Submitted for publication on August 27, 2008.

Hierarchical Models for the Effect of Spatial Interpolation Error on the Inferred Relationship between Ambient Particulate Matter Exposure and Cardiovascular Health

JAMES L. CROOKS*

Atmospheric Model Division
National Exposure Research Laboratory
U.S. Environmental Protection Agency, Research Triangle Park, NC 27701, USA
crooks.james@epa.gov

ERIC A. WHITSEL

Departments of Epidemiology and Medicine
Schools of Public Health and Medicine
The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-8050, USA

DIANE J. CATELLIER

Department of Biostatistics
School of Public Health
The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-8030, USA

DUANPING LIAO

Division of Epidemiology
Department of Public Health Sciences
Penn State University College of Medicine, Hershey, PA 17033, USA

P. MIGUEL QUIBRERA

Department of Epidemiology
School of Public Health
The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-8050, USA

RICHARD L. SMITH

Department of Statistics and Operations Research
The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3260, USA

SUMMARY

*To whom correspondence should be addressed.

There are numerous sources of uncertainty when estimating individual air pollution exposures based on data from a dispersed network of monitors. One of the most important concerns the fact that personal exposures to air pollution are at best only crudely approximated by measurements from the nearest air pollution monitor. A major component of the discrepancy can be addressed by interpolating ambient air pollution data from monitors to the individual addresses of participants, and in a previous study, this has been done by lognormal kriging. The primary purpose of the present study is to show how error in kriging air pollution variables affects the regression coefficients for the epidemiological effect under study. We address these issues in the context of a Women's Health Initiative (WHI) ancillary study, the Environmental Epidemiology of Arrhythmogenesis in WHI (EEAWHI). We are specifically interested in whether EEAWHI findings are robust to ignoring log-normal spatial interpolation error from kriging. We use a multivariable adjusted, Bayesian hierarchical model in which the response is the logarithm of the median RR interval (ms) from 53,494 clinical trial participants' first-recorded, resting standard twelve-lead electrocardiograms (ECGs) between 1999 and 2002. The exposures are daily mean concentrations (μg/m^3) of ambient particulate matter < 2.5 and 10 μm in diameter ($PM_{2.5}$; PM_{10}) spatially interpolated at participants' geocoded addresses and averaged over the day of and before the ECGs (lag 0-1). We perform within-site regressions involving only those women examined at each of 57 exam sites in the contiguous U.S.; however, the site-specific exposure-response effects are drawn from a distribution centered on a national effect that is our main epidemiologic focus. When the uncertainty in PM measurements is modeled explicitly by placing a log-normal prior on the exposures the posterior distribution of the national effect is shifted in an unpredictable direction and the width of the distribution either increases only slightly or decreases. However, doubling the kriging error causes all the posteriors to move toward zero and narrow. This result contrasts with the well-known effect of Gaussian measurement error in a standard linear model: a broadening, not a narrowing of the posterior, and a shift of the mean toward zero. Simulations are presented to demonstrate that our result is not spurious, but is in fact a natural outcome when predictor error is log-normally distributed. Consequently, the inverse PM-RR association observed among non-smokers without chronic lung disease remains discernible when kriging error is accounted for (the posterior probabilities remain greater than 0.95), though the effect sizes change. We therefore conclude that the significance of the observed association is robust against ignoring kriging uncertainty, though the magnitude of the association is not.

Keywords: Measurement Error, Bayesian Hierarchical Model, Particulate Matter, Kriging

1. INTRODUCTION

Mounting evidence suggests that the effect of ambient particulate matter (PM) air pollution on acute coronary heart disease (CHD) events depends in part on autonomic, i.e. "involuntary" nervous system mechanisms (NCEA (2006)). Since autonomic nerve fibers control the firing rate of the sinoatrial node – specialized cardiac muscle cells that serve as the heart's pacemaker – much of this evidence is based on measures of autonomic status that are readily obtained from the electrocardiogram (ECG) (Zareba *and others* (2001)). These ECG measures include the time elapsed between successive waves of ventricular depolarization, aka the RR interval (RR); a well-known measure derived from RR, heart rate (HR); as well as time- and frequency-domain measures of their beat-to-beat variability over the long and short term (Task Force of the ESC and the NASPE (1996); Schroeder *and others* (2004)).

Decreases in RR (ms), increases in HR (beat/sec), and decreases in their variability are consistent with heightened activity of the sympathetic division of the autonomic nervous system. Although numerous studies have linked increased PM concentrations and CHD incidence to these changes (NCEA (2006); Greenland *and others* (1999); Liao *and others* (1997); Gillum *and others* (1991); Kannel *and others*

(1987)), none have examined this autonomic mechanism in an epidemiologic study of susceptibility to the adverse cardiovascular effects of ambient PM exposures using hierarchical models that account for measurement error inherent in the spatial interpolation of PM concentrations. Such studies remain uncommon despite calls for methodologically sound epidemiologic research designed to help elucidate pathophysiologic mechanisms underlying PM-CHD associations in human populations (HEI (2002); Brook *and others* (2004); NRC (2004)).

This paper examines the effects of log-normally distributed PM interpolation error from kriging on Bayesian hierarchical models of the PM-RR association and its effect modifiers. We conducted this examination in an ancillary study of a large, geographically diverse population of U.S. women enrolled in the Women's Health Initiative (WHI) clinical trials (WHI Study Group (1998)), The Environmental Epidemiology of Arrhythmogenesis in WHI (EEAWHI). In Section 2, we describe pertinent ancillary study data. In Sections 3-5, we compare results from our simple linear, meta-, and Bayesian hierarchical regression models. In Sections 6 and 7, we discuss and summarize our findings, then in Appendix A, provide additional details. In Appendix B we present convergence diagnostics for the Bayesian hierarchical model. Finally, in Appendix C we provide simulations to support our findings.

2. THE EEAWHI DATASET

2.1 *Setting, Design & Study Population*

The WHI clinical trials were designed to allow randomized, controlled evaluation of estrogen $\pm$ progestin treatment, calcium / vitamin D supplementation, and dietary modification on risk of breast and colorectal cancer, cardiovascular disease, and bone fractures (WHI Study Group (1998)). Between 1993 and 1998, the trials enrolled 68,132 postmenopausal women aged 50 to 79 years. Interested and eligible enrollees were invited to follow-up exams at one, three, six and nine years after their baseline exam in one of 57 U.S. locations (including satellite clinics and their changes in location). Rigorous quality assurance programs were in place through close-out (September 2004-March 2005).

2.2 *RR Interval*

Centrally trained and officially certified technicians recorded resting, supine standard twelve-lead ECGs at the baseline and year three, six and nine exams using MAC PC electrocardiographs (GE Marquette, Inc.) (WHI Study Group (1998)). Upon successful recording, they transmitted ECGs by telephone modem to the Epidemiological Cardiology Research (EPICARE) Center for visual inspection, error / missing lead detection, quality grading, and electronic reading by the Marquette 12SL program (GE Marquette, Inc.). Electronic reading produced several measures, one of which is the focus of the present report: the median RR across all twelve leads. All analysis methods in this paper used logarithm of RR as the response variable.

2.3 *PM Concentrations*

Participant addresses were collected at each exam, updated at least biannually, and cleaned following a standardized protocol. We submitted them en bloc to a single geocoding vendor selected from four candidates on the basis of the accuracy of coordinates (latitudes; longitudes) assigned to the addresses (Whitsel *and others* (2004, 2006)). We then obtained ambient PM concentration data recorded at monitors operating during the study period from the U.S. Environmental Protection Agency Air Quality System (EPA Air Quality System (1994)). The data included the longitude and latitude of each monitor. We cleaned the

data and estimated daily mean PM concentrations (standard errors) at each geocoded address in the contiguous U.S. from baseline through closeout using a spherical model to perform national-scale, lognormal ordinary kriging and the weighted least-squares method to estimate cross-validated semivariograms (Liao *and others* (2006)). For simplicity, the present report focuses on concentrations of PM < 2.5 and 10 μm ($PM_{2.5}$; PM_{10}) averaged over the day of and before each ECG (lag 0-1).

2.4 *Temperature*

We obtained and cleaned temperatures, longitudes, and latitudes recorded at meteorological stations operating in the contiguous U.S. during the study period from the National Climatic Data Center (U.S. NCDC). We computed station-specific daily mean temperature at all stations with $\leqslant$ six consecutive hours (25%) of missing data. We estimated daily mean temperature (^{o}C) at each geocoded address from baseline to closeout by averaging these daily means across all stations within 50 km, a distance over which their station-to-station correlations exceed 0.90 (Ito *and others* (2001)). Temperatures at lag 1 were used in the analysis.

2.5 *Participant Characteristics*

Attributes of participants were determined at each exam by standardized interview and examination. Interim health events also were identified via standardized medical record review and physician adjudication. For the present report, age (yr) is defined at the time of the ECG; self-reported ethnicity as White/Non-Hispanic, Black/African-American, Hispanic/Latino, Asian/Pacific Islander or other; education as $\geqslant$ college graduate; diabetes by anti-diabetic medication use or history; hypertension by anti-hypertensive medication use, systolic blood pressure $\geqslant$ 140 mm Hg, diastolic blood pressure $\geqslant$ 90 mm Hg, or history; beta-blocker as use of a beta-adrenergic receptor antagonist (a medication class associated with well-known increases in RR); hypercholesterolemia by anti-hyperlipidemic medication use or history; smoking as current smoker or not; chronic lung disease by history of asthma, emphysema or lung cancer; coronary heart disease (CHD) by anti-anginal medication use, history of angina or myocardial infarction, or medical record review / adjudication; revascularization by history of coronary artery angioplasty, stent or bypass, or medical record review / adjudication; and congestive heart failure (CHF) by cardiac glycoside and diuretic use, history, or medical record review / adjudication.

2.6 *Exclusions*

Of the 58,705 participants with an ECG recorded between 1999 and 2002, we excluded 5,211 (8.9%) with conditions that affect the availability or accuracy of PM or RR: foreign, U.S. military, U.S. protectorate, Hawaiian, Alaskan or missing addresses; poor ECG quality grades; < 5 or 50% normally conducted RR intervals; atrioventricular conduction defects; electronic pacing; frequent premature ventricular beats; arrhythmias; anti-arrhythmic medication use; and some other individuals for whom relevant covariate data was missing. We conducted all analyses on the first ECG recorded during the study period among the remaining 52,805 participants.

3. SIMPLE REGRESSION ANALYSES

This section presents simple linear regression analyses of the data described in Section 2. This will serve as preliminary to the more complicated analyses described in Sections 4 and 5. The analysis used logarithm of RR as the response variable, and the following covariates: exam site as a factor variable with 57

Variable	Lag(s)	Estimate ($\times 10^4$)	SE ($\times 10^4$)	t statistic	2-sided p-value
PM_{10}	0	–1.58	0.55	–2.89	0.0039
PM_{10}	1	–1.51	0.57	–2.65	0.0080
PM_{10}	2	–0.26	0.60	–0.43	0.66
PM_{10}	3	0.43	0.59	0.72	0.47
PM_{10}	4	0.36	0.57	0.63	0.52
PM_{10}	5	0.26	0.55	0.47	0.64
PM_{10}	01	–2.06	0.64	–3.20	0.0014
$PM_{2.5}$	0	–2.42	0.85	–2.86	0.0043
$PM_{2.5}$	1	–2.23	0.87	–2.57	0.010
$PM_{2.5}$	2	–1.22	0.85	–1.43	0.15
$PM_{2.5}$	3	–0.48	0.83	–0.58	0.56
$PM_{2.5}$	4	–0.04	0.82	0.04	0.97
$PM_{2.5}$	5	0.69	0.82	0.83	0.40
$PM_{2.5}$	01	–2.89	0.96	–3.03	0.0025

Table 1. Simple linear regression models for 12 particulate matter measures.

levels; temperature; time of day, day of week, and season of ECG recording; plus each of the participants characteristics listed in Section 2.5.

In addition to the above, we included (one at a time) concentrations of PM_{10} and $PM_{2.5}$ at each of lags 0-5 days, as well as the average of lags 0 and 1. Results are in Table 1. From this we make two conclusions. First, the strongest effects are at lags 0 and 1, with the strongest of all at the average of lags 0 and 1. Second, for most lags, the effect due to PM_{10} is stronger than the effect due to $PM_{2.5}$, if measured in standardized form (via the t statistic).

The next step was to examine putative interactions between PM_{10} or $PM_{2.5}$ and participant characteristics. For example, we might expect the response to PM to be different in women with and without chronic lung disease. This possibility was studied by splitting the PM covariate in two, one restricted to individuals without chronic lung disease and the other restricted to individuals with chronic lung disease ("LD"). In this way we measured the effect of PM on RR separately among women with and without chronic lung disease. A number of other splits were also made: those corresponding to smoking status ("SM" = current smoker) and beta-blocker use ("BB") as well as various combinations of these variables. In all cases, the PM variable was taken to be the mean of lags 0 and 1. Results are in Table 2 for PM_{10} and Table 3 for $PM_{2.5}$.

