Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort: Final Report

Principal Investigator:

Michael Jerrett, PhD

Co-Investigators:

Richard T. Burnett, PhD
Arden Pope III, PhD
Daniel Krewski, PhD
George Thurston, ScD
George Christakos, PhD, ScD
Edward Hughes, PhD
Zev Ross, MS
Yuanli Shi, MD
Michael Thun, MD

Student and Postdoctoral Co-Investigators:

Bernardo Beckerman, MS
Michelle Catherine Turner, MS
Jason Su, PhD
Seung-Jae Lee, PhD

Prepared for:

Contract # 06-332
State of California Air Resources Board
Research Division
PO Box 2815
Sacramento CA95812

REVIEW DRAFT

CONTENTS

ABSTRACT **4**

LIST OF ACRONYMS AND ABBREVIATIONS **8**

LIST OF FIGURES **10**

MAIN TEXT 10
APPENDIX B 10

LIST OF TABLES **11**

MAIN TEXT 11

SIGNIFICANCE **14**

BACKGROUND AND RATIONALE **14**

THE HARVARD SIX CITIES STUDY AND AMERICAN CANCER SOCIETY STUDY OF PARTICULATE AIR POLLUTION AND MORTALITY 15
THE REANALYSIS PROJECT: OBJECTIVES AND FINDINGS 15
Geographic Scale of Analysis *16*
Refinement of Exposure Estimates *17*
Limitations of the Standard Cox Model *18*
POST-REANALYSIS STUDIES OF THE ACS COHORT 19
Exposure Time Windows *19*
DESCRIPTION OF COHORT STUDIES REPORTING $PM_{2.5}$ MORTALITY RISKS 20
American Cancer Society Cancer Prevention II Study (ACS CPS II) *20*
The National Medicare Cohort (MET) *21*
The Six Cities Study (SCS) *21*
The Six-City National Medicare Cohort (SCMC) *21*
Netherlands Cohort Study on Diet and Cancer (NLCS) *21*
Summary of Cohort Studies for All Causes of Death *22*
Summary of Cohort Studies for Specific Causes of Death *22*

MATERIALS AND METHODS **23**

EXPOSURE MODELING 23
Available Pollution Monitors *24*
Methodology for Exposure Modeling *25*
Inverse Distance Weighted (IDW) Interpolation 25
Kriging $PM_{2.5}$ 26
Data 26
Land Use Regression 27
Statistical Methods for Land Use Regression 28
Land Use and Traffic Data as Covariates 29
Roads and Traffic 30
$PM_{2.5}$ estimates derived from remotely sensed data 30
Residual Error Modeling - Bayesian Maximum Entropy 31
GEOCODING METHODS USED FOR CONVERTING INDIVIDUAL ADDRESSES TO GEOGRAPHIC COORDINATES FOR EXPOSURE ASSIGNMENT 32
Equipment/Software Details *32*
Geocoding Locators *32*
Cohort Raw Text Files from the American Cancer Society *32*
Reading Text Files and Preparing for Address Correction *32*
Address Correction *33*
Clean Up Address Correction Output and Prepare File for Geocoding *33*
Geocoding *33*
Compile Final Geocoded Data *34*

REVIEW DRAFT

STUDY POPULATION 34
Data Collection and Assembly *35*
STATISTICAL METHODS AND DATA ANALYSIS 35
The Spatial Cox Model *36*
Auxiliary Random Effects Poisson Models *37*
Orthodox Best Linear Unbiased Predictor Approach *38*
Prediction of random effects 38
Estimation of regression parameters 39
Estimation of dispersion parameters 40
Individual and Ecologic Covariates *40*
Control for Urban-Rural Differences 41
The Association between Time-Dependent Air Pollution Exposure and Mortality *41*

RESULTS 43

EXPOSURE MODELING RESULTS 43
Inverse Distance Weighted Interpolation *43*
Kriging $PM_{2.5}$ *43*
Land Use Regression *45*
Bayesian maximum entropy interpolation of land use regression residuals *51*
Combined spatiotemporal BME-LUR model *54*
GEOCODING RESULTS 60
HEALTH EFFECTS MODELING RESULTS 66
Description of the Final Analytical Dataset *66*
Descriptive Results for the Analytic Cohort 66
Causes of Death Investigated 68
ZIP Code Exposure Assignment *68*
Results from Exposure Assignments to the Geocoded Residential Address *69*
Risk Estimates for Other Pollutants *70*
Ozone 70
NO_2 71
PM_{10} and Sulfate in PM_{10} 71
Road Proximity 71
Time-Dependent Analysis Results *72*
ADDITIONAL FIGURES REFERENCED IN RESULTS 100

DISCUSSION 108

KEY FINDINGS 108
NARRATIVE 110

APPENDICES 114

APPENDIX A: ZP4 OUTPUT FIELDS 114
APPENDIX B: DERIVING ESTIMATES OF TRAFFIC FOR THE LAND USE REGRESSION MODELS 115
Summary *115*
Traffic Count Data Used to Generate Traffic Weights *115*
Temporal Trends in Traffic Volumes *115*
Assigning Traffic Values to Unmeasured Road Links *116*
Final Traffic Weight Assignment Method *118*
Specific Steps in Detail 119
Results *120*
Issues and Concerns *120*
APPENDIX C: MATHEMATICAL FOUNDATIONS OF BAYESIAN MAXIMUM ENTROPY ESTIMATION 138

REFERENCES 141

REVIEW DRAFT

ABSTRACT

Problem: Studies using the American Cancer Society (ACS) Cancer Prevention II (CPS II) cohort to assess the relation between particulate air pollution and mortality rank among the most influential and widely cited. The original study, a reanalysis that introduced new random effects methods and spatial analytic techniques, and recent studies with longer follow-up and improved exposure assignment, have all demonstrated statistically significant and substantively large air pollution effects on all-cause and cause-specific mortality. Due to this robust association and a lack of other large cohort studies on the long-term effects, the ACS studies have proven important to government regulatory interventions and health burden assessments.

At present there are no ACS CPS II statewide studies in California that investigate whether the risks are similar to or different from those reported in the above-mentioned reports. Existing estimates come from either national-level ACS studies, in which the California subjects comprise less than 10% of the total sample, or from select metropolitan or county areas of California, where questions remain about their applicability to the rest of the state. A need therefore exists to investigate whether the results hold for other parts of California or are sensitive to alternative measures of exposure assessment and confounding control. In addition, none of the existing ACS studies have used high-resolution exposure assignment or investigated the temporal dimensions of the dose-response relationship. In this study we used advanced exposure modeling to reduce problems of measurement error, and we investigated time windows of exposure.

Previous Work: Our previous work includes the original ACS study of particulate air pollution and mortality, the reanalysis of the ACS study, analytic extensions to both studies using new spatial models, and the first assessment of particulate air pollution at the within-city or "intraurban" scale. In this last investigation, our team documented the first assessment of mortality and particulate air pollution estimates for the within-city scale in Los Angeles. Our results suggest the chronic health effects associated with intraurban gradients in exposure to fine particulate matter ($PM_{2.5}$) are even larger than previously reported across metropolitan areas in the original study by Pope et al. [1] and the reanalysis by Krewski et al. [2]. For the within-city models, we observed effects nearly three times greater than in models relying on between-community exposure contrasts. These were confirmed using more refined exposure models in a subsequent Health Effects Institute report [3]. In that report, we also found risks for the national study that were greater than in earlier studies for deaths due to cardiovascular causes.

Objectives: In this context, we pursued the following research objectives: (1) to derive detailed assessments of the health effects from particulate and gaseous air pollution on all-cause and cause-specific mortality in California based on the ACS CPS II cohort, (2) to investigate whether specific particle characteristics associate with larger health effects through examination of intraurban gradients in exposure to different particle constituents and sources, and (3) to determine whether critical exposure time windows exist in the relationship between air pollution and mortality in California.

Description: We identified more than 76,000 California subjects in the ACS cohort who serve as the study population (20,432 deaths with an 18 year follow-up ending in 2000). These subjects are widely distributed across California, giving comprehensive coverage for much of the

REVIEW DRAFT

population of the state (i.e., 54 of 58 California counties have ACS subjects). For the first time in the ACS studies, we have geocoded subjects to their home address to refine our exposure assignment.

As a basis for exposure assessment, we utilized interpolation estimates derived by Air Resources Board staff for the California Teachers Cohort Study led by Dr. Michael Lipsett, with Dr. Jerrett as co-investigator. We also implemented geostatisical kriging, advanced remote sensing coupled with atmospheric modeling, land-use regression, and Bayesian models capable of assessing space-time patterns in exposure to improve exposure assignment.

We employed a comprehensive set of 20 individual risk factor variables similar to those used in previous ACS studies. These variables control for lifestyle, dietary, demographic, occupational, and educational influences that may confound the air pollution-mortality association. We used ecological variables in the neighborhoods of residence to control for "contextual" neighborhood confounding (e.g., unemployment). Although we used similar variables as in previous analyses to promote comparison to earlier results, we also tested other model specifications.

We assessed the association between air pollution and death using standard and multilevel Cox proportional hazards models. Control was also applied for residence in the five largest urban conurbations, which likely have different mortality rates than nonmetropolitan areas. We also assessed spatial autocorrelation in the health effect estimates.

Key Results: Below we summarize the key results from our investigation.

1. Cardiovascular disease (CVD) deaths, especially those from ischemic heart disease (IHD), are consistently and robustly associated with measures of fine particulate and traffic-related air pollution. The effects on CVD and IHD in California are virtually identical to those of the national ACS study (**see Abstract Table 1**).

Abstract Table 1: Comparison of Relative Risk Estimates from the California and National American Cancer Society Cohorts for $PM_{2.5}$ using a 10 µg/m^3 Exposure Increment

	California*		National Level**	
	Hazard Ratio	95% CI	Hazard Ratio	95% CI
All Cause	1.08	(1.00, 1.15)	1.08	(1.04, 1.11)
CVD	1.15	(1.04, 1.28)	1.17	(1.11, 1.24)
IHD	1.28	(1.12, 1.47)	1.29	(1.18, 1.40)

* California study uses residential address with a Land Use Regression estimate of exposure
**National level study uses metropolitan area of residence with the average of all $PM_{2.5}$ monitors within the metropolitan area as the exposure estimate; source for the National estimate for all-cause and IHD from Krewski et al. 2009 [3] Table 9; CVD estimate produced for this report for comparison with the California using the same model and sample as in the Krewski report (i.e., two level random effects, with no spatial autocorrelation – referred to as MSA and DIFF in Table 9). Note numbers slightly differ from the Krewski report due to rounding.

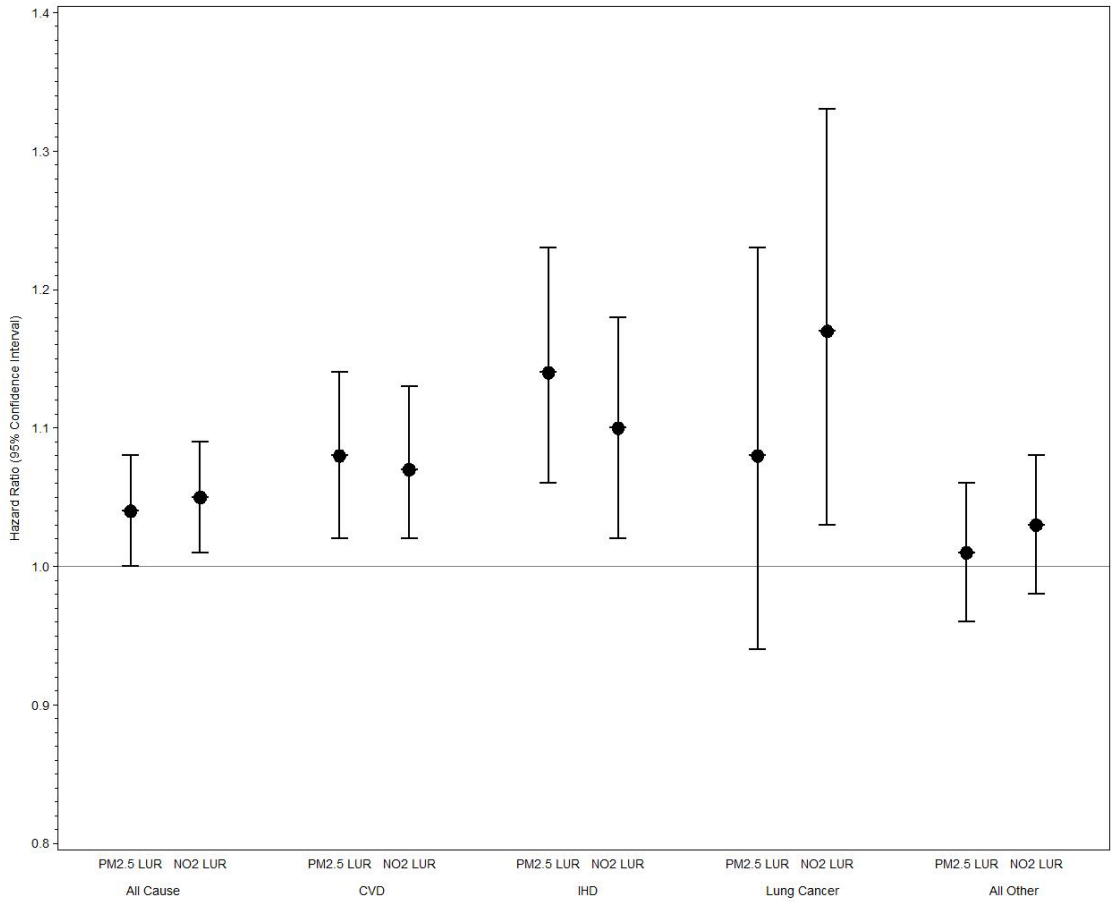
Models for both risk estimates control for individual risk factors (e.g., smoking),

REVIEW DRAFT

contextual risk factors (e.g., unemployment in area of residence) and are stratified by age, race and sex. Results for the California cohort are also additionally adjusted for place of residence in five major urban conurbations. Follow up period for both studies was from 1982-2000.

2. All-cause mortality is significantly associated with $PM_{2.5}$ exposure, but the results are sensitive to statistical model specification and to the exposure model used to generate the estimates. When we applied control for residence in the largest urban conurbations, and we employed the land use regression (LUR) model, we found significantly elevated effects on all cause mortality. The effect sizes are the same as those in the national study (see **Abstract Table 1** for a comparison). We observed effects that were of similar size, but of borderline significance when using other exposure models.

3. The strongest and most consistent effects are observed when there is finer-scale spatial resolution in the exposure predictions. All-cause mortality is associated with $PM_{2.5}$ in Los Angeles where there is a dense monitoring network with reasonably good spatial resolution and in models using the LUR estimate that serve as markers of relatively local variation in pollution (see **Abstract Figure 1** for a comparison of the risks from statewide LUR models of $PM_{2.5}$ and NO_2 for various causes of death).



Abstract Figure 1: Summary of key results for $PM_{2.5}$ and NO_2 with all-cause and cause specific death. Estimates derived from single pollutant models and calibrated to the inter-quartile range of exposure for each pollutant

4. The strongest evidence of mortality effects is with exposure models that are markers of traffic-related air pollution. The NO_2 LUR estimate has significant associations with all cause, CVD, IHD, and lung cancer deaths. Exposure estimates based on roadway proximity had elevated, but insignificant risks, suggesting weaker effects than with the NO_2 model, probably due to increased exposure measurement error.

5. With regard to other causes of death, there is no evidence of an air pollution effect. In fact for some regional $PM_{2.5}$ exposure there is some evidence of negative association, but when residence in the five largest urban conurbations was accounted for in the model, the effects become positive, but insignificant.

6. Other pollutants − namely PM_{10}, sulfate derived from PM_{10} filters, NO_2, and ozone estimates from interpolation models − all show consistent associations with CVD that are similar in size to those observed for $PM_{2.5}$. In general, the interpolation estimates of these pollutants were highly correlated with each other and with $PM_{2.5}$. Therefore caution must be exercised in interpreting effects from any single pollutant when the exposure estimate relies solely on interpolation.

CONCLUSION

Taken together, the results from this investigation indicate consistent and robust effects of $PM_{2.5}$ − and other pollutants commonly found in the combustion-source mixture with $PM_{2.5}$ − on deaths from CVD and IHD. We also found significant associations between $PM_{2.5}$ and all causes of death, although these findings were sensitive to model specification. In Los Angeles, where the monitoring network is capable of detecting intraurban variations in $PM_{2.5}$, we observed large effects on death from all causes, CVD, IHD, and respiratory disease. These results are consistent with past ACS analyses and with findings from other national or international studies reviewed in this report. Our strongest results are from a land use regression estimate of NO_2, which represents traffic sources, indicating significantly elevated effects on deaths from all causes, CVD, IHD, and lung cancer. **We conclude that combustion-source air pollution, especially from traffic, is significantly associated with premature death in this large cohort of Californians.**

REVIEW DRAFT

List of Acronyms and Abbreviations

A4 – local roads data value
ACS – American Cancer Society (or American Cancer Society Cohort)
AHSMOG – Adventist Health Study of Smog
BME – Bayesian Maximum Entropy, *BME*lib is the toolbox software used to run a BME (see section on "Residual Error Modeling")
BMI – body mass index
C/O – "Care Of", address error
CARB – California Air Resources Board
CER – cerebrovascular (hazard ratio for related death)
CPS II – Cancer Prevention Study II
CV – cross-validation
CVD – cardiovascular disease
DSA – Deletion/Substitution/Addition algorithm (see section on "Statistical Methods for Land Use Regression)
EPA or USEPA – U.S. Environmental Protection Agency
ESRI – Environmental Systems Research Institute (developers of GIS software, ArcInfo)
FCC – Feature Class Code
GEOS-Chem – chemical transport model
GINI – statistic for measuring dispersion (coefficient of income inequality)
GIS – Geographic Information System(s)
GPS – Global Positioning System
HEI – Health Effects Institute
HR – hazard ratio
IDW – inverse distance weighting
IHD – ischemic heart disease
IQR – interquartile range statistic
KRG – kriging
LA MSA – Los Angeles Metropolitan Statistical Areas
LACS – software add-on for ZP4 which converts rural- to street-style addresses
LUR – land-use regression (model)
LUR-BME – combined Land Use Regression - Bayesian Maximum Entropy model
MAUP – modifiable areal unit problem
MET – The National Medicare Cohort
MSA – metropolitan statistical area
NA – not available (no data accessible)
NDI – National Death Index
NHS – The Nurses Health Study
NLCD – USGS National Land Use Cover Dataset (2001)
NLCS – Netherlands Cohort Study on Diet and Cancer
NO_2 – nitrogen dioxide
O_3– ozone
PM – particulate matter of any diameter
PM_{10} – other (not fine) particulate matter
$PM_{2.5}$–fine particulate matter

REVIEW DRAFT

ppb – parts per billion (unit)
QA – quality assurance (audit)
RR – relative risk
S/TRF – Space/Time Random Field
SCMC – The Six City Medicare Cohort
SCS – The Six Cities Study
SES – socioeconomic status
SM – streetmap
SO_4^{-2} – sulfate
Stata 10 – statistical software, Stata Corp.
TRAPCA – Traffic Related Air Pollution and Childhood Asthma Study
USGS – United States Geological Survey
ZCA – ZIP code area
ZP4 – software that includes official United States Postal Service data (ZIP + 4 code)
µm – micro-meter unit

REVIEW DRAFT

List of Figures

Main Text

FIGURE 1: DISTRIBUTIONS OF OBSERVED POLLUTANTS: NO_2, PM_{10} AND $PM_{2.5}$ 25
FIGURE 2: HISTOGRAM OF OBSERVED $PM_{2.5}$ AVERAGES FOR 2000 27
FIGURE 3: ILLUSTRATION OF LAND USE REGRESSION METHODS USING GIS DATA 28
FIGURE 4: SEMI-VARIOGRAM USED FOR $PM_{2.5}$ KRIGING MODEL 44
FIGURE 5: KRIGING $PM_{2.5}$ ESTIMATES PLOTTED AGAINST CORRESPONDING OBSERVATIONS 45
FIGURE 6: HISTOGRAMS OF RAW AND LOG-TRANSFORMED POLLUTION OBSERVATIONS 47
FIGURE 7: CV RISK PLOTS VERSUS MODEL SIZE FOR $PM_{2.5}$ MODEL 48
FIGURE 8: CV RISK PLOTS VERSUS MODEL SIZE FOR PM_{10} MODEL 49
FIGURE 9: CV RISK PLOTS VERSUS MODEL SIZE FOR NO_2 MODEL 49
FIGURE 10: RESIDUAL PLOTS AND HISTOGRAMS OF RESIDUALS FOR SELECTED MODELS 51
FIGURE 11: SPACE/TIME EXPERIMENTAL COVARIANCE VALUES (CIRCLES) AND A COVARIANCE MODEL FITTED TO THE EXPERIMENTAL VALUES (CURVES) – NO_2 CASE 53
FIGURE 12: SPACE/TIME EXPERIMENTAL COVARIANCE VALUES (CIRCLES) AND A COVARIANCE MODEL FITTED TO THE EXPERIMENTAL VALUES (CURVES) – PM_{10} CASE 54
FIGURE 13: SPACE/TIME EXPERIMENTAL COVARIANCE VALUES (CIRCLES) AND A COVARIANCE MODEL FITTED TO THE EXPERIMENTAL VALUES (CURVES) – $PM_{2.5}$ CASE 54
FIGURE 14: COMPARISON OF PREDICTED ON OBSERVED PLOTS FOR LUR ONLY AND LUR-BME MODELS FOR NO_2, PM_{10} AND $PM_{2.5}$, RESPECTIVELY. 56
FIGURE 15: COMPARATIVE RESIDUAL PLOTS FOR LUR ONLY AND LUR-BME MODELS FOR NO_2, PM_{10} AND $PM_{2.5}$, RESPECTIVELY 58
FIGURE 16: COMPARATIVE RESIDUALS HISTOGRAMS FOR LUR ONLY AND LUR-BME MODELS FOR NO_2, PM_{10} AND $PM_{2.5}$,RESPECTIVELY 59
FIGURE 17: $PM_{2.5}$ MAP FROM THE KRIGING INTERPOLATION MODEL FOR 2000 100
FIGURE 18: OZONE MAP FROM THE INVERSE DISTANCE WEIGHTING MODEL INTERPOLATION MODEL FOR 1988-2002 AVERAGE 101
FIGURE 19: SULFATE MAP FROM THE INVERSE DISTANCE WEIGHTING MODEL INTERPOLATION MODEL FOR 1996 102
FIGURE 20: PM_{10} MAP FROM THE INVERSE DISTANCE WEIGHTING INTERPOLATION MODEL FOR 1988-2000 AVERAGE 103
FIGURE 21: NO_2 MAP FROM THE INVERSE DISTANCE WEIGHTING INTERPOLATION MODEL FOR 1988-2000 AVERAGE 104
FIGURE 22: HAZARD RATIOS AND 95% CONFIDENCE INTERVALS FOR THE ASSOCIATION BETWEEN DIFFERENT $PM_{2.5}$ INDICATORS (EACH 10 UG/M3) AT BOTH THE INDIVIDUAL AND ZIP CODE-LEVEL AND ALL CAUSE MORTALITY, FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL COVARIATES (1990), STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP) 105
FIGURE 23: HAZARD RATIOS AND 95% CONFIDENCE INTERVALS FOR THE ASSOCIATION BETWEEN DIFFERENT $PM_{2.5}$ INDICATORS (EACH 10 UG/M3) AT BOTH THE INDIVIDUAL AND ZIP CODE-LEVEL AND CARDIOVASCULAR MORTALITY, FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL COVARIATES (1990), STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP) 106
FIGURE 24: HAZARD RATIOS AND 95% CONFIDENCE INTERVALS FOR THE ASSOCIATION BETWEEN DIFFERENT $PM_{2.5}$ INDICATORS (EACH 10 UG/M3) AT BOTH THE INDIVIDUAL AND ZIP CODE-LEVEL AND ISCHEMIC HEART DISEASE MORTALITY, FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL COVARIATES (1990), STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP) 107
FIGURE 25: SUMMARY OF KEY RESULTS FOR $PM_{2.5}$ AND NO_2 WITH ALL-CAUSE AND CAUSE SPECIFIC DEATH. ESTIMATES DERIVED FROM SINGLE POLLUTANT MODELS AND CALIBRATED TO THE INTER-QUARTILE RANGE OF EXPOSURE FOR EACH POLLUTANT 110

Appendix B

APPENDIX B - FIGURE 1: HISTOGRAM OF NUMBER OF AVAILABLE TRAFFIC COUNTS YEAR 121
APPENDIX B - FIGURE 2: SCATTERPLOT OF TRAFFIC VOLUMES BY YEAR WITH OVERLAID STATISTICAL MEASURES FOR ALL ROAD TYPES 122
APPENDIX B - FIGURE 3: SCATTERPLOT OF TRAFFIC VOLUMES BY YEAR WITH OVERLAID STATISTICAL MEASURES FOR PRIMARY INTERSTATE HIGHWAYS 123

REVIEW DRAFT

Appendix B - Figure 4: Scatterplot of traffic volumes by year with overlaid statistical measures for Primary US and states highways 124
Appendix B - Figure 5: Within location averages versus the most recent counts. 125
Appendix B - Figure 6: County level urban mean/median traffic volumes for primary interstates 126
Appendix B - Figure 7County level non-urban mean/median traffic volumes for primary interstates 127
Appendix B - Figure 8: County level mean/median traffic volumes for primary US and states highways 128
Appendix B - Figure 9: County level mean/median traffic volumes for secondary state and county highways 129
Appendix B - Figure 10: County level mean/median traffic volumes for local, neighborhood, rural road, and citystreets 130
Appendix B - Figure 11: Boxplots of traffic count by road type comparing urban to non-urbanized areas 131
Appendix B - Figure 12: Categorization of urbanized areas based on population with small, medium, large, very large - excluding Los Angeles 132
Appendix B - Figure 13: Boxplots of traffic counts by categorized urban area - Los Angeles is a singular category 133
Appendix B - Figure 14: Traffic volume against population category 134
Appendix B - Figure 15: Traffic volume against population potential category 135
Appendix B - Figure 16: The plot shows how the counties were broken down in terms of population size. 136
Appendix B - Figure 17: Traffic Assignment Decision Making Tree 137

