having been applied to the communities at random. For related discussions on the general topic of randomization and causal inference, refer to Greenland (1990) and references therein.

We make three brief further comments about this view. First, we recognize that the independence of community aggregate air pollution levels from other unmeasured variables that affect a child's disease risk is not a testable assumption in our strictly observational setting, and so it is particularly important that efforts to identify, measure and adjust for other variables be continued and improved upon in this and in future studies. Second, taken literally, this view of our study design leads naturally to permutation-based testing of the significance of regression estimates at the between-community level of analysis, by developing a permutation distribution of the regression estimate under the null hypothesis of no influence of air pollution on risk. However, it seems unlikely to us that important differences would arise between the results of such permutation-based tests and those based on using regression-based inference that allows for random effects for community. Third, a community level view of our analyses is required because we are certain that other unmeasured risk factors that cluster by community exist, so the children in our study cannot be assumed to have been sampled at random with respect to their sensitivity to the effects of air pollution. Ultimately, however, it is individual children rather than communities that make up the fundamental population at which that inference is aimed.

In conclusion, we are encouraged by this discussion that our aim to stimulate further methodologic research in environmental epidemiology has been fruitful and we hope to see further developments in this field.

ADDITIONAL REFERENCES

ACHEN, C. H. and SHIVELY, W. P. (1995). *Cross-Level Inference*. Univ. Chicago Press.

BARNETT, V. and O'HAGAN, A., eds. (1997). *Setting Environmental Standards*: *The Statistical Approach to Handling Uncertainty and Variation*. Chapman and Hall, London.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–252.

BRUMBACK, B. A. and RICE, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *J. Amer. Statist. Assoc.* **93** 961–994.

BRUNEKREEF, B. and HOLGATE, S. T. (2002). Air pollution and health. *Lancet* **360** 1233–1242.

CHRISTENSEN, O. F. and WAAGEPETERSEN, R. (2002). Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics* **58** 280–286.

CLYDE, M. (2000). Model uncertainty and health effect studies for particulate matter. *Environmetrics* **11** 745–763.

CONTI, D. V., CORTESSIS, V., MOLITOR, J. and THOMAS, D. C. (2003). Bayesian modeling of complex metabolic pathways. *Human Heredity* **56** 83–93.

COPAS, J. B. and LI, H. G. (1997). Inference for non-random samples (with discussion). *J. Roy. Statist. Soc. Ser. B* **59** 55–95.

DAVIDIAN, M. and GALLANT, A. R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika* **80** 475–488.

DIGGLE, P. J. (1997). Spatial and longitudinal data analysis: Two histories with a common future? *Modelling Longitudinal and Spatially Correlated Data*: *Methods, Applications and Future Directions*. *Lecture Notes in Statist.* **122** 387–402. Springer, Berlin.

DIGGLE, P. J., TAWN, J. A. and MOYEED, R. A. (1998). Model-based geostatistics (with discussion). *Appl. Statist.* **47** 299–350.

DOMINICI, F., MCDERMOTT, A. and HASTIE, T. (2003). Improved semi-parametric time series models of air pollution and mortality. Available at http://www-stat.stanford.edu/~hastie/pub.htm.

DOMINICI, F., MCDERMOTT, A., ZEGER, S. L. and SAMET, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American J. Epidemiology* **156** 193–203.

DOMINICI, F., SHEPPARD, L. and CLYDE, M. (2003). Health effects of air pollution: A statistical review. *Internat. Statist. Rev.* **71** 243–276.

EBELT, S. T., PETKAU, A. J., VEDAL, S., FISHER, T. V. and BRAUER, M. (2000). Exposure of chronic obstructive pulmonary disease patients to particulate matter: Relationships between personal and ambient air concentrations. *J. Air and Waste Management Association* **50** 1081–1094.

GILLILAND, F. D., BERHANE, K. T., LI, Y.-F. and KIM, D. H. (2002a). Dietary magnesium, potassium, sodium and children's lung function. *American J. Epidemiology* **155** 125–131.

GILLILAND, F. D., BERHANE, K., RAPPAPORT, E., THOMAS, D. C., AVOL, E., GAUDERMAN, W. J., LONDON, S. J., MARGOLIS, H. G., MCCONNELL, R., ISLAM, K. T. and PETERS, J. M. (2001). The effects of ambient air pollution on school absenteeism due to respiratory illnesses. *Epidemiology* **12** 43–54.

GILLILAND, F. D., GAUDERMAN, W., VORA, H., RAPPAPORT, E. and DUBEAU, L. (2002b). Effects of glutathione-S-transferase M1, T1, and P1 on childhood lung function growth. *American J. Respiratory and Critical Care Medicine* **166** 710–716.

GILLILAND, F. D., LI, Y.-F. and PETERS, J. M. (2001). Effects of low vitamins A, E, and C intake on children's lung function. *American J. Respiratory and Critical Care Medicine* **163** A174.