The results showed that PM significantly affects RR only among non-smokers without chronic lung disease. This suggested defining two risk groups, one consisting of non-smokers without chronic lung disease ("No SM, No LD") and the other of everyone else ("SM or LD"). To examine the effect of beta-blocker use, each of the prior groups was split into two further subgroups, labeled "No BB" or "BB". Within the "No SM, No LD" group where the effect of PM is significant, this further subdivision into BB and no BB showed about the same level of statistical significance as measured by the t statistic, but a much larger magnitude of regression coefficient in the BB group, suggesting that PM has a greater effect among beta-blocker users than non-users. The results are quite consistent between PM_{10} and $PM_{2.5}$ but, again, generally stronger for PM_{10} if characterized by the t statistic. The results for the non-smoking/no lung disease group were statistically significant even when using the conservative Bonferroni correction for the 22 comparisons. When further split by beta-blocker use the PM_{10} results were of borderline statistical

Subgroup	Percent of total	Estimate ($\times 10^4$)	SE ($\times 10^4$)	t statistic	2-sided p-value
No LD	89.8	–2.41	0.67	–3.58	0.00034
LD	10.2	0.96	1.81	0.53	0.60
No BB	87.2	–1.72	0.68	–2.53	0.011
BB	12.8	–4.45	1.65	–2.70	0.0069
No SM	94.0	–2.29	0.66	–3.48	0.00051
SM	6.0	1.60	2.37	0.67	0.50
NoBB NoLD	77.9	–2.02	0.71	–2.85	0.0043
NoBB, LD	9.3	0.84	1.83	0.46	0.65
BB, NoLD	11.9	–5.00	1.66	–3.01	0.0026
BB, LD	1.0	2.05	3.05	0.67	0.50
NoBB, NoSM	81.8	–1.92	0.69	–2.76	0.0058
NoBB, SM	5.40	1.28	2.40	0.53	0.59
BB, NoSM	12.3	–4.91	1.66	–2.96	0.0031
BB, SM	0.6	3.63	3.71	0.98	0.33
NoSM, NoLD	84.6	–2.62	0.69	–3.83	0.00013
NoSM, LD	9.5	0.75	1.84	0.41	0.68
SM, NoLD	5.2	1.33	2.42	0.55	0.58
SM, LD	0.7	3.31	3.41	0.97	0.33
NoBB,NoSM,NoLD	73.1	–2.12	0.71	–2.98	0.0029
NoBB,SMorLD	14.0	0.48	1.37	0.35	0.73
BB,NoSM,NoLD	11.4	–5.35	1.67	–3.21	0.0013
BB,SMorLD	1.4	1.99	2.54	0.78	0.43

Table 2. Regression analysis in which the population is split into subgroups in various ways. Variable of interest is PM_{10}, mean of lags 0 and 1.

significance after Bonferroni correction, though their clinical significance was such that we retained the split in further analyses. We also retained this split for $PM_{2.5}$ for the sake of comparison.

For subsequent analysis, we concentrate on the split defined by the last box in Tables 2 and 3, dividing the population first into two groups, one consisting of non-smokers without chronic lung disease, and the other consisting of everyone else. Then, each of those subgroups was further split according to beta-blockers use.

4. NMMAPS-STYLE ANALYSIS

The preceding analysis could misrepresent the overall effect of PM on RR if there were widespread differences in some regression coefficients across sites. To take an obvious example, the effect of temperature is presumably different in Phoenix than Minneapolis.

To improve the robustness of the results against this heterogeneity, an analysis was performed similar to that used in the well-known NMMAPS study (e.g. Dominici *and others* (2000a, 2003); Samet *and others* (2000)). The same regression model was fitted *separately* to data from each of the 57 exam sites, with individual counts ranging from 75 to 1972. For each site the coefficient and standard error corresponding to the particulate matter effect was retained. The results were combined across sites using the `tlnise` program based on Everson and Morris (2000), implemented in R by Dr. Roger Peng (http://cran.r-

Subgroup	Percent of total	Estimate ($\times 10^4$)	SE ($\times 10^4$)	t statistic	2-sided p-value
No LD	89.8	–3.20	1.00	–3.20	.0014
LD	10.2	–0.19	2.70	–0.07	0.94
No BB	87.2	–2.47	1.01	–2.44	0.015
BB	12.8	–5.77	2.41	–2.40	0.017
No SM	94.0	–3.02	0.98	–3.08	0.0021
SM	6.0	–0.96	3.44	–0.28	0.78
NoBB NoLD	77.9	–2.64	1.05	–2.50	0.012
NoBB, LD	9.3	–1.06	2.75	–0.39	0.70
BB, NoLD	11.9	–6.91	2.44	–2.83	0.0047
BB, LD	1.0	5.70	4.88	1.17	0.24
NoBB, NoSM	81.8	–2.54	1.04	–2.44	0.015
NoBB, SM	5.40	–1.40	3.46	–0.41	0.69
BB, NoSM	12.3	–6.16	2.42	–2.55	0.011
BB, SM	0.6	6.43	6.73	0.96	0.34
NoSM, NoLD	84.6	–3.34	1.02	–3.27	0.0011
NoSM, LD	9.5	–0.09	2.75	–0.03	0.97
SM, NoLD	5.2	–0.95	3.52	–0.27	0.79
SM, LD	0.7	–0.93	5.64	–0.17	0.87
NoBB, NoSM,NoLD	73.1	–2.75	1.07	–2.56	0.10
NoBB, SMorLD	14.0	–1.01	2.11	–0.48	0.63
BB, NoSM,NoLD	11.4	–7.24	2.45	–2.96	0.0031
BB, SMorLD	1.4	5.57	4.22	1.32	0.19

Table 3. Regression analysis in which the population is split into subgroups in various ways. Variable of interest is $PM_{2.5}$, mean of lags 0 and 1.

project.org/web/packages/tlnise/index.html). The analysis was restricted to the data split mentioned at the end of Section 3 (based on chronic lung disease, current smoking status and betablocker use), and to PM concentrations averaged across lags 0 and 1. Both PM_{10} and $PM_{2.5}$ were considered. Results are shown in Table 4.

The results reinforce the findings of Section 3. The PM-RR associations are statistically significant only among the non-smokers without chronic lung disease, but among that group, the effect is much stronger among beta-blocker users than among non-users.

5. Fully Bayesian Analysis

This section presents a fully Bayesian analysis, combining data across sites as in Section 4, but also including the effects of measurement error in PM_{10} or $PM_{2.5}$. The analysis relies on a Bayesian hierarchical model, using Gibbs sampling to update the regression parameters and a Metropolis-Hastings update for the particulate matter variable. Full details of the statistical model and MCMC algorithm are included in Appendix A.

The Bayesian hierarchical analysis was run using the average for lags 0 and 1 of either PM_{10} or $PM_{2.5}$, with a multiplier $M = 0$, 1 or 2 applied to the kriging error. Since the concentrations were kriged on a daily basis we calculated the error for this average using the kriging errors from lags 0 and 1 and the lag-

Subgroup	Pollutant	Posterior Mean ($\times 10^4$)	95% CI ($\times 10^4$)
NoBB,NoSM,NoLD	PM_{10}	–1.77	(–3.40,–0.14)
NoBB,SMorLD	PM_{10}	0.40	(–3.25,4.05)
BB,NoSM,NoLD	PM_{10}	–5.63	(–9.80,–1.46)
BB,SMorLD	PM_{10}	2.08	(–3.33,7.49)
NoBB,NoSM,NoLD	$PM_{2.5}$	–2.76	(–5.27,–0.25)
NoBB,SMorLD	$PM_{2.5}$	–0.92	(–6.27,4.43)
BB,NoSM,NoLD	$PM_{2.5}$	–5.57	(–11.63,0.49)
BB,SMorLD	$PM_{2.5}$	6.92	(–2.57,16.41)

Table 4. Posterior means and 95% credible intervals (CIs) for combined coefficient of PM_{10} or $PM_{2.5}$, subdivided into four groups of participants, based on hierarchical analysis across sites but without taking into account PM measurement error.

1 autocorrellation of PM_{10} (0.57) and $PM_{2.5}$ (0.67) calculated from our data. The introduction of M is intended to allow examination of different levels of measurement error. $M = 0$ ignores the measurement error and gives results fairly comparable to Section 4 (see Table 5). $M = 1$ is the case of primary interest, while $M = 2$ is included to allow examination of the effect of under-estimating the true PM measurement error. As before, the analysis included the participant characteristics listed in Section 2.5.

The participants were divided into the same four subgroups, as discussed in Section 3, based on chronic lung disease, current smoking status and beta-blocker use. As defined in Appendix A, α_k is the regression coefficient of PM_{10} or $PM_{2.5}$ on log RR in subgroup k, $k = 1, 2, 3, 4$, where, for example, α_1 refers to the "no BB, no SM, no LD" subgroup. For each analysis, a total of 50,000 MCMC iterations was run, divided into 1000 loops of size 50. The first 20,000 iterations were discarded as burn-in, the rationale for this choice being discussed further in Section B.

5.1 *Results*

Figures 1 and 2 show the posterior density estimates of $\alpha_k,\ k = 1, ..., 4$. The comparison between the posterior density curves for $M = 0$ and $M = 1$ show that in all cases the density is shifted as a result of taking account of measurement error, though not necessarily towards 0. It appears that the shift in the posterior pdf as a result of taking account of measurement error is greater with PM_{10} than with $PM_{2.5}$. In all cases, however, doubling the measurement error standard deviation results in a much more marked shift, towards a posterior density that is highly peaked near 0. This feature of the results was initially unexpected, but there is a natural explanation for it, detailed further in Appendix C.

Table 5 shows the posterior means and 95% credible intervals corresponding to the posterior density curves in Figures 1 and 2. The $M = 0$ results here should be compared with those of Table 4 and show the effect of using a fully Bayesian approach approach to the hierarchical analysis as compared with the NMMAPS approach. In Table 4 the credible intervals were calculated assuming a normal distribution based on the posterior mean and posterior standard deviation that are produced by the `tlnise` program; in Table 5 they are based directly on the MCMC output with boundaries at the 2.5 and 97.5 percentiles of the posterior distribution. The results clearly show some differences between the two approaches, though they are not so great as to affect the epidemiological interpretation of the results. Table 6 shows the posterior probabilities of $\alpha_k < 0,\ k = 1, ..., 4$, for three values of the error multiplier M, and two pollutants. These probabilities have a similar interpretation to that of a p-value in classical statistics: a

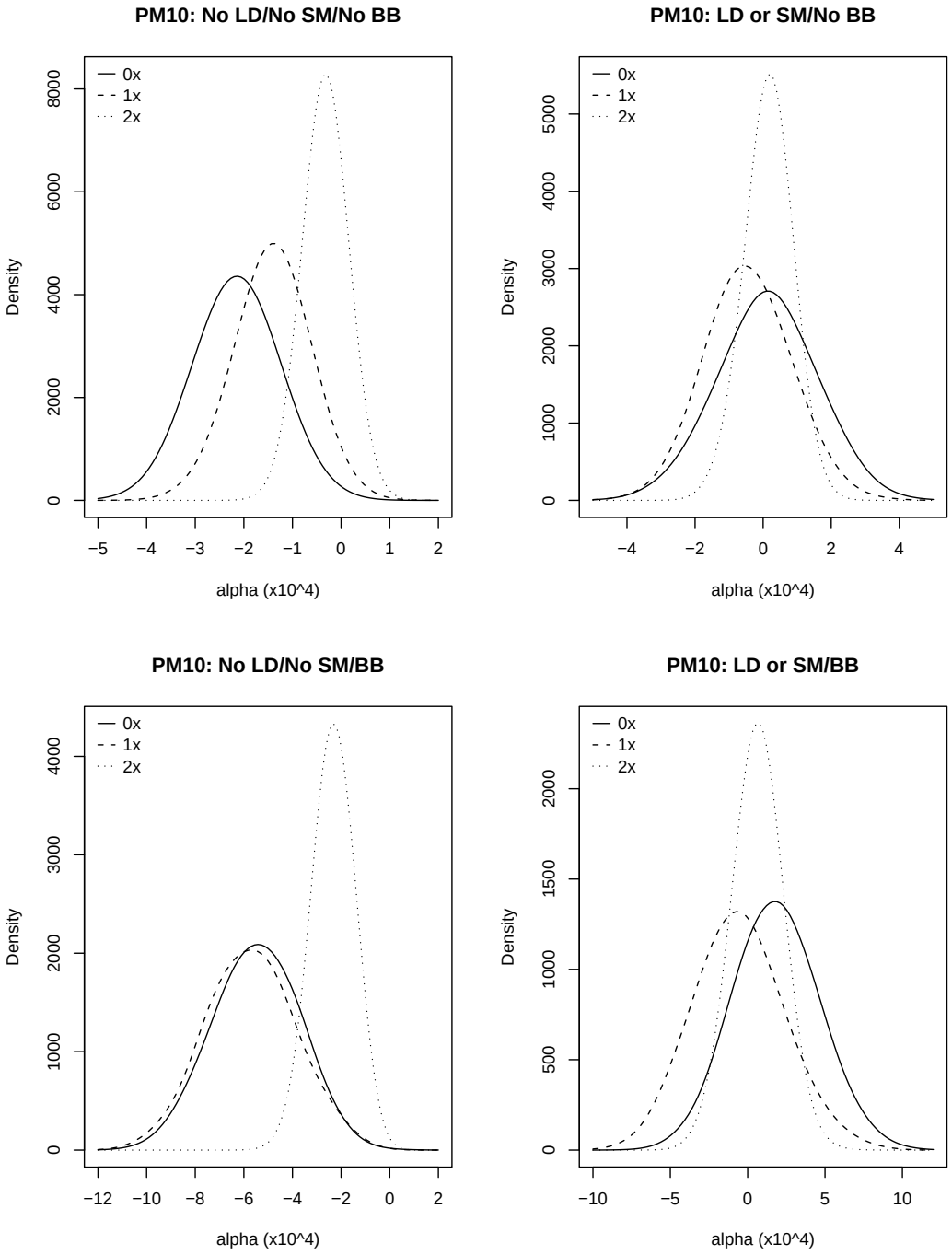

Fig. 1. Posterior densities for PM_{10} regression coefficient by subgroup.