List of Tables

Main Text

Table 1: Hazard ratios from cohort studies examining risk of death from all causes associated with long-term particulate matter exposure based on 10 mg/m^3 change in $PM_{2.5}$. 23
Table 2: Cause-specific mortality hazard ratios associated with a 10 mg/m^3 change in $PM_{2.5}$ for selected cohort studies (95% confidence intervals given in parentheses). 23
Table 3: Summary statistics of number of observations per monitor 24
Table 4: Summary statistics for mean $PM_{2.5}$ for 2000 27
Table 5: $PM_{2.5}$ Kriging parameters used for data fitting and prediction 43
Table 6: Shapiro-Wilk test for normality 45
Table 7: LUR model parameters for NO_2, $PM_{2.5}$ and PM_{10} models (see Table 8 below for description of pollutants and variables) 47
Table 8: Description of pollutants and variable short-hand names in regression output table (above) 47
Table 9: Comparison of average cross-validation risks for selected models 49
Table 10: Pseudo-R^2 for selected models compared to maximum capable fit 50
Table 11: Space/time covariance parameters for three air pollutants in California 52
Table 12: Slope of mixed model regression of predicted on observed data. 56
Table 13: Summary statistics of models fits. 60
Table 14: Match rate for corrected addresses using the TeleAtlas (unstandardized) and StreetMap (composite) locators. 61
Table 15: Overall match rates by ACS file. Note that these numbers include non-geocodeable addresses like PO Boxes etc. 62
Table 16: Approximate match numbers by locator* 62
Table 17: Breakdown of records with a "C/O" representing Care Of 62
Table 18: Counts by ACS file by state. Note that the California file has participant addresses at the beginning of the study. 63
Table 19: Reasons for exclusion of CPS-II participants from final California analytic cohort, follow-up 1982-2000 73
Table 20: Participant characteristics in the Nationwide Study Compared to the California Cohort. 74
Table 21: Distribution of air pollutants at individual level. 75
Table 21a: Distribution of air pollutants at ZIP level. 76
Table 22: Distribution of Ecological Covariates based on 1990 Census data and defined at the ZIP Code Area. 77

REVIEW DRAFT

TABLE 23A: PEARSON CORRELATIONS (X100) BETWEEN AIR POLLUTANTS (CALIFORNIA OVERALL). 78
TABLE 23B: PEARSON CORRELATIONS (X100) BETWEEN AIR POLLUTANTS (CALIFORNIA MINUS LA). 79
TABLE 23C: PEARSON CORRELATIONS (X100) BETWEEN AIR POLLUTANTS (LA ONLY). 80
TABLE 24: HAZARD RATIOS OF $PM_{2.5}$ IDW ZIP CODE-LEVEL (IQR=9.37) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL COVARIATES (1990), STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 70,084. 81
TABLE 25: HAZARD RATIOS OF $PM_{2.5}$KRG ZIP CODE-LEVEL (IQR=8.4735) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL COVARIATES (1990), STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 73,609. 82
TABLE 26: HAZARD RATIOS OF $PM_{2.5}$IDW INDIVIDUAL-LEVEL (IQR=8.74) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL COVARIATES (1990), STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 76,041. 83
TABLE 27: HAZARD RATIOS OF $PM_{2.5}$KRG INDIVIDUAL-LEVEL (IQR=8.52902) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL COVARIATES (1990), STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 76,135. 84
TABLE 28: HAZARD RATIOS OF $PM_{2.5}$LUR INDIVIDUAL-LEVEL (IQR= 5.35) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL COVARIATES (1990), STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 76,105. 85
TABLE 29: HAZARD RATIOS OF $PM_{2.5}$RS INDIVIDUAL-LEVEL (IQR= 5.39) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL COVARIATES AND ECOLOGIC LEVEL (1990) COVARIATES, STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 76,105. 87
TABLE 30: HAZARD RATIOS OF OZONE IDW (IQR=24.07) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL AND ECOLOGIC COVARIATES, STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 76,135. 88
TABLE 31: HAZARD RATIOS OF OZONE BME AT ZIP-LEVEL (IQR=24.241) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL LEVEL AND ECOLOGIC COVARIATES, STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 69,551. 89
TABLE 32: HAZARD RATIOS OF NO_2 IDW (IQR=17.17) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL AND ECOLOGIC LEVEL COVARIATES, STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 75,364. 90
TABLE 33: HAZARD RATIOS OF NO_2LUR (IQR= 4.12) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL AND ECOLOGIC LEVEL COVARIATES, STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE (1-YEAR GROUPINGS), GENDER AND RACE USING THE STANDARD COX MODEL AND THE RANDOM EFFECTS MODEL, 1 CLUSTER LEVEL (ZIP), (95% CONFIDENCE INTERVALS GIVEN IN PARENTHESIS), ZIP VARIANCE (10^{-3}) BENEATH, N = 76,105. 91
TABLE 34: HAZARD RATIOS OF PM_{10} (IQR=17.3150) FOR SELECTED CAUSES OF DEATH IN THE ACS COHORT WITH FOLLOW-UP FROM 1982 TO 2000, ADJUSTING FOR INDIVIDUAL ECOLOGIC LEVEL COVARIATES, STRATIFYING THE BASELINE HAZARD FUNCTION BY AGE

REVIEW DRAFT

(1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,135. 93
Table 35: Hazard ratios of Sulfate (IQR=2.6705) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level covariates and Ecologic 1990, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,135. 94
Table 36: Hazard ratios of Road Buffers (EXP100 or MJR50) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105. 95
Table 37: Hazard ratios of Road Buffer MJR50 for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and Ecologic covariates level, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105. 96
Table 38: Hazard ratios of Road Buffer EXP100 for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105. 97
Table 39: Hazard ratios of $PM_{2.5}$ KRG with Road Buffer (EXP100 and MJR50) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105. 98
Table 40: Long-term and sub-acute cardiovascular mortality log-hazard ratios per 1 ppb of nitrogen dioxide or ozone using a follow-up period of January 1989 to December 2000, adjusted for individual and ecological risk factors using the Standard Cox survival model by definition of exposure time: long-term, 12 month moving average, or joint exposure (long-term and sub-acute). Standard errors are given in parenthesis. 99
Table 41: Comparison of Relative Risk Estimates from the California and National American Cancer Society Cohorts for $PM_{2.5}$ using a 10 $\mu g/m^3$ Exposure Increment 108

REVIEW DRAFT

SIGNIFICANCE

In a recent analysis commissioned by the U.S. Environmental Protection Agency seeking an expert consensus on risk of mortality from exposure to $PM_{2.5}$ [4], every expert panel member cited studies based on the American Cancer Society (ACS) cohort as influential. A study by Jerrett et al. [5], the first to examine the risks with improved exposure models based on gradients across Los Angeles, was continuously noted as one of the most important determinants of the risk from PM by the majority of the experts. The consistent results of higher risk estimates from this Los Angeles study, probably resulting from reduced exposure measurement error, have raised questions about which risk estimate should be used for assessing the benefits of air quality regulations in California. The California population of the cohort analyzed by Pope et al. [6] comprises about 15% of the total study population (about 76,000 of 500,000 subjects). While the Jerrett et al. [5] study is based solely on Los Angeles residents, questions about the applicability of the results for the remainder of California persist, given the generally higher levels of pollution in the Los Angeles region and the different population mixture there.

Two other studies have examined multiple counties in California. One study by Ostro et al. [7, 8] includes subjects in eight counties and another by Enstrom (2005) [9, 10] has subjects in 11 counties. The Ostro study is restricted to female subjects who were employed as teachers, and the Enstrom study relies on an earlier ACS cohort with many elderly subjects. The special populations and constrained geographic coverage in these studies limit their applicability for health benefits estimation in California. In our study we have subjects and exposures assigned in 54 of the 58 California counties. Although the ACS CPS II is not necessarily representative of the California population, this cohort does include men and women of various ages and has a wide range of socio-demographic characteristics reported on the survey responses. ***Our study therefore supplies the first California-wide estimates of mortality associated with $PM_{2.5}$ exposure and other criteria co-pollutants, thus supplying policymakers with a valuable resource for deriving benefits estimates.***

Beyond the immediate contributions to benefits estimation, this study resolves key uncertainties in the science of air pollution health effects. For the first time we geocoded the ACS subjects to their home addresses, as compared to the previous studies that have used either metropolitan area of residence or the home ZIP code to assign exposure. This improved locational accuracy is combined with the advanced models of exposure. The integration of numerous land use, traffic, physiographic, and remotely sensed data into a rigorous mathematical model capable of estimating exposures in time and space extends the science of exposure assessment and gives results that are less prone to measurement error.

BACKGROUND AND RATIONALE

In this section we review key studies that inform our investigation. We emphasize prior ACS studies, but also cover many other studies that contribute to knowledge on the chronic effects of air pollution on mortality.

REVIEW DRAFT

The Harvard Six Cities Study and American Cancer Society Study of Particulate Air Pollution and Mortality

Epidemiologic studies conducted over several decades have provided evidence to suggest that long-term exposure to elevated ambient levels of particulate air pollution is associated with increased mortality. Two US cohort studies, the Harvard Six Cities Study [11], a 20 year prospective cohort study, and the American Cancer Society (ACS) Study [1], a larger retrospective cohort study involving 156 cities, estimated that annual average all-cause mortality increased in association with an increase in sulfate fine particles (all particles less than 2.5 µm in median aerodynamic diameter or $PM_{2.5}$) and to undifferentiated $PM_{2.5}$ in some 50 cities.

Both studies came under intense scrutiny in 1997 when the results were used by the US Environmental Protection Agency (EPA) to support new National Ambient Air Quality Standards for $PM_{2.5}$ and to maintain the standards for particles less than 10 µm in median aerodynamic diameter (PM_{10}) already in effect. The findings were the subject of debate regarding the following factors: possible residual confounding by individual risk factors (e.g., sedentary lifestyle, active or passive cigarette smoke exposure) or ecologic risk factors (e.g., aspects of climate or social milieu); inadequate characterization of the long–term exposure of study subjects; different kinds of bias in allocating exposure to separate cities; and robustness of the results to changes in the specification of statistical models [12, 13]. To address growing public controversy concerning the studies' methods and results, Harvard University and the ACS requested that the Health Effects Institute (HEI) organize an independent reanalysis of these studies.

Through a competitive process, a reanalysis team led by Dr. Daniel Krewski of the McLaughlin Centre for Population Health Risk Assessment at the University of Ottawa was selected by an independent Expert Panel appointed by the HEI Board of Directors, with support from the EPA, industry, Congress, and other stakeholders. The reanalysis study was overseen by the Expert Panel, chaired by Dr. Arthur Upton from the University of Medicine and Dentistry of New Jersey and former Director of the National Cancer Institute, with assistance by a broad-based Advisory Board of stakeholders and scientists. The findings of the reanalysis (Phase I and Phase II) were published in 2000 [2, 14]. The final results were extensively peer reviewed by an independent Special Panel of the HEI Review Committee, which was chaired by Dr. Millicent Higgins of the University of Michigan.

The Reanalysis Project: Objectives and Findings

The overall objective of the Particle Epidemiology Reanalysis Project was to conduct a rigorous and independent assessment of the findings of the Harvard Six Cities and ACS Studies of air pollution and mortality. Phase 1: Replication and Validation involved a quality assurance (QA) audit of a sample of the original data and validation of the original numeric results. Phase II: Sensitivity Analysis tested the robustness of the original analyses to alternate risk models and analytics.

Overall, the reanalysis assured the quality of the original data, replicated the original results, and tested those results against alternative risk models of the Cox proportional-hazards family and other analytic approaches without substantively altering the original findings of an association

REVIEW DRAFT

between indicators of PM air pollution and mortality [15]. Phase II of the reanalysis made innovative contributions to the understanding of the air pollution-mortality association by developing new methods of spatial analysis for cohort studies involving both individual and ecologic covariates. Key findings from the reanalysis indicated that (1) educational status has a significant modifying effect with risk of mortality associated with fine particles declining with increasing educational attainment, (2) sulfur dioxide may exert a more robust effect on mortality than sulfates, (3) other possible ecologic confounders have no significant effect in models that control for spatial autocorrelation, and (4) spatial risk models attenuate the air pollution effect, both in terms of size and certainty.

The implications of the findings for air quality risk management were significant and pointed to the vital need for further study to better understand the role of ecologic covariates in the association between air pollution and mortality. Although the new methods developed in the reanalysis were useful for exploring the spatial structure of the data and the impact of spatial autocorrelation on estimates of risk associated with exposure to particulate air pollution, it was recognized that further work is required to determine how robust the results are to more sophisticated spatial models.

Geographic Scale of Analysis

As an initial step toward understanding the effects of ecologic variables in confounding or modifying the relationship between particulate air pollution and mortality, the reanalysis relied on the metropolitan statistical area (MSA) scale to match the one used by the original investigators [1, 2, 14]. The reanalysis demonstrated that several ecologic covariates were significant when incorporated into the standard Cox regression model, assuming independent observations. One of the more surprising results was the lack of confounding effect ecologic covariates exerted on the air pollution-mortality relationship in models that controlled for spatial autocorrelation. Sulfate and sulfur dioxide were the only significant ecologic variables in the spatial regression models.

The reanalysis relied on cross-level or multi-level data, and the extensive battery of individual variables used in the first stage of the random effects model may have removed most of the potential confounding effects before the ecologic variables were tested in the second stage. Yet this seems unlikely because of compelling literature which points to the importance of 'contextual' or community level effects on mortality [16-18]. Other methodological limitations probably contributed to the unexpected results. At the MSA scale of aggregation, many ecologic variables may display too much intra-urban variation to represent the socioeconomic or environmental phenomenon of interest without large measurement error. A growing literature points to neighborhood scale ecologic effects [18-21]. In many cases, variation within large metropolitan areas is greater than variation between these areas [22, 23] for comparable analyses at the county and census tract scales).

Another aggregation issue known as 'modifiable areal unit problem' (MAUP) emphasizes the need for correct scale. This problem arises due to the uncertainty induced by the aggregation process. Observed spatial patterns might be a function of the zones chosen for analysis rather than the underlying spatial pattern in attribute values. Spatially aggregated data display higher

REVIEW DRAFT

levels of uncertainty than the individual data on which those aggregations are based, and observed patterns may result from artifacts of aggregation [24].

Aggregation produces changes in values of statistics computed on the variables in two ways. First, there is the 'scale effect' that results from a loss of information that occurs when individual data are aggregated to ecologic zones, and fewer observations exist in the model [25]. The scale effect also suggests that some changes in statistical results occur because the aggregate data refer to different levels in the geographic hierarchy (e.g., states, metropolitan areas, cities, ZIP code areas (ZCA)) and each of these contains different information about the geographic process of interest [26]. Each scale can have a different spatial pattern in both mortality experience and the ecologic variables that influence mortality.

Second, the MAUP occurs when the boundaries of the unit of analysis affect variation in statistical values derived from these units. This is referred to as the 'zoning effect.' Even variables measured at the same scale may display different spatial patterns because of the zonal boundaries chosen for analysis rather than the underlying spatial pattern. For example, if the boundary of an ecologic unit includes a neighborhood with high poverty, changes to this boundary that exclude the poor neighborhood would reduce the poverty rate for the entire ecologic unit.

To minimize these aggregation problems, some researchers suggest that the smallest available unit of analysis should be used unless prior evidence indicates larger units will reveal more about the effect in question [27]. Testing ecologic variables at either the scale that previous studies have shown to be important or at the smallest available scale could alter the results of the reanalysis and show that ecologic variables confound the air pollution-mortality relationship. In related literature on environmental justice that investigates whether disadvantaged and minority groups suffer greater pollution exposure than wealthier groups and whites, empirical evidence and compelling conceptual arguments suggest geographic scale affects the outcome of the analysis [23, 28-30]. Likewise, some of the observed air pollution-health relationship may be reduced or modified by the contextual effects of ecologic variables measured at finer scales than metropolitan areas or at multiple scales. Further analyses of ecologic covariates in the ACS study at multiple scales would answer lingering questions about whether these variables exert a significant influence and provide important guidance for location-specific air quality management policies.

Refinement of Exposure Estimates

Previous ACS studies have relied on between-community central monitor estimates that assign entire MSAs the same level of exposure. Recent studies have recognized that exposure to air pollution vary spatially within a city [22, 31-34],and these variations may follow social gradients that influence susceptibility to environmental exposures [35]. Residents of poorer neighborhoods may live closer to point sources of industrial pollution or roadways with higher traffic density [36]. Health effects may be larger around such sources, and these effects are diminished when using average concentrations for the entire community. Recent studies of $PM_{2.5}$ have shown that within-city or intraurban exposure gradients can be associated with atherosclerosis [37] and high risks of premature mortality [38]. These studies have used

REVIEW DRAFT

geostatistical interpolation models that capture regional patterns of pollution well, but often fail to account for near-source impacts from local traffic and industry. Given the large health effects reported in these and other European studies [39, 40] a need exists to refine these estimates of intra-urban exposure to reduce uncertainties potentially associated with measurement error.

Several recent studies have demonstrated the potential of land use regression (LUR) to supply accurate, small area estimates of air pollution concentrations without the expense of dispersion or exposure modeling [41, 42]. The goal of LUR is to explain as much of the variation in existing air quality data for a given pollutant using data on nearby traffic, land use and other variables. Using multiple linear regression, a model is developed with existing monitors that can then be applied to unmonitored locations provided the appropriate geographic data is available.

Ross et al. (2006)[43] developed LUR models using traffic, distance to the coast and road length measures that explained nearly 80% of the variation in nitrogen dioxide levels in San Diego, California and were able to predict validation locations – locations that were not included in the modeling – to within, on average, 2.1 ppb. LUR models predicting nitrogen dioxide using traffic and other variables in Montréal and several European cities also produced accurate predictions [44].

In contrast to more locally varying gases such as nitrogen dioxide, however, PM mass has a significant regional secondary component with smaller contributions from local sources [45]. This complicates the estimation of intraurban exposure with LUR. LUR models have been used with some success to predict fine particle concentrations in Europe as part of the Traffic Related Air Pollution and Childhood Asthma Study (TRAPCA) [31]. In California, one study has used LUR to predict fine particle concentrations in Los Angeles. When applied in health effects modeling, the LUR exposure assessment was found to associate with mortality, with less effect attenuation from confounders, than with a geostatistical model[3]. This finding suggests that LUR models might reduce exposure measurement error and result in better health effects assessment.

Limitations of the Standard Cox Model

While the reanalysis made progress toward understanding the influence of spatial autocorrelation on the sulfate effect, the methods utilized were criticized on a number of grounds [46]. In particular, all methods relied on a fixed relationship over space. For example, the filtering method used a 600-km buffer to remove significant spatial autocorrelation prior to estimation with weighted least squares. Yet the relationship between air pollution and mortality may display non-stationarity behavior over space, meaning the air pollution effect differs depending on the location within the U.S. A more flexible modeling strategy is needed to assess non-stationary relationships. Also, reliance on one autocorrelation parameter may have effectively filtered variables that operate at the broad regional scale such as sulfate, but it may not have controlled autocorrelation from pollutants such as sulfur dioxide [46], which has a more spatially concentrated or local distribution [2, 14]. The inability of the spatial regression methods to deal simultaneously with variables that exhibit different spatial patterns may have contributed to the second key finding of the reanalysis, namely, that the effect of sulfur dioxide was more robust to control for spatial autocorrelation and other ecologic covariates than the sulfate effect [46]. Models capable of adapting to the available data and observed empirical relationships may alter

REVIEW DRAFT

the results, and the implications for health benefits assessment are considerable. In Phase II of the reanalysis a Poisson modeling approach for Cox models was developed with two levels of nested random effects [47]. This method allows us to characterize the clustering of spatial effects at two geographic levels. To accommodate more than two levels of geographic nesting (e.g., neighborhood within county, county within MSA, MSA within state), our random effects Cox model needs to be extended to fully describe complex spatial patterns in the ACS data.

Post-Reanalysis Studies of the ACS Cohort

Following the reanalysis, Pope et al.[6] undertook a subsequent analysis using an additional 10 years of data that doubled the follow-up time to more than 16 years and tripled the number of deaths. Exposure data were expanded to include gaseous co-pollutant data and new $PM_{2.5}$ data, which had been collected since the enactment of the new air quality standards. Recent advances in statistical modeling were incorporated in the analyses, including the introduction of random effects and nonparametric spatial smoothing components in the Cox proportional hazards model. The findings provided strong evidence that long-term exposure to fine particulate air pollution common to many metropolitan areas is an important risk factor for lung cancer and cardiopulmonary mortality. Each 10 $\mu g/m^3$ increase in long-term average $PM_{2.5}$ ambient concentrations was associated with approximately a 4%, 6% and 8% increased risk of all-cause, cardiopulmonary, and lung cancer mortality respectively. There was no evidence of statistically significant spatial autocorrelation in the survival data after controlling for fine particulate air pollution and the various individual risk factors. Graphical examination of the correlations of the residual mortality with distance between metropolitan areas revealed no significant spatial autocorrelation.

Pope et al. [48] have examined the pathways by which particles may increase cardiopulmonary mortality in the ACS cohort. While the challenges associated with making empirical observations relating to potential mechanistic pathways of disease from epidemiological studies are recognized, the results of the analysis are largely consistent with the proposition that the general physiological pathways that link long-term PM exposure and cardiopulmonary mortality risk include pulmonary and systemic inflammation, accelerated atherosclerosis, and altered cardiac autonomic function. Künzli et al. [49] have published the first epidemiologic evidence in support of the idea of a chronic vascular response to respiratory and system effects of PM exposure.

Exposure Time Windows

While the Six Cities study and ACS study have demonstrated an association between long-term exposure to particulate air pollution and mortality [1, 6, 11], none of these studies provided an indication of the critical period of exposure responsible for the observed association (Goddard et al., 1995)[50]. Investigations by Zeger et al. [51] and Schwartz [52] have shown that mortality cannot be attributed entirely to the effects of short-term peak exposures, which may affect sensitive individuals with pre-existing conditions [53-56]. Krewski et al. [2, 14] developed individual temporal exposure profiles for subjects in the Harvard Six-Cities Study by coding the residence histories of those subjects. However, limited population mobility and limited variation

REVIEW DRAFT

in individual time-dependent exposure profiles precluded identification of critical exposure-time windows [57].

The identification of critical exposure-time windows has important implications for establishing time lines for public health policy interventions. The identification of critical exposure-time windows requires information on temporal patterns of exposure at the individual level. Given the regulatory importance of the results, further work to develop individual time-dependent exposure profiles for ACS cohort participants is needed.

Description of Cohort Studies Reporting $PM_{2.5}$ Mortality Risks

There are now several more population-based cohort studies of air pollution and mortality, which report $PM_{2.5}$ risk estimates. These studies are briefly described below and the results are summarized in tabular form to offer a basis of comparison for our subsequent results. We have included studies that do not focus on special or limited populations such as veterans or women only.

American Cancer Society Cancer Prevention II Study (ACS CPS II)

About 1.2 million cohort participants were enrolled by American Cancer Society volunteers between September 1982 to February 1983 across all 50 states, the District of Columbia, and Puerto Rico. Enrolment was restricted to persons who were at least 30 years of age in households with at least one individual 45 years of age or older. Participants completed a confidential questionnaire that included demographic characteristics, smoking history, alcohol use, diet, and education. The study population included only those who resided in U.S. metropolitan statistical areas (MSAs) within the 48 contiguous states and the District of Columbia with available pollution data. Mortality was ascertained by volunteers in 1984, 1986, and 1988, and subsequently using the National Death Index. In addition, several ecological risk factors representing social (education, race, air conditioning, percent non-white) and economic (household income, unemployment, income disparity) conditions were obtained from the 1979 U.S. Census. The $PM_{2.5}$ measure used here was the average of all $PM_{2.5}$ concentrations obtained from fixed site monitors within an MSA for the years 1999 and 2000. This resulted in 116 MSAs being included in the analysis. Cohort follow up was from 1982 to 2000 [3].

Multi-level random effects Cox proportional hazards models were used to assess the risk of mortality in relation to pollution exposures, while stratifying for age (single-year groupings), sex, and race in the baseline hazard. A total of 20 variables were used to control for individual characteristics that may confound or modify the air pollution-mortality association. These individual variables included occupation, education, diet, smoking, and other lifestyle characteristics. In addition to the individual level covariates, seven ecological covariates defined at the ZIP code level were included in the model. Each subject within the same ZIP code was assigned the same value of these ecological variables. Each ecological covariate was represented by two separate variables, the MSA average and the difference between the ZIP code area value and the MSA average. Community-level random effects allowed for an assessment of residual variation in mortality among communities.

REVIEW DRAFT

The National Medicare Cohort (MET)

The cohort [51]consists of all those 65 years or older, who were enrolled in Medicare between the years 2000 and 2005 with ZIP code centroids within 6 miles of an EPA $PM_{2.5}$ monitoring station. The study includes more than 13 million subjects. The outcome measure is the six-year (2000-2005) mortality rate for persons residing within each of 4,568 ZIP codes for each of three age stratum: 65-74, 75-84, and 85 years and older. The $PM_{2.5}$ data were obtained from the U.S. Environmental Protection Agency, which included 1,006 monitors for the period 2000-2005. A ZIP code six-year average of $PM_{2.5}$ was used as a measure of the long-term exposure to $PM_{2.5}$ for an individual living within a ZIP code. To account for socioeconomic status (SES) at the ZIP-code level, age-specific SES variables from the 2000 US Census were used. Results are reported for three regions of the United States (East, Central, and West). We also report a summary-pooled estimate of these three regions.

The Six Cities Study (SCS)

The study population consisted of 8,096 white participants residing in Watertown, MA; Kingston and Harriman, TN; St. Louis,MO; Steubenville, OH; Portage, Wyocena, and Pardeeville,WI; and Topeka, KS [58]. Participants were recruited between 1974 and 1977. Follow-up was terminated for this analysis in 1998. Daily outdoor $PM_{2.5}$ concentrations were measured at a centrally located air-monitoring station in each community from 1979 to 1987. Daily $PM_{2.5}$ concentrations were estimating from 1985 to 1998 by city-specific regression equations based on the extinction coefficient from visibility data obtained at local airports and routinely collected PM_{10} concentrations by the U.S. EPA from representative monitors within 80 km of each city. Proportional hazards Cox survival regression models were used to relate long-term city-specific $PM_{2.5}$ concentrations to survival, controlling for individual risk factors such as smoking, education and body mass index. The baseline hazard function was stratified by 1-year age groups and gender.

The Six-City National Medicare Cohort (SCMC)

The study design of this cohort is similar to that of the National Medicare Cohort (NMC) except subject's enrolment was restricted to the counties corresponding to the six cities examined in the SCS[59]. The follow-up period of this study, 2000 to 2002, was shorter than the NMC (2000 to 2005). $PM_{2.5}$ concentrations were obtained from EPA monitoring data for 2000 to 2002 in each of the six counties.

Netherlands Cohort Study on Diet and Cancer (NLCS)

The NLCS study is a study among about 120,000 subjects who were between 55 and 69 years old at enrolment in 1986 [60]. The study is primarily on nutrition and cancer, following a case-cohort design with a random subgroup of ~5,000 subject characterized in detail to estimate person time, as reference for cancer cases obtained from the cohort as a whole from cancer registries. Using a variety of modeling approaches, air pollution exposure to Black Smoke and NO_2 was estimated at the home address of all study participants based on measurement data obtained in the period of follow up from routine monitoring stations. Exposure to PM_{10} was estimated using more recent routine measurements, and from the PM_{10} measurements, $PM_{2.5}$ was

REVIEW DRAFT

estimated using a fixed $PM_{2.5}/PM_{10}$ ratio. After ascertaining that individual level risk factors did not confound associations between air pollution and mortality in the case-cohort analysis, the data from the full cohort were analyzed using a limited set of potential confounders which included age, gender, smoking status and area-level socio economic status Specifically, the effect of living near busy roads on mortality was also estimated.

Summary of Cohort Studies for All Causes of Death

The hazard ratios from these studies are reported in Table 1 for all causes of death. There is a range in risks, ranging from 1.06 to 1.21.

Summary of Cohort Studies for Specific Causes of Death

There are a few of cohort studies that report hazard ratios for selected specific causes of death. (see Table 2).