GILLILAND, F. D., LI, Y.-F., TSAI, W., DUBEAU, L., AVOL, E. and PETERS, J. M. (2003). TNF-308 genotype and ozone effects on asthma and wheezing: Results from the Children's Health Study (CHS). *American J. Respiratory and Critical Care Medicine* **167** A831.

GILLILAND, F. D., MCCONNELL, R., PETERS, J. and GONG, H., JR. (1999). A theoretical basis for investigating ambient air pollution and children's respiratory health. *Environmental Health Perspectives* **107**(suppl.) 403–407.

GILLILAND, F. D., RAPPAPORT, E., BERHANE, K., ISLAM, K. T., DUBEAU, L., GAUDERMAN, W. J. and MCCONNELL, R. (2002c). Effects of glutathione-S-transferase P1, M1, and T1 on acute respiratory illness in school children. *American J. Respiratory and Critical Care Medicine* **166** 346–351.

GOODMAN, L. (1959). Some alternatives to ecological correlation. *American J. Sociology* **64** 610–625.

GREENLAND, S. (1990). Randomization, statistics and causal inference. *Epidemiology* **1** 421–429.

GREENLAND, S. (1992). Divergent biases in ecologic and individual-level studies. *Statistics in Medicine* **11** 1209–1223.

GU, C. (2002). *Smoothing Spline ANOVA Models.* Springer, Berlin.

GUO, W. (2002). Functional mixed effects models. *Biometrics* **58** 121–128.

GUTHRIE, K., SHEPPARD, L. and WAKEFIELD, J. (2002). A hierarchical aggregate data model with spatially correlated disease rates. *Biometrics* **58** 898–905.

GUTTORP, P., MEIRING, W. and SAMPSON, P. D. (1994). A space–time analysis of ground-level ozone data. *Environmetrics* **5** 241–254.

HECKMAN, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47** 153–161.

KAMMANN, E. E. and WAND, M. P. (2003). Geoadditive models. *Appl. Statist.* **52** 1–18.

KARCHER, P. and WANG, Y. D. (2001). Generalized nonparametric mixed effects models. *J. Comput. Graph. Statist.* **10** 641–655.

KE, C. L. and WANG, Y. D. (2001). Semiparametric nonlinear mixed-effects models and their applications (with discussion). *J. Amer. Statist. Assoc.* **96** 1272–1298.

KÜNZLI, N., MCCONNELL, R., BATES, D., BASTAIN, T. M., HRICKO, A., LURMANN, F., AVOL, E., GILLILAND, F. D. and PETERS, J. M. (2003). Breathless in Los Angeles: The exhausting search for clean air. *American J. Public Health* **93** 1494–1499.

KÜNZLI, N., MEDINA, S., KAISER, R., QUÉNEL, P., HORAK, F. and STUDNIKA, M. (2001). Assessment of deaths attributable to air pollution: Should we use risk estimates based on time series or on cohort studies? *American J. Epidemiology* **153** 1050–1055.

LANGE, N. and RYAN, L. (1989). Assessing normality in random effects models. *Ann. Statist.* **17** 624–642.

LE, N. D., MAO, Y., SUN, L. and ZIDEK, J. V. (2004). Air pollution exposure and lung cancer incidence risk. Unpublished manuscript.

LIPPMANN, M. et al. (2003). The U.S. Environmental Protection Agency Particulate Matter Health Effects Research Centers Program: A midcourse report of status, progress and plans. *Environmental Health Perspectives* **111** 1074–1092.

LITTLE, R. J. A. (1985). A note about models for selectivity bias. *Econometrica* **53** 1469–1474.

LIU, A., MEIRING, W. and WANG, Y. D. (2005). Testing generalized linear models using smoothing spline methods. *Statistica Sinica* **15** 235–256.

MCCONNELL, R., BERHANE, K., GILLILAND, F., MOLITOR, J., THOMAS, D., LURMANN, F., AVOL, E., GAUDERMAN, W. J. and PETERS, J. M. (2003). Prospective study of air pollution and bronchitic symptoms in children with asthma. *American J. Respiratory and Critical Care Medicine* **168** 790–797.

MCCULLOCH, C. E. (2003). *Generalized Linear Mixed Models*. IMS, Hayward, CA.

MCCULLOCH, C. E. and SEARLE, S. R. (2001). *Generalized, Linear and Mixed Models.* Wiley, New York.

MENG, Z., DABDUB, D. and SEINFELD, J. H. (1997). Chemical coupling between atmospheric ozone and particulate matter. *Science* **277** 116–119.

NATIONAL RESEARCH COUNCIL (1998). *Research Priorities for Airborne Particulate Matter* **1**, **2**. National Academy Press, Washington.

NOVARTIS FOUNDATION (1999). *Environmental Statistics*: *Analysing Data for Environmental Policy.* Wiley, Chichester.

OTT, W., WALLACE, L. and MAGE, D. (2000). Predicting particulate (PM_{10}) personal exposure distributions using a random component superposition statistical model. *J. Air and Waste Management Association* **50** 1390–1406.

PALTA, M. and YAO, T.-J. (1991). Analysis of longitudinal data with unmeasured confounders. *Biometrics* **47** 1355–1369.