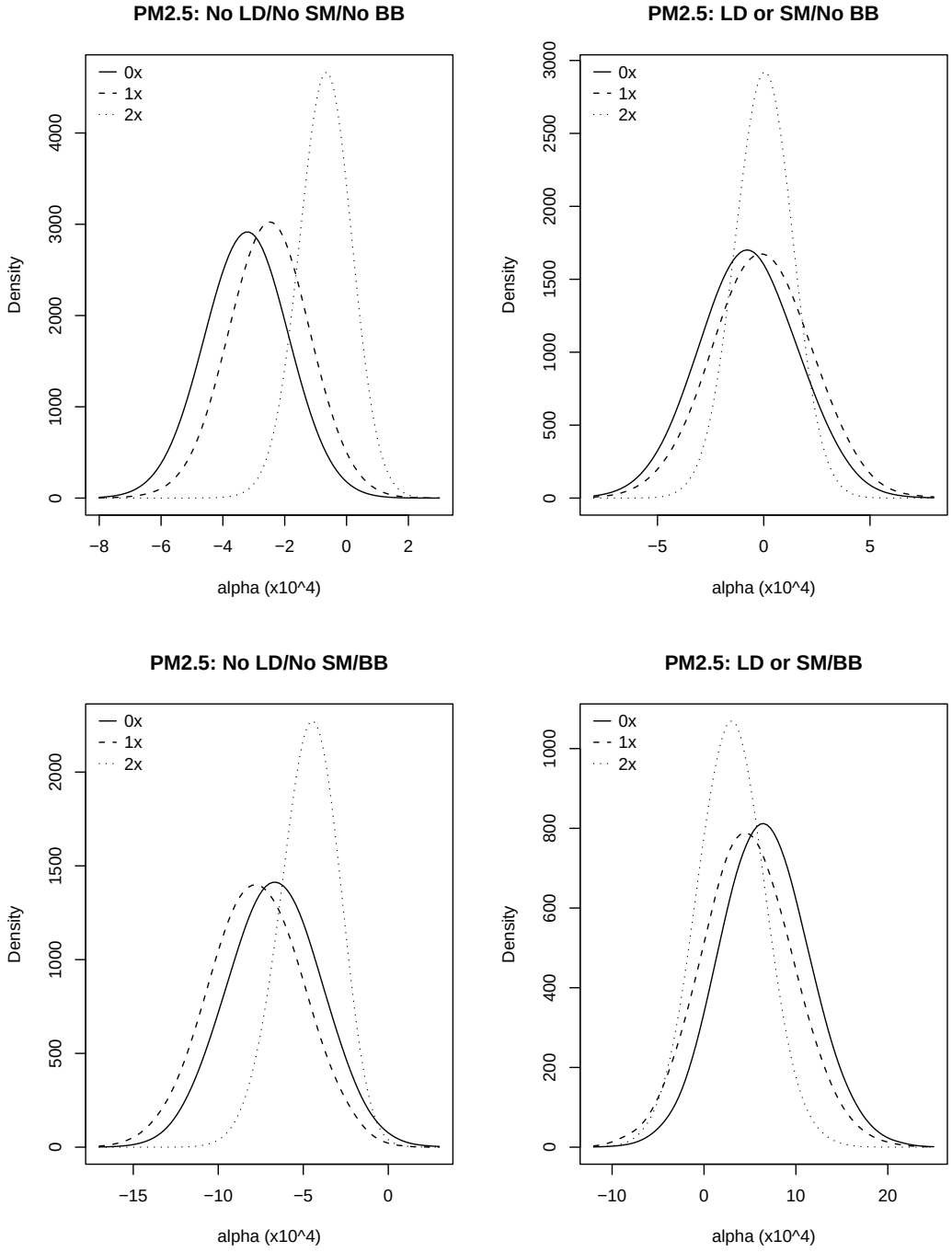

Fig. 2. Posterior densities for $PM_{2.5}$ regression coefficient by subgroup.

		$M = 0$		$M = 1$		$M = 2$	
Subgroup	PM	Mean ($\times 10^4$)	95% CI ($\times 10^4$)	Mean ($\times 10^4$)	95% CI ($\times 10^4$)	Mean ($\times 10^4$)	95% CI ($\times 10^4$)
NoBB,NoSM,NoLD	10	–2.15	(–3.93,–0.37)	–1.41	(–3.03,0.14)	–0.33	(–1.29,0.61)
NoBB,SMorLD	10	0.12	(–2.88,3.02)	–0.44	(–2.98,2.20)	0.15	(–1.35,1.56)
BB,NoSM,NoLD	10	–5.48	(–9.13,–1.96)	–5.72	(–9.45,–1.92)	–2.34	(–4.17,–0.62)
BB,SMorLD	10	1.83	(–3.75,7.56)	–0.54	(–6.30,5.73)	0.62	(–2.76,3.95)
NoBB,NoSM,NoLD	2.5	–3.24	(–4.84,–0.50)	–2.51	(–4.64,0.10)	–0.69	(–2.43,0.96)
NoBB,SMorLD	2.5	–0.72	(–4.42,3.68)	–0.08	(–4.16,4.08)	–0.01	(–2.74,2.66)
BB,NoSM,NoLD	2.5	–6.68	(–11.23,–1.29)	–7.78	(–11.57,–2.31)	–4.67	(–8.32,–1.42)
BB,SMorLD	2.5	6.59	(–3.18,11.52)	4.73	(–5.25,11.25)	2.93	(–4.52,9.84)

Table 5. Posterior means and 95% credible intervals, for combined coefficient of mean PM_{10} or $PM_{2.5}$ at lag 0-1, subdivided into four groups of participants, based on fully Bayesian model for cases $M = 0$, $M = 1$, $M = 2$.

Variable	M	NoBB,NoSM,NoLD ($k = 1$)	NoBB,SMorLD ($k = 2$)	BB,NoSM,NoLD ($k = 3$)	BB,SMorLD ($k = 4$)
PM_{10}	0	0.991	0.465	0.999	0.266
PM_{10}	1	0.962	0.637	0.999	0.580
PM_{10}	2	0.750	0.409	0.997	0.354
$PM_{2.5}$	0	0.991	0.622	0.992	0.088
$PM_{2.5}$	1	0.971	0.515	0.998	0.176
$PM_{2.5}$	2	0.789	0.498	0.998	0.218

Table 6. Posterior probabilities of $\alpha_k < 0,\ k = 1, 2, 3, 4$, for two pollutants ($PM_{10}$ and $PM_{2.5}$) and three noise multiplies ($M = 0$, 1 and 2). Based on 50000 iterations, the first 20000 iterations discarded as burn-in.

posterior probability close to 1 indicates a high level of confidence that the true value of α_k is negative. The results demonstrate that for the "No SM, No LD" groups, the posterior probability that $\alpha_k < 0$ is near 1 in every case for which $M = 0$ or $M = 1$; in fact, with one exception (NoBB,NoSM,NoLD; PM_{10}) the same is also true when $M = 2$. We therefore see that for this dataset, the inference that there is an inverse PM-RR association is quite robust against measurement error.

5.2 *Subgroup Differences*

Though we originally split the study population into four subgroups based on scientific plausibility it is worth revisiting whether the differences are borne out in the analysis. In a frequentist context, a test of the null hypothesis $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ would be interpreted as a test of interaction between subgroup number and PM. In the present Bayesian context, if at least one of the posterior probabilities that $\alpha_{k_1} - \alpha_{k_2} < 0$ (for difference $k_1, k_2 \in \{1, 2, 3, 4\}$) is very close to 0 or 1, we conclude that interactions exist.

For example, our results suggest that α_1 and α_3 are both negative but $\alpha_3 < \alpha_1$, implying that PM has a greater relative effect on beta-blocker users than non-users among non-smokers who do not have chronic lung disease. In fact, for the results based on PM_{10}, the posterior probability that $\alpha_3 < \alpha_1$ is 0.96, 0.98, and 0.98 for $M = 0, 1, 2$ respectively. For $PM_{2.5}$ the corresponding probabilities are 0.87, 0.96, and 0.98. In all cases there appears to be strong evidence that $\alpha_3 < \alpha_1$, but interestingly, the evidence is stronger

when measurement error is explicitly modeled ($M = 1$ or 2) that when it is ignored ($M = 0$).

On the other hand, among participants who do not use beta-blockers, the posterior probabilities that the PM effect is stronger for non-smokers who do not have chronic lung disease than for others are 0.92, 0.75, and 0.72 for PM_{10} and 0.84, 0.83, and 0.67 for $PM_{2.5}$. In this case there appears to be strong evidence for a difference between groups only for PM_{10} and only if kriging error is ignored.

6. DISCUSSION

Resting heart rate ($\propto RR^{-1}$) is directly associated with incident CHD morbidity and mortality in population-based studies including the Framingham Heart Study (Kannel *and others* (1987)), National Health and Nutrition Examination Survey (Gillum *and others* (1991)), and Chicago Heart Association Project in Industry (Greenland *and others* (1999)). Thus, factors that increase heart rate (decrease RR) may be important risk factors in CHD. In the present study we found an inverse association between short-term PM exposure and RR interval among the 84.6% of the study population consisting of non-smoking women without chronic lung disease. Furthermore, within this subgroup we found a marked difference in effect size between beta-blocker users and non-users (Table 5).

Explicitly modeling the intrinsic kriging uncertainty caused the means and variances of the posterior distributions of the association to change in unpredictable ways. However, when the kriging errors were doubled (to assess the impact of underestimating them) the posterior distributions all shifted toward the null and narrowed. By comparison, in a standard univariate linear model the conventional wisdom is that incorporating Gaussian measurement error results in a shift toward the null and a broadening of the posterior distribution instead of a narrowing (though this intuition can be misleading in some common situations, Gustafson (2004)). We performed a simulation study (Appendix C) to determine whether these differences could be explained by the error distributions alone. Our simulations showed that the qualititative changes we observed as a result of varying M occured naturally in a simple univariate setting with log-normal measurement error.

We furthermore found strong evidence of a differential PM effect between beta-blocker users and non-users (specifically among non-smokers who did not have chronic lung disease), and indeed this evidence was *strengthened* by including measurement error in the model. On the other hand, among non-users we could find at best weak evidence of a differential PM effect between the non-smokers without chronic lung disease and everyone else. The moderate evidence for PM_{10} we originally observed was weakened considerably by including measurement error.

Given the observational nature of this study it was impossible to determine why we did not observe a PM-RR association among current smokers or those with chronic lung disease. There are a number of possible reasons. First, the effects of smoking or chronic lung disease on RR may have obscured relatively weak effects of ambient PM. Second, there may have been a censoring effect in that among women who smoke or who have chronic lung disease those with higher sensitivity to the adverse effects of PM may have been more likely to die before their WHI exam site visit than those with lower sensitivity. Finally, the true effect may simply have been unobservable due to the small number of smokers and participants with chronic lung disease (%15.4 of the total study population). It should be stressed that our results do not imply that PM has no effect on RR among current smokers or women with chronic lung disease. Likewise, the observational nature of the study also made it impossible to determine the extent to which the inverse PM-RR association among non-smokers without chronic lung disease was driven by the direct effects of beta-blockers or the indirect, confounding effects of the medical conditions for which they were prescribed.

7. Summary and Conclusions

The fact that epidemiological study results can be affected by measurement error has been recognized for some time (Jurek *and others* (2006)), but progress in accounting for this error has been slowed by the difficultly in characterizing measurement error in air pollution fields. However, there are several statistical methods (such as kriging and its variants) that produce a probability distribution for the measurement error. Despite this, there are still relatively few studies where this distribution has been directly incorporated into the inferential process, and when it has it has usually been based on the linear Gaussian measurement error model. A few recent studies that do in fact incorporate measurement error are described below, and differences with the present study are highlighted.