The ACS, SCS, and NLCS all report hazard ratios for specific causes of death although each study reports results for a different selection of causes. Hazards ratios from these studies are not all directly comparable since some studies incorporate both within-community and between-community exposure contrasts, while other studies only used between-community exposure gradients.

In general, the lowest hazard ratios are reported for "Other Causes of Death," which are non-cardiopulmonary or lung cancer. This suggests that the majority of evidence supports a restriction of $PM_{2.5}$ effects on mortality to heart, circulation, or lung diseases.

There was some variation in the hazard ratios observed for respiratory deaths (1.02 to 1.08). The hazard ratios for Ischemic/Coronary Heart Disease (IHD) deaths are generally larger (1.26 to 1.29) than those from other specific causes of death examined except for the hazard ratio reported from the Netherlands study (0.96). Positive associations between $PM_{2.5}$ and lung cancer deaths are observed for all studies (1.06 to 1.27).

REVIEW DRAFT

Table 1: Hazard ratios from cohort studies examining risk of death from all causes associated with long-term particulate matter exposure based on 10 µg/m^3 change in $PM_{2.5}$.

Cohort Study/ Characteristics	Number of Subjects	Follow-up Period	Geographic Coverage	Hazard Ratio (95% Confidence Interval)
American Cancer Society (ACS) - National	486,133	1982-2000	United States (116 MSAs)	1.08 (1.04, 1.11)
Medicare (National)	13,200,000	2000-2005	United States (668 Counties)	1.06 (1.04, 1.07)
Netherlands Study on Diet and Cancer (NLCS)	120, 852	1987-1996	Netherlands	1.06 (0.97, 1.16)
Six Cities (SCS)	8,111	1979-1998	Northeast and Midwest US (6 Cities)	1.16 (1.07, 1.26)
Medicare (SCS)	341,099	2000-2002	Northeast and Midwest U.S. (6 Counties)	1.21 (1.15, 1.27)

Table 2: Cause-specific mortality hazard ratios associated with a 10 µg/m^3 change in $PM_{2.5}$ for selected cohort studies (95% confidence intervals given in parentheses).

Cause of Death /Cohort	ACS (National)	SCS	NLCS
Cardio-Vascular	1.17 (1.11, 1.24)	1.28 (1.13, 1.44)	1.11 (0.93, 1.33)
Ischemic /Coronary Heart Disease	1.29 (1.18, 1.40)	1.26 (1.08, 1.47)	0.96 (0.75, 1.22)
Respiratory	1.02 (0.93, 1.13)	1.08 (0.79, 1.49)	1.07 (0.87, 1.52)
Lung Cancer	1.14 (1.06, 1.23)	1.27 (0.96, 1.69)	1.06 (0.82, 1.38)
Others*	0.98 (0.94, 1.03)	1.02 (0.90, 1.17)	1.08 (0.96, 1.23)

*: All other causes of death not including cardiovascular, respiratory, and lung cancer.

Materials and Methods

Exposure Modeling

This section describes the available monitoring data, methods and results of the exposure assessments that were used to assign exposures to the ACS cohort used in the health effects models. In order of the least to the most sophisticated methods employed here, we report methods and results for the modeling of exposures using inverse distance weighted (IDW) interpolation, kriging interpolation, land use regression (LUR) and Bayesian Maximum Entropy (BME) interpolation of LUR residuals.

REVIEW DRAFT

Available Pollution Monitors

Data employed for the land use regression analysis was monthly average data collected at numerous governmental sites operated by the California Air Resources Board (CARB) over the period 1988-2002. Over that time period there were 138 sites – with land use data support – that collected NO_2 measurement. Over the same time period, there were 223 sites measuring PM_{10} with land use data support. $PM_{2.5}$ data were only available between 1998 and 2002 with a total of 112 active monitoring sites. Table 3 and Figure 1 describe the distribution number of available monitors by respective pollutant.

Table 3: Summary statistics of number of observations per monitor

Pollutant	No. of Sites	Mean No. of Obs	Min No. of Obs	Max No. of Obs
NO2	138	110.9	3	180
PM10	223	98.4	3	180
PM2.5	112	42.4	2	60

REVIEW DRAFT

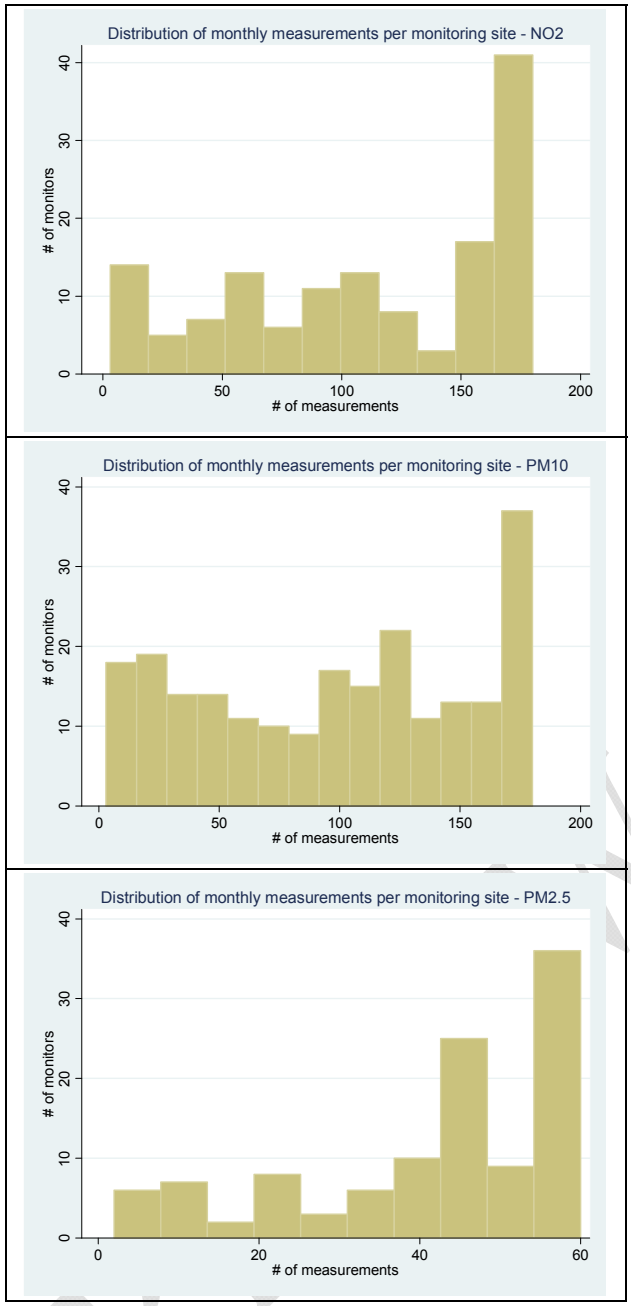


Figure 1: Distributions of observed pollutants: NO_2, PM_{10} and $PM_{2.5}$

Methodology for Exposure Modeling

Inverse Distance Weighted (IDW) Interpolation

Inverse distance weighting estimates were supplied to us by the California Air Resources Board Staff. This effort built on a major investigation led by Dr. Michael Lipsett on the California

DRAFT

Teachers cohort (with Dr. Jerrett as a Co-I). More detail is given in Dr. Lipsett's report (Lipsett et al. 2011 in review).

Briefly, routine ambient air monitoring data for $PM_{2.5}$ were available from 1998 through 2002. For our study data for ozone, PM_{10}, NO_2 were available for the years 1988-2002. Dr. Lipsett's team validated monitor locations and spatial files were produced. Specifically, they had received from ARB staff a file with monitor addresses and geocoded coordinates. In addition, the California Air Resources Board (ARB) website listed GPS coordinates for a large number of monitors. These were re-geocoded to the monitor addresses in the file from ARB and compared these coordinates with those already in the file on the ARB website. In a few instances, there were non-matching coordinates that ultimately were addressed by comparing them with aerial photographs to identify the most likely locations.

The inverse distance weighting (IDW) estimates were done the same way for both projects. Estimates were restricted to areas within 50 km of a monitoring location. Monitors with no measurement data during the period of interest (1988-2002) were also excluded. The IDW was done to the power of 2 using the Arc GIS 9.0 Spatial Analyst extension. This was done for each month for all pollutants for all the years the IDW was calculated using 250 m grid to join to the residential geocoded address.

For sulfate, we computed the IDW estimates ourselves using ArcGIS 9.2. Because of sparseness in some of the areas state with regard to monitoring and the more regional pattern that sulfate follows because it is a secondary pollutant, we did not restrict the interpolation to within 50 km of a monitor, but used all available monitors and assigned to all locations, regardless of their distance to a monitor. There were several years with the sulfate data that had smaller counts, so we used the year with the maximum amount of monitoring data (i.e., 1996).

Kriging $PM_{2.5}$

Kriging is a geostatistical technique belonging to the group of linear least squares estimators. Its purpose is to interpolate values a spatial random field at unobserved locations based on previously acquired data. Basically, kriging estimates are weighted averages in a similar way to the IDW interpolation except that the weights have been optimized through a least squares operation – based on the spatial covariance – to minimize the variance of the prediction error. The spatial covariance is estimated from the empirical distribution and requires that the model describing the empirical distribution result in a covariance matrix that is symmetric and positive-semidefinite.

Data

We created kriging estimates for an annual average (2000) $PM_{2.5}$ statewide exposure surface to analogously replicate the ACS Los Angeles analysis at the statewide level; this is in keeping with our objectives specified in the contract. As monitors were not necessarily operated over the entire 12 months of the year 2000 we restricted the input data to those monitors that where operated for the entire 12-month period. This was done as the missing data could potentially bias the estimate of the observed annual mean $PM_{2.5}$ concentrations at the monitors. Of the 112

REVIEW DRAFT

monitoring that measured PM2.5 from 1998-2002, only 72 of these monitors had observations for all 12 months of the year 2000. Table 4 summarizes some basic descriptive statistics of the observed $PM_{2.5}$ data and Figure 2 illustrates the distribution of the data of the 72 monitors that were used for the kriging exposure assessment.

Table 4: Summary statistics for mean $PM_{2.5}$ for 2000

	Obs	Mean	Std. Dev.	Min	Max
$PM_{2.5}$	72	15.16493	5.827055	3.7775	28.4516

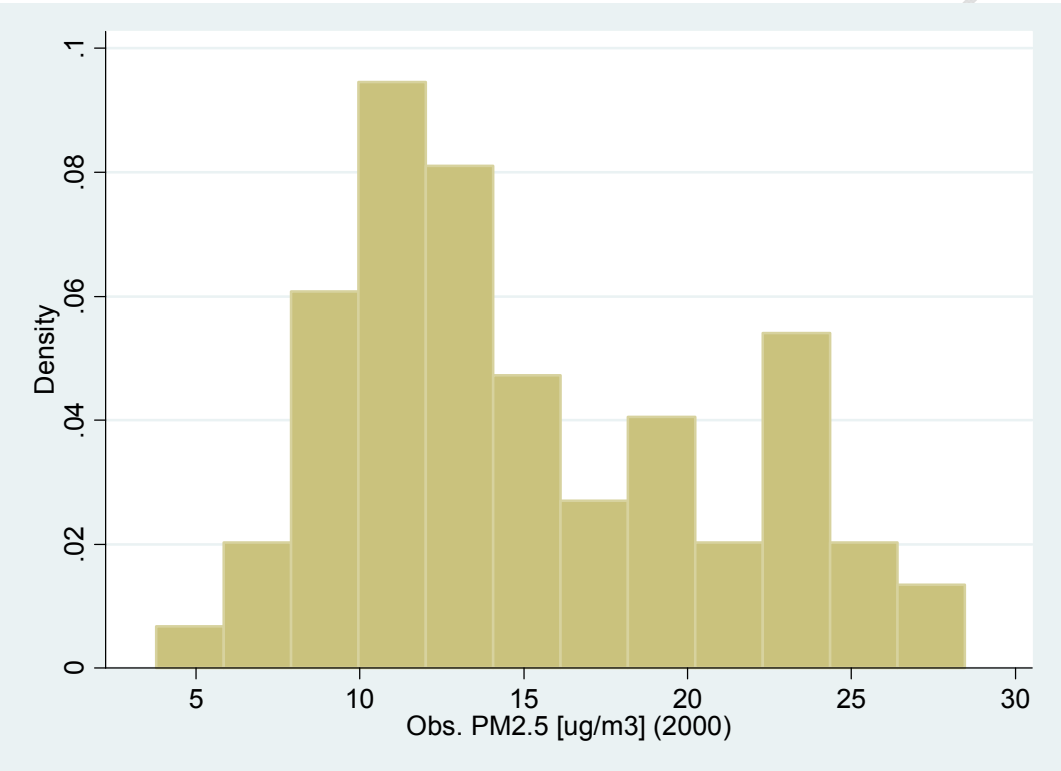


Figure 2: Histogram of observed $PM_{2.5}$ averages for 2000

Land Use Regression

Land use regression (LUR) is a statistical technique that employs the pollutant of interest as the dependent variable and proximate land use, traffic, and physical environmental variables as independent predictors. Thus the methodology seeks to predict pollution concentrations at a given site based on surrounding land use, traffic characteristics and other identified auxiliary data which may act as a pollution predictor. Surrounding land use and traffic characteristics are measured in circular areas (buffers) that surround a given location. During the model selection/building phase of the LUR technique, buffers of varying sizes are tested for their ability to predict the data. Specifically, this method uses measured pollution concentrations (**y**) at locations (**s**) as the response variable and measures of land-use/traffic (**x**) within circular areas

around **s** as predictors of the measured concentrations. Figure 3 illustrates the LUR method in relation to potential land use and road predictor datasets.

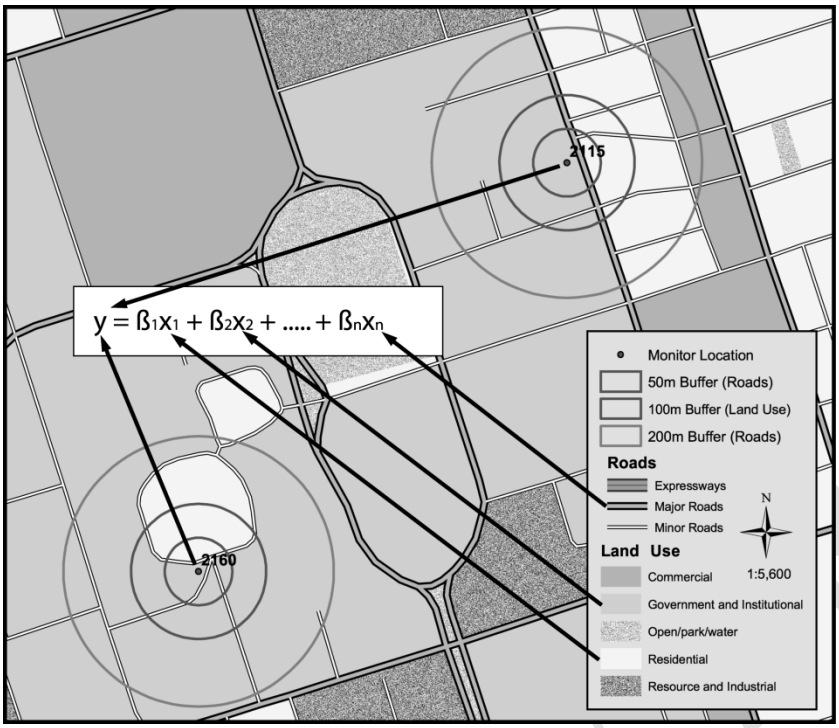


Figure 3: Illustration of land use regression methods using GIS data

We have used the LUR method to predict the multiple year mean exposure model over the full dataset, i.e., 1998-2002 for $PM_{2.5}$ and 1988-2002 for NO_2 and PM_{10}. As the observed data is at the monthly scale and the output model of the LUR methods is overall mean model describing the "chronic" expected exposure over the respective periods, there is expected to be spatially correlated temporal residuals of the regression model. We employ the Bayesian Maximum Entropy (BME) technique (described in a following section) to model these temporal residuals and apply these BME estimates to the LUR estimates to obtain a predicted monthly exposure.

Statistical Methods for Land Use Regression

Under the framework of the LUR methodology, a machine learning approach using the Deletion/Substitution/Addition algorithm (DSA) − pioneered by Sinisi and van der Laan[61] − was employed when selecting predictive models for $PM_{2.5}$, PM_{10}, and NO_2. DSA uses an aggressive covariate search algorithm predicated entirely on the power of cross-validation (CV) to select the best predictive model. The algorithm fits a generalized linear model that accounts for non-independence in the observations – based on monitoring location – and assesses the cross-validation performance using the L2 loss function. The L2 loss function, otherwise known as the risk, is defined as the expectation of the square of the CV error. The approach tests all possible covariate combinations with both polynomial and interaction terms. A number of parameters are required to implement the DSA algorithm and they specify the size, the level of interaction, the polynomial fit of the resultant model, and the number 'v' of cross-validation sets

REVIEW DRAFT

for the v-fold cross-validation. We allowed models to have a squared polynomial terms and first-order interactions. We tested models using 5-fold CV. This means that each candidate model was fit and cross-validated 5 times at every step in the model building. The selection of the "best" model is done with the assistance of the CV plot, which plots the average CV risk as a function of the size of the model. As the model grows in size there is typically diminishing return in the amount of additional predictive power that is added with each additional variable. The CV plot gives us a visual tool to select a parsimonious model, i.e., a model that balances the need to minimizing prediction error while minimizing model complexity.

Land Use and Traffic Data as Covariates

We used ArcGIS 9.2 (ESRI, Redlands) to calculate measures of land use, roads and traffic for the land use regression (LUR) analysis. Each of the covariates listed in the following sections were calculated using circular buffers of the following radii (in meters): 50, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000.

Data for land use covariates were created from the USGS National Land Use Cover Dataset (NLCD) 2001. The NLCD 2001 is a land cover classification dataset, which is derived from GIS and remotely sensed data. The data create 29 classifications for land use, as defined by the USGS for the NLCD 2001 dataset:

11 - Open water
12 - Perennial Ice/Snow
21 - Developed, Open Space
22 - Developed, Low Intensity
23 - Developed, Medium Intensity
24 - Developed, High Intensity
31 - Barren Land
32 - Unconsolidated Shore
41 - Deciduous Forest
42 - Evergreen Forest
43 - Mixed Forest
51 - Dwarf Scrub
52 - Scrub/Shrub
71 - Grassland/Herbaceous
72 - Sedge Herbaceous
73 - Lichens
74 - Moss
81 - Pasture/Hay
82 - Cultivated Crops
90 - Woody Wetlands
91 - Palustrine Forested Wetland
92 - Palustrine Scrub/Shrub
93 - Estuarine Forested Wetlands
94 - Estuarine Scrub/Shrub
95 - Emergent Herbaceous Wetland
96 - Palustrine Emergent Wetland (Persistent)

REVIEW DRAFT

97 - Palustrine Emergent Wetland
98 - Palustrine Aquatic Bed
99 - Estuarine Aquatic Bed

We consolidated land uses into natural groups corresponding with one group for open/undeveloped land, one for agricultural land and four groups of differing developed land. The developed land categories corresponded with land use classes 21-24 (above); classes 81 and 82 are agricultural land; we categorized all other land uses as open space.

Additionally, we tested variables describing surrounding population density and proxies of industrial activity as potential covariates in the regression analysis. These two classes of covariates were found not to be associated with air pollution in our analysis. Zev Ross of ZevRoss Spatial supplied both these data.

Roads and Traffic

Measures of quantities of various road types and estimates of vehicle miles traveled – based on a spatially comprehensive estimated traffic dataset – where calculated to be used as covariates in the LUR analysis. The traffic volume dataset compiled counts that had been collected between 1990-2001. We used the TeleAtlas GIS road network shapefile provided in the data package supplied with ESRI's ArcGIS Business Analyst extension. We calculated roads for various buffer distances based on road types. In the TeleAtlas road dataset, road links were characterized by their Feature Class Code (FCC). The U.S. Census Bureau created the FCC classification system. The system uses a multipart classification where the first character is a letter that characterizes the parent object type, i.e., road, railway, point of interest, etc and subsequent numbers break the parent type into sub-types (second character) and further sub-types (third character). For this analysis, we used FCC classifications A1, A2, and A4. A1 are limited access expressways, A2 are primary roads without limited access and A4 are local/neighborhood roads.

Zev Ross of ZevRoss Spatial supplied the traffic covariate data at various buffering distances; see Appendix for a full description of the data and its development.

$PM_{2.5}$ estimates derived from remotely sensed data

In our LUR model we aimed to assess the marginal benefit of including an estimate of $PM_{2.5}$ derived from remotely-sensed satellite data in our model. Ground-level concentrations of $PM_{2.5}$ were estimated using satellite atmospheric composition data combined with local, coincident scaling factors from a chemical transport model (GEOS-Chem, 2011). Specifically, $PM_{2.5}$ estimates were derived from aerosol optical depth (AOD) data from the Terra satellite, in combination with output from GEOS-Chem simulations to estimate the relationship between aerosol optical depth over the atmospheric column and ground-level $PM_{2.5}$[62]. $PM_{2.5}$ estimated at a 0.1x 0.1-degree resolution (~10x10km). Estimates for $PM_{2.5}$ are from data for 2001 to 2006 to ensure sufficient observations for stable estimation.

REVIEW DRAFT

Residual Error Modeling - Bayesian Maximum Entropy

This section describes our proposed estimation method of air pollutants that melds the Bayesian Maximum Entropy of modern geostatistics [79-81] (also see its numerical implementation *BMElib* from (http://www.unc.edu/depts/case/BMELIB) with the aforementioned LUR approach. BME not only requires mean trend values to translate a non-homogeneous/non-stationary air pollution field to a homogeneous/stationary field, but copes with the random fluctuations of the space/time distribution of air pollution arising from uncertainty in the LUR estimates. Specifically the LUR approach provides its fitted values $m_Z(\boldsymbol{p})$ at space/time points $\boldsymbol{p}$ which is then used as mean trend in BME to derive residuals $X'(\boldsymbol{p})$ through Eq. (0):

$$X'(\boldsymbol{p})=X(\boldsymbol{p})-m_Z(\boldsymbol{p}), \qquad (0)$$

where $X(\boldsymbol{p})$ is defined as air pollution measurements representing a spatially non-homogeneous and temporally non-stationary Space/Time Random Field (S/TRF) (as defined in Appendix C). The S/TRF is used to account for the random fluctuation in the LUR-driven residuals (Eq. C1)[1]. We assume that the $X'(\boldsymbol{p})$ is a homogeneous/stationary random field in the sense that the expected value of $X'(\boldsymbol{p})$ is constant (fluctuated around zero), and the covariance of $X'(\boldsymbol{p})$ can generally be expressed solely as a function of spatial and temporal lags. While with the $X(\boldsymbol{p})$ one faces with numerical difficulty to seek solutions desired, BME depends on $X'(\boldsymbol{p})$ for a series of its estimation processes involving 1) covariance characterization to quantify space/time variability in monthly air pollution of the geographical phenomena (Eq. C4), and 2) space/time interpolation of air pollution at unsampled points (C8). The mean trend and covariance constitutes the general knowledge base that is processed at the first stage of BME to construct the prior probability density function (pdf) (Eq. C3). The prior pdf is enriched by assimilating the site-specific measurements (Eqs. C6-7) through Bayesian statistics (Eq. C8). This final solution provides the posterior pdf$f_{\mathcal{K}}(\chi_k)$ (Eq. C8) for the pollutant level at each estimation point (i.e., 180 individual months in 1988-2002 for NO_2 and PM_{10}, each of 60 months in 1998-2002 for $PM_{2.5}$, and 10 km-gridded locations over the state of California for all of the pollutants).

The first order statistical moment of the posterior pdf is related to the estimate regarding air pollution. The estimate could be a mean, mode, or percentile value depending on study aims. For instance, if an estimate minimizing mean square errors is needed, the best choice would be the mean value. In the case where the most probable value is needed, the mode value is a better choice. The second order statistical moment of the posterior pdf is associated with estimation uncertainty. The estimation uncertainty is increasing as an estimation point is away from the presence of data. The BME estimation surface is then assigned to the specific subjects in California of the ACS database. The LUR values at the subject points are calculated and added to the assigned estimates in the end.

[1] All equations referenced in this section that are prefaced with "C", i.e., Eq. C1, can be found in Appendix C.

REVIEW DRAFT

Geocoding Methods Used for Converting Individual Addresses to Geographic Coordinates for Exposure Assignment

Equipment/Software Details

Geocoding of American Cancer Society data was conducted between March 24, 2008 and March 28, 2008 in Atlanta, Georgia, at the American Cancer Society offices using a Dell Precision M65 laptop with 2 GB of memory and a dual core processor. The laptop included ArcGIS 9.2 and ZP4 (Semaphore Corporation) software for use in the geocoding. All road data and geocoding locators were prepared in advance and brought to Atlanta. TeleAtlasDynamap 2000 road data were used for all geocoding in California and StreetMap (2006 based on Census TIGER files) for historical address information that included data outside of California for which TeleAtlas road data was not available.

Geocoding Locators

In preparation for the geocoding, three different address locators were prepared, two based on TeleAtlas data and one based on StreetMap. All three locators used a side offset of 15 meters. In addition, a 3m end offset, 80% spelling sensitivity, minimum candidate score of 10 and minimum match score of 60 was used. Addresses that matched more than one address erroneously were assigned a "non-match."

The two TeleAtlas locators were identical with one exception. One locator used the default street name available through TeleAtlas and one used a new street name created by applying ESRI address standardization to the street names themselves.[2] The TeleAtlas locators were based on the US Streets with Zone format with ZIP code as the zone. StreetMap data was used in a composite locator made up of a "Streets With Zone" locator as well as a locator using hyphenated addresses and an alphanumeric locator. As mentioned above, while this StreetMap composite locator was applied to all the data its main value was in geocoding non-California addresses for which TeleAtlas data was not available.

Cohort Raw Text Files from the American Cancer Society

Three text files were prepared by the American Cancer Society for geocoding: the main California file (n=105217), the nutritional sub-cohort (n=23065) and an address history file (n=17733) which includes all of the historical addresses for nutritional sub-cohort participants. In order to protect the ACS unique ID staff at ACS generated new unique identifier ("id1").

Reading Text Files and Preparing for Address Correction

All three files were read into R statistical software and appended together to create a single, larger text file to allow for quicker geocoding. In general the fields were the same except that the

[2] The concern was that, as part of the geocoding process, ArcGIS does its own address standardization and might, for example, convert W Main St to West Main Street. We thought that potentially running the street addresses from the GIS files through the same address standardization (e.g., W Main St -> West Main Street) would improve match rates.

REVIEW DRAFT

historical file had three date fields representing day, month and year. To append the files into one large file, a day, month and year field was generated for the main California file and the Nutrition sub-cohort file and populated with NA values. Regular expressions were used to identify and remove errant punctuation in the addresses (i.e. *, % and "). Several addresses were preceded by "C/O" which would hinder address standardization and likely need to be filtered from the modeling. Therefore a new field of TRUE/FALSE values was created representing the existence of a "C/O" in the address and then "C/O" was removed from the addresses. Lastly, a new unique ID was creating by concatenating the new ID, the archive day, month, year, the file that a record came from (historical, California, nutrition sub-cohort) and the street [For more detail on the issue of IDs see the section "An Important Note About IDs"]. This file was exported to the external drive for further address processing.