PIEGORSCH, W. W., SMITH, E. P., EDWARDS, D. and SMITH, R. L. (1998). Statistical advances in environmental science. *Statist. Sci.* **13** 186–208.

RABL, A. (2003). Interpretation of air pollution mortality: Number of deaths or years of life lost? *J. Air and Waste Management Association* **53** 41–50.

RAMSAY, J. O. and SILVERMAN, B. W. (2002). *Applied Functional Data Analysis*: *Methods and Case Studies.* Springer, Berlin.

RAMSAY, T., BURNETT, R. and KREWSKI, D. (2003). Exploring bias in a generalized additive model for spatial air pollution data. *Environmental Health Perspectives* **111** 1283–1288.

RONDEAU, V., BERHANE, K. and THOMAS, D. C. (2004). A three-level model for binary time-series data: The effects of air pollution on school absences in the Southern California Children's Health Study. *Statistics in Medicine*. To appear.

SALWAY, R. and WAKEFIELD, J. C. (2004). A comparison of approaches to ecological inference in epidemiology, political science, and sociology. In *Ecological Inference*: *New Methodological Strategies* (G. King, O. Rosen and M. Tanner, eds.) 303–332. Cambridge Univ. Press, Cambridge, UK.

SAMPSON, P. D., DAMIAN, D. and GUTTORP, P. (2001). Advances in modeling and inference for environmental processes with nonstationary spatial covariance. In *GeoENV III*: *Geostatistics for Environmental Applications* (P. Monestiez, D. Allard and R. Froidevaux, eds.) 17–32. Kluwer, Dordrecht.

SAMPSON, P. D. and GUTTORP, P. (1999). Operational evaluation of air quality models (with discussion). In *Environmental Statistics*: *Analysing Data for Environmental Policy* 33–51. Wiley, Chichester.

SCHAUER, J. J., ROGGE, W. F., HILDERMAN, L. M., MAZUREK, M. A., CASS, G. R. and SIMONEIT, B. R. T. (1996). Source apportionment of airborne particulate matter using organic compounds as tracers. *Atmospheric Environment* **30** 3837–3855.

SEIGNEUR, C. (2001). Current status of air quality models for particulate matter. *J. Air and Waste Management Association* **51** 1508–1521.

SHEPPARD, L. (2002). Ecologic study design. *Encyclopedia of Environmetrics* **2** 602–606. Wiley, New York.

SHEPPARD, L. and DAMIAN, D. (2000). Estimating short-term PM effects accounting for surrogate exposure measurements from ambient monitors. *Environmetrics* **11** 675–687.

SMITH, R. L., DAVIS, J. M. and SPECKMAN, P. (1999). Human health effects of environmental pollution in the atmosphere. In *Statistical Aspects of Health and the Environment* (V. Barnett, A. Stein and K. F. Turkman, eds.) 91–115. Wiley, New York.

SPIEGELHALTER, D., THOMAS, A. and BEST, N. (1998). *WinBUGS*: *Bayesian inference using Gibbs Sampling, Manual v1.2*. Medical Research Council Biostatistics Unit, Cambridge Univ. Available at www.mrc-bsu.cam.ac.uk/bugs.

STONE, C. J., HANSEN, M. H., KOOPERBERG, C. and TRUONG, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion). *Ann. Statist.* **25** 1371–1470.

WAHBA, G., WANG, Y. D., GU, C., KLEIN, R. and KLEIN, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy. *Ann. Statist.* **23** 1865–1895.

WAKEFIELD, J. C. (2003). Sensitivity analyses for ecological regression. *Biometrics* **59** 9–17.

WAKEFIELD, J. C. (2004). Ecological inference for 2×2 tables (with discussion). *J. Roy. Statist. Soc. Ser. A* **167** 385–445.

WAND, M. P. (2003). Smoothing and mixed models. *Comput. Statist.* **18** 223–249.

WANG, Y. D. (1998). Mixed effects smoothing spline analysis of variance. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 159–174.

WIKLE, C. K. (2003). Hierarchical models in environmental science. *Internat. Statist. Rev.* **71** 181–199.

WILSON, W. E., MAGE, D. T. and GRANT, L. D. (2000). Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: Why and how. *J. Air and Waste Management Association* **50** 1167–1183.

ZHANG, D. W. and LIN, X. H. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* **4** 57–74.

ZIDEK, J. V. and LE, N. D. (1999). Measuring and modelling pollution for risk analysis (with discussion). In *Environmental Statistics*: *Analysing Data for Environmental Policy* 105–121. Wiley, Chichester.

ZIDEK, J. V., MELOCHE, J., SHADDICK, G., CHATFIELD, C. and WHITE, R. (2003). A computational model for estimating personal exposure to air pollutants with application to London's PM_{10} in 1997. Technical Report 205, Dept. Statistics, Univ. British Columbia.

ZIDEK, J., SUN, L., LE, N. and ÖZKAYNAK, H. (2002). Contending with space–time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM_{10} field. *Environmetrics* **13** 595–613.