Gryparis *and others* (2006) examined the association between northeastern U.S. ambient PM_{10} concentration and various health outcomes taken from the Nurses' Health Study (Belanger *and others* (1978)). They assumed Gaussian residuals, but unlike the present study they interpolated PM using splines with Gaussian measurement errors. The fully Bayesian model they presented implemented the exposure model and the health model simultaneously, but they also noted that a two-stage Bayesian approach that uses the posterior of the former as the prior of the latter yielded almost identical results. This latter approach was both computationally simpler and eliminated arguably spurious feedback from the health data to the spatial interpolation. It is also similar to the present approach, though our exposure model was not Bayesian. Furthermore, their health model was not hierarchical and included far fewer covariates than ours.

Künzli *and others* (2005) studied the relationship between $PM_{2.5}$ exposure and corotid artery intimal-medial thickness (CIMT) among residents of the Los Angeles basin, using a geostatistical analysis that was similar to ours in a number of ways. In particular, they used geocoding to determine the latitude and longitude of individual participants' addresses, and kriging from existing $PM_{2.5}$ monitors to estimate the long-term mean at each address. Unlike our study, they did not attempt to construct a $PM_{2.5}$ field for each day. In incorporating the results of their kriging analysis into the epidemiological analysis, they proposed (among other analyses) a weighted least squares (WLS) approach in which the weight associated with a given observation is inversely proportional to the standard error of the kriging estimate, claiming that the estimates computed this way are little different from those in which the weights are ignored. However, this approach was incorrect because WLS presumes that the weights are related to the errors in the dependent variable (CIMT in this case), not the independent variable ($PM_{2.5}$ exposure). Perversely, using their approach one could multiply all the kriging errors by an arbitrary positive constant without changing the resulting estimates of the CIMT-$PM_{2.5}$ association.

Also of relevence to the present work, Künzli *and others* (2005) found that the $PM_{2.5}$-CIMT association was stronger among subjects who had never smoked than among former smokers. Furthermore, in the former group the effect was significant while in the latter it was not. Though the $PM_{2.5}$-smoking interaction was not observed in a subsequent paper on CIMT (Diez Roux *and others* (2008)), effect modification by smoking status in the direction of Künzli *and others* (2005) was later described by Hoffmann *and others* (2006), albeit in a study of traffic-related exposures and coronary heart disease.

The present study differs from previous work by assuming a log-normal model, which is usually a better fit than the linear model in air pollution contexts. However, under this model, measurement error cannot be accounted for through simple adjustments to the regression coefficients. We have demonstrated a method that accounts for this measurement error using a computationally intensive but conceptually straightforward Bayesian technique.

Our results contradict some common conceptions about this kind of analysis. First, including measurement error did not necessarily cause the means of the posterior distributions to shift toward zero. Indeed, at $M = 1$ there was no discernable pattern to the direction of "bias" compared to $M = 0$. However, when the kriging error was doubled all the posteriors shifted toward zero. Second, our posteriors did not widen systematically after accounting for measurement error, and in fact *narrowed* in the $M = 2$ scenario. We

presented evidence from a simulation study that these results are due to the log-normal measurement error distribution and not some other aspect of our model (Appendix C).

Our primary aim was to determine whether the epidemiological results were robust to ignoring kriging uncertainty. We have shown that, because of the way the posterior distributions change under the effect of log-normal emeasurement error, the credibility of the result is robust, though the effect magnitudes generally are not. We stress, though, that this final conclusion should not be taken to be generally true of log-normal measurement error models though it may be true in individual studies.

APPENDIX

A. DETAILS OF MODEL AND MCMC UPDATING EQUATIONS

Let y_{ij} denote the observed response of individual j at exam site i, and let $\{x_{ijk}\}$ denote the individual-level covariates. We assume site-level regression coefficients $\{\beta_{ik}\}$ and precisions $\{\kappa_i\}$ to have prior normal and gamma distributions respectively, where the prior mean of β_{ik} depends on site-level covariates $\{z_{it},\ t = 1, ..., T\}$. Parameters α_{kt} and ψ_k define the prior distributions of $\{\beta_{ik}\}$. This is the general formulation of the model, that allows for both site-level and individual-level covariates, but for the applications in this paper, we take $T = 1$, $z_{i1} = 1$ for all i, and write α_k instead of $\alpha_{kt},\ t = 1, ..., T$. We write P_{ij} for the true value of the pollution variable (PM_{10} or $PM_{2.5}$) for individual j in site i, and we assume that the log-normal kriging analysis specifies the prior mean and standard deviation, V_{ij} and s_{ij}, of P_{ij}; the standard deviation was converted back to a logarithmic scale by a delta-function argument. In cases where the coefficient of P_{ij} is the same for everyone in the population, we just take $x_{ij1} = P_{ij}$ and then α_1 is the coefficient. For subgroup analyses (e.g. four subgroups, group 1 consisting of individuals who do not take beta-blockers and do not have chronic lung disease or smoking, etc.) we define $x_{ijk} = P_{ij}\delta_{ijk},\ k = 1, ..., 4$ where δ_{ijk} is the indicator for individual j at site i to be in subgroup k, and in that case α_k for $k = 1, ..., 4$ is the regression coefficient of P_{ij} in subgroup k. The full model is

$$
\begin{aligned}
y_{ij} &\sim N\left[\sum_k x_{ijk}\beta_{ik}, \kappa_i^{-1}\right],\ \ j = 1, ..., n_i,\ i = 1, ..., C,\\
\kappa_i &\sim \Gamma[\gamma_\kappa, \delta_\kappa],\ i = 1, ..., C,\\
\beta_{ik} &\sim N\left[\sum_{t=1}^{T} z_{it}\alpha_{kt}, \psi_k^{-1}\right],\ \ k = 1, ..., K,\\
\psi_k &\sim \Gamma[\gamma_\psi, \delta_\psi],\ k = 1, ..., K,\\
\alpha_{kt} &\sim U[-\infty, \infty],\ k = 1, ..., K,\ t = 1, ..., T,\\
\gamma_\kappa, \delta_\kappa &\sim \Gamma[a_0, b_0],\ \ (\text{e.g. } a_0 = b_0 = 0.001)\\
\log P_{ij} &\sim N\left[\log V_{ij}, u_{ij}\right]\ \ (u_{ij} \text{ known}, = \frac{s_{ij}^2}{V_{ij}^2}),\ j = 1, ..., n_i,\ i = 1, ..., C.
\end{aligned}
$$

The joint density of all observations is:

$$\prod_{i=1}^{C}\left[\kappa_i^{n_i/2}\exp\left\{-\frac{\kappa_i}{2}\sum_j\left(y_{ij}-\sum_k x_{ijk}\beta_{ik}\right)^2\right\}\cdot\left\{\frac{\delta_\kappa^{\gamma_\kappa}}{\Gamma(\gamma_\kappa)}\kappa_i^{\gamma_\kappa-1}e^{-\delta_\kappa\kappa_i}\right\}\cdot\right.$$
$$\cdot\prod_{k=1}^{K}\left\{\psi_k^{1/2}\exp\left(-\frac{\psi_k}{2}\left(\beta_{ik}-\sum_t z_{it}\alpha_{kt}\right)^2\right)\right\}\cdot$$
$$\left.\cdot\prod_{j=1}^{n_i}\left\{\frac{1}{P_{ij}}\exp\left(-\frac{1}{2u_{ij}}(\log P_{ij}-\log V_{ij})^2\right)\right\}\right]\cdot$$
$$\cdot\gamma_\kappa^{a_0-1}e^{-b_0\gamma_\kappa}\cdot\delta_\kappa^{a_0-1}e^{-b_0\delta_\kappa}.$$

Define $\boldsymbol{\alpha}_k$ to be the vector of $\alpha_{kt},\ t=1,...,T$, $\boldsymbol{\beta}_{i\cdot}$ to be the vector of $\beta_{ik},\ k=1,...,K$ for fixed i, $\boldsymbol{\beta}_{\cdot k}$ to be the vector of $\beta_{ik},\ i=1,...,C$ for fixed k, $\mathbf{z}_i$ to be the vector of $z_{it},\ t=1,...,T$, and $\mathbf{x}_{ij}$ to be the vector of $x_{ijk},\ k=1,...,K$ for fixed i and j. All these are column vectors. Also let Z be the matrix of z_{it} ($C\times T$), A be the matrix of α_{kt} ($K\times T$), and let Ψ be the $K\times K$ diagonal matrix with diagonal entries $(\psi_1,...,\psi_K)$.

The conditional distributions are:

$$\boldsymbol{\beta}_{i\cdot}\mid\text{rest}\sim N\left[\left(\kappa_i\sum_j\mathbf{x}_{ij}\mathbf{x}_{ij}^T+\Psi\right)^{-1}\left(\kappa_i\sum_j\mathbf{x}_{ij}y_{ij}+\Psi A\mathbf{z}_i\right),\left(\kappa_i\sum_j\mathbf{x}_{ij}\mathbf{x}_{ij}^T+\Psi\right)^{-1}\right],$$
$$\boldsymbol{\alpha}_k\mid\text{rest}\sim N\left[(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{\cdot k},(\psi_kZ^TZ)^{-1}\right],$$
$$\kappa_i\mid\text{rest}\sim\Gamma\left[\gamma_\kappa+\frac{n_i}{2},\delta_\kappa+\frac{1}{2}\sum_j\left(y_{ij}-\sum_k x_{ijk}\beta_{ik}\right)^2\right],$$
$$\psi_k\mid\text{rest}\sim\Gamma\left[\gamma_\psi+\frac{C}{2},\delta_\psi+\frac{1}{2}\sum_i\left(\beta_{ik}-\sum_t z_{it}\alpha_{kt}\right)^2\right],$$
$$\delta_\kappa\mid\text{rest}\sim\Gamma\left(a_0+C\gamma_\kappa,b_0+\sum_i\kappa_i\right).$$

This covers the "Gibbs sampling" part of the solution. The other elements must be updated by Metropolis sampling. In the case of P_{ij}, we propose the following algorithm. First note that in the subgroups analysis with four subgroups, each P_{ij} is one of $x_{ijk},\ k=1,...,4$ depending on which subgroup contains individual j of site i; without loss of generality, let us assume $P_{ij}=x_{ij1}$ for a particular (i,j). Then

$$f(x_{ij1})=\exp\left\{-\frac{\kappa_i}{2}\left(y_{ij}-\sum_k x_{ijk}\beta_{ik}\right)^2\right\}\cdot\exp\left\{-\frac{1}{2u_{ij}}(\log x_{ij1}-\log V_{ij})^2\right\}.$$

At step (i,j) define x'_{ij1} by replacing x_{ij1} with $x'_{ij1}=x_{ij1}e^{\Delta(U-\frac{1}{2})}$ where U is uniform on $[0,1]$ and Δ is arbitrary; accept x'_{ij1} with probability $\min\left\{\frac{f(x'_{ij1})}{f(x_{ij1})},1\right\}$, otherwise keep x_{ij1} at its present value until the next iteration.

For γ_κ, define

$$g(\gamma_\kappa) = \frac{\left(\prod_i \kappa_i\right)^{\gamma_\kappa} \delta_\kappa^{C\gamma_\kappa}}{\Gamma(\gamma_\kappa)^C} \cdot \gamma_\kappa^{a_0} e^{-b_0\gamma_\kappa}.$$

Based on current γ_κ define new $\gamma'_\kappa = \gamma_\kappa e^{\Delta'(U-\frac{1}{2})}$ where U is uniform on $[0,1]$ and Δ' is arbitrary; accept γ'_κ with probability $\min\left\{\frac{g(\gamma'_\kappa)}{g(\gamma_\kappa)}, 1\right\}$.

We could treat γ_ψ, δ_ψ in a similar manner to $\gamma_\kappa, \delta_\kappa$ (i.e. defining a prior distribution using hyperparameters) but we prefer not to for the following reason: the κ_i are exchangeable (they represent equivalent measurements taken in different exam sites, but we believe the sites are similar) so it makes sense to estimate a distribution across sites that could, for example, be used to predict responses at a new site should one ever be added to the dataset. The same argument does not apply to the ψ_k parameters, which represent qualitatively different covariates. It is doubtful that this distinction has much impact on the results, but we did find in preliminary analysis that some care was needed in handling γ_ψ and δ_ψ. For the following analysis we took $\gamma_\psi = 0.01, \delta_\psi = 10^{-6}$. The value of γ_ψ ensures that the prior distribution is highly diffuse, while our choice of δ_ψ then ensures that the prior mean of the ψ_k is about 10^4. That would fit in with the fact that the standard errors of the α parameters, estimated through a conventional regression approach, are of the order of 0.01 ($= 1/\sqrt{10^4}$).

The values of Δ and Δ' are arbitrary but we took $\Delta = 5$ for the measurement error analysis with $M = 1$ and $\Delta = 10$ for the measurement error analysis with $M = 2$. These choices were guided by the criteria of Gelman *et al.* (1996) for optimal acceptance rates in Hasting-Metropolis sampling. We took $\Delta' = 1$.