Address Correction

The addresses were then run through an up-to-date version of the address correction software ZP4 along with the add-on LACS (LACS attempts to convert the rural-style addresses to street-style addresses). The output fields are listed in Appendix A. ZP4 is software that includes official United States Postal Service data files on a single DVD-ROM that provides a tool for automatically determining the correct mailing address, ZIP + 4 code, and mail carrier route number for any location in the United States. Running addresses through ZP4 has shown to dramatically improve match success rates, often by as much as 10 or 20%.

Clean Up Address Correction Output and Prepare File for Geocoding

After running the addresses through ZP4 address standardization the resulting text file was read into R where field names were modified so that they would be acceptable to a GIS system. The extra unnecessary fields with detail on the performance of the address correction software were outputted to a separate file called ZP4_extra_output.txt. The records for geocoding were exported to a file called acsComplete_forGeocode_ZP4.txt.

Geocoding

For comparison purposes both the ZP4-corrected addresses as well as the raw addresses were geocoded using all three of the geocoding locators producing six different shapefiles – two geocoded files based on the TeleAtlas un-standardized locator, two geocoded files based on the TeleAtlas standardized locator and two geocoded files based on the StreetMap composite locator. All shapefiles were unprojected with a geographic coordinate system of North American Datum 1983.

The geocoding results were imported to R, modified and exported as the following text files:

acsComplete_TA_noAlt_B.txt[3] – Uncorrected addresses using the locator based on unaltered TeleAtlas

[3]The suffix "_B" represents files meant for Berkeley

REVIEW DRAFT

acsComplete_TA_Std_B.txt – Uncorrected addresses using the locator based on standardized TeleAtlas
acsComplete_StMap_composite_B.txt – Uncorrected addresses using the locator based on StreetMap
acsComplete_TA_noAlt_ZP4_B.txt – Corrected addresses using the locator based on unaltered TeleAtlas
acsComplete_TA_Std_ZP4_B.txt – Corrected addresses using the locator based on standardized TeleAtlas
acsComplete_StMap_composite_ZP4_B.txt – Corrected addresses using the locator based on StreetMap

Compile Final Geocoded Data

Based on our review of the results it was clear that, as expected, the standardized addresses resulted in much better matches. It was also clear that the geocoding based on standardized TeleAtlas street names failed to improve match results. As a result, the geocoding results based on corrected ACS addresses and the un-standardized TeleAtlas locator were used as the "gold standard" geocodes based on StreetMap were used to fill in missing data.

The final geocoded results file was assembled as follows: The geocodes based on the ZP4-corrected ACS data based on the un-standardized TeleAtlas locator represented the starting point. For records with a TeleAtlas match with a score greater than 80 this was the geocode used. For non-matches or low-score matches a StreetMap geocoded (where available) was used. The vast majority of matches (98%) were attributed to TeleAtlas. The final X, Y coordinates along with a final geocoding score, final status, final determination of side of street and final locator name were added as new fields that all begin with the letter "f" to distinguish the fields as final (fX, fY, fScore, fStatus, fSide, fLoc_name, respectively). The final file is called: final_geocode_results.txt.

Study Population

The original analysis [1], the HEI sponsored reanalysis[2, 14] and our new research reported here all have relied on data from the ACS CPS-II, an ongoing prospective mortality study of approximately 1.2 million adults. Cohort participants were enrolled by ACS volunteers beginning in the fall of 1982. The participants resided in all 50 states, the District of Columbia, and Puerto Rico. Most were friends, neighbors, or acquaintances of the ACS volunteers. Enrollment was restricted to persons who were at least 30 years of age and who were members of households with at least one individual 45 years of age or more. Participants completed a confidential questionnaire, which included questions about a variety of demographic characteristics, smoking history, alcohol use, and other lifestyle factors, occupational exposures, and other characteristics. As with the previous analysis, the analytic cohort must be restricted to include those who resided in US metropolitan areas within the 48 contiguous states (including the District of Columbia) with available pollution data. We then selected those subjects who were resident in California at the time of enrollment.

REVIEW DRAFT

Mortality of the study participants was ascertained by volunteers in 1984, 1986, and 1988, and subsequently using the National Death Index[63]. Death certificates or multiple cause-of-death codes were obtained for participants known to have died.

Data Collection and Assembly

A major component of this research dealt with improving control for confounding variables over time and in space, particularly at the 'within community' or intraurban scale for this latter objective. We obtained information on neighborhood social confounders from the 1990 US Census. With follow up from 1982-2000, the 1990 census provides an approximate mid-point for characterizing likely conditions in the neighborhoods that ACS subjects resided.

Several variables are examined in this analysis, including: median household income; 125% of poverty line; percentage of unemployed persons over the age of 16 years; percentage of adults with less than grade 12 education; percentage of homes with air conditioning; the GINI Coefficient of income inequality, and percentage of population that are not white. We only included those ZCAs that contained ACS subjects in order to more accurately represent the social environment of the ACS participants.

Statistical Methods and Data Analysis

To examine the association between ambient concentrations of air pollution and longevity on a national scale, we initially used the standard Cox model to link pollution levels to survival adjusting for potentially confounding risk factors measured at the individual level[6]. We included eight variables representing active smoking habits including non-linear terms for cigarettes per day and number of years smoked, seven variables characterising former smoking habits, one variable for exposure to passive smoke, two variables representing marital status, two variables representing a linear and squared term for body mass index, six variables characterising consumption of beer, wine, and alcohol, seven variables characterising the subject's main lifetime occupation and their potential exposure to PM in the workplace, one variable representing self reported exposure to dust and fumes in the workplace, eight variables representing diet and two variables characterising education, totalling 20 variables. We also examined seven additional risk factors measured at the five-digit ZIP code level. The baseline hazard function was stratified by 1-year age groups, gender, and race. This model assumes that all observations are statistically independent, an assumption that is relaxed in subsequent analyses.

The Cox survival model assumes that the response variable is statistically independent over space and time. However, there is a concern that survival experience may cluster by community or neighborhood. That is, longevity among subjects within the same community or neighborhood will be more similar than subjects in different geographic locales even after controlling for all known and available risk factor information, such as smoking habits, diet, education, and occupation. Furthermore, subjects who live closer together may also share more similar longevity patterns. Lack of statistical control for these factors can both bias the estimate of air pollution effects on health and their associated standard errors. To characterize the statistical error structure of survival data, statistical methodology and computer software that

REVIEW DRAFT

incorporates two levels of spatial clustering (e.g. MSA and ZIP code area within MSA) have been developed. At each of the two cluster levels incorporate a spatial autocorrelation structure such that the correlation in survival after adjusting for known risk factors such as smoking depends on the distance between clusters. This distance can be defined as Euclidian, adjacency or based on other notions of distance in economic or social terms. In our analyses we used adjacency or nearest neighbor.

The Spatial Cox Model

The original version of the Cox regression model[64] assumes that the survival times for individuals are statistically independent. In our earlier analyses of the ACS cohort, we found evidence of spatial autocorrelation in the data[2, 14], which needs to be considered in the analysis. Ma et al.[47] developed a modification of the Cox regression model incorporating random effects to represent spatial patterns in the data, and established large sample properties of the maximum likelihood estimates of the model parameters.

Here, we consider a Cox model with two levels of spatially correlated random effects. Suppose that the cohort of interest is composed of m spatially correlated clusters indexed by i. Within the ith cluster, there are J_i spatially correlated subclusters indexed by (i, j). Specifically, we assume that the cluster-level random effects $U_1, \ldots, U_m$ are positive random effects with

$$E(U_i) = 1 \text{ and } cov(U_s, U_i) = \sigma^2 \rho_1^{d(s,i)}. \tag{1}$$

where $0<\rho_1<1$ and $d(s, i)$ indicates the distance between clusters indexed by s and i. The distance between two independent clusters indexed by s and i is defined as $d_1(s, i) = \infty$. Negative $\rho_1$can be estimated if the distances are integers.

We further assume that, given the cluster-level random effects $U_* = u_* = (u_1, \ldots, u_m)$, the subcluster-level random effects $U_{11}, \ldots, U_{mJm}$ are positive and spatially dependent with

$$E(U_{ij}|U_*) = U_i \text{ and } cov(U_{st}, U_{ij}|U_*) = \delta(s, i)\, v^2 \rho_2^{r\{(s,t),(i,j)\}}. \tag{2}$$

where $0<\rho_2<1$ and r indicates the distance between sub clusters indexed by (s, t) and (i, j). The Kronecker notation $\delta(s, i)$ is 1 if $s = i$ and 0 otherwise. In addition, the conditional distribution of U_{ij}, given $U_* = u_*$, is assumed to depend on u_i only.

Furthermore, within each subcluster(i, j) there are n_{ij} individuals. Suppose that the cohort is stratified on the basis of one or more relevant covariates and these strata are indexed by $s = 1, 2, \ldots, a$. The (i, j, k) and s notation does not imply any inclusion relation between strata and clusters, e.g. males and females could correspond to different strata, whereas clusters and subclusters could be communities and families including both males and females. Let the hazard function for individual (i, j, k) from stratum s at time t be denoted by $\lambda_{ijk}^{(s)}(t)$. Given the random effects, we assume that the individual hazard functions are conditionally independent, with

$$\lambda_{ijk}^{(s)}(t) = \lambda_0^{(s)}(t) u_{ij} \exp(\beta^{\mathrm{T}} x_{ijk}^{(s)}). \tag{3}$$

Clearly, the survival times, either observed or censored, are spatially correlated. The distribution of random effects is assumed not to depend on the regression parameter β. Without loss of generality, we assume that the design matrix $X = (x_{111}^{(1)}, \ldots, x_{mJ_m n_{mJ_m}}^{(a)})^{\mathrm{T}}$ is of full rank. A Cox proportional hazards model with a single level of spatially correlated random effects is obtained

REVIEW DRAFT

as a special case of the Cox model with two levels of random effects by setting $v^2 = 0$ and $J_i = 1$ for all i.

Our assumptions (1) and (2) on random effects concern the spatial dependence and the first two moments only. This is desirable since the random mechanism by which the unobserved random effects were generated is usually not completely known[47].

Auxiliary Random Effects Poisson Models

As in Ma et al. [47], we made inferences on the random effects Cox models by fitting random effects Poisson models. Let $\tau_{s1}, \ldots, \tau_{sq_s}$ denote the distinct failure times in the sth stratum, with m_{sh} indicating the multiplicity of failures occurring at time τ_{sh} $(s = 1, \ldots, a)$. The risk set at time τ_{sh} is a subset of stratum s: $R(\tau_{sh}) = \{(i, j, k): t_{ijk} \geq \tau_{sh}\}$ where t_{ijk} is the observed survival time for individual (i, j, k) from the sth stratum. In addition, let $Y_{ijk,h}^{(s)}$ be 1 if a failure occurs for individual (i, j, k) from the sth stratum at time τ_{sh} and 0 otherwise. Let Y and U denote the vectors of the $Y_{ijk,h}^{(s)}$ and the random effects U_i and U_{ij}, respectively. Given the random effects $U = u$, Peto's version of the conditional partial likelihood [65, p.103] is

$$\ell_p(\beta; Y \mid u) = \prod_{s=1}^{a} \prod_{h=1}^{q_s} \frac{\prod_{(i,j,k)\in\Re(\tau_{sh})} u_{ij}^{Y_{ijk,h}^{(s)}} \{\exp(x_{ijk}^{\mathrm{T}}\beta)\}^{Y_{ijk,h}^{(s)}} (m_{sh}!)}{\left\{\sum_{(i,j,k)\in\Re(\tau_{sh})} u_{ij} \exp(x_{ijk}^{\mathrm{T}}\beta)\right\}^{m_{sh}}}. \tag{4}$$

We now define an auxiliary random effects Poisson regression model. Assume that the components of Y are conditionally independent, given random effects $U = u$, with P

$$Y_{ijk,h}^{(s)} \sim \mathrm{Po}\{u_{ij} \exp(\alpha_{sh} + x_{ijk}^{\mathrm{T}}\beta)\} (i, j, k) \in \Re(\tau_{sh}). \tag{5}$$

Given the random effects, the conditional likelihood for the random effects Poisson model is

$$\ell(\alpha, \beta; Y \mid u) = \prod_{s=1}^{a} \prod_{h=1}^{q_s} \frac{\prod_{(i,j,k)\in R(\tau_{sh})} u_{ij}^{Y_{ijk,h}^{(s)}} \{\exp(\alpha_{sh} + x_{ijk}^{\mathrm{T}}\beta)\}^{Y_{ijk,h}^{(s)}}}{\exp\left\{\sum_{(i,j,k)\in R(\tau_{sh})} u_{ij} \exp(\alpha_{sh} x_{ijk}^{\mathrm{T}}\beta)\right\}}. \tag{6}$$

Since the random effects vector does not depend on the regression parameter vector, as in Ma et al.[47], we can show that

$$\ell(\hat{\alpha}, \hat{\beta}; Y, U) = \prod_{s=1}^{a} \left\{ \prod_{h=1}^{q_s} \frac{m_{sh}^{m_{sh}} \exp(-m_{sh})}{m_{sh}!} \right\} \ell_p(\hat{\beta}; Y, U),$$

regardless of covariance structures assumed for random effects. The term in parentheses on the right-hand side does not depend on the parameters of interest. This demonstrates that the maximum joint Poisson likelihood estimators for the regression parameter vector β from (6) are the maximum joint partial likelihood estimators for the regression parameter vector β from (4). In addition, the nonparametric estimator of the cumulative baseline hazard function remains the same as given in Ma et al.[47].

REVIEW DRAFT

Orthodox Best Linear Unbiased Predictor Approach

Prediction of random effects

As in Ma et al. [47], we predicted the random effects by the following orthodox best linear unbiased predictor of U given Y.

$$\hat{U} = E(U) + \mathrm{cov}(U,Y)\mathrm{cov}^{-1}(Y,Y)\,(Y - EY), \tag{7}$$

where cov(Y,Y) denotes the marginal covariance of Y instead of the conditional covariance of Y given U. This is the linear unbiased predictor of U given Y which minimizes the mean squared distance between the random effects U and their predictor within the class of linear functions of Y.

Unlike the case of nested random effects in Ma et al.[47], the explicit expression for the inverse of var(Y) is no longer available with spatially correlated random effects. We therefore compute the random effects predictor of U given Y given in (7) through numerically inverted var(Y). However, the order of covariance matrix var(Y) is ACS data of air pollution and mortality between 1982-2000, the order of this matrix is over 47 million or over 6 million for California. Since the matrices Cov(Y,Y) and Cov(U,Y) are dense matrices, even the computer memory required by these matrices may create a serious problem for the ACS study data.

To facilitate the computation, we have derived the following sparse representations for Cov(U,Y) and Cov(Y,Y) after some algebra.

$$\mathrm{cov}(U,Y) = \mathrm{cov}(U, U)\, B^{\mathrm{T}} \text{ and } \mathrm{cov}(Y,Y) = \mathrm{diag}(EY) + B\,\mathrm{cov}(U,U)B^{\mathrm{T}}, \tag{8}$$

where diag (EY)denotes the diagonal matrix with EY on its diagonal. The matrix B is a sparse matrix of the same order as that of cov(Y,U)= cov(U,Y)^{T} with column i of B corresponding to cluster i. The elements of column i of B are zeros except being $\mu_{ijk,h}^{(s)} = \exp(\alpha_{sh} + \beta^{\mathrm{T}} x_{ijk}^{(s)})$ at the positions corresponding to those of $\mathrm{cov}(Y_{ijk,h}^{(s)}, U_i)$ in the matrix cov(Y,U).

These sparse representations not only make the computer memory feasible, but also make the inverting problem of cov(Y,Y) tractable as follows. Let matrices diag(EY)and cov(U,U) be denoted by A and D, we have

$$\mathrm{cov}^{-1}(Y,Y) = (A + BDB^{\mathrm{T}})^{-1}$$

$$= A^{-1} - A^{-1}B(B^{\mathrm{T}}A^{-1}B + D^{-1})^{-1}B^{\mathrm{T}}A^{-1},$$

where A^{-1}=diag$(EY)^{-1}$and D = cov(U,U)is generally small enough to be inverted numerically. In fact, there is a similar sparse representation of cov(U,U); therefore the inverse of cov(U,U)can be obtained through inverting numerically the much smaller covariance matrix of cluster level random effects.

The mean squared distances between the random effects U and its predictor can now be evaluated through the following equation.

$$\mathrm{cov}(\hat{\mathbf{U}} - \mathbf{U}, \hat{\mathbf{U}} - \mathbf{U}) = \mathrm{cov}(\mathbf{U}, \mathbf{U}) - \mathrm{Cov}(\mathbf{U}, \mathbf{Y})\mathrm{cov}^{-1}(\mathbf{Y})\mathrm{Cov}(\mathbf{Y}, \mathbf{U}).$$

In addition, we have the following two desirable orthogonality properties concerning the orthodox BLUP predictor:

$$\mathrm{cov}(\hat{\mathbf{U}} - \mathbf{U}, \hat{\mathbf{U}}) = \mathbf{0} \text{ and } \mathrm{cov}(\hat{\mathbf{U}} - \mathbf{U}, \mathbf{Y}) = \mathbf{0}.$$

REVIEW DRAFT

Estimation of regression parameters

Consider first estimation of the regression parameters in the case of known dispersion parameters.

As in Ma et al. [47], we can estimate the regression parameters through an optimal estimating function. Differentiating the joint log-likelihood of the auxiliary model for the data and random effects yields the joint score function. Replacing the random effects with their predictors, we have an unbiased estimating function for the regression parameters $\gamma = (\alpha^{\mathrm{T}}, \beta^{\mathrm{T}})^{\mathrm{T}}$:

$$\psi(\gamma) = \sum_{s=1}^{a} \sum_{h=1}^{q_s} \sum_{(i,j,k)\in\Re(\tau_{sh})} \left\{ Y_{ijk,h}^{(s)} - \hat{U}_{ij}(\gamma)\mu_{ijk,h}^{(s)}(\gamma) \right\} \tag{9}$$

The sensitivity matrix $S(\gamma)$ and the variability matrix $V(\gamma)$ are defined by

$$S(\gamma) = E_{\gamma}\left\{ \frac{\partial \psi(\gamma)}{\partial \gamma^{\mathrm{T}}} \right\},$$

$$V(\gamma) = E_{\gamma}\left\{ \psi(\gamma)\psi^{T}(\gamma) \right\}.$$

According to Ma et al. [47], we have the following global matrix expression for estimating function $\psi(\gamma)$ as their proof holds regardless of the covariance structure assumed for random effects.

$$\psi(\gamma) = X^{\mathrm{T}}\mathrm{diag}(EY)\ \mathrm{cov}^{-1}(Y,Y)\ (Y - EY)\ , \tag{10}$$

Similarly, we have

$$S(\gamma) = -\ V(\gamma) = 0 - X^{\mathrm{T}}\mathrm{diag}(EY)\ \mathrm{cov}^{-1}(Y,Y)\mathrm{diag}(EY)\ X. \tag{11}$$

With an appropriate partition of matrix $C = X^{\mathrm{T}}\mathrm{diag}(EY)\ \mathrm{cov}^{-1}(Y,Y) = (C_1, \ldots, C_m)$, it follows from (10) that

$$\psi(\gamma) = \sum_{i=1}^{m} C_i(Y_i - EY_i) = \sum_{i=1}^{m} \psi_i(\gamma),$$

where the unbiased estimating function $\psi_i(\gamma) = \mathbf{C}_i\ (Y_i - EY_i)$ corresponds to the ith cluster. The estimating function $\sum_{i=1}^{m} \psi_i(\gamma)$ can easily be shown to be optimal in the sense that it attains the minimum asymptotic covariance for the estimator $\hat{\gamma}$ among a certain class of linear functions of Y [66, 67].

The solutions $\hat{\gamma}$ of the estimating equation $\sum_{i=1}^{m} \psi_i(\gamma) = 0$ provide estimators of the regression parameters; however, the unbiased estimating functions $\psi_1(\gamma), \ldots, \psi_m(\gamma)$ are no longer independent because of spatial dependence. Under mild regularity conditions, it can be shown the component-wise asymptotic normality of parameter estimator $\hat{\gamma}$ [68, 69]. Specifically, for any constant vector of appropriate dimension b, $b^{\mathrm{T}}\hat{\gamma}$ is asymptotically normal with asymptotic mean $b^{\mathrm{T}}\gamma$ and asymptotic variance given by $-b^{\mathrm{T}}S^{-1}(\gamma)b$ as $m \to \infty$.

REVIEW DRAFT

The Newton scoring algorithm introduced by Jørgensen et al. [70] can be used to solve this estimating equation $\sum_{i=1}^{m} \psi_i(\gamma) = 0$. This algorithm, which is the Newton algorithm, but with the derivative of $\sum_{i=1}^{m} \psi_i(\gamma) = 0$ replaced by its expectation $S(\gamma)$, gives the following updated value for γ:

$$\gamma^* = \gamma - S^{-1}(\gamma)\psi(\gamma).$$

The computation of $S(\gamma)$ can be realized through (11).

Estimation of dispersion parameters

With the case of dispersion parameters unknown, we use moment estimates for the dispersion parameter with bias correction to give unbiased estimating function for dispersion parameters. The detailed process of estimating dispersion parameters is given in Appendix B and thus is omitted here. Unlike in previous approaches in the literature, the asymptotic variance of our regression parameter estimator is not affected by variability in the dispersion parameter estimators due to insensitivity of our estimation function to dispersion parameters[71]. The computational procedures regarding the initial values and iteration steps are exactly as outlined in Ma et al.[47].

Individual and Ecologic Covariates

A total of 20 variables were used to control for individual characteristics that might confound or modify the association between air pollution and death. These individual variables included individual-level risk factors for which data had been collected in the CPS II questionnaire, and were considered to be of potential importance on the basis of previous studies, as per our recent analyses of this cohort *vis-a-vis* ozone and $PM_{2.5}$ mass air pollution[72]. Variables considered included for tobacco smoking, education, marital status, body mass index (BMI) and BMI squared, alcohol consumption variables, occupational exposure, and diet indices (accounting for fat consumption and consumption of vegetables, citrus, and high-fiber grains).

The sensitivity of the $PM_{2.5}$ mortality risk estimates to modeling approaches and assumptions was evaluated by estimating Cox proportional hazard models both: (1) with and without the random effects; and, (2) with and without socioeconomic contextual variables.

Specifically, as mentioned earlier, six ecologic covariates obtained from the 1990 U.S. Census (median household income, the proportion of persons living in households with an income below 125% of the poverty line, the percentage of persons over the age of 16 years who were unemployed, the percentage of adults with less than a high-school [12th-grade] education, the Gini coefficient of income inequality [ranging from 0 to 1, with 0 indicating an equal distribution of income and 1 indicating that one person has all the income and everyone else has no income] and the percentage of persons who were white) were also included. These variables were included at ZIP code level, but sensitivity analyses were also conducted with random effects at the county and air basin levels as the average for the county or air basin level and as the difference between the average for the ZIP code of residence and the average for the metropolitan statistical area. Because results were insensitive to inclusion of these larger clusters, our results are reported with and without clusters at the ZIP code area.

REVIEW DRAFT

Control for Urban-Rural Differences

Initial evaluation of mortality patterns indicated likely differences between urban and rural areas. This pattern is consistent with what has been termed the "non-metropolitan mortality penalty," which was present throughout the entire study period and increased in size from the start to the end of follow up by more than a factor of 11. Specifically across the United States, in the 1980s there were on average 6.2 excess deaths per 100,000 in non-metropolitan areas compared to metropolitan areas, and this number increased to 71.7 excess deaths for the period 2000-2004 [73]. We therefore conducted analyses by including indicator variables representing the five largest metropolitan conurbations in California. Four of these conurbations were classified as "combined statistical areas": (1) Fresno-Madera; (2) Los Angeles-Long Beach-Riverside; (3) Sacramento-Arden-Arcade-Truckee-Nevada; and (4) San Jose-San Francisco-Oakland. One area, San Diego County, was classified as a primary statistical area, but we included this as a fifth conurbation given its size and urban structure. These are the largest urban conurbations in California and were selected to differentiate from rural and metropolitan areas that have different determinants of health and illness.

Because we had prior evidence of effects on all cause and cause-specific mortality in the Los Angeles Metropolitan Statistical Area (LA MSA) and examination of the spatial patterns in unpredicted mortality suggested the residual mortality there was particularly low for all causes and high for cardiovascular disease, we included an indicator for the LA MSA as a sensitivity test. Los Angeles is also different from the other conurbations because it is much larger than any of the other areas, with a population about 2.5 times greater than the next largest area in California. We also included an interaction between the LA MSA and pollution to determine whether the effects in the LA MSA were different from the rest of the state.

The Association between Time-Dependent Air Pollution Exposure and Mortality

Previous analyses in this report have assigned each cohort member an air pollution exposure value based on long-term concentration averages. Using this exposure assignment approach, we used ambient concentrations to create long-term averages that include values recorded both prior to and after a subject's event time or time to death or censoring. The estimated hazard ratios are interpreted as the summary effect of a subject living in an environment with an assigned pollution value on their mortality rate, regardless of the temporal sequence of exposure. However, changes in exposure over time may be related to risk of death in addition to their long-term location-specific exposure profile. To examine this temporal association with mortality we took advantage of the monthly estimates of both NO_2 and O_3 from 1988 to 2000.

For each subject we created an estimate of their sub-acute exposure by constructing 12-month moving averages from January 1988 to December 2000. For example, the exposure assigned to a subject in January 1989 was based on the average of the 12 monthly estimates of exposure from February 1988 to January 1989. We call this "sub-acute" because we are not using the entire exposure time averaged as in other analyses, but this is longer than the acute effects that we would normally see in a lag of 1-5 days before the death event. We assumed that each subject

REVIEW DRAFT

resided at their home address in 1982 throughout the follow-up period to December 2000. Let $x_i(t)$ represent the average pollution concentration assigned to the ith subject in the previous 12-month time period (in months) t and let $\bar{x}_i$ be the long-term exposure assigned to the ith subject based on the average concentration from 1988 to 2000.

We selected a 12-month averaging period for the sub-acute exposure assignment for two reasons. First, using a complete calendar year avoids further adjustment for seasonal cycles in both mortality and air pollution. Second, we wanted the clearest possible differentiation between the sub-acute and long-term exposure measures to have the greatest possibility to disentangle the effects of these two exposure measures on mortality. The longer the sub-acute averaging time the more correlated the two measures, so we selected the shortest exposure period that covered all seasons – the 12-month moving average. We also selected cardiovascular deaths to illustrate the effects of exposure assignment on mortality since this group of underlying causes of death was most strongly associated with long-term air pollution related mortality. We restricted the follow-up time of the cohort members to January 1989 to December 2000 since we required a 12-month historical average as one of the exposure assignments with the first month of pollution data being January 1988.

We then examined three survival models. Each model contained our 20 mortality risk factor variables measured at the individual level and six ecological covariates measured at the ZIP code area level. The standard Cox model was used for these analyses since there was little difference between the Cox and Random Effects survival models in our previous analyses. The first model contained the long-term exposure measure $\bar{x}_i$, the second model contained the 12 month moving averages of pollution exposure $x_i(t)$, and the third and final model contained two exposure measures; $\bar{x}_i$ and $x_i(t)-\bar{x}_i$. The corresponding log-relative risk parameters are denoted as: β, γ, ϕ and λ respectively.