B. CONVERGENCE DIAGNOSTICS FOR MCMC

In order to test the convergence of the MCMC algorithms, the preceding analysis was repeated four times, with 50,000 MCMC iterations in each trial. To save overall computing time, this was not done for all the different analyses, but only for the one that seems of greatest interest: taking PM_{10} as the pollution covariate of interest, and $M = 1$.

The test procedures of Gelman and Rubin (1992) and Brooks and Gelman (1998) require that multiple runs be conducted from starting values that are overdispersed relative to the stationary distribution of the Markov chain. In the present analysis, taking into account that the greatest uncertainty seems to be in the posterior distributions of the PM_{10} values themselves, this was achieved by multiplying the initial PM_{10} estimates by four multiplicative factors — 0.2, 1, 5 and 25 — as starting values for the MCMC procedure, and also varying the seed of the random number generator. In all other respects, the four MCMC simulations, of 50,000 iterations each, were identical. All computations were calculated using CODA (Plummer *and others* (2007)), which is available as a downloadable package for the R programming language (R Development Core Team (2007)).

The parameters of primary interest are α_k for $k = 1, 2, 3, 4$, which were generated through the sequence of conditional distributions

$$\alpha_k \mid \text{rest} \sim N\left[(Z^T Z)^{-1} Z^T \boldsymbol{\beta}_{\cdot k}, (\psi_k Z^T Z)^{-1}\right], \tag{B.1}$$

where $\boldsymbol{\beta}_{\cdot k}$ is the set of $\beta_{i,k},\ i = 1, ..., C$ coefficients associated with the k'th covariate and ψ_k is the corresponding scale parameter.

The calculations were subdivided as follows:

	$k=1$	$k=2$	$k=3$	$k=4$
Run 1, Estimate	–1.45	–0.50	–5.45	–0.30
Run 2, Estimate	–1.41	–0.44	–5.72	–0.54
Run 3, Estimate	–1.36	–0.57	–5.88	–0.92
Run 4, Estimate	–1.41	–0.51	–5.70	–0.55
Run 1, 95% CI	(–1.51,–1.39)	(–0.77,–0.22)	(–5.78,–5.12)	(–0.76,0.17)
Run 2, 95% CI	(–1.48,–1.35)	(–0.65,–0.24)	(–6.11,–5.32)	(–1.11,0.03)
Run 3, 95% CI	(–1.43,–1.28)	(–0.80,–0.34)	(–6.18,–5.57)	(–1.38,–0.47)
Run 4, 95% CI	(–1.48,–1.34)	(–0.75,–0.26)	(–6.05,–5.35)	(–1.10,–0.01)

Table 7. Estimates and 95% confidence intervals for the posterior means of each of the α_k parameters, $k = 1, 2, 3, 4$, based on four parallel MCMC runs of length 50,000, discarding the first 20,000 iterations as burn-in, and using the Heidelberger-Welch procedure to take account of autocorrelation. All results have been multiplied by 10^4 for ease of numerical display.

B.1 *Posterior mean of* α_k

According to Equation (B.1), the conditional posterior mean of α_k, given all the values of β_{ik}, is $(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{.k}$. Unconditionally, the posterior mean may be calculated by averaging this quantity over all MCMC iterations, after discarding initial burn-in iterations. Therefore, for each of $k = 1, 2, 3, 4$, $(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{.k}$ was generated for each iteration of the MCMC and for each of the four replications, and subsequent analyses was based on the resulting time series.

The diagnostic of Gelman and Rubin (1992) is based on a "potential scale reduction factor" R, and the CODA package calculates both a median value and an upper confidence limit for R — in the present discussion, the confidence limit has been calculated as the 97.5% quantile of the distribution. A value of R that is much above 1 is taken to indicate non-convergence of the algorithm.

In Figure 3, we show a plot of the Gelman-Rubin diagnostic (function `gelman.plot` in CODA) as a function of the number of iterations for estimating the posterior means of each of the parameters $\alpha_k,\ k = 1, 2, 3, 4$. In each case, we note that there is a sharp reduction in R over the first 3,000 iterations and that it settles down to a value less than 1.2 by about the 15,000'th iteration. However, the convergence is not uniformly fast over all four parameters — in particular, for $k = 2, 3, 4$, it seems to be slower than for $k = 1$. This may reflect the inherent uncertainty of these parameters (of the four subgroups of the population, group 1 is by far the largest, therefore the group about which one would expect the most precise inferences). Guided by these results, in subsequent analyses we have discarded the first 20,000 iterations as burn-in, and calculate output statistics based on iterations 20,001 through 50,000.

When reporting results from a simulation, it is conventional to calculate both the mean of the quantity of interest (in this case, the posterior mean of α_k) and a standard error. Since simulation outputs are autocorrelated, however, it is necessary to correct for the autocorrelation. We have used the method of Heidelberger and Welch (1981), which is based on a nonparametric estimate of the spectral density at zero frequency, and is implemented in CODA through the function `spectrum0`, after discarding the first 20,000 iterations as previously described. This procedure leads to the point estimates and 95% confidence intervals for the posterior means given in Table 7.

From the table, we can draw some conclusions about how accurately the posterior means have been estimated. Clearly α_1 is the best estimated, all four confidence intervals within the range –1.51 to –1.28. For α_3, the range is wider (–6.18 to –5.12) also much further from 0 — in both cases, it is clear that the posterior mean is less than 0 (consistent with an adverse effect of PM on RR). For α_2 and α_4, the confidence intervals are much wider. Indeed for α_4, we cannot even state the sign of the posterior mean.

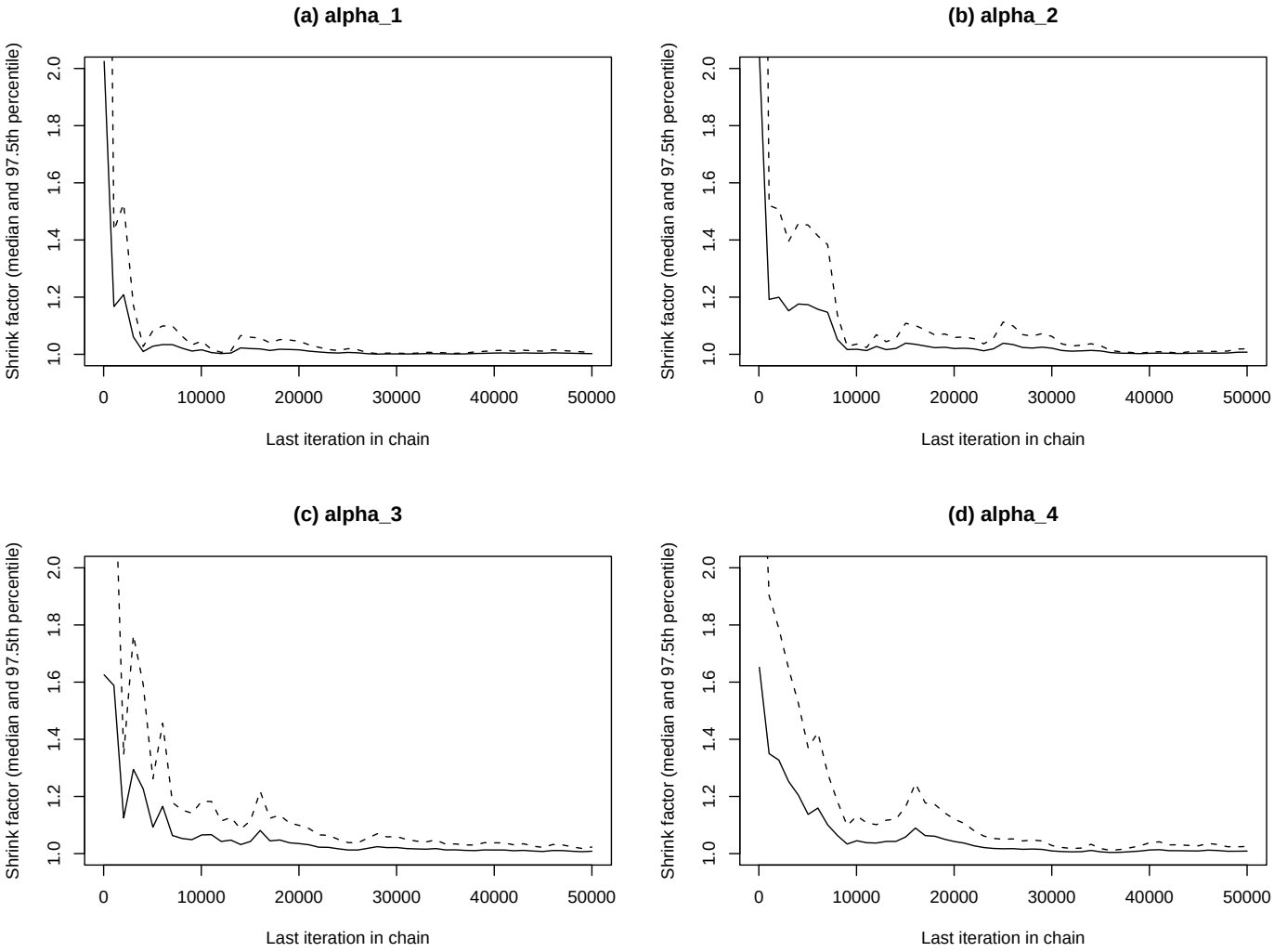

Fig. 3. Gelman-Rubin diagnostics for convergence of MCMC. The solid curve represents the median value of Gelman-Rubin's R statistic; the dashed curve is the 97.5% quantile. All curves are for $M = 1$ and PM_{10} as the pollutant of interest.

	$k=1$	$k=2$	$k=3$	$k=4$
Run 1, Estimate	0.963	0.640	0.9958	0.542
Run 2, Estimate	0.962	0.637	0.9993	0.580
Run 3, Estimate	0.953	0.659	0.9996	0.629
Run 4, Estimate	0.957	0.631	0.9988	0.576
Run 1, 95% CI	(0.955,0.972)	(0.573,0.707)	(0.987,1.004)	(0.481,0.603)
Run 2, 95% CI	(0.955,0.969)	(0.571,0.703)	(0.999,1)	(0.510,0.649)
Run 3, 95% CI	(0.943,0.962)	(0.592,0.726)	(0.999,1)	(0.567,0.691)
Run 4, 95% CI	(0.949,0.965)	(0.562,0.701)	(0.997,1.001)	(0.507,0.644)

Table 8. Estimates and 95% confidence intervals for the posterior probability that $\alpha_k < 0$ for $k = 1, 2, 3, 4$.

B.2 *Posterior probability that* $\alpha_k < 0$

According to Equation (B.1), conditionally on ψ_k and all the values of β_{ik}, the posterior probability that $\alpha_k < 0$ is

$$\Phi\left(-\frac{(Z^TZ)^{-1}Z^T\boldsymbol{\beta}_{\cdot k}}{\sqrt{(Z^TZ)^{-1}/\psi_k}}\right) \tag{B.2}$$

with $\Phi(\cdot)$ the standard normal distribution function (note that Z^TZ is scalar in this instance, so Equation (B.2) makes sense). Therefore, we can calculate the unconditional posterior probability that $\alpha_k < 0$ by averaging Equation (B.2) over all MCMC runs, discarding burn-in.

For the analysis, the quantity Equation (B.2) was calculated for each iteration, and the resulting values analyzed using the same tests as in Section B.1. The Gelman-Rubin test easily confirmed that a burn-in of 20,000 iterations is adequate for convergence of the MCMC. The Heidelberger-Welch test was applied to find confidence intervals for the desired probabilities, with results in Table 8.

As a result of this calculation, we can see that the posterior probabilities that $\alpha_k < 0$ are close to 1 in the cases $k = 1$ and 3. (The slightly anomalous confidence intervals for $k = 3$ are probably explained by the non-normality of the posterior distribution of $\Pr\{\alpha_k < 0\}$.) Conversely, for $k = 2$ and $k = 4$, none of the boundaries of the confidence intervals are very close to 0 or 1. Despite the obvious ambiguities that remain, it seems safe to say that α_1 and α_3 are statistically significant (< 0), while α_2 and α_4 are not.

B.3 *Posterior densities*

Figure 4 shows the posterior density of $\alpha_k,\ k = 1, 2, 3, 4$ computed from each of the four runs. The four runs are in excellent agreement for $k = 1$; less so for $k = 2, 3, 4$, though the agreement is still good. Based on these posterior densities, from each run a 95% equal-tailed credible interval is computed for each α_k; see Table 9. These results reinforce that there is strong evidence that both $\alpha_1 < 0$ and $\alpha_3 < 0$, but for both α_2 and α_4, all four intervals cover 0 and therefore do not indicate a significant result. On the other hand, even for α_1 and α_3, the 95% prediction intervals are wide relative to the magnitudes of the posterior means, implying that even though we have high confidence that these parameters are < 0, it would still not be possible to make precise risk calculations based on their numerical values.

B.4 *Summary*

The Gelman-Rubin diagnostics indicate that reasonable convergence has been achieved by iteration 20,000 at latest. The other statistics show generally good but not perfect agreement across the four runs of the

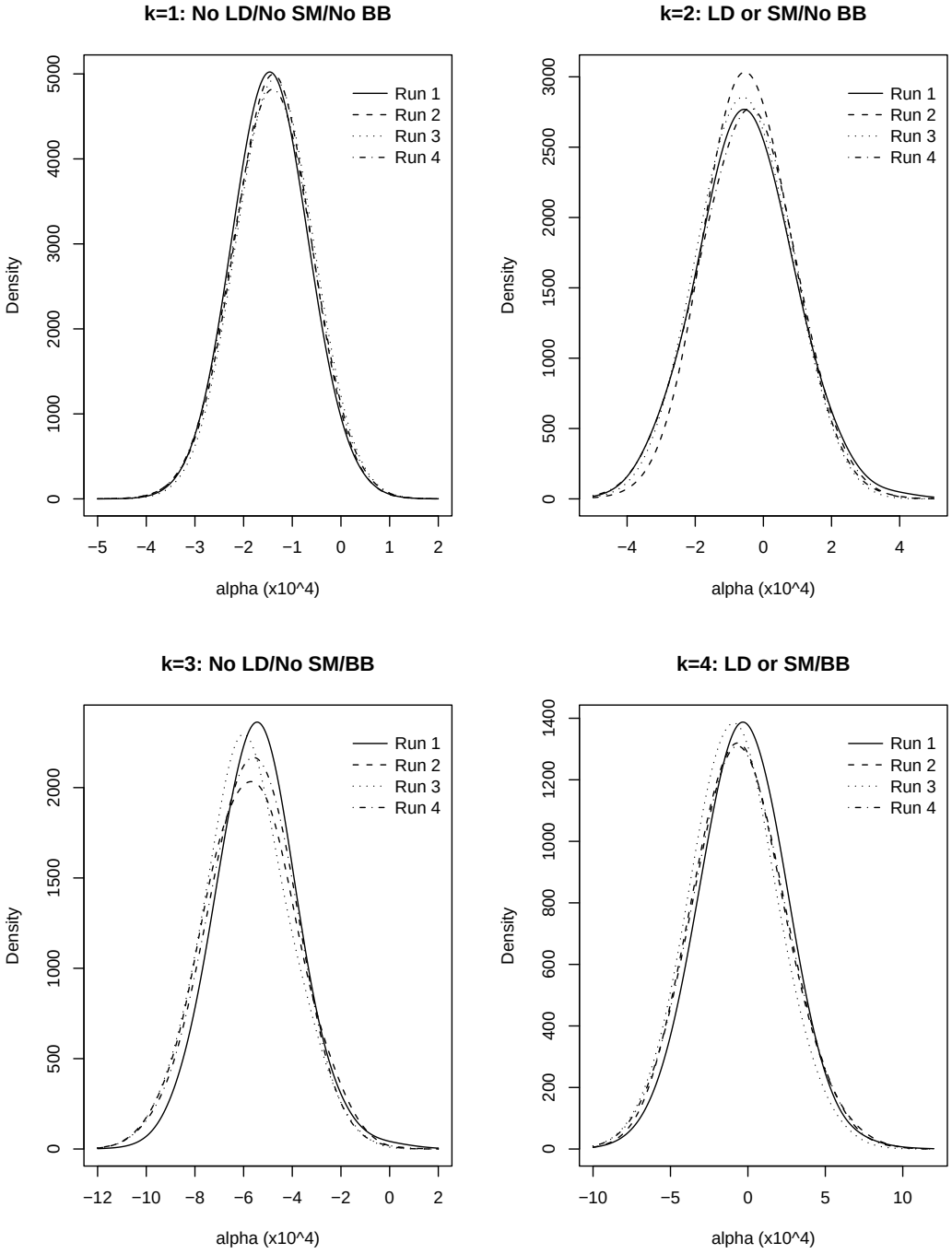

Fig. 4. Posterior densities for PM_{10} regression coefficient by subgroup: 4 runs of MCMC.

	$k=1$	$k=2$	$k=3$	$k=4$
1	(–3.02,0.14)	(–3.36,2.48)	(–8.85,–1.85)	(–6.02,5.47)
2	(–3.03,0.14)	(–2.98,2.20)	(–9.45,–1.92)	(–6.30,5.73)
3	(–2.94,0.23)	(–3.23,2.10)	(–9.44,–2.23)	(–6.61,4.94)
4	(–3.05,0.20)	(–3.37,2.27)	(–9.38,–2.20)	(–6.52,5.58)

Table 9. 95% credible intervals for $\alpha_k,\ k=1,2,3,4$, computed from each of four MCMC runs.

MCMC. We have very strong reason to believe that both α_1 and α_3 are negative (indicating a detrimental effect of PM_{10} in the two subgroups consisting of non-smokers without chronic lung disease) but the prediction intervals for the parameters themselves are still wide relative to the respective posterior means.

C. SIMULATIONS

Our results raise the questions of whether the narrowing and movement toward the origin of the slope parameter with increasing M is a general feature of Bayesian lognormal measurement error models, and if so, how much measurement error is required to create the effect. To explore these questions we have built an extremely simple univariate linear regression model with predictor point estimates of $V = \{1, ..., 100\}$ and response values generated from $y_i \sim \mathcal{N}[V_i, 1],\ i = 1, ..., 100$. We assume for simplicity that measurement error is of the form $s_i = M \times V_i,\ i = 1, ..., 100$, where M is a constant chosen to parameterize the magnitude of measurement error. We choose this form because it closely approximates the relationship between the real point estimates and standard errors in the EEAWHI data ($M \approx 0.54$).

The statistical model is chosen to be a simplified version of the one used body of this paper:

$$\begin{aligned} y_i &\sim \mathcal{N}\left[\beta_0 + x_i\beta_1, \tau^{-1}\right],\ i = 1, ..., 100 \\ \log(x_i) &\sim \mathcal{N}\left[\log(V_i), s_i/V_i\right],\ i = 1, ..., 100, \end{aligned} \tag{C.1}$$

with prior distributions chosen roughly as before:

$$\begin{aligned} \pi(\beta_j) &\propto 1,\ j = 0, 1 \\ \tau &\sim \mathcal{G}\left[0.01, 0.01\right]. \end{aligned} \tag{C.2}$$

We are most interested in how the posterior distributions of the slope and precision parameters (β_1 and τ) are affected by increasing M, though for completeness we will not ignore the intercept β_0. When measurement error is ignored (*i.e.*, $M = 0$) we have the standard linear model with the posterior distributions of the slope and intercept tightly focused around 1 and 0 respectively. As shown in figure 5, deviation from this "ignored-error model" becomes apparent around $M = 0.5$ as the distributions widen, then move toward the "intercept-only model" values of 0 and $\bar{y} \equiv 100^{-1}\sum_{i=1}^{100} y_i = 50.5$ respectively, and then finally narrow again. The distributions are most diffuse around $M = 1.25$. By $M = 2.5$ the slope posterior is very sharply peaked around 0 and the intercept has posterior mean of 47.8 and standard deviation of 3.4. These distributions do not change appreciably for $M > 2.5$. The precision parameter τ behaves in a similar way; when M is low it is distributed around 0.8^{-2}, but by $M = 2.5$ it becomes very tightly focused near the inverse of the sample variance of the y_i's. When viewed on the log-scale, it too passes through a wide intermediate phase around $M \approx 1$.

So far it is unclear if these results are specific to this particular simulated data set or if they are more general. Therefore we are interested in what happens when the predictor point estimates are subject to a

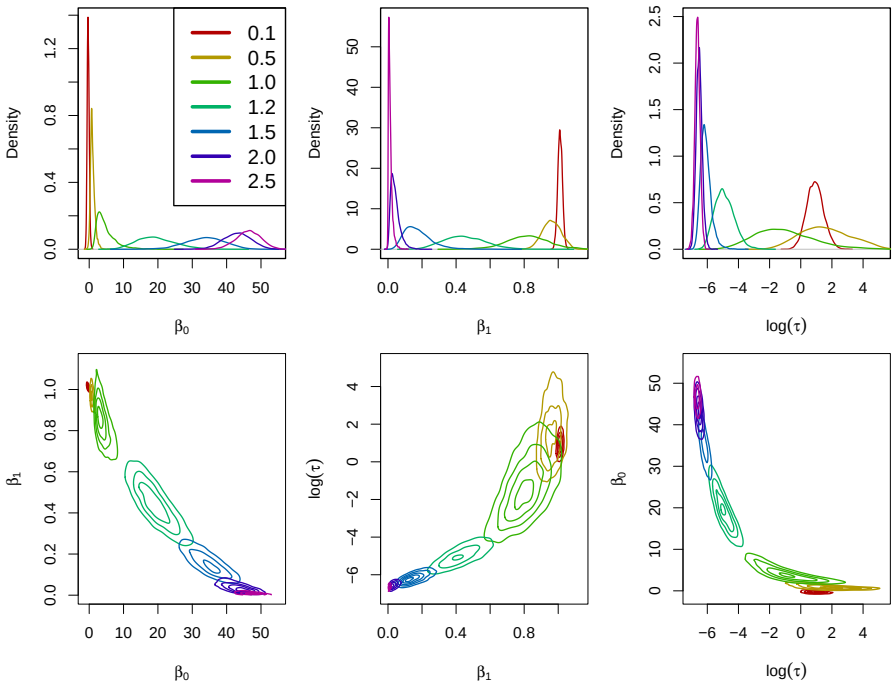

Fig. 5. The effect of varying M from 0.1 to 2.5

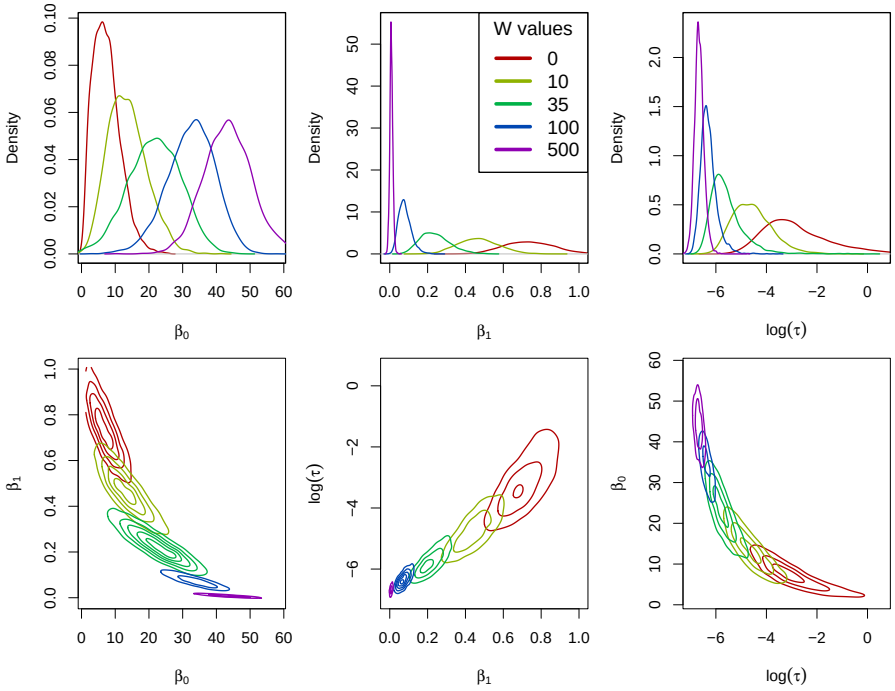

Fig. 6. The effect of moving x values to the right by a given amount W when $M = 1$ while holding s/V constant

linear transformation while we hold either the s_i values or the s_i/V_i ratios constant. Holding the latter constant while multiplying the x_i-values by a scalar has no effect on the posterior distributions besides a trivial rescaling of the slope parameter. Other transformations, though, are more interesting.

Figure 6 illustrates what happens when the V_i's are shifted to the right by an amount W (so that they take the values $\{W+1, ..., W+100\}$) while s_i/V_i is held constant. Holding s_i/V_i constant causes the variance of the lognormal distribution to scale like V_i^2, and increasing the error variance means the predictor values are less informative. Thus, as W is increased the posterior distributions approach their intercept-only forms. Looked at another way, the posterior estimates are more sensitive to measurement error when W is large. Indeed, this effect is so strong that even overwhelms the tendency for the intercept distribution to move with $-W$.