These parameters can be interpreted in the following manner. If there is no sub-acute association with mortality and only long-term effects, then

$\beta = \gamma$ and $\beta = \phi$ with $\lambda = 0$.

If there is both a long-term association with mortality and a sub-acute effect then

$\gamma > \beta$ with $\phi > 0$ and $\lambda > 0$.

The relative size of the parameters ϕ and λ represent the impact of long-term and sub-acute effects on mortality respectively. Within this model structure, if there exists a sub-acute exposure association with mortality it will be reflected in the estimate of the long-term effect β as long as there is some variation in exposure over space. Only the third model can appropriately separate the effects of long-term and sub-acute exposure on mortality.

REVIEW DRAFT

Results

Exposure Modeling Results

Inverse Distance Weighted Interpolation

Inverse distance weighted (IDW) exposure estimates were successfully assigned to the ACS cohort using ArcGIS 9.2 software (ESRI, Redlands, CA). For the main analyses, we averaged all available monthly estimates to form one estimate of exposure. For the time windows analyses, we used the monthly data and 12 month averages to examine the periods of exposure mostly likely to be associated with death. Further descriptive statistics and maps illustrating the IDW surfaces are provided below in the health effects results section.

Kriging $PM_{2.5}$

As the data were not intrinsically stationary we used a universal kriging model to first fit a second order polynomial surface to detrend the data. Once the data were reasonably stationary, we fit the empirical spatial covariance with the parametric K-Bessel function. Table 5 lists the parameters and procedural fitting techniques used in ArcGIS Geostatistical Analyst to derive the $PM_{2.5}$ kriging surface.

Table 5: $PM_{2.5}$ Kriging parameters used for data fitting and prediction

Modeling Procedure	*Class of Parameter*	Parameter Value
Transformation	*Yes*	Log
Detrending	*Polynomial*	2nd order
	Poly. weighting	560189.3
	Smoothing factor	0.5
	Major/minor semiaxis	343485.9
Semi-variogram	*Function*	K-Bessel
	# of Lags	8
	Lag size	18152
	Nugget	0.01473
	Measurement Error	0
	K-Bessel Parameter	2.1263
	Range	137642
	Anisotropy	None
	Partial Sill	0.0618
Searching Neighborhood	*Neighbors to include*	3<x<15
	Sector Type	Full
	Angle	90
	major/minor semiaxis	343485.9

Figure 4 illustrates the empirical and fitted model semi-variogram using the K-Bessel function. The semi-variogram is one of a number of graphical means by which to illustrate the spatial covariance. The diagram shows a good deal of scatter but the model does a good job of fitting through the center of mass in the data.

REVIEW DRAFT

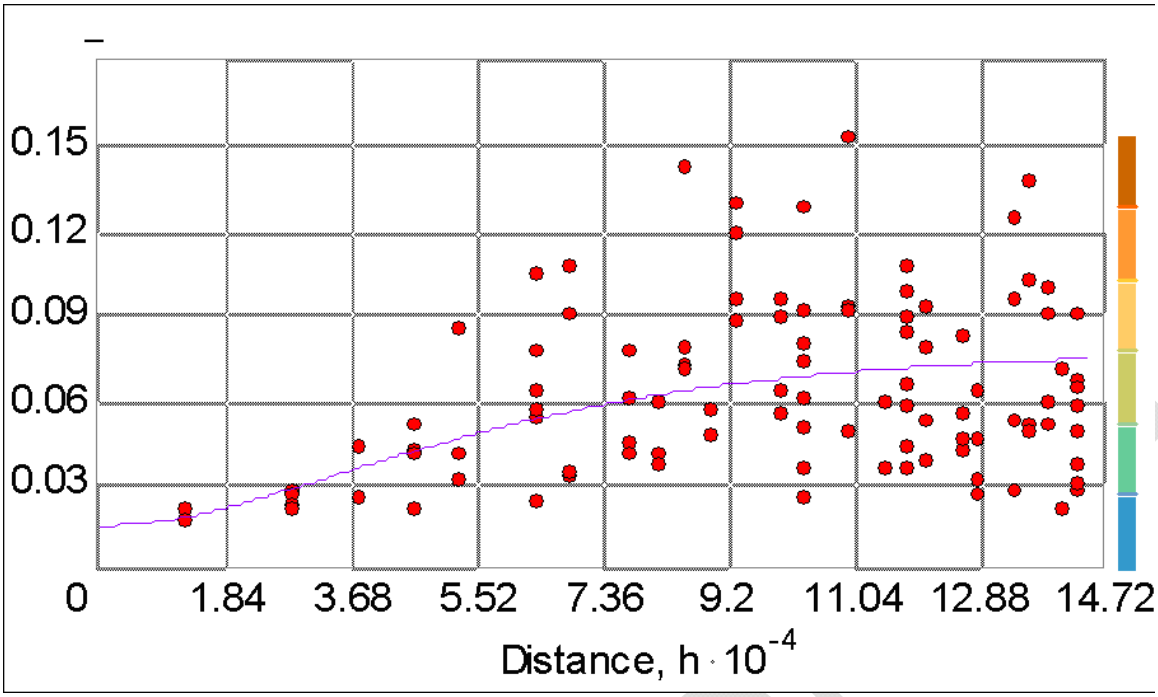


Figure 4: Semi-variogram used for $PM_{2.5}$ kriging model

Fitted estimates appear to show that there is no serious bias in the estimation. A simple regression (see footnote below[4]) of the predicted on observed values shows a coefficient of 1.01 and intercept of what is essentially zero indicates that there is no overall bias on the estimation. If these values deviated from one and zero, respectively, this could indicate bias in the estimate. Figure 5 illustrates the correspondence between predicted and observed values.

[4]Regression output from a regression of predicted values on observed values

```
      Source|       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  1,    70) =  211.49
       Model |   1811.2634     1   1811.2634           Prob> F       =  0.0000
    Residual |  599.510763    70  8.56443948           R-squared     =  0.7513
-------------+------------------------------           Adj R-squared =  0.7478
       Total |  2410.77416    71  33.9545657           Root MSE      =  2.9265

------------------------------------------------------------------------------
    observed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   predicted |   1.010312   .0694727    14.54   0.000     .8717527     1.14887
       _cons |  -.1069909   1.105337    -0.10   0.923    -2.311517    2.097535
------------------------------------------------------------------------------
```

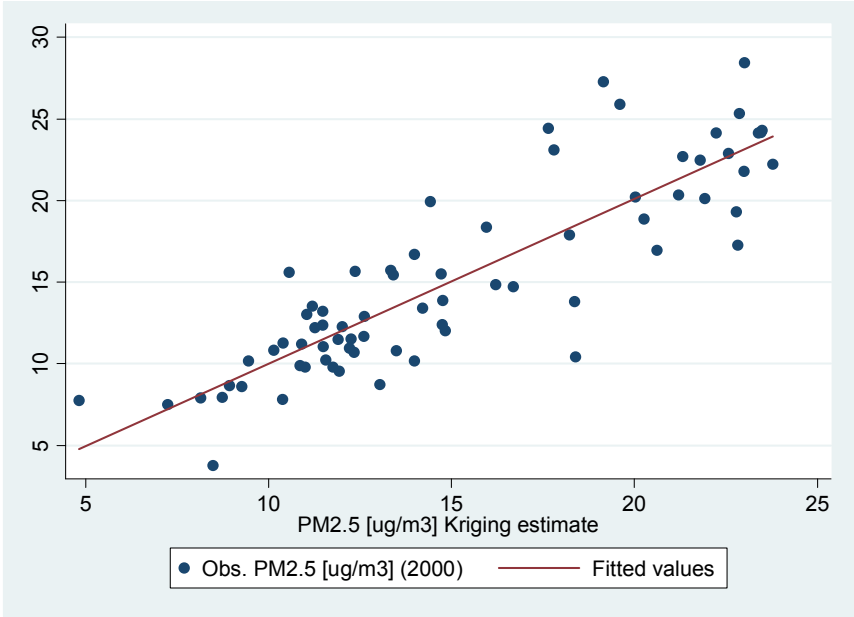


Figure 5: Kriging $PM_{2.5}$ estimates plotted against corresponding observations

Tests of the normality of the distributions of both the observed data and the residual kriging errors show that the log-transformation was important in ensuring efficient estimation of the predicted values. Using the Shapiro-Wilks test for normality, we can see from the output attached below that prior to modeling there is sufficient evidence to suggest that the observed data are different from normally distributed while the errors do not appear to be different than normally distributed at an alpha-level of 0.05 (Table 6). In the output below, **W** and **V** are scale parameters used to assess the statistical significance of the test using the Wald statistics **z**.

Table 6: Shapiro-Wilk test for normality

Shapiro-Wilk W test for normal data					
Variable	**Obs**	**W**	**V**	**z**	**Prob>z**
Measured	72	0.94158	3.679	2.837	0.00228
Error	72	0.97392	1.642	1.081	0.13993

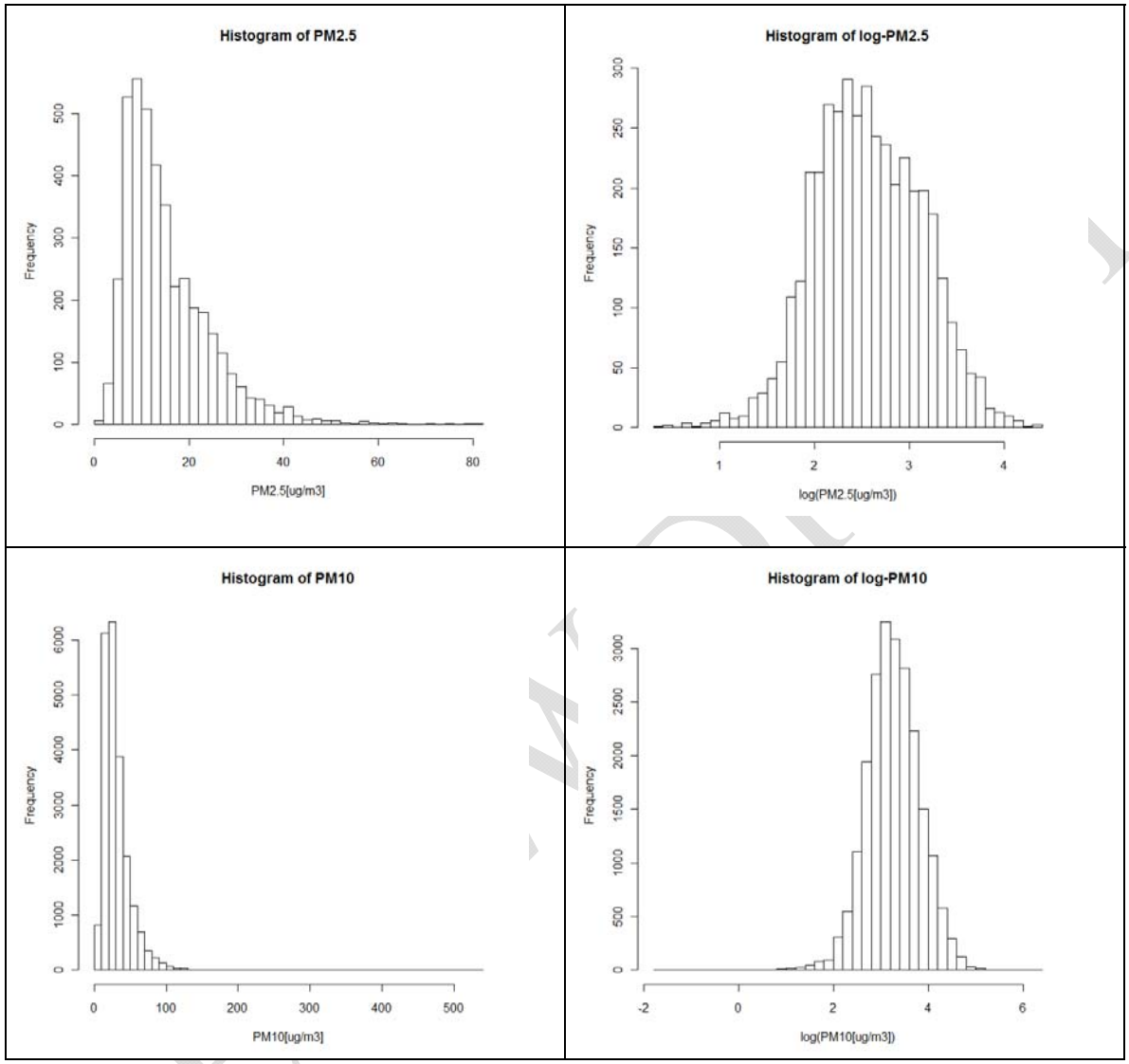
We can conclude that the $PM_{2.5}$ kriging model provides us with an efficient estimator of ambient annual $PM_{2.5}$ exposure throughout California for the year 2000 as the estimates appear unbiased, with normally distributed errors and a small root mean square error of 2.93.

Land Use Regression

As the distributions of $PM_{2.5}$, PM_{10} and $NO_2$are left skewed, a natural logarithm transformation was applied to them (Figure 6). This transformed the data so as to make it conform more to a normal distribution and therefore decrease the likelihood that a violation of the assumption of homoskewdastic errors did not occur with the regression models. Below are the distributions of $PM_{2.5}$, PM_{10}, and NO_2 before and after transformation. The log distribution did a reasonable job of fixing the skew in the data for $PM_{2.5}$ and PM_{10}. The transformation of NO_2 was not as nearly

REVIEW DRAFT

as effective in creating a normal distribution; it resulted in a right skew. However, the transformation did remove the problem of having a hard lower bound to the data distribution at a value of zero.



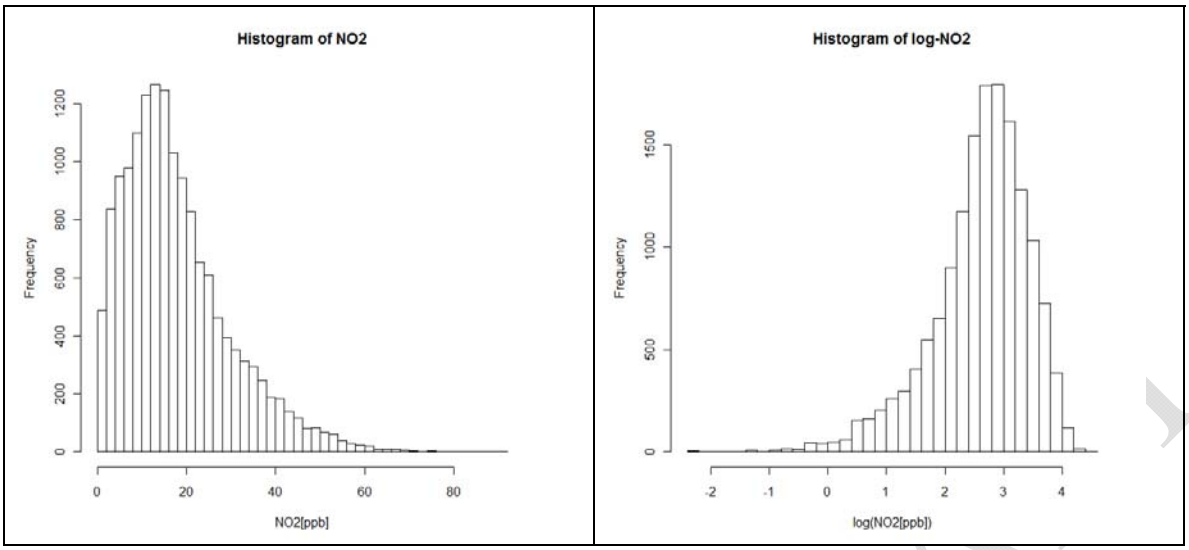


Figure 6: Histograms of raw and log-transformed pollution observations

After processing the observed pollution data through the model selection algorithm, the following three LUR models were selected and are outlined in Table 7 and 8, below.

Table 7: LUR model parameters for NO_2, $PM_{2.5}$ and PM_{10} models (see Table 8 below for description of pollutants and variables)

Pollutant	Variables	Beta	Std. Err	z	P>z	[95% Conf. Interval]	
logNO2	FCC12_5000	0.007317	0.0017	4.37	0.00	0.0040	0.0106
	FCC4_5000	0.001408	0.0002	8.02	0.00	0.0011	0.0018
	Open400^2	-0.000549	0.0001	-7.54	0.00	-0.0007	-0.0004
	Intercept	2.069401	0.0801	25.85	0.00	1.9125	2.2263
logPM2.5	logPM25_RS	0.612978	0.0639	9.59	0.00	0.4877	0.7383
	Open500	-0.006717	0.0031	-2.18	0.03	-0.0128	-0.0007
	Devlow5000	0.000107	2.71E-05	3.93	0.00	0.0001	0.0002
	Intercept	1.121769	0.1452	7.72	0.00	0.8371	1.4064
logPM10	Open5000	-0.000091	1.03E-05	-8.87	0.00	-0.0001	-0.0001
	Intercept	3.549622	0.0400	88.64	0.00	3.4711	3.6281

Table 8: Description of pollutants and variable short-hand names in regression output table (above)

Variable Name	Variable Description
logNO2	The logarithm transformation of observed monthly NO_2 (1988-2002)
logPM2.5	The logarithm transformation of observed monthly $PM_{2.5}$ (1998-2002)
logPM10	The logarithm transformation of observed monthly PM_{10} (1988-2002)
FCC12_5000	Length of expressways and highways within a 5000 meter buffer (km)
FCC4_5000	Length of residential roads within a 5000 meter buffer (km)
Open400^2	Area (in hectares) of open space within a 400 meter buffer to the second power
logPM25_RS	Remote sensing estimate of $PM_{2.5}$
Open500	Area (in hectares) of open space within a 500 meters buffer
Devlow5000	Area (in hectares) of low density developed land
Open5000	Area (in hectares) of open space within a 5000 meters buffer

REVIEW DRAFT

The cross-validation risk plots, Figures 7, 8, and 9 show the expected prediction error for each of the best models chosen by the DSA algorithm as a function model size during the model building/selection process. These plots show that the functional form of the models did not include any interaction terms, hence 'Interaction order 1' having been reported in the plots. In keeping with the DSA algorithm, we chose the model with size that produced the lowest CV-risk. Table 9 outlines the average cross-validated risk (CV risk) of the chosen models. Additionally, the minimum possible average prediction error for each of the datasets was reported in the table. As this was a longitudinal dataset, with unbalanced measurements at each monitoring location, and modeling was conducted with a set of explanatory variables that were not time varying, the theoretical maximum explanation of the observed variance would be to predict the mean value of the observed pollution levels per monitoring location. If this had been possible for all monitors the residual variance would equal the minimum possible CV risk under asymptotic conditions.

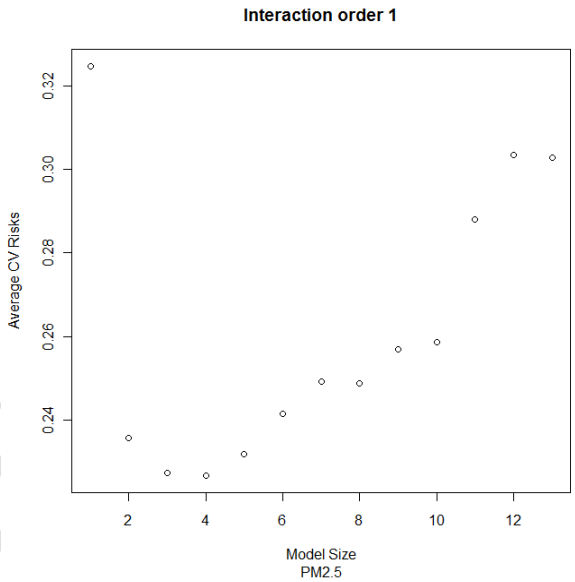


Figure 7: CV risk plots versus model size for $PM_{2.5}$ model

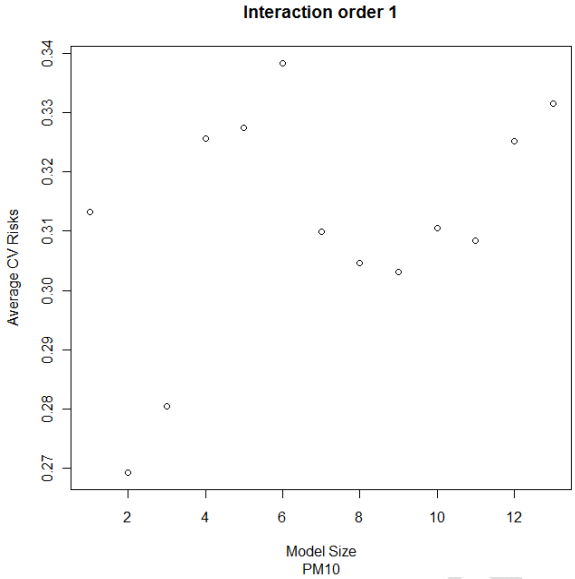


Figure 8: CV risk plots versus model size for PM_{10} model

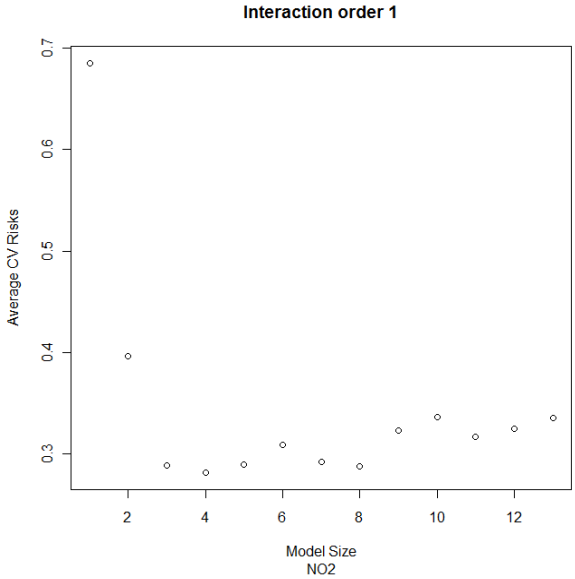


Figure 9: CV risk plots versus model size for NO_2 model

Table 9: Comparison of average cross-validation risks for selected models

Average CV Risk	NO_2	PM_{10}	$PM_{2.5}$
DSA	0.275	0.266	0.218
Minimum Possible Risk	0.118	0.156	0.163

The pseudo cross-validated risk R^2 values described below help to draw a parallel with the more familiar R^2 measure used in a cross-sectional OLS regression. The values reported in Table 9 are equal to:

$$pseudo\ R^2 = \frac{(\text{Variance of dataset - Average CV Risk})}{\text{Variance of dataset}}$$

Under asymptotic conditions, the pseudo R^2 represents the expected proportion of the variance in the observed data described by the model. We see the estimated maximum attainable R^2 given our modeling constraints reported under the random effects model (Zu), where Zu represent the monitor averages. When using the random effects model pseudo R^2 to normalize the R^2 by the DSA model, the DSA model explained 71%, 30% and 65% of the mean variability in NO_2, PM_{10}, and $PM_{2.5}$, respectively (Table 10).

Table 10: Pseudo-R^2 for selected models compared to maximum capable fit

Psuedo R^2	**NO_2**	**PM_{10}**	**$PM_{2.5}$**
DSA	0.59	0.15	0.32
Minimum Possible Risk	0.83	0.50	0.49
Normalized Psuedo R^2	0.71	0.30	0.65

Figure 10 illustrates the observed residuals on predicted values next to their empirical distribution of the residuals. The residuals of the models do not appear to violate the homoskewdasticity assumption and are nearly normally distributed; aside from a few outliers observed with each of the pollutants. Also, the errors do not appear correlated. PM_{10} appears to have a few very extreme residuals associated with the few extreme observed values. There are 174 residual values between 150 $\mu g/m^3$ and 539 $\mu g/m^3$ for PM_{10} out of a total of 21974 observations, which is less than 0.1% of the total observations. The remaining pollutants – NO_2 and $PM_{2.5}$ – do not appear to have such extreme prediction errors

.

REVIEW DRAFT

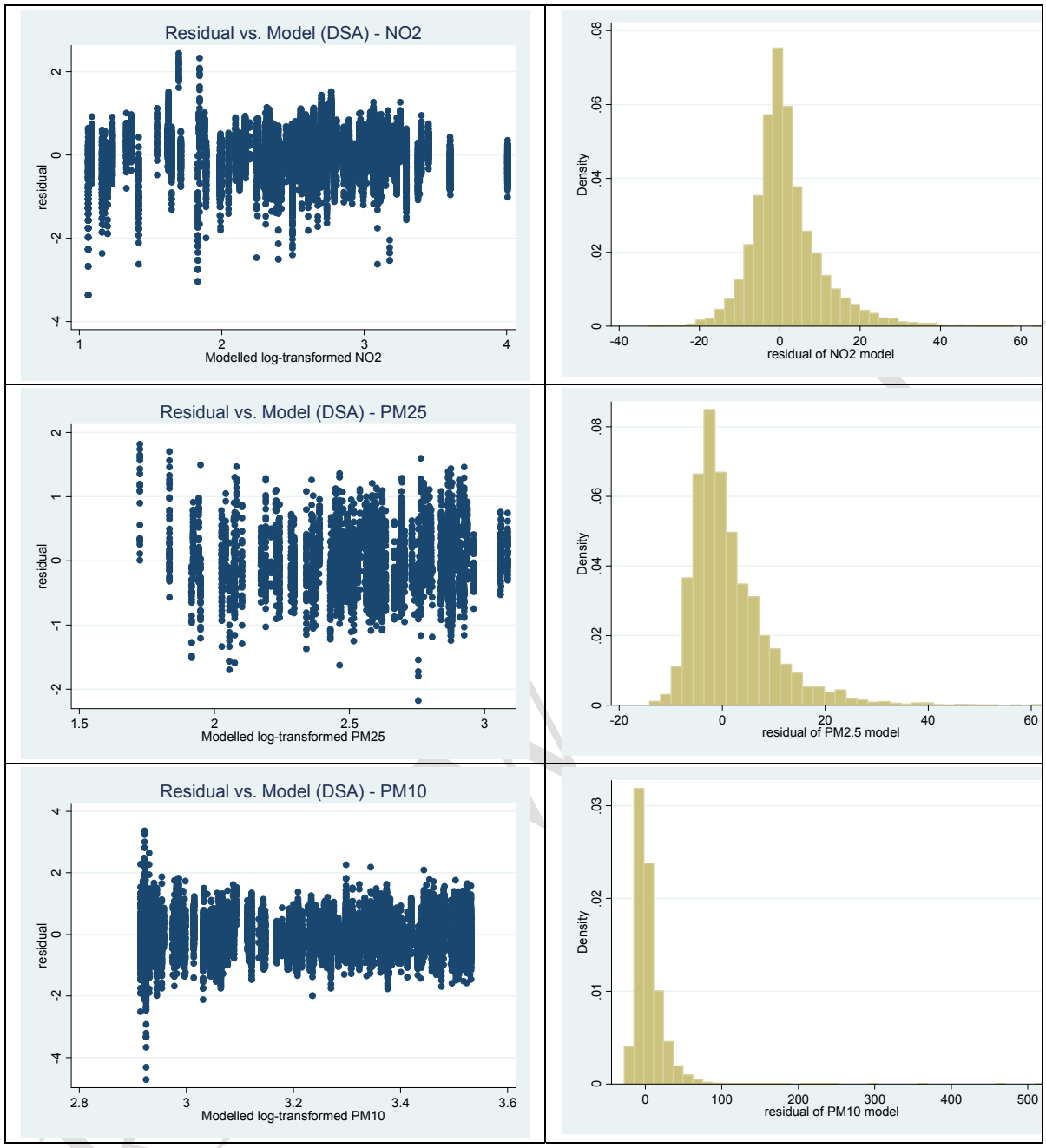


Figure 10: Residual plots and histograms of residuals for selected models

Bayesian maximum entropy interpolation of land use regression residuals

After the extracting the residuals of the LUR models for each of the three pollutant models, NO_2, PM_{10} and $PM_{2.5}$, respectively, we fit a BME spatiotemporal model to the resulting space-time model residuals.

With the residuals (also $X'(\boldsymbol{p})$ in Eq. 0 above) we may estimate space/time variability (experimental covariance, also see the circles in Figures 11-13) for a given spatial lag r and temporal lag τ using Eq. A4.

The experimental covariance reveals the temporal cyclic characteristics of the data as demonstrated in Figures 11-13, and it may be parameterized by sill ($c_{01} - c_{04}$ in Eqs. 12-13) and range ($a_{r1} - a_{r4}$, $a_{t1} - a_{t4}$ in Eqs. 10-11) in a covariance model that fits the experimental covariance. The covariance models (also red curves in Figures. 11-13) we selected are Eq. (12) for NO_2 and Eq. (13) for PM_{10} and $PM_{2.5}$, and their parameters are summarized in Table 11.

NO_2

$$c_X(r=|\boldsymbol{s}-\boldsymbol{s}'|,\tau=|t-t'|) = c_{01}\delta(r)\delta(\tau) + c_{02}\exp\left(\frac{-3r}{a_{r2}}\right)\exp\left(\frac{-3\tau}{a_{t2}}\right)$$
$$+ c_{03}\exp\left(\frac{-3r^2}{a_{r3}}\right)\cos\left(\frac{2\pi\tau}{a_{t3}}\right) + c_{04}\exp\left(\frac{-3r}{a_{r4}}\right)\sin\left(\frac{2\pi\tau}{a_{t4}}\right) \quad (12)$$

where $\delta(r)=1$ if $r=0$, $\delta(r)=0$ if $r>0$, $\delta(\tau)=1$ if $\tau=0$, $\delta(\tau)=0$ if $\tau>0$

PM_{10} and $PM_{2.5}$

$$c_X(r=|\boldsymbol{s}-\boldsymbol{s}'|,\tau=|t-t'|) = c_{01}\exp\left(\frac{-3r}{a_{r1}}\right)\exp\left(\frac{-3\tau}{a_{t1}}\right) + c_{02}\exp\left(\frac{-3r}{a_{r2}}\right)\exp\left(\frac{-3\tau}{a_{t2}}\right)$$
$$+ c_{03}\exp\left(\frac{-3r^2}{a_{r3}}\right)\cos\left(\frac{2\pi\tau}{a_{t3}}\right) + c_{04}\exp\left(\frac{-3r}{a_{r4}}\right)\sin\left(\frac{2\pi\tau}{a_{t4}}\right) \quad (13)$$

Table 11: Space/time covariance parameters for three air pollutants in California

Space/time covariance model parameters	NO_2 (Eq. 10)	PM_{10} (Eq. 11)	$PM_{2.5}$ (Eq. 11)
c_{01}	6.6086 (ppb)^2	125.3575 (μg/m^3)^2	22.5148 (μg/m^3)^2
c_{02}	48.7259 (ppb)^2	118.5815 (μg/m^3)^2	9.6492 (μg/m^3)^2
c_{03}	11.4649 (ppb)^2	47.4326 (μg/m^3)^2	16.0820 (μg/m^3)^2
c_{04}	11.4649 (ppb)^2	47.4326 (μg/m^3)^2	16.0820 (μg/m^3)^2
a_{r1}	N/A	80 km	38 km
a_{r2}	110 km	350 km	5 km
a_{r3}	600 km	280 km	550 km
a_{r4}	3500 km	1500 km	2500 km
a_{t1}	N/A	1 month	1 month
a_{t2}	8000 months	6000 months	3500 months
a_{t3}	12 months	12 months	12 months
a_{t4}	12 months	12 months	12 months

The sum of the sill components ($c_{01} - c_{04}$) for a pollutant becomes the overall variance for the pollutant, and each of the sill components explains how much it contributes to the total variation

REVIEW DRAFT

in the measurements. The space/time ranges ($a_{r1} - a_{r4}$, $a_{t1} - a_{t4}$) in each term of Eqs. 12-13 define spatiotemporal distances at which spatiotemporal covariance drops down to 5% of the variance. For the case of PM_{10} the first covariance element includes a spatial range of 80km and a temporal range of 1 month while the second covariance part has a spatial range of 350km and a temporal range of 6000 months. The first component may correspond to correlation structure within urban areas with high vehicle traffic and large number of industrial facilities, and monthly fluctuations in the data, while the second indicates regional variation of PM_{10} due to secondary formation organic carbon, sulfates and nitrates over a relatively long-term time period. Likewise the spatial ranges in the third and fourth components are associated with the regional variation. In particular the combination of the temporal ranges in the third and fourth components explains the temporal cyclic characteristics of the data, and both a_{t3} and a_{t4} denote the periodicity of the seasonal variation. In addition each of c_{03} and c_{04} indicates half amplitude of the seasonal variation.

The top and bottom plots in each of Figures. 11-13 denote purely spatial (when temporal lag τ=0) and purely temporal (when spatial lag r=0) covariances, respectively. For NO_2 the spatial piece (top plot in Figure 11) is a linear combination of the nugget effect (i.e., variance arising from measurement uncertainty and/or inherent variability over short space/time distances), exponential, and gaussian functions in *BMElib*, whereas the temporal portion (bottom plot in Figure 11) includes a linear combination of nugget effect, exponential function together with cosinusoidal and sinusoidal functions associated with the seasonal effects of NO_2 attribute. The covariance models of PM_{10} and $PM_{2.5}$ are similar to that of NO_2, but exclude the nugget effect from their models.

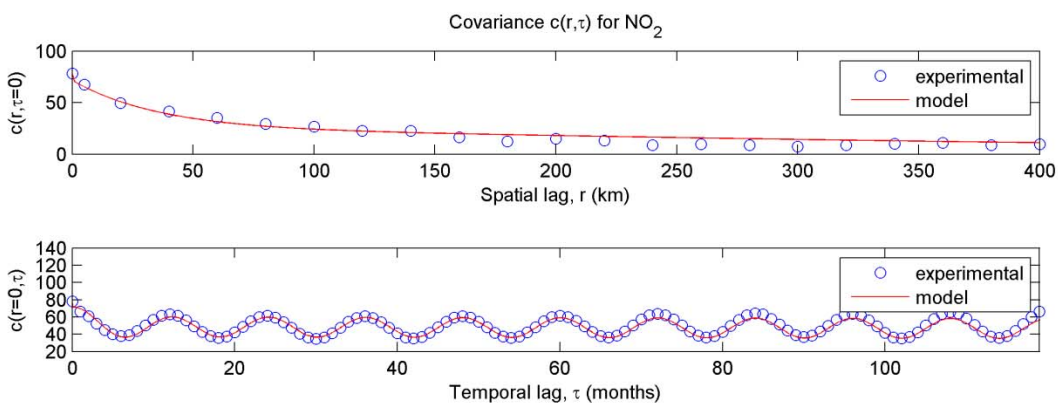


Figure 11: Space/time experimental covariance values (circles) and a covariance model fitted to the experimental values (curves) – NO_2 case

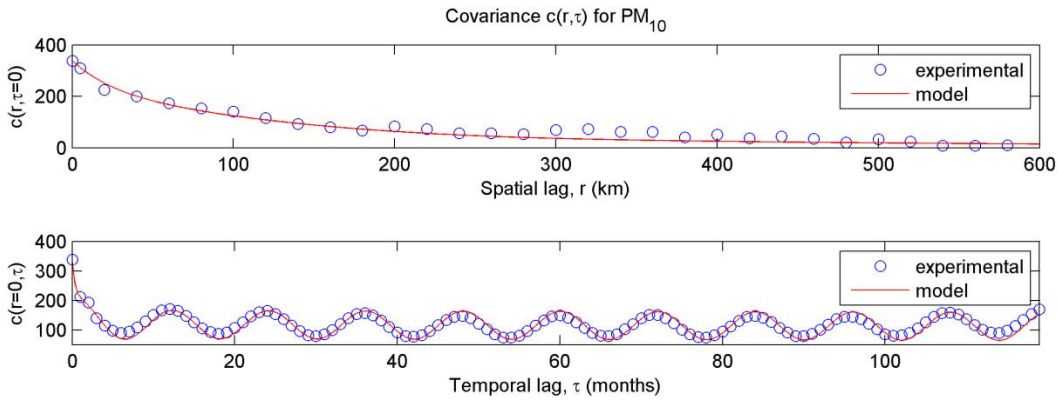


Figure 12: Space/time experimental covariance values (circles) and a covariance model fitted to the experimental values (curves) – PM_{10} case

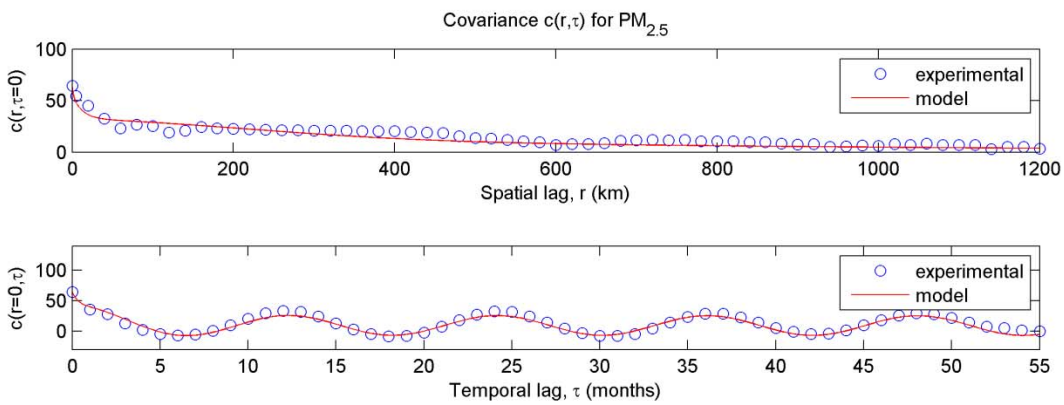


Figure 13: Space/time experimental covariance values (circles) and a covariance model fitted to the experimental values (curves) – $PM_{2.5}$ case

The covariance models obtained are used to construct an initial probability regarding air pollution distribution over space and time (Eq. C3 in Appendix C). The initial probability is then updated to the final probability of air pollution at spatiotemporal estimation points (i.e., 180 individual months in 1988-2002 for NO_2 and PM_{10}, each of 60 months in 1998-2002 for $PM_{2.5}$, and 10 km-gridded locations over the state of California for all of the pollutants). The final probability is equivalent to the posterior probability density function (Eq.C9) representing estimation outputs. The results of the BME models were then assigned to the monitoring locations to assess the combined LUR-BME model fit.

Combined spatiotemporal BME-LUR model

After the monthly BME estimates were assigned to the monitoring locations, they were added to the LUR estimates to get the spatiotemporal BME-LUR estimate that combined the overall mean model from the LUR and the temporal fluctuations of the pollutants for every month.

Figure 14, shows a comparison between the observed value on the predicted values for both the LUR only models and combined LUR-BME models for NO_2, PM_{10} and $PM_{2.5}$. These plots

illustrate that the combined LUR-BME model does a superior job in fitting the observed data in all cases. The temporal artifacts present in the data that could not be modeled with the LUR framework alone - as indicated by the vertical lines in the first column of plots - have been accounted for by the BME models. We see much less scatter about the one-to-one correspondence line indicating a better fit to the data. There appears to be little change in the one-to-one correspondence relationship as indicated by the slope of a mixed model regression of fitted on observed values. Table 12 reports these slopes – a slope of 1 indicates no bias in the estimate - and we see that the slope increases slightly above the LUR-only value for the BME-LUR models; however the magnitude of the difference between unity and the comparison of the slopes is small.

REVIEW DRAFT

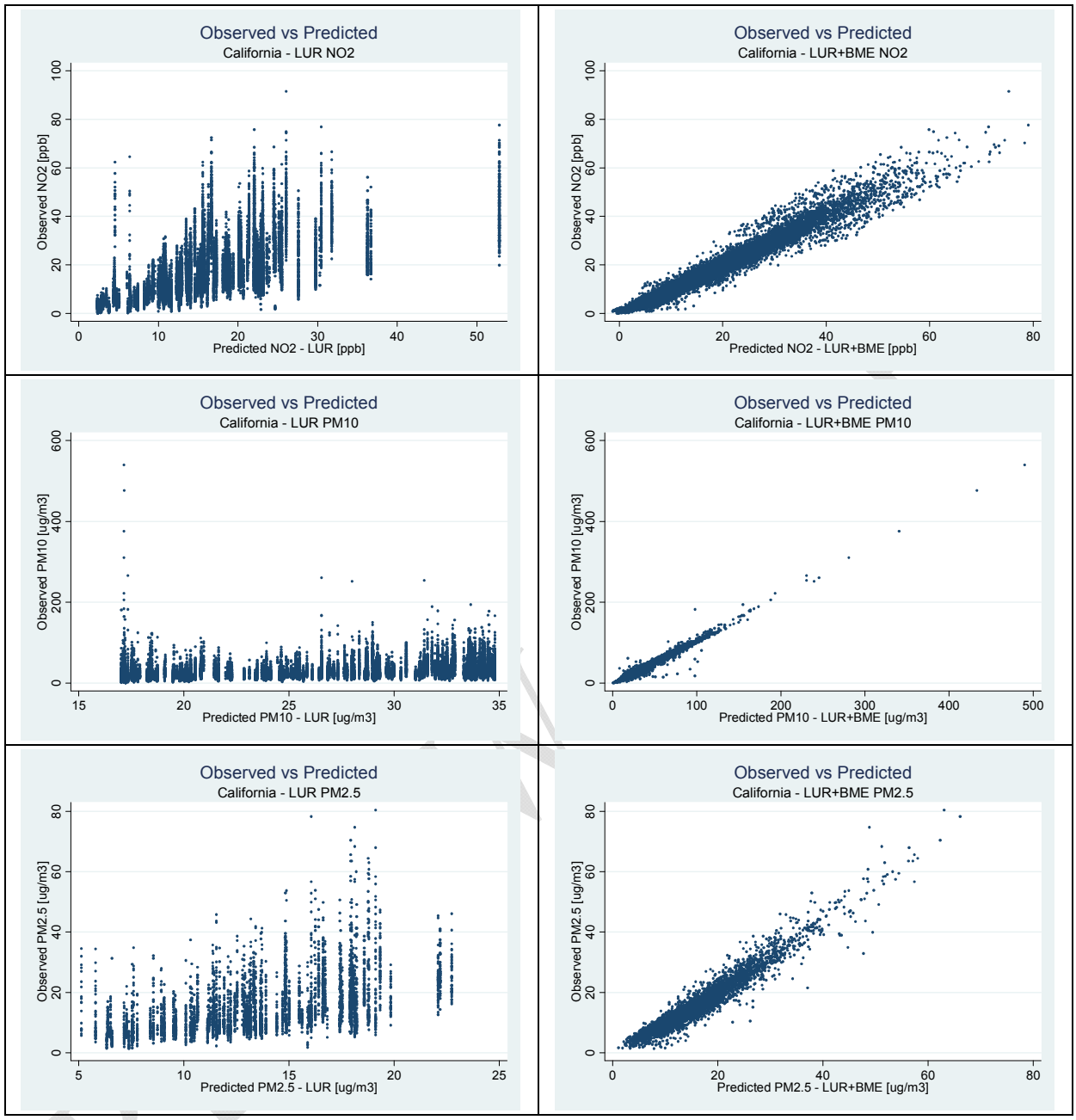


Figure 14: Comparison of predicted on observed plots for LUR only and LUR-BME models for NO_2, PM_{10} and $PM_{2.5}$, respectively.

Table 12: Slope of mixed model regression of predicted on observed data.

Pollutant	Slope LUR	Slope BME-LUR
NO_2	0.87	1.08
PM_{10}	0.88	1.05
PM_{25}	1.07	1.11

Figure 15 illustrates a comparison between the LUR only model residuals and the residuals of the combined LUR-BME model. In all three cases, the residuals of the LUR-BME model are smaller and less prone to extreme outliers due to high pollution events due to temporal considerations as are apparent in the LUR only models. The monitor specific error structures present in the LUR only model appear to be accounted for by the BME model with no obvious violation of the functional form or heteroskedasticity assumptions. Histogram comparison in Figure 16 further illustrates a better fit of the data that the BME-LUR models have compared to the LUR only model. The LUR-BME models do not show the skewed distribution of the LUR only models and have a tighter distribution when compared to their LUR only counterparts, i.e., the variance of the residuals is smaller. Table 13 illustrates summary statistics of the models' fit. The reported R^2 is the "overall R^2" from the random-effects generalized least squares (mixed model) output in Stata 10 (Stata Corp, College Station, TX). The R^2 for the LUR only model cannot be compared to the pseudo-R^2 reported earlier as the pseudo-R^2 was calculated on the CV-Risks and these are determined strictly by model fit.

REVIEW DRAFT

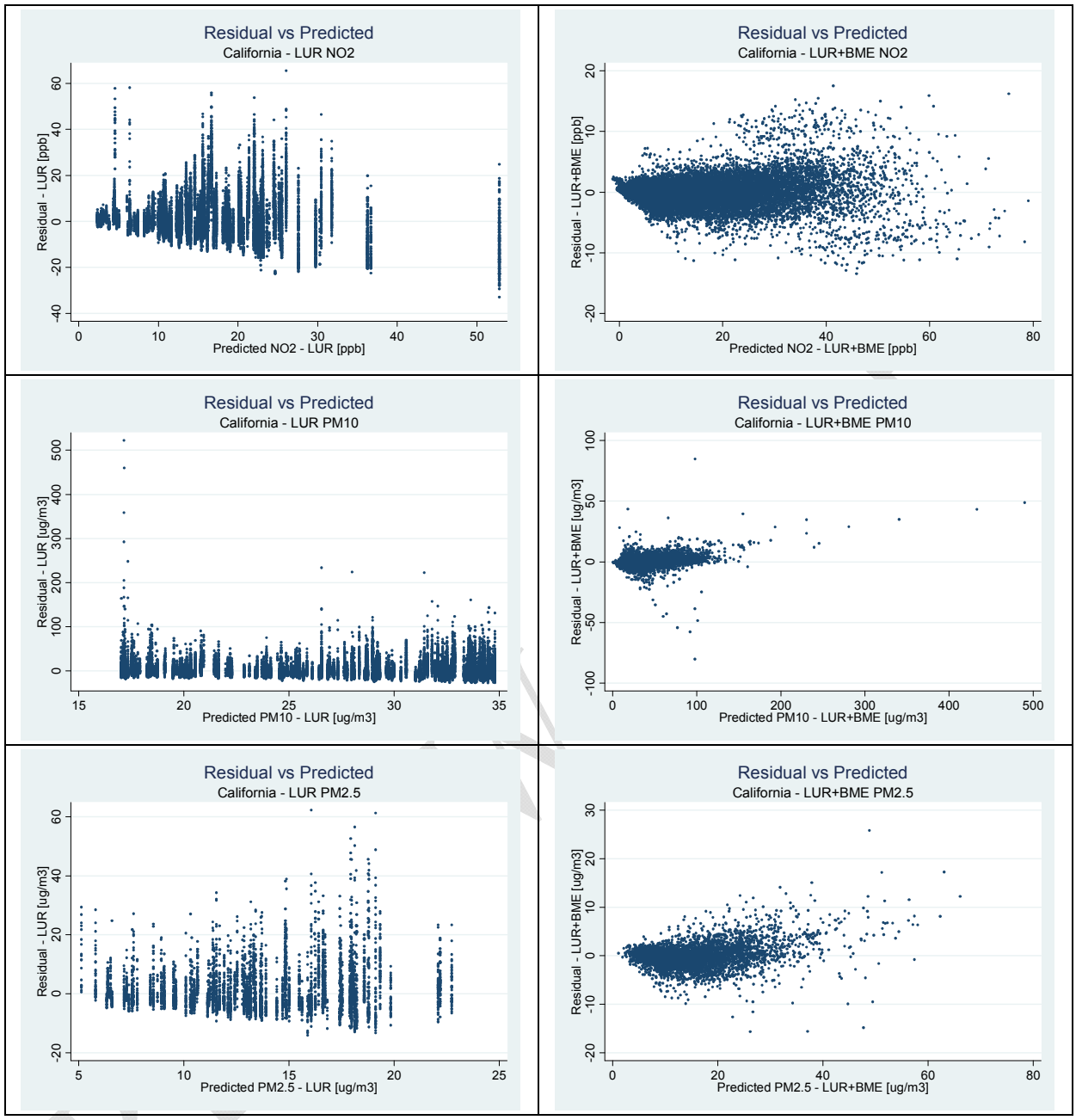


Figure 15: Comparative residual plots for LUR only and LUR-BME models for NO_2, PM_{10} and $PM_{2.5}$, respectively

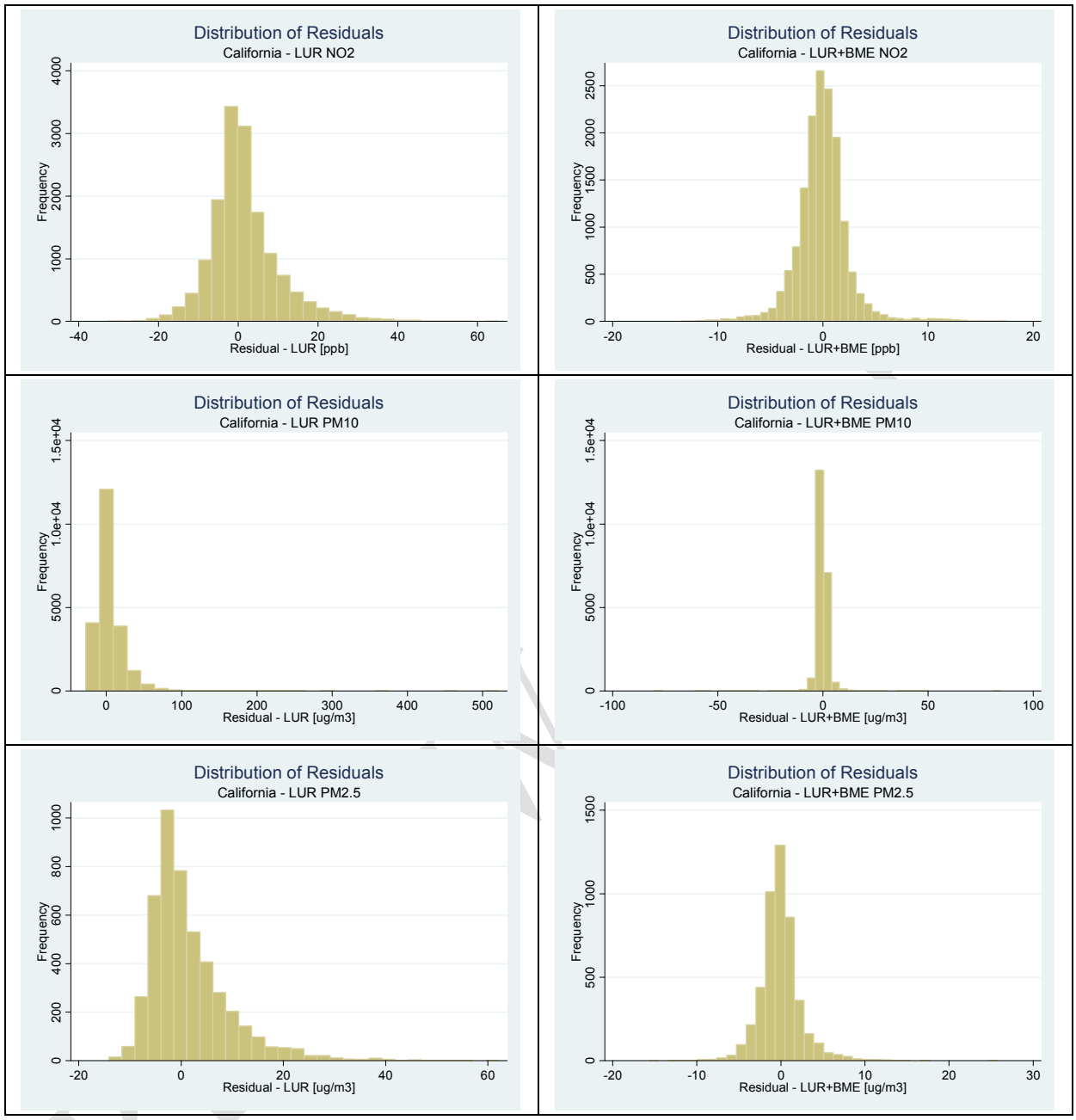


Figure 16: Comparative residuals histograms for LUR only and LUR-BME models for NO_2, PM_{10} and $PM_{2.5}$,respectively

Table 13: Summary statistics of models fits.

Pollutant	Obs	Std. Dev. Model		Std. Dev. Residuals		R^2	
		LUR only	LUR-BME	LUR only	LUR-BME	LUR only	LUR-BME
NO_2	15310	8.45	11.47	8.85	2.50	45%	96%
PM_{10}	21947	5.71	18.33	18.40	2.68	9%	98%
$PM_{2.5}$	4748	3.95	8.17	8.01	2.46	25%	94%

Table 13 shows that the data were better fit by the BME model for all three pollutants. When compared to the LUR only model, the LUR-BME model was better able to approximate the variability in the observed data as evident by both the larger standard deviation of the model and smaller variance on the residual. The R^2 for the LUR-BME models is 96%, 98% and 94% for NO_2, PM_{10} and $PM_{2.5}$, respectively; this represents a minimum two-fold improvement in the fit of the data for all pollutants and as much as a ten-fold improvement for PM_{10}. Some caution should be exercised when interpreting these results as the possibility exists that these results are indicative of some degree of over-fitting. In addition, the BME approach reproduces the monitoring values nearly exactly, with the difference being the 10 km grid. So these results are not surprising and cannot be used to determine how well the models fits data not used in the original prediction. Because of the paucity of monitoring sites we were unable to leave out cross-validation locations as we would customarily do in a stronger assessment of cross-validation. This suggests, however, that all pollutants have some element of local variation that is predicted well by the land use regression, but also a broader spatial pattern of regional differences in space and time that appear to be well captured by the BME estimates.

Geocoding Results

Overall match rates were excellent. More than 90% of geocodable addresses overall and 98% of the nutritional sub-cohort addresses were geocoded. Further details are included in the tables below.

REVIEW DRAFT

Table 14: Match rate for corrected addresses using the TeleAtlas (unstandardized) and StreetMap (composite) locators.

	TeleAtlas	StreetMap
Total Observations	146015	146015
Total Obs IN CA	142155	142155
California Obs*	105188	105217
Nutritional Obs**	21143	23065
Archived Obs	15824	17733
M, score=100	87.2	76.1
M, Score>80	87.7	84
M, score=100, CA	85.5	75.7
M, Score>80, CA	86.1	83.4
M, score=100, Nutri	94.4	81.3
M, Score>80, Nutri	94.8	82.4
M, score=100, Arch	88.6	71.9
M, Score>80, Arch	89.3	82.9
Missing address	1249	1249
Missing ZIP and City	309	309
No Numbers in Address	682	1938
"PO BOX"	3899	4213
"P O BOX"	807	814
" BOX "**	5310	5645
Approx Geocodeable***	134763	138232
M, score=100	91.9	80.3
M, Score>80	92.4	88.6
M, score=100, CA	90.1	79.8
M, Score>80, CA	90.6	87.9
M, score=100, Nutri	98.4	85.1
M, Score>80, Nutri	98.8	91.4
M, score=100, Arch	94.6	77.2
M, Score>80, Arch	95.3	89

*Note that 29 observations were coded in the SAS file as "CA" but the city and ZIP placed them outside of CA, these 29 are excluded in percentages

**Note that this has a space before and after to distinguish between addresses with the word "BOX" inside another word

*** Eliminate missing addresses, the different PO Box variants and space Box space, eliminate addresses with no numbers, eliminate those with no city AND no ZIP

PREVIEW DRAFT

Table 15: Overall match rates by ACS file. Note that these numbers include non-geocodeable addresses like PO Boxes etc.

Matches at Score of 100	0.87
Matches at Score of >80	0.88
CA Matches at Score of 100	0.86
CA Matches at Score of >80	0.87
Nutrition Matches at Score of 100	0.92
Nutrition Matches at Score of >80	0.93
Add Hist Matches at Score of 100	0.86
Add Hist Matches at Score of >80	0.88

Table 16: Approximate match numbers by locator*

	Locator			
	SM_Alpha	SM_Hyphen	SM_Streets	TeleAtlas
Addresshx	2	12	1,442	16,277
California	0	4	777	104,436
Nutrition	6	2	1,346	21,711

*Note we tabulated these numbers based on the full file rather than limiting to the matches so these numbers are close but approximate. SM is streetmap, alpha is alphanumeric and hyphen is hyphenated.

Table 17: Breakdown of records with a "C/O" representing Care Of

FALSE	TRUE
145,887	128

REVIEW DRAFT

Table 18: Counts by ACS file by state. Note that the California file has participant addresses at the beginning of the study.

File	State	Count
Addresshx	AE	5
Addresshx	AK	14
Addresshx	AL	10
Addresshx	AP	1
Addresshx	AR	35
Addresshx	AZ	240
Addresshx	CA	15824
Addresshx	CN	1
Addresshx	CO	83
Addresshx	CT	2
Addresshx	FL	92
Addresshx	GA	27
Addresshx	GU	1
Addresshx	HI	37
Addresshx	IA	13
Addresshx	ID	73
Addresshx	IL	21
Addresshx	IN	21
Addresshx	KS	9
Addresshx	KY	6
Addresshx	LA	4
Addresshx	MA	9
Addresshx	MD	13
Addresshx	ME	8
Addresshx	MI	13
Addresshx	MN	40
Addresshx	MO	28
Addresshx	MS	1
Addresshx	MT	17
Addresshx	NC	30
Addresshx	ND	1
Addresshx	NE	6
Addresshx	NH	4
Addresshx	NJ	10
Addresshx	NM	56
Addresshx	NV	147
Addresshx	NY	30
Addresshx	OH	17
Addresshx	OK	21
Addresshx	OR	202
Addresshx	PA	20
Addresshx	RI	10
Addresshx	SC	21
Addresshx	SD	4
Addresshx	TN	13
Addresshx	TX	91

REVIEW DRAFT

Addresshx	UT	100
Addresshx	VA	36
Addresshx	VT	9
Addresshx	WA	222
Addresshx	WI	22
Addresshx	WV	6
Addresshx	WY	7
California	CA	105188
California	CO	5
California	FL	1
California	GA	1
California	GU	1
California	ID	1
California	KY	1
California	LA	2
California	MA	1
California	MN	2
California	NC	1
California	NM	1
California	NV	4
California	PA	2
California	TX	1
California	VA	2
California	WI	3
Nutrition	AE	1
Nutrition	AK	5
Nutrition	AL	9
Nutrition	AP	2
Nutrition	AR	20
Nutrition	AZ	281
Nutrition	CA	21143
Nutrition	CO	98
Nutrition	CT	1
Nutrition	DC	1
Nutrition	FL	76
Nutrition	GA	23
Nutrition	HI	29
Nutrition	IA	17
Nutrition	ID	64
Nutrition	IL	17
Nutrition	IN	15
Nutrition	KS	9
Nutrition	KY	6
Nutrition	LA	6
Nutrition	MA	9
Nutrition	MD	7
Nutrition	ME	13
Nutrition	MI	19
Nutrition	MN	31
Nutrition	MO	32

REVIEW DRAFT

Nutrition	MS	4
Nutrition	MT	19
Nutrition	NC	23
Nutrition	ND	3
Nutrition	NE	7
Nutrition	NH	3
Nutrition	NJ	8
Nutrition	NM	42
Nutrition	NV	177
Nutrition	NY	16
Nutrition	OH	20
Nutrition	OK	16
Nutrition	OR	251
Nutrition	PA	24
Nutrition	RI	4
Nutrition	SC	14
Nutrition	SD	4
Nutrition	TN	7
Nutrition	TX	105
Nutrition	UT	104
Nutrition	VA	26
Nutrition	VT	4
Nutrition	WA	210
Nutrition	WI	29
Nutrition	WV	1
Nutrition	WY	10

REVIEW DRAFT

Health Effects Modeling Results

In this section we present detailed results of the health effects modeling. We begin with descriptions of the analytical cohort, the exposures, and the confounding variables. We then turn to reporting the mortality effects from the ZIP code and individual exposure assignments.

Description of the Final Analytical Dataset

From a total of 1,184,588 ACS CPS-II participants, there were 104,277 who resided at enrollment in the state of California (See Table 19; note: all tables for these results appear at the end of this section). Out of the 104,277 California participants, there were 97,165 for whom complete ZIP code data was available; for 5,640 participants a ZIP code was assigned according to available street address information. A total of 1,472 participants were excluded where no ZIP code could be assigned. Participants were also excluded from the final analytic cohort if their residential address contained a 'PO Box' or 'c/o' field (17,061) or if they could not be assigned to a latitude and longitude with a high degree of certainly based on available data (233). A total of 9,168 participants were excluded due to missing data on individual-level covariates of interest. In total 76,343 participants were retained for the analysis with some variation depending on the exposure model among which 20,432 deaths were observed through the year 2000. Various other exposures models used here had slightly more or less similar exclusion numbers based in part on the method used to assign exposures.

As shown in Table 20, which compares key individual variables that we have included in our statistical models, the composition of the California cohort is fairly similar to the National ACS cohort. The California cohort has a lower proportion of whites than the national cohort, but is still predominantly white. In California there are considerably more subjects with postsecondary education. California has a lower proportion of current smokers. More Californians are divorced than in the national cohort. On diet, Californians tend to consume less fat, and they report drinking more wine and liquor.

Descriptive Results for the Analytic Cohort

Tables 21a and 21b show the number of subjects with assigned exposures, the mean and variance of each exposure model, and the percentile distribution for each exposure model. These statistics are given for assignments at both the individual home address and the ZIP code of residence. In some cases there are differences in the underlying sample size available due to variations in the available monitors, the ZIP codes used to assign exposures, or the statistical procedure used to generate the exposure model. For example, some of the exposures were assigned in 2007-8, when a less complete ZIP code file that matched the ACS subjects was available in GIS format. Subsequent updates to this file were made and used for several of the estimates, but due to resource and time constraints, we were unable to update all estimates with the more recent and more comprehensive ZIP files.

In comparing estimates assigned at the individual address level for $PM_{2.5}$, the inverse distance weighting (IDW) exposure estimates tend to have a higher mean and variance than those estimated by the kriging model. The IDW estimates also have a much higher maximum value than the kriging models. The same estimates, when assigned to the ZIP codes rather than the

REVIEW DRAFT

home address, have closer values with the means being more similar. The variance of the IDW is still larger as is the maximum value, but the two estimates are much closer in their distributions when assigned to the ZIP code.

For ozone, the individual assignment using the IDW exposure estimate has about the same mean and variance as the BME model, but the BME model tends to have a smaller range. The same comparison for PM_{10} reveals that the BME model has about the same mean, but a much smaller variance and range than the IDW model. With NO_2, the mean and variances are substantially reduced with the BME compared to the IDW and the values are much lower overall for the BME.

Table 22 includes distribution of the ecologic covariates used in the analysis. These are based on the 1990 ZIP code tabulations. Not surprisingly given the predominantly white composition of the cohort, most live in neighborhoods that have a preponderance of whites. Many also live in neighborhoods with high proportions of postsecondary graduate earners, with relatively high incomes. The Gini coefficient suggests that even at the 50th percentile, there is often considerably income inequality in the neighborhoods. Notably the unemployment rates are fairly low compared to the study enrollment period in 1982, when there was a severe national recession [3].

Figures 17 - 21 show maps that illustrate the general spatial patterns present in the air pollution exposures used in our epidemiological analyses. In Figure 17, we see that $PM_{2.5}$ varies smoothly across the state, with the highest concentrations being observed in Los Angeles and the Central Valley. The lower concentrations tend to occur around the San Francisco Bay area and in rural areas of state. PM_{10} follows a similar spatial pattern, although the spatial pattern appears to have slightly more spatial variability (Figure 20). For sulfate, areas along the US-Mexico border have the highest concentrations, while in the remainder of the state there appears to be a trend toward lower values in the north (Figure 19). For ozone, the highest values are observed in the inland areas of southern California and to a lesser extent in the central valley (Figure 18). Ozone tends to be lower in coast regions. With NO_2, monitoring coverage is more limited, but here we observed again higher levels in southern California, particularly in Los Angeles (Figure 21). The central valley and the South San Francisco Bay area also display elevated levels of NO_2. Of importance to subsequent interpretation, many of the major metropolitan areas have higher levels of pollution.

Table 23a-c show the correlations of the exposure surfaces used in the analysis. Correlations are high, with Pearson's R values generally over 0.7. Because our analyses and past evidence indicated that the Los Angeles Metropolitan Statistical Areas (LA MSA) might be different than the rest of the state, we present correlations between the various exposure estimates for the entire state, for the state without LA included, and for LA alone. Statewide results are shown in Table 23a. $PM_{2.5}$ estimates from the IDW and kriging models are highly correlated regardless of whether they are assigned at the individual address or ZIP code level. $BME_{2.5}$ estimates by contrast tend to be less correlated with the other $PM_{2.5}$ estimates. In general this pattern holds for correlations between IDW or kriging estimates and BME estimates. $PM_{2.5}$ kriging estimates tend have high or very high correlations with PM_{10}, sulfate and NO_2, with slightly lower correlations with ozone. The exception is the NO_2 LUR and BME models where the correlations are overall much lower with the other exposure estimates. Similarly some of the correlations between the

REVIEW DRAFT

PM_{10} BME model and other pollutants, notably ozone, are lower as well. A similar pattern holds for the IDW $PM_{2.5}$estimates. In general, all estimates display positive correlations. BME estimates among the pollutants tend to have lower correlations, with ozone being particularly low in its correlation with NO_2 and $PM_{10.}$

Results for the state minus the LA MSA are shown in Table 23b. Here the patterns are similar in terms of correlations among the various IDW and kriging models, although for sulfate and NO_2 correlations with $PM_{2.5}$ are lower than in the statewide analysis. Again the BME and LUR models appear to follow a different spatial pattern and do not correlate as highly with either the other exposure estimates of the same pollutants or with the other pollutants that were estimated with the BME models. Here most pollutants show a positive correlation, with the exception being ozone and the NO_2 LUR model.

In LA the patterns of correlation are quite different than in the statewide analyses. Here correlations between estimates tend to be much lower, with the exception of those correlating ZIP to the individual address, which remain fairly high. Correlations between the kriging estimate and PM_{10} and NO_2 continue to be high with the IDW estimates. Notably the correlations with the IDW estimates of $PM_{2.5}$ and ozone, PM_{10} and sulfate all seem to be lower, although NO_2 retains a high correlation. Here correlations between the BME models and the other exposure estimates tend to be even lower. Another important difference is with ozone that tends to show very low or negative correlations with nearly all the other pollutants except for PM_{10} where there is a moderate, positive correlation. The strongest negative correlation is between sulfate and ozone. Again correlations among the BME estimates tend to be low to moderate.

Causes of Death Investigated

For all exposure models presented in subsequent tables we investigated the association between all causes of mortality (all International Classifications of Diseases 9th Revision or ICD9 codes) and air pollution exposures. Several other cause-specific categories were also investigated, including cardiovascular disease (CVD = ICD9 Codes 390-459), ischemic heart disease (IHD = ICD9 Codes 410-414), respiratory disease (RESP= ICD9 Codes 460-519), lung cancer (LC = ICD 9 Code 162), all cancers (AlC = IC9 codes 140-239), other causes of death ((RS others = All ICD9 – (CV+RESP)), and other causes of death not due to cancer ((RSNC others non-cancer = All ICD9 – (CV+RESP+ALC)).

ZIP Code Exposure Assignment

Tables 24 and 25 present the results of three different exposure models with exposures assigned at the residential ZIP code of the subject. We conducted this analysis to determine whether the effects observed in earlier studies of Los Angeles [3, 5] would replicate across California. In these earlier studies we also used the ZIP code of residence to assign exposures.

The two models for $PM_{2.5}$ include: (1) inverse distance weighting (IDW); and (2) kriging; (see exposure methods for more detail). The tables show results for various model specifications all of which stratify the baseline hazard function on age, race, and gender. In the table, results are presented for both standard Cox proportional hazards models (SC) and random effects Cox models (RE). Specific models include: (a) those that control only for individual level covariates,

REVIEW DRAFT

and (b) those that control for individual covariates and six ecologic covariates measured in 1990. We had prior evidence of effects on all cause and cause-specific mortality in the Los Angeles Metropolitan Statistical Area (LA MSA) and an examination of the spatial patterns in unexplained or unpredicted mortality indicated the residual mortality there was particularly low for all causes and high for cardiovascular disease. For these reasons, we also include results with an indicator for the LA MSA as a sensitivity test. We also included an interaction between the LA MSA and pollution to determine whether the effects in the LA MSA were different from the rest of the state.

In these statewide models we find no effect of $PM_{2.5}$ on all cause mortality; however, there are significantly elevated effects for cardiovascular deaths (CVD) and for deaths from ischemic heart disease (IHD). For both IDW and kriging, the results for both CVD and IHD are highly consistent among exposure models and statistical model specification. Effects sizes for CVD indicate relative risks or RRs ~ 1.11 over the interquartile range of exposure in each model. For IHD, effect sizes range from RR ~ 1.14-1.16, with the kriging estimates being slightly larger than those with the IDW models. None of the other causes of death show elevated risks in relation to $PM_{2.5}$ exposure, and in few instances, they show mildly negative risks, except when the LA indicator is included. When the LA indicator is included, all of the significantly negative risks become null.

For the models containing the LA indicator, LA has a lower death rate overall than the rest of the state, as shown by the indicator RRs, which are generally below 1, except for CVD and IHD, which are above 1. In the interaction models using kriging and IDW exposure estimates, LA displays a consistently lower death rate, but a much higher dose-response to air pollution with most causes of death showing elevated risks in relation to $PM_{2.5}$. Although not significant in most instances the point estimates are large overall. With all cause mortality and respiratory deaths, the RRs are significantly elevated. For all cause the RR ~ 1.25 for LA, similar in size to earlier reports [5]and for respiratory deaths, there is a RR is ~ 1.91.

Consistent with the results that do not use the LA interaction, we also observe significantly elevated risks for CVD and for IHD across all of California that are similar in size when the interaction term is included, but these are slightly smaller, than those presented from the non-interaction models. In these cases, too, the confidence intervals overlap and the estimates are probably not significantly different from one another.

Results from Exposure Assignments to the Geocoded Residential Address

In models that control for residence in the five large conurbations, there are significantly elevated risks for all-cause (RR ~ 1.04), CVD (RR ~ 1.08), and IHD (RR ~ 1.14). Lung cancer and respiratory deaths were also elevated but were not significantly so. There was little evidence of association with other causes of death, which were slightly elevated but highly insignificant, or with other cancers. The model controlling for residence in the five major urban conurbations probably supplies the most valid results because this accounts for the non-metropolitan mortality penalty discussed earlier. Within the indicator variables, there is evidence of a negative association between other causes, including some of the cancers, and air pollution, which is

REVIEW DRAFT

likely indicative of negative confounding due to the non-metropolitan elevation in mortality and the fact that many of the non-metropolitan areas also have lower levels of pollution. Failure to control for this effect probably introduces a negative bias into the air pollution coefficient.

Results for the $PM_{2.5}$ exposures assigned at the home address were similar to those assigned at the ZIP code (see Tables 26-29 for the IDW, kriging, land use regression, and remote sensing estimates, respectively). Both CVD and IHD have significantly elevated risks, regardless of which exposure models were used. For other causes of death, there were no significantly elevated effects.

Effect sizes were slightly larger when using the kriging model than the IDW models. For CVD, effect sizes for the IDW are RR ~ 1.08 and for IHD the range from RR ~ 1.11 to 1.13. In the kriging models the CVD effects have RR ~ 1.1 and for IHD they are ~ 1.15. There was very little sensitivity to the model specification, the confounders included, or including ozone as a co-pollutant.

In comparing results obtained with the different $PM_{2.5}$ indicators (Figures 22-24), the largest HRs per each 10 µg/m3 increase in $PM_{2.5}$ were observed for $PM_{2.5}$ Remote Sensing (RS) and $PM_{2.5}$ LUR. In contrast, somewhat lower HRs were observed for $PM_{2.5}$ KRG and $PM_{2.5}$ IDW. Results were comparable for $PM_{2.5}$ KRG and $PM_{2.5}$ IDW at both the individual and ZIP code –levels, although slightly lower for the individual-level assignments. Results for the BME models were very similar to those of kriging and IDW.

The pattern of effects with the LA interaction was similar to those reported in the ZIP code analysis, with generally lower death rates indicated by the binary variable for Los Angeles and positive interactions between LA and $PM_{2.5}$, suggesting a higher dose response there. Effects on CVD and IHD deaths were similar to those in models without the interaction term, although in these models we also observe elevated but insignificant effects of $PM_{2.5}$ on all causes and respiratory deaths. The models using the kriging exposure estimate produce similar results to those with the IDW when ZIPs are included, although there are not elevated risks on all causes of death in these models and the positive, insignificant effects on respiratory death are larger.

Risk Estimates for Other Pollutants

Although not the primary motivation underlying this research, we also tested the effects of ozone (O_3), nitrogen dioxide (NO_2), other particulate matter (PM_{10}), sulfate (SO_4^{-2}), and of proximity to highways and major roads.

Ozone

We assessed ozone effects on the same causes of death and with the same model specifications as the $PM_{2.5}$models. For ozone we assigned exposure based on the IDW method at the individual scale and on the BME method at the ZIP code of residence (see Tables 30 and 31). There were no effects on all cause mortality, but in all model specifications, we found significant effects on CVD and IHD deaths. These were similar in size to those observed for $PM_{2.5}$, and they were only slightly reduced by inclusion of ecologic confounders. In random effects models including

REVIEW DRAFT

interactions for Los Angeles, there were no significant interactions, although the risks of ozone tended to be somewhat lower in LA than in the rest of the state. Effects in the IDW and BME models were similar in size, but the inclusion of the LA indicator exerted a stronger confounding effect on the BME estimate than on the IDW exposure model. None of the other cause-specific deaths had a positive association with ozone, although in the interaction models there was an insignificantly elevated risk for respiratory deaths as well. In some instances, there were negative risks for other cancers, lung cancer and other causes of death, but these became insignificantly elevated or were of borderline significance when the Los Angeles indicator and interaction terms were included. There were elevated risks in the interaction models for lung cancer, all cancers and for other causes of death, both with and without cancers included, although these were insignificant.

NO_2

Effects for NO_2 were significantly elevated for several causes of death. Effects on CVD and IHD were similar to those of $PM_{2.5}$ across the IQR range (see Table 32). For CVD the RR was ~ 1.07-1.08, and for IHD RR ~ 1.12-1.13. We conducted sensitivity analyses using the BME-LUR model and found these to be similar to the kriging and inverse distance weighting estimates (results not shown).

With the land use regression model, effects for NO_2 were significantly elevated for several causes of death (see Table 33). Effects on CVD and IHD were similar to those of $PM_{2.5}$ across the interquartile range. For CVD the RR was ~ 1.07-1.08, and for IHD RR ~ 1.12-1.13. There were significantly elevated risks of for all causes of death. Relative risks were in the range of 1.03-1.05 for all causes across the relatively small interquartile range of 4.12 ppb. NO_2 exposure was also significantly associated with deaths from lung cancer. Relative risks were in the range of 1.10-1.17, with the larger risks being observed in models that controlled for clustering in the ZIP codes and for ecologic influences on death.

PM_{10} and Sulfate in PM_{10}

Results for PM_{10} were very similar to those observed from the $PM_{2.5}$ models (see Table 34). There were no elevated risks for all causes of death, but significantly elevated risks for deaths from CVD and from IHD. Across the IQR of exposure, the RR was about 1.08 for CVD and 1.14 for IHD. The patterns of interaction with Los Angeles were similar to those observed for $PM_{2.5}$. We conducted sensitivity analyses using the BME-LUR for PM_{10} model and found these to also be similar to the kriging and inverse distance weighting estimates (results not shown).

Effects for the sulfate models were nearly the same as those for PM_{10}, which is not surprising given that the sulfate was derived from the PM_{10} filters (see Table 35).

Road Proximity

For road proximity, we observed elevated relative risks of death, but these were not significantly different than the null effect (Tables 36-38). This general pattern held regardless of whether we

REVIEW DRAFT

used 50m to major roads, 100m to major highways, or a variable combining both 50m to major roads or 100m to major highways. Because residential mobility could substantially influence the proximity to road variable potentially more than for the other pollutants, we examined a shorter follow up period ending in 1988 when residential mobility would be less of a factor. In these analyses, we found larger effects, but they were still not significantly elevated. In Table 39, we included $PM_{2.5}$ and found that the road buffers did not appreciably affect the $PM_{2.5}$ risks.

Time-Dependent Analysis Results

The estimates of the log-hazard ratio parameters β, γ, ϕ and λ are given in Table 40 for both nitrogen dioxide and ozone. For both pollutants $\beta \cong \gamma$ suggesting that all of the association between air pollution exposure and cardiovascular mortality in this cohort is spatial in nature and not temporal. This conclusion is reinforced by the results from the joint exposure model in which there is no evidence of an association between the time varying exposure variable $x_i(t) - \bar{x}_i$ and cardiovascular mortality (p>0.05).

The standard error of the estimate of λ is much larger than those of the other parameters. This is due to the much smaller variation in $x_i(t) - \bar{x}_i$ compared to $\bar{x}_i$. For nitrogen dioxide, the variance of $\bar{x}_i$ is 110.25 ppb^2 and that of $x_i(t) - \bar{x}_i$ is 4.84 ppb^2. Thus 96% of the combined variation in these two exposure measures is associated with spatial variability and only 4% with variation over time. For ozone, the percentage of spatial variation is smaller than nitrogen dioxide at 83% but the vast majority of variation in exposure is associated with changes in ozone concentrations over space compared to time. We conclude that the exposure patterns are much more variable spatially over California than over the time period from 1988 to 2000 within any locale. Due to this lack of temporal exposure variability we were not able to adequately investigate the temporal association between exposure to either nitrogen dioxide or ozone in this study.

REVIEW DRAFT

Table 19: Reasons for exclusion of CPS-II participants from final California analytic cohort, follow-up 1982-2000

1. Total CPS-II Population	1,184,588
2. California participants	104,277
3. Participants with complete zip code data	102,805
4. Participants without 'PO Box/co' in address	85,744
5. Participants geocoded	85,511
6. Participants with missing/erroneous covariate data	
Race	391
Education	953
Marital Status	528
Body Mass Index	1,664
Passive Smoking	755
Cigarette Smoking	4,877
TOTAL	9,168
6. Participants with air pollution data	76,343
$PM_{2.5}$ KRG individual-level	76,105
$PM_{2.5}$ BME individual-level	76,105
$PM_{2.5}$ IDW individual-level	73,609
$PM_{2.5}$ KRG zip-code level	74,029
$PM_{2.5}$ IDW zip-code level	70,084
PM_{10} KRG individual-level	76,260
PM_{10} BME individual-level	76,105
PM_{10} BME zip-code level	74,029
Ozone IDW individual-level	76,222
Ozone BME zip-code level	69,551
Sulfate IDW individual-level	76,260
NO_2 IDW individual-level	75,364
NO_2 BME individual-level	76,105
NO_2 LUR individual-level	76,105
Road buffers	76,105

REVIEW DRAFT

Table 20: Participant characteristics in the Nationwide Study Compared to the California Cohort.

Variable	Nationwide	California
Participants (n)	488,370	76,343
Participants died from (%)		
All Causes	26.4	26.8
CPD	13.1	14.1
Lung Cancer	2.0	2.0
All other causes	11.3	10.6
Age	56.6 (10.5)	57.3 (10.6)
Female (%)	56.5	56.2
White (%)	94.0	91.5
Education (%)		
< High School	12.1	8.7
High School	31.2	23.0
>High School	56.7	68.3
Current Smoker (%)	21.9	19.4
Cigarettes per day	22.0 (12.4)	21.5 (12.6)
Years of smoking	33.6 (11.0)	34.1 (11.4)
Former Smoker (%)	30.3	28.9
Cigarettes per day	21.6 (14.6)	20.8 (14.7)
Years of smoking	22.2 (4.1)	22.1 (12.7)
Age when started smoking (%)		
< 18 yras (current smoker)	9.3	7.7
< 18 yras (former smoker)	11.8	10.3
Hours per day exposed to smoking	3.2 (4.4)	2.7 (4.1)

REVIEW DRAFT

Table 21: Distribution of air pollutants at individual level.

				Percentiles								
Air Pollution	**# Subjects**	**Mean**	**Variance**	**0**	**5**	**10**	**25**	**50**	**75**	**90**	**95**	**100**
$PM_{2.5}$ KRG	76,343	15.25	22.28	5.14	9.26	10.59	11.56	13.29	20.09	22.13	22.54	24.73
$PM_{2.5}$ IDW	76,041	21.95	35.12	3.77	11.52	15.16	18.23	20.42	26.97	29.33	31.35	36.26
$PM_{2.5}$ BME	76,105	16.13	25.60	4.30	8.86	9.88	12.14	14.93	21.08	22.73	23.71	28.50
Ozone	76,222	50.54	210.60	17.11	28.84	31.25	36.93	50.89	61.00	68.59	74.50	96.20
PM_{10}	76,260	29.83	106.6	8.21	16.48	18.08	21.04	28.67	38.44	43.63	47.50	84.27
PM_{10} BME	76,105	32.35	138.24	8.53	16.23	18.37	23.00	30.22	41.59	49.10	52.36	66.37
Sulphate IDW	76,260	2.822	2.026	0.395	0.977	1.201	1.607	1.933	4.277	4.692	4.820	5.072
NO_2 IDW	75,364	25.22	102.73	3.17	11.08	13.85	17.85	22.47	35.00	40.37	42.03	45.51
NO_2 BME	76,105	14.21	50.23	0.46	4.80	6.23	9.13	12.31	19.24	25.22	27.63	36.01
$PM_{2.5}$ LUR	76,105	14.14	12.45	4.24	8.28	9.47	11.63	14.09	16.98	18.43	19.35	25.08
$PM_{2.5}$ RS	76,105	9.81	11.11	2.00	5.03	5.70	7.34	8.82	12.73	14.46	15.36	17.89
PM_{10} LUR	76,105	28.78	21.77	15.51	19.59	21.75	25.62	30.08	32.81	33.70	33.94	34.18
NO_2 LUR	76,105	12.28	8.50	3.04	7.93	8.82	10.23	12.15	14.35	16.20	17.07	21.93

REVIEW DRAFT

Table 22: Distribution of air pollutants at ZIP level.

				Percentiles								
Air Pollution	**928 ZIPs # Subjects**	**Mean**	**Variance**	**0**	**5**	**10**	**25**	**50**	**75**	**90**	**95**	**100**
$PM_{2.5}$ KRG	73,609	15.16	22.14	5.277	9.152	10.50	11.58	12.81	20.05	22.06	22.57	24.70
$PM_{2.5}$ IDW	70,084	15.98	26.32	3.777	9.752	10.69	11.56	13.79	20.93	23.29	24.17	27.08
Ozone BME	69,551	50.50	210.8	22.65	28.98	31.17	37.02	50.69	61.26	68.47	74.99	89.13

REVIEW DRAFT

Table 23: Distribution of Ecological Covariates based on 1990 Census data and defined at the ZIP Code Area.

				Percentiles								
Ecological Covariate	**# Subjects**	**Mean**	**Variance**	**0**	**5**	**10**	**25**	**50**	**75**	**90**	**95**	**100**
White	76,343	0.7733	0.0265	0.0442	0.4419	0.5507	0.6976	0.8092	0.8965	0.9338	0.9579	1.0000
Black	76,343	0.0463	0.0076	0.000	0.0011	0.0034	0.0101	0.0201	0.0468	0.0944	0.1700	0.8608
Hispanic	76,343	0.1702	0.0240	0.000	0.0292	0.0386	0.0678	0.1183	0.2143	0.3707	0.5343	0.9749
Post Secondary Ed.	76,343	0.6103	0.0235	0.000	0.3510	0.4086	0.5041	0.6208	0.7142	0.8181	0.8512	1.0000
Mean Income	76,343	42,472	$2.58*10^{8}$	4,999	23,075	25,979	32,202	40,488	48,432	61,686	77,669	129,654
Gini Coefficient	76,343	0.3746	0.0036	0.000	0.3080	0.3237	0.3492	0.3803	0.4071	0.4297	0.4445	0.5990
Unemployment	76,343	0.0389	0.0021	0.000	0.0139	0.0168	0.0242	0.0325	0.0434	0.0564	0.0680	0.5208

REVIEW DRAFT

Table 24: Pearson correlations (x100) between air pollutants (California overall).

Air Pollutants	$PM_{2.5}$ KRG	$PM_{2.5}$ IDW	$PM_{2.5}$ BME	$PM_{2.5}$ LUR	$PM_{2.5}$ RS	Ozone	PM_{10} KRG	PM_{10} BME	Sulfate KRG	NO_2 IDW	NO_2 BME	NO_2 LUR	$PM_{2.5}$ KRG (ZIP)	$PM_{2.5}$ IDW (ZIP)	$PM_{2.5}$ BME (ZIP)	Ozone (ZIP)	PM_{10} BME (ZIP)	NO_2 BME (ZIP)
$PM_{2.5}$ KRG	--	86.53	90.16	77.53	81.29	71.38	89.69	89.39	72.7	81.9	81.57	35.39	98.83	95.44	64.15	71.43	36.12	36.88
$PM_{2.5}$ IDW		--	85.45	74.74	72.93	56.71	78.63	78.01	72.34	87.66	76.8	41.78	85.88	90.19	65.36	59.04	36.11	36.31
$PM_{2.5}$ BME			--	94.04	87.1	64.07	86.37	88.16	76.47	82.8	82.99	50.56	89.47	89.75	77.26	66.36	48.17	47.48
$PM_{2.5}$ LUR				--	87.17	54.67	76.38	80.46	66.16	69.87	73.68	55.46	76.81	75.28	81.19	56.8	54.46	48.12
$PM_{2.5}$ RS					--	50.9	80.22	76.44	70.27	67.58	62.93	37.51	80.83	81.48	72.31	51.47	49.4	39.71
Ozone						--	75.24	78.02	45.67	43.1	59.76	-1.27	70.93	69.14	40	98.88	15.25	15.25
PM_{10} KRG							--	93.52	69.06	69.36	75.76	26.07	89.27	87.13	60.39	79.06	36.16	30.19
PM_{10} BME								--	60.78	68.1	80.33	37.57	88.49	85.14	65.06	79.66	45.12	39.03
Sulfate KRG									--	78.69	73.45	35.23	73	74.42	58.45	47.72	28.47	35.45
NO_2 IDW										--	81.73	50.46	81.31	81.71	62.96	45.66	31.57	40.78
NO_2 BME											--	55.42	80.49	74.35	57.82	61.31	37.02	40.87
NO_2 LUR												--	35.17	28.98	50.02	-2.11	56.2	67.58
$PM_{2.5}$ KGR (ZIP)													--	96.24	65.5	71.45	37.34	38.5
$PM_{2.5}$ IDW(ZIP)														--	64.64	69.89	31.69	34.51
$PM_{2.5}$ BME(ZIP)															--	40.91	65.86	68.18
Ozone BME (ZIP)																--	12.38	12.89
PM_{10} BME(ZIP)																	--	74.66
NO_2 BME(ZIP)																		--

See footnote for years represented by air pollution exposure[5]

[5] Years represented by air pollution exposures

Pollutant	$PM_{2.5}$ KRG	$PM_{2.5}$ IDW	$PM_{2.5}$ BME	$PM_{2.5}$ LUR	$PM_{2.5}$ RS	Ozone	PM_{10} KRG	PM_{10} BME	Sulfate IDW
Years	2000	1998-2002	1998-2002	1998-2002	2001-2006	1988-2002	1988-2002	1988-2002	1996
Pollutant	**NO_2 IDW**	**NO_2 BME**	**NO_2 LUR**	**$PM_{2.5}$ KGR (ZIP)**	**$PM_{2.5}$ IDW (ZIP)**	**$PM_{2.5}$ BME (ZIP)**	**Ozone BME (ZIP)**	**PM_{10} BME (ZIP)**	**NO_2 BME (ZIP)**
Years	1988-2002	1988-2002	1988-2002	2000	1998-2002	1998-2002	1988-2002	1988-2002	1988-2002

REVIEW DRAFT

Table 25: Pearson correlations (x100) between air pollutants (California minus LA).

Air Pollutants	$PM_{2.5}$ KRG	$PM_{2.5}$ IDW	$PM_{2.5}$ BME	$PM_{2.5}$ LUR	$PM_{2.5}$ RS	Ozone	PM_{10} KRG	PM_{10} BME	Sulfate KRG	NO_2 IDW	NO_2 BME	NO_2 LUR	$PM_{2.5}$ KRG (ZIP)	$PM_{2.5}$ IDW (ZIP)	$PM_{2.5}$ BME (ZIP)	Ozone (ZIP)	PM_{10} BME (ZIP)	NO_2 BME (ZIP)
$PM_{2.5}$ KRG	--	80.28	85.20	73.50	81.90	71.33	88.28	87.80	63.49	63.49	72.21	16.48	98.78	95.72	55.13	70.93	33.38	18.23
$PM_{2.5}$ IDW		--	79.93	72.00	72.76	48.52	74.11	72.33	57.78	82.41	69.99	32.05	79.46	86.46	59.19	50.79	38.45	21.28
$PM_{2.5}$ BME			--	94.98	87.92	60.45	84.69	86.98	61.94	71.38	80.21	42.29	84.00	86.10	72.24	63.04	51.35	36.49
$PM_{2.5}$ LUR				--	87.94	47.76	72.94	77.74	55.15	64.26	74.11	52.63	72.11	72.62	76.81	49.97	57.70	41.99
$PM_{2.5}$ RS					--	52.86	81.40	78.85	61.14	61.01	67.55	28.60	80.82	80.66	68.33	54.11	48.17	28.55
Ozone						--	72.69	75.46	34.69	27.65	49.08	-16.49	71.05	71.95	30.31	98.94	14.13	6.88
PM_{10} KRG							--	92.84	60.11	58.31	70.45	12.33	87.76	88.11	54.06	77.81	32.87	18.41
PM_{10} BME								--	47.21	54.18	73.11	23.12	87.11	87.07	59.27	77.44	43.34	26.97
Sulfate KRG									--	61.11	67.67	21.06	51.62	54.68	42.41	35.34	24.88	17.36
NO_2 IDW										--	72.15	43.89	60.38	65.36	54.61	28.69	31.10	22.03
NO_2 BME											--	50.76	69.42	69.59	53.00	49.47	40.05	30.67
NO_2 LUR												--	14.70	11.36	46.82	-18.99	54.82	62.20
$PM_{2.5}$ KGR (ZIP)													--	96.23	56.13	71.07	34.45	19.45
$PM_{2.5}$ IDW(ZIP)														--	56.11	72.62	31.10	15.59
$PM_{2.5}$ BME(ZIP)															--	31.21	70.39	64.85
Ozone BME (ZIP)																--	12.20	4.08
PM_{10} BME(ZIP)																	--	75.58
NO_2 BME(ZIP)																		--

See footnote for years represented by air pollution exposure [6]

[6] Years represented by air pollution exposures

Pollutant	$PM_{2.5}$ KRG	$PM_{2.5}$ IDW	$PM_{2.5}$ BME	$PM_{2.5}$ LUR	$PM_{2.5}$ RS	Ozone	PM_{10} KRG	PM_{10} BME	Sulfate IDW
Years	2000	1998-2002	1998-2002	1998-2002	2001-2006	1988-2002	1988-2002	1988-2002	1996
Pollutant	**NO_2 IDW**	**NO_2 BME**	**NO_2 LUR**	**$PM_{2.5}$ KGR (ZIP)**	**$PM_{2.5}$ IDW (ZIP)**	**$PM_{2.5}$ BME (ZIP)**	**Ozone BME (ZIP)**	**PM_{10} BME (ZIP)**	**NO_2 BME (ZIP)**
Years	1988-2002	1988-2002	1988-2002	2000	1998-2002	1998-2002	1988-2002	1988-2002	1988-2002

REVIEW DRAFT

Table 26: Pearson correlations (x100) between air pollutants (LA only).

Air Pollutants	$PM_{2.5}$ KRG	$PM_{2.5}$ IDW	$PM_{2.5}$ BME	$PM_{2.5}$ LUR	$PM_{2.5}$ RS	Ozone	PM_{10} KRG	PM_{10} BME	Sulfate KRG	NO_2 IDW	NO_2 BME	NO_2 LUR	$PM_{2.5}$ KRG (ZIP)	$PM_{2.5}$ IDW (ZIP)	$PM_{2.5}$ BME (ZIP)	Ozone (ZIP)	PM_{10} BME (ZIP)	NO_2 BME (ZIP)
$PM_{2.5}$ KRG	--	64.14	68.09	30.34	39.34	5.14	83.43	73.80	35.58	87.36	57.39	49.06	90.25	63.38	8.22	-1.41	46.36	37.20
$PM_{2.5}$ IDW		--	51.90	10.21	5.67	8.72	39.53	43.76	24.70	74.91	37.49	24.38	52.52	65.70	5.21	3.82	2.83	25.30
$PM_{2.5}$ BME			--	78.30	62.20	-1.28	54.20	60.17	36.45	58.94	38.89	42.35	58.99	54.54	47.37	-5.80	45.55	35.44
$PM_{2.5}$ LUR				--	58.48	11.91	34.87	46.81	9.57	12.60	21.34	28.82	23.58	9.63	64.91	8.88	39.51	21.00
$PM_{2.5}$ RS					--	-52.04	26.63	10.22	60.53	33.22	-16.01	22.36	37.83	47.39	42.65	-56.55	49.99	29.05
Ozone						--	3561	51.61	-69.30	-17.18	56.72	-12.84	-2.99	-41.61	-8.63	95.69	-22.85	-29.21
PM_{10} KRG							--	78.66	4.54	59.57	62.30	19.61	79.76	36.90	4.94	28.81	38.26	7.53
PM_{10} BME								--	14.46	51.08	82.78	48.60	66.99	19.44	19.38	47.29	47.61	29.48
Sulfate KRG									--	51.13	31.57	21.81	24.78	54.60	38.36	-24.30	83.67	55.05
NO_2 IDW										--	40.50	41.28	79.13	74.58	2.32	-21.41	36.09	40.40
NO_2 BME											--	43.94	50.39	0.85	-8.83	54.81	28.23	13.31
NO_2 LUR												--	45.85	21.29	19.04	-13.03	55.66	68.98
$PM_{2.5}$ KGR (ZIP)													--	70.86	13.57	-4.73	55.30	44.81
$PM_{2.5}$ IDW(ZIP)														--	16.74	-43.93	28.00	38.64
$PM_{2.5}$ BME(ZIP)															--	-10.34	55.05	55.76
Ozone BME (ZIP)																--	-25.16	31.53
PM_{10} BME(ZIP)																	--	71.78
NO_2 BME(ZIP)																		--

See footnote for years represented by air pollution exposure[7]

[7] Years represented by air pollution exposures

Pollutant	$PM_{2.5}$ KRG	$PM_{2.5}$ IDW	$PM_{2.5}$ BME	$PM_{2.5}$ LUR	$PM_{2.5}$ RS	Ozone	PM_{10} KRG	PM_{10} BME	Sulfate IDW
Years	2000	1998-2002	1998-2002	1998-2002	2001-2006	1988-2002	1988-2002	1988-2002	1996
Pollutant	**NO_2 IDW**	**NO_2 BME**	**NO_2 LUR**	**$PM_{2.5}$ KGR (ZIP)**	**$PM_{2.5}$ IDW (ZIP)**	**$PM_{2.5}$ BME (ZIP)**	**Ozone BME (ZIP)**	**PM_{10} BME (ZIP)**	**NO_2 BME (ZIP)**
Years	1988-2002	1988-2002	1988-2002	2000	1998-2002	1998-2002	1988-2002	1988-2002	1988-2002

REVIEW DRAFT

Table 27: Hazard ratios of $PM_{2.5}$ IDW ZIP code-level (IQR=9.37) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level covariates (1990), stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 70,084.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=18,744)	(n=7,649)	(n=4,346)	(n=1,871)	(n=1,414)	(n=6,171)	(n=8,871)	(n= 2,700)
Individual covariates	SC	0.989 (0.963, 1.016)	1.107 (1.061, 1.155)	1.143 (1.081, 1.208)	0.964 (0.883, 1.051)	0.941 (0.852, 1.039)	0.915 (0.873, 0.960)	0.905 (0.870, 0.942)	0.883 (0.822, 0.949)
	RE	0.991 (0.962, 1.020) 4.92	1.108 (1.060, 1.157) 5.33	1.142 (1.078, 1.210) 7.47	0.964 (0.883, 1.052) 2.48	0.941 (0.852, 1.040) 0.53	0.916 (0.872, 0.962) 4.33	0.906 (0.870, 0.944) 2.90	0.885 (0.822, 0.954) 13.18
+7 1990 Ecological	SC	0.995 (0.967, 1.025)	1.112 (1.062, 1.164)	1.154 (1.086, 1.225)	0.992 (0.903, 1.090)	0.951 (0.854, 1.059)	0.922 (0.876, 0.970)	0.911 (0.872, 0.950)	0.886 (0.820, 0.958)
	RE	0.996 (0.966, 1.026) 1.28	1.112 (1.060, 1.166) 4.57	1.153 (1.084, 1.226) 2.74	0.992 (0.903, 1.090) 0.87	0.951 (0.854, 1.059) 0.47	0.922 (0.876, 0.971) 0.99	0.911 (0.872, 0.950) 0.35	0.887 (0.819, 0.961) 9.51
+ LA indicator	SC LA Ind. $PM_{2.5}$	0.967 (0.925, 1.011) 1.012 (0.975, 1.049)	1.036 (0.968, 1.110) 1.092 (1.031, 1.156)	0.997 (0.911, 1.091) 1.155 (1.072, 1.244)	0.827 (0.717, 0.954) 1.086 (0.968, 1.218)	1.026 (0.870, 1.209) 0.939 (0.821, 1.073)	0.957 (0.885, 1.036) 0.942 (0.883, 1.004)	0.938 (0.878, 1.002) 0.940 (0.891, 0.991)	0.898 (0.796, 1.012) 0.934 (0.848, 1.029)
	RE LA Ind. $PM_{2.5}$	0.967 (0.924, 1.012) 1.012 (0.975, 1.051) 1.10	1.038 (0.967, 1.114) 1.091 (1.028, 1.158) 4.57	0.997 (0.911, 1.093) 1.154 (1.070, 1.245) 2.73	0.827 (0.716, 0.955) 1.086 (0.968, 1.219) 0.80	1.026 (0.870, 1.209) 0.939 (0.821, 1.074) 0.47	0.957 (0.884, 1.036) 0.942 (0.883, 1.005) 0.93	0.938 (0.878, 1.002) 0.940 (0.891, 0.991) 0.27	0.900 (0.796, 1.017) 0.934 (0.845, 1.031) 7.72
+ LA * $PM_{2.5}$	SC LA Ind. $LA*PM_{2.5}$ $PM_{2.5}$	0.680 (0.522, 0.885) 1.173 (1.043, 1.321) 0.996 (0.958, 1.035)	0.729 (0.488, 1.090) 1.173 (0.980, 1.403) 1.074 (1.011, 1.141)	0.652 (0.380, 1.119) 1.211 (0.953, 1.539) 1.134 (1.048, 1.226)	0.247 (0.098, 0.625) 1.722 (1.143, 2.594) 1.038 (0.920, 1.172)	0.503 (0.189, 1.338) 1.384 (0.893, 2.146) 0.909 (0.789, 1.048)	0.832 (0.527, 1.315) 1.066 (0.867, 1.310) 0.936 (0.875, 1.001)	0.781 (0.531, 1.148) 1.087 (0.914, 1.293) 0.932 (0.881, 0.986)	0.669 (0.326, 1.370) 1.143 (0.829, 1.576) 0.922 (0.833, 1.021)
	RE LA Ind. $LA*PM_{2.5}$ $PM_{2.5}$	0.681 (0.521, 0.889) 1.173 (1.041, 1.321) 0.996 (0.958, 1.036) 0.62	0.736 (0.487, 1.113) 1.168 (0.971, 1.405) 1.074 (1.009, 1.143) 3.92	0.654 (0.380, 1.127) 1.209 (0.949, 1.540) 1.133 (1.047, 1.226) 1.96	0.247 (0.097, 0.626) 1.722 (1.142, 2.595) 1.038 (0.920, 1.172) 0.70	0.503 (0.189, 1.338) 1.384 (0.892, 2.147) 0.909 (0.789, 1.048) 0.47	0.831 (0.525, 1.318) 1.066 (0.867, 1.312) 0.936 (0.875, 1.002) 0.93	0.781 (0.530, 1.150) 1.087 (0.913, 1.293) 0.932 (0.881, 0.986) 0.26	0.675 (0.326, 1.400) 1.139 (0.821, 1.579) 0.922 (0.831, 1.023) 7.23

REVIEW DRAFT

Table 28: Hazard ratios of $PM_{2.5}$KRG ZIP code-level (IQR=8.4735) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level covariates (1990), stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 73,609.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=19,734)	(n=8,046)	(n=4,536)	(n=1,971)	(n=1,483)	(n=6,500)	(n=9,340)	(n= 2,840)
Individual covariates	**SC**	0.993 (0.967, 1.020)	1.111 (1.066, 1.158)	1.156 (1.094, 1.221)	0.964 (0.885, 1.051)	0.931 (0.845, 1.027)	0.917 (0.876, 0.961)	0.912 (0.877, 0.948)	0.898 (0.837, 0.964)
	RE	0.994 (0.966, 1.024) 4.92	1.111 (1.064, 1.160) 5.34	1.155 (1.091, 1.223) 7.91	0.964 (0.885, 1.051) 4.02	0.931 (0.845, 1.027) 0.52	0.918 (0.875, 0.963) 3.67	0.912 (0.877, 0.949) 2.55	0.900 (0.837, 0.969) 13.98
+7 1990 Ecological	**SC**	0.994 (0.965, 1.024)	1.110 (1.060, 1.161)	1.162 (1.094, 1.235)	0.988 (0.899, 1.086)	0.939 (0.843, 1.045)	0.918 (0.872, 0.966)	0.912 (0.874, 0.951)	0.898 (0.831, 0.971)
	RE	0.994 (0.965, 1.025) 1.15	1.109 (1.058, 1.163) 4.70	1.161 (1.092, 1.235) 3.14	0.988 (0.899, 1.086) 0.98	0.939 (0.843, 1.046) 0.45	0.918 (0.872, 0.966) 0.65	0.912 (0.874, 0.951) 0.24	0.899 (0.831, 0.974) 9.94
+ LA indicator	**SC** **LA Ind.** **$PM_{2.5}$**	0.966 (0.922, 1.012) 1.012 (0.974, 1.052)	1.030 (0.958, 1.106) 1.092 (1.028, 1.159)	0.985 (0.896, 1.082) 1.171 (1.083, 1.267)	0.833 (0.716, 0.968) 1.088 (0.963, 1.228)	1.051 (0.885, 1.248) 0.915 (0.795, 1.053)	0.963 (0.887, 1.046) 0.936 (0.876, 1.002)	0.938 (0.876, 1.005) 0.943 (0.892, 0.997)	0.887 (0.783, 1.005) 0.957 (0.865, 1.059)
	RE **LA Ind.** **$PM_{2.5}$**	0.967 (0.922, 1.014) 1.012 (0.973, 1.053) 0.98	1.032 (0.958, 1.111) 1.090 (1.024, 1.161) 4.77	0.985 (0.896, 1.084) 1.170 (1.080, 1.267) 3.05	0.833 (0.716, 0.968) 1.088 (0.963, 1.229) 0.85	1.051 (0.884, 1.248) 0.915 (0.795, 1.053) 0.45	0.963 (0.886, 1.046) 0.937 (0.875, 1.002) 0.65	0.938 (0.876, 1.005) 0.943 (0.892, 0.997) 0.21	0.889 (0.783, 1.010) 0.957 (0.862, 1.061) 7.98
+ LA * $PM_{2.5}$	**SC** **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.566 (0.415, 0.772) 1.254 (1.102, 1.428) 0.995 (0.956, 1.035)	0.770 (0.483, 1.228) 1.131 (0.930, 1.375) 1.081 (1.016, 1.150)	0.523 (0.275, 0.993) 1.306 (1.000, 1.705) 1.148 (1.058, 1.245)	0.177 (0.058, 0.539) 1.917 (1.210, 3.037) 1.043 (0.919, 1.184)	0.417 (0.132, 1.317) 1.483 (0.915, 2.404) 0.886 (0.765, 1.027)	0.680 (0.398, 1.163) 1.159 (0.925, 1.454) 0.925 (0.863, 0.992)	0.545 (0.345, 0.859) 1.260 (1.041, 1.526) 0.926 (0.874, 0.982)	0.312 (0.131, 0.741) 1.556 (1.085, 2.231) 0.927 (0.834, 1.030)
	RE **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.566 (0.414, 0.773) 1.255 (1.101, 1.430) 0.995 (0.955, 1.036) 0.49	0.769 (0.475, 1.244) 1.133 (0.926, 1.386) 1.079 (1.011, 1.151) 4.66	0.522 (0.273, 0.998) 1.307 (0.998, 1.711) 1.147 (1.056, 1.246) 2.46	0.177 (0.058, 0.539) 1.918 (1.210, 3.040) 1.043 (0.918, 1.184) 0.79	0.417 (0.132, 1.318) 1.483 (0.914, 2.405) 0.886 (0.0.765, 1.027) 0.43	0.680 (0.397, 1.165) 1.160 (0.924, 1.455) 0.925 (0.863, 0.993) 0.61	0.545 (0.345, 0.860) 1.260 (1.040, 1.526) 0.926 (0.874, 0.982) 0.17	0.316 (0.132, 0.757) 1.548 (1.076, 2.228) 0.927 (0.833, 1.032) 5.22

REVIEW DRAFT

Table 29: Hazard ratios of $PM_{2.5}$IDW individual-level (IQR=8.74) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level covariates (1990), stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,041.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,351)	(n=8,292)	(n=4,692)	(n=2,035)	(n=1,536)	(n=6,711)	(n=9,639)	(n= 2,928)
Individual covariates	**SC**	0.990 (0.968, 1.013)	1.076 (1.038, 1.116)	1.113 (1.061, 1.168)	0.950 (0.883, 1.023)	0.964 (0.886, 1.048)	0.948 (0.911, 0.987)	0.934 (0.904, 0.966)	0.904 (0.851, 0.961)
	RE	0.991 (0.967, 1.016) 4.94	1.076 (1.036, 1.118) 6.25	1.112 (1.058, 1.169) 7.92	0.950 (0.883, 1.023) 2.46	0.964 (0.886, 1.048) 0.56	0.949 (0.911, 0.989) 5.51	0.935 (0.904, 0.969) 4.13	0.905 (0.850, 0.964) 14.33
+7 1990 Ecological	**SC**	1.003 (0.978, 1.028)	1.084 (1.043, 1.127)	1.132 (1.075, 1.191)	0.975 (0.901, 1.055)	0.982 (0.898, 1.073)	0.962 (0.922, 1.004)	0.948 (0.915, 0.983)	0.918 (0.860, 0.979)
	RE	1.003 (0.978, 1.029) 1.31	1.085 (1.042, 1.129) 5.24	1.131 (1.074, 1.191) 2.82	0.975 (0.901, 1.055) 0.87	0.982 (0.898, 1.074) 0.49	0.963 (0.922, 1.005) 1.74	0.948 (0.915, 0.983) 0.81	0.919 (0.859, 0.982) 11.02
+ LA indicator	**SC** **LA Ind.** **$PM_{2.5}$**	0.962 (0.922, 1.004) 1.017 (0.988, 1.048)	1.052 (0.985, 1.124) 1.063 (1.014, 1.113)	1.004 (0.920, 1.096) 1.130 (1.063, 1.201)	0.856 (0.746, 0.983) 1.032 (0.940, 1.132)	0.992 (0.848, 1.160) 0.984 (0.885, 1.095)	0.921 (0.855, 0.993) 0.992 (0.943, 1.045)	0.908 (0.853, 0.967) 0.983 (0.942, 1.026)	0.880 (0.784, 0.987) 0.962 (0.891, 1.040)
	RE **LA Ind.** **$PM_{2.5}$**	0.963 (0.922, 1.005) 1.017 (0.987, 1.048) 1.04	1.053 (0.984, 1.128) 1.063 (1.012, 1.115) 5.09	1.005 (0.920, 1.097) 1.129 (1.061, 1.201) 2.83	0.856 (0.746, 0.983) 1.032 (0.940, 1.132) 0.76	0.992 (0.848, 1.160) 0