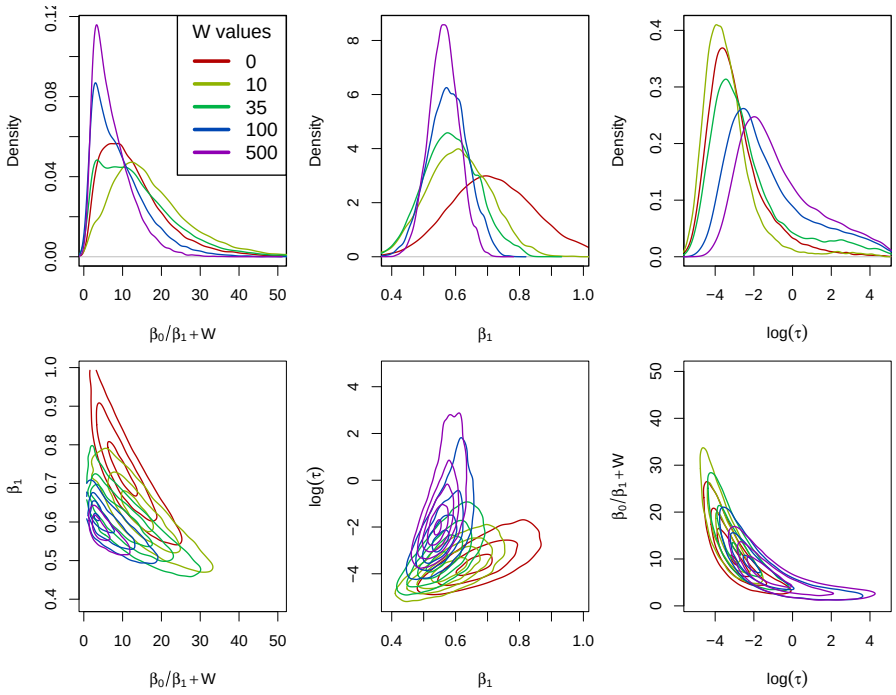

Fig. 7. The effect of shifting the V values by a given factor W while holding s constant

When the s_i's alone are held constant, on the other hand, we get quite different behavior, as shown in figure 7. As W grows the slope and rescaled x-intercept $\beta_0/\beta_1 + W$ (which should be invariant under the ignored-error model) decrease and sharpen while the precision increases and widens. However, these changes are relatively small, and they also saturate as W grows larger than the width of the data. They are due to the changing shape of the lognormal distribution as the shape parameter (here s/V) varies.

Figure 8 displays the results when the predictor point estimates are scaled by a constant Z (yielding $V = \{Z, ..., 100Z\}$) while the s_i's are left unchanged. The slope is shown after being scaled by Z to divide out the effect of stretching alone. Unlike the shift example with the s_i's constant, here the transformation has a large impact. The parameters converge to their ignored-error values as Z increases and to their intercept-only values when Z decreases. Therefore multiplying the V_i's by Z causes the same changes (given a slope rescaling) as the first example above with $M = Z^{-1}$.

These examples taken together imply that a linear transformation of the V_i's causes qualitative changes

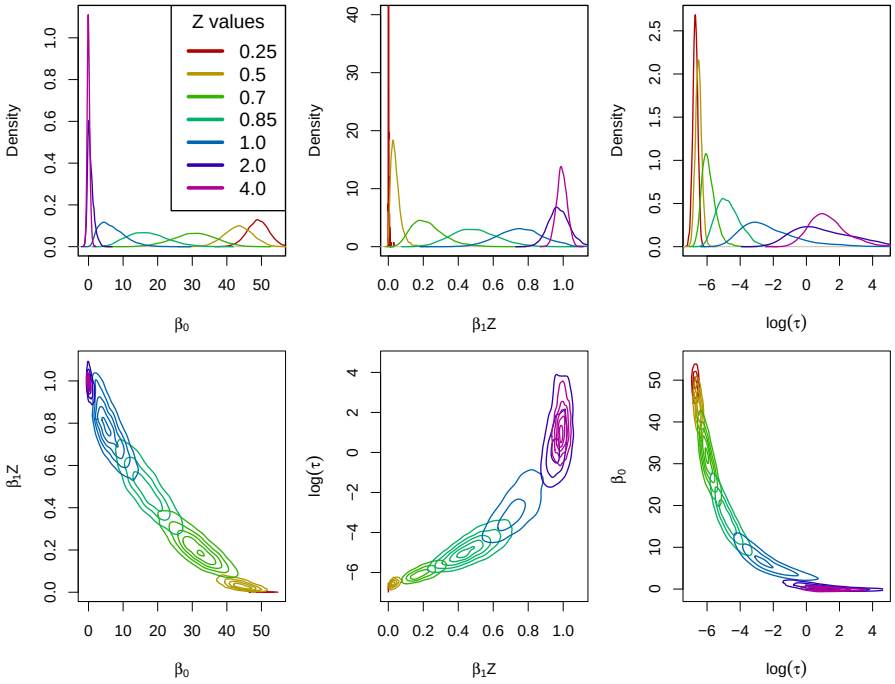

Fig. 8. The effect of stretching and contracting the V values by a given factor Z while holding s constant

in posterior distributions when the s_i's are changed relative to the width of the whole dataset, not to the V_i's. In addition, minor changes can occur when the point estimates are shifted but the standard errors are fixed.

Finally, we examine the effect of stretching and contracting the y_i-values by a factor Q while leaving the s_i and V_i values unchanged. We need not consider the effect of an additive vertical change because the absolute position of the y_i's does not affect the posterior distributions except for a trivial displacement of the intercept. Figure 9 displays the results; the parameters have been scaled by powers of Q to separate the effect of stretching alone from the interaction of stretching with measurement error. Increasing Q decreases the relative vertical errors, but, modulo rescaling of the slope and intercept, growing it past 1 has no effect except a trivial shift of the precision. However, decreasing Q far below unity causes the vertical error to overwhelm the effects of both the trend in the data and the measurement error. Thus, the slope and intercept become less certain and the precision cannot grow above 1.

ACKNOWLEDGEMENTS

Conflict of Interest: None declared.

The authors published their preliminary findings as an abstract Crooks *and others* (2008) and acknowledge the contributions of WHI Investigators in the:

Program Office (National Heart, Lung, and Blood Institute, Bethesda, Maryland) Elizabeth Nabel, Jacques Rossouw, Shari Ludlam, Linda Pottern, Joan McGowan, Leslie Ford, and Nancy Geller.

Clinical Coordinating Center (Fred Hutchinson Cancer Research Center, Seattle, WA) Ross Prentice, Garnet Anderson, Andrea LaCroix, Charles L. Kooperberg, Ruth E. Patterson, Anne McTiernan; (Wake

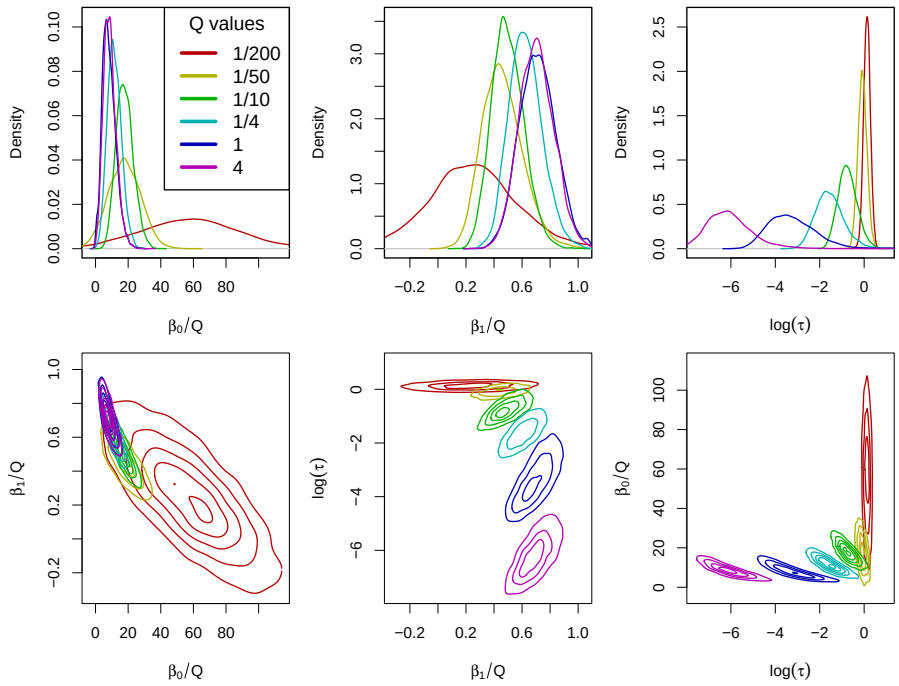

Fig. 9. The effect of stretching and contracting the y values by a given factor while holding s and V constant

Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker; (Medical Research Labs, Highland Heights, KY) Evan Stein; (University of California at San Francisco, San Francisco, CA) Steven Cummings.

Clinical Centers (Albert Einstein College of Medicine, Bronx, NY) Sylvia Wassertheil-Smoller; (Baylor College of Medicine, Houston, TX) Aleksandar Rajkovic; (Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn Manson; (Brown University, Providence, RI) Annlouise R. Assaf; (Emory University, Atlanta, GA) Lawrence Phillips; (Fred Hutchinson Cancer Research Center, Seattle, WA) Shirley Beresford; (George Washington University Medical Center, Washington, DC) Judith Hsia; (Harbor-UCLA Research and Education Institute, Torrance, CA) Rowan Chlebowski; (Kaiser Permanente Center for Health Research, Portland, OR) Evelyn Whitlock; (Kaiser Permanente Division of Research, Oakland, CA) Bette Caan; (Medical College of Wisconsin, Milwaukee, WI) Jane Morley Kotchen; (MedStar Research Institute/Howard University, Washington, DC) Barbara V. Howard; (Northwestern University, Chicago/Evanston, IL) Linda Van Horn; (Rush Medical Center, Chicago, IL) Henry Black; (Stanford Prevention Research Center, Stanford, CA) Marcia L. Stefanick; (State University of New York at Stony Brook, Stony Brook, NY) Dorothy Lane; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Alabama at Birmingham, Birmingham, AL) Cora E. Lewis; (University of Arizona, Tucson/Phoenix, AZ) Tamsen Bassford; (University at Buffalo, Buffalo, NY) Jean Wactawski-Wende; (University of California at Davis, Sacramento, CA) John Robbins; (University of California at Irvine, CA) F. Allan Hubbell; (University of California at Los Angeles, Los Angeles, CA) Lauren Nathan; (University of California at San Diego, LaJolla/Chula Vista, CA) Robert D. Langer; (University of Cincinnati, Cincinnati, OH) Margery Gass; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Hawaii, Honolulu, HI) David Curb; (University of Iowa, Iowa City/Davenport, IA) Robert Wal-

lace; (University of Massachusetts/Fallon Clinic, Worcester, MA) Judith Ockene; (University of Medicine and Dentistry of New Jersey, Newark, NJ) Norman Lasser; (University of Miami, Miami, FL) Mary Jo O'Sullivan; (University of Minnesota, Minneapolis, MN) Karen Margolis; (University of Nevada, Reno, NV) Robert Brunner; (University of North Carolina, Chapel Hill, NC) Gerardo Heiss; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (University of Tennessee, Memphis, TN) Karen C. Johnson; (University of Texas Health Science Center, San Antonio, TX) Robert Brzyski; (University of Wisconsin, Madison, WI) Gloria E. Sarto; (Wake Forest University School of Medicine, Winston-Salem, NC) Mara Vitolins; (Wayne State University School of Medicine/Hutzel Hospital, Detroit, MI) Susan Hendrix.

FUNDING The National Institute of Environmental Health Sciences funded this ancillary study (5-R01-ES012238). The National Heart, Lung and Blood Institute, U.S. Department of Health and Human Services funded the WHI program.

REFERENCES

BELANGER, C.F., HENNEKENS, C.H., ROSNER, B., AND SPEIZER, F.E. (1978). The nurses' health study. *The American Journal of Nursing* **78**, 1039-1040.

BROOK, R. D., FRANKLIN, B., CASCIO, W., HONG, Y. L., HOWARD, G., LIPSETT, M., LUEPKER, R., MITTLEMAN, M., SAMET, J., SMITH, S., TAGER, I.2004 Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation***109**, 2655-2671.

BROOKS, S.P., AND GELMAN, A.(1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics***7**, 434-455.

COZZI, E., HAZARIKA, S., STALLINGS, H.W., CASCIO, W.E., DEVLIN, R.B., LUST, R.M., WINGARD, C.J., AND VAN SCOTT, M.R. (2006). Ultrafine particulate matter exposure augments ischemia-reperfusion injury in mice. *American Journal of Physiology - Heart and Circulatory Physiology* **291**, 894-903.

CROOKS, J. L., WHITSEL, E. A., QUIBRERA, P. M., CATELLIER, D. J., LIAO, D., SMITH, R. L..2008 The effect of ignoring interpolation error on the inferred relationship between ambient particulate matter exposure and median RR interval in post-menopausal women. *Epidemiology*, in press.

DEVLIN, R.B., GHIO, A.J., KEHRL, H., SANDERS, G., AND CASCIO, W. (2003). Elderly humans exposed to concentrated air pollution particles have decreased heart rate variability. *European Respiratory Journal* **21**, 78s–80s.

DIEZ ROUX, A. V., A. H. AUCHINCLOSS, T. G. FRANKLIN, T. RAGHUNATHAN, R. G., BARR, J. KAUFMAN, B. ASTOR, AND J. KEELER.(2008) Long-term exposure to ambient particulate matter and prevalence of subclinical atherosclerosis in the Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol* **167**, 667-675.

DOMINICI, F., ZEGER, S.L. AND SAMET, J.M. (2000). Combining evidence on air pollution and daily mortality from the largest 20 US cities: A hierarchical modeling strategy (with discussion). *Journal of the Royal Statistical Society, Series A* **163**, 263-302.

DOMINICI, F., ZEGER, S.L., AND SAMET, J.M. (2000). A measurement error model for time-series studies of air pollution and mortality. *Biostatistics* **1**, 157-175.

DOMINICI, F., MCDERMOTT, A., DANIELS, M., ZEGER, S.L., AND SAMET, J.M. (2003). Mortality among residents of 90 cities. *Revised Analyses of the National Morbidity, Mortality and Air Pollution Study, Part II* Health Effects Organization, Cambridge, MA, 9-24.

DOMINICI, F., MCDERMOTT, A., ZEGER, DANIELS, M., S.L., AND SAMET, J.M. (2005). Revised Analysis of the National Morbidity, Mortality and Air Pollution Study: Mortality Among Residents of 90 Cities. *Journal of Toxicologicy and Environmental health* **68**, 1071–1092.

DOMINICI, F., PENG, D., BELL, M., PHAM, M., MCDERMOTT, A., ZEGER, S.L., AND SAMET, J.M. (2006). Fine Particulates Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. *Journal of the American Medical Association* **295**, 1127–1135.

DONALDSON, K., STONE, V., SEATON, A., AND MACNEE, W. (2001). Ambient Particle Inhalation and the Cardiovascular System: Potential Mechanisms . *Environmental Health Perspectives* **109**, 523–527.

U.S. ENVIRONMENTAL PROTECTION AGENCY. TECHNOLOGY TRANSFER NETWORK. AIR QUALITY SYSTEM (AQS). Women's Health Initiative (WHI) Study Group. Volume 2 (Section 13): ECG procedures. WHI Clinical Coordinating Center, Fred Hutchinson Cancer Research Center, Seattle WA, 1994. Available at `http://www.epa.gov/ttn/airs/airsaqs/`.

EVERSON, P.J. AND MORRIS, C.N. (2000). Inference for multivariate normal hierarchical models. *Journal of the Royal Statistical Society Series B*, **62**, 399-412.

GELMAN, A. AND RUBIN, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457-511.

GELMAN, A., ROBERTS, G.O. AND GILKS, W.R. (1996). Efficient Metropolis jimping rules. In *Bayesian Statistics 5*, Bernardo J.M., Berger J.O., Dawid A.P. and Smith A.F.M. (eds.), Oxford University Press, New York, pp. 599-607.

GILLUM, R.F., MAKUC, D.M. AND FELDMAN, J.J.(1991). Pulse Rate, Coronary Heart Disease, and Death: the NHANES I Epidemiologic Follow-up Study. *American Heart Journal* **121(1)**, 172–177.

GREENLAND, P., DAVIGLUS, M. L., DYER, A. R., LIU, K., HUANG, C. F., GOLDBERGER, J. J., STAMLER, J. (1999) Resting heart rate is a risk factor for cardiovascular and noncardiovascular mortality: the Chicago Heart Association Detection Project in Industry. *Am J Epidemiol***149(9)**, 853-862.

GRYPARIS, A., PACIOREK, C.J., AND COULL, B.A. (2006). Measurement Error Caused by Spatial Misalignment in Environmental Epidemiology. *Harvard University Biostatistics Working Paper Series* **2006-59**.

GUSTAFSON, P. (2004). Measurement Error and Misclassification in Statistics and Epidemiology. Boca Raton, Chapman and Hall-CRC.

HEALTH EFFECTS INSTITUTE (HEI). Health Review Committee and scientific staff. Understanding the health effects of components of the particulate matter mix: progress and next steps. Perspectives 2. Health Effects Institute, Boston, MA, 2002. Available at `http://pubs.healtheffects.org/getfile.php?u=244`.

HEIDELBERGER, P. AND WELCH, P.D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM* **24**, 233-245.

HOFFMANN, B., S. MOEBUS, A. STANG, E. M. BECK, N. DRAGANO, S. MOHLENKAMP, A. SCHMERMUND, M. MEMMESHEIMER, K. MANN, R. ERBEL, AND K. H. JOCKEL.2006 Residence close to high traffic and prevalence of coronary heart disease. *Eur Heart J***27**, 2696-2702.

ITO, K., THURSTON, G.D., NÁDAS, A., LIPPMANN, M. Monitor-to-monitor temporal correlation of air pollution and weather variables in the North-Central U.S. *J Expo Anal Environ Epidemiol* **11(1)**, 21-32.

ITURRIA, S., CARROLL, R.J., AND FIRST, D. (1999). Polynomial regression and estimating funtions in the presence of multiplicative measurement error. *Journal of the Royal Statistical Society* **B 61**, 547–561.

JUREK, A.M., MALDONADO, G., GREENLAND, S., CHURCH, T.R.(2006). Exposure measurement error is frequently ignored when interpreting epidemiologic study results. *European Journal of Epidemiology***21**, 871–876.

KANNEL, W.B., KANNEL, C., PAFFENBARGER, R.S., CUPPLES, L.A.(1987). Heart Rate and Cardiovascular Mortality: the Framingham Study. *American Heart Journal* **113(6)**, 1489–1498.

KAROLY, E.D., LI, Z., DAILEY, L.A., HYSENI, X., AND HUANG, Y.T. (2007). Up-regulation of Tissue Factor in Human Pulmonary Artery Endothelial Cells after Ultrafine Particle Exposure. *Environmental Health Perspectives* **115**, 535–540.

KÜNZLI, N., JERRETT, M., MACK, W.J., BECKERMAN, B., LABREE, L., GILLILAND, F., THOMAS, D., PETERS, J., AND HODIS, H.N. (2005). Ambient Air Pollution and Atherosclerosis in Los Angeles. *Environmental Health Perspectives* **113**, 201–205.

LIAO, D., CAI, J., ROSAMOND, W. D., BARNES, R. W., HUTCHINSON, R. G., WHITSEL, E. A., RAUTAHARJU, P., HEISS, G. Cardiac autonomic function and incident coronary heart disease: a population-based case-cohort study. The ARIC Study. Atherosclerosis Risk in Communities Study. *Am J Epidemiol***145(8)**, 696-706.

LIAO, D., CREASON, J., SHY, C., WILLIAMS, R., WATTS, R., AND ZWEIDINGER, R. (1999). Daily Variation of Particulate Air Pollution and Poor Autonomic Control in the Elderly. *Environmental Health Perspectives* **115**, 521–527.

LIAO, D., DUAN, Y., WHITSEL, E.A., ZHENG, Z., HEISS, G., CHINCILLI, V.M., AND LIN, H. (2004). Association of Higher Levels of Ambient Criteria Pollutants with Impaired Cardiac Autonomic Control: A Population-based Study. *American Journal of Epidemiology* **159**, 768–777.

LIAO, D., PEUQUET, D. J., DUAN, Y., WHITSEL, E. A., DOU, J., SMITH, R. L., LIN, H. M., CHEN, J. C.,

HEISS, G. GIS approaches for the estimation of residential-level ambient PM concentrations. *Environ Health Perspect***114(9)**, 1374-1380.

LIPPMANN, M., HWANG, J., MACIEJCZYK, P., AND CHEN, L. (2005). PM Source Apportionment for Short-Term Cardiac Function Changes in ApoE$^{-/-}$ Mice. *Environmental Health Perspectives* **113**, 1575–1579.

LYLES, R.H., AND KUPPER, L.L. (1997). A Detailed Evaluation of Adjustment Methods for Multiplicative Measurement Error in Linear Regression with Applications in Occupational Epidemiology. *Biometrics* **53**, 1008–1025.

NATIONAL CENTER FOR ENVIRONMENTAL ASSESSMENT (NCEA). OFFICE OF RESEARCH AND DEVELOPMENT. U.S. ENVIRONMENTAL PROTECTION AGENCY. Provisional Assessment of Recent Studies on Health Effects of Particulate Matter Exposure. Research Triangle Park, NC. July 2006. Available at `http://www.epa.gov/particles/pdfs/ord_report_20060720.pdf`.

NATIONAL RESEARCH COUNCIL (NRC). Committee on Research Priorities for Airborne Particulate Matter. Research priorities for airborne particulate matter: IV. Continuing research progress. National Academies Press, Washington, DC, 2004.

NURKIEWICZ, T.R., PORTER, D.W., BARGER, M., MILLICCHIA, L., RAO, K.M.K, MARVAR, P.J., HUBBS, A.F., CASTRANOVA, V., AND BOEGEHOLD, M.A. (2006). Systemic Microvascular Dysfunction and Inflammation after Pulmonary Particulate Matter Exposure. *Environmental Health Perspectives* **114**, 412-419.

PARK, S.K., O'NEILL, M.S., VOKONAS, P.S., SPARROW, D., AND SCHWARTZ, J. (2005). Effects of Air Pollution on Heart Rate Variability: The VA Normative Aging Study. *Environmental Health Perspectives* **113**, 304–309.

PLUMMER, M., BEST, N., COWLES, K., VINES, K.(2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R NEWS***6**, 7-11

R DEVELOPMENT CORE TEAM (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. Available from `http://www.R-project.org`.

SAMET, J.M., DOMINICI, F., CURRIERO, F., COURSAC, I., AND ZEGER, S.L. (2000). Particulate air pollution and mortality: Findings from 20 U.S. cities. *New England Journal of Medicine* **343**, 1742-1757.

SANDHU, R., PETRONI, D., AND GEORGE, W. (2005). Ambient Particulate Matter, C-Reactive Protein, and Coronary Artery Disease. *Inhalation Toxicology* **17**, 409-413.

SCHROEDER, E. B., WHITSEL, E. A., EVANS, G. W., PRINEAS, R. J., CHAMBLESS, L. E., HEISS, G. Repeatability of heart rate variability measures. *J Electrocardiol***37(3)**, 163-172.

SCHWARTZ, J., LITONJUA, A., SUH, H., VERRIER, M., ZANOBETTI, A., SYRING, M., NEARING, B., VERRIER, R., STONE, P., MACCALLUM, G., SPEIZER, F.E., AND GOLD, D.R. (2005). Traffic related pollution and heart rate variability in a panel of elderly subjects. *Thorax* **60**, 455–461.

ŞENTÜRK, D., AND MÜLLER, H. (2005). Covariate Adjusted Regression. *Biometrika* **92**, 75–89.

TASK FORCE OF THE EUROPEAN SOCIETY OF CARDIOLOGY AND THE NORTH AMERICAN SOCIETY OF PACING AND ELECTROPHYSIOLOGY. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation***93(5)**, 1043-1065.

THOMAS, D., STRAM, D., AND DWYER, J. (1993). Exposure Measurement Error: Influence on Exposure-Disease Relationships and Methods of Correction. *Annuals Reviews of Public Health* **14**, 69–93.

WHITSEL, E. A., ROSE, K. M., WOOD, J. L., HENLEY, A. C., LIAO, D., HEISS, G. Accuracy and Repeatability of Commercial Geocoding. *Am J Epidemiol***160(10)**, 1023-1029.

WHITSEL, E.A., QUIBRERA, P.M., SMITH, R.L., CATELLIER, D.J., LIAO, D., HENLEY, A.C. AND HEISS, G. (2006). Accuracy of commercial geocoding: assessment and implications. *Epidemiologic Perspectives & Innovations* **3:8**, doi:10.1186/1742-5573-3-8.

WHI STUDY GROUP. Design paper: design of the Women's Health Initiative clinical trial and observational study.

*Controlled Clin Trials***19**, 61-109.

YATERA, K., HSIEH, J., HOGG, J.C., TRANFIELD, E., SUZUKI, H., SHIH, C.-H., BEHZAD, A.R., VINCENT, R., AND VAN EEDEN, S.F. (2007). Particulate matter air pollution exposure promotes recruitment of monocytes into atherosclerotic plaques. *American Journal of Physiology - Heart and Circulatory Physiology* **294**, 944-953.

YEATTS, K., SVENDSEN, E., CREASON, J., ALEXIS, N., HERBST, M., SCOTT, J., KUPPER, L., WILLIAMS, R., NEAS, L., CASCIO, W., DEVLIN, R.B., AND PEDEN, D.B. (2007). Coarse Particlate Matter ($PM_{2.5-10}$) Affects Heart Rate Variability, Blood Lipids, and Circulating Eosinophils in Adults with Asthma. *Environmental Health Perspectives* **115**, 709–714.

ZAREBA, W., NOMURA, A., COUDERC, J. P. Cardiovascular effects of air pollution: what to measure in ECG? *Environ Health Perspect***109(S4)**, 533-538.

[*Received ??; accepted for publication ??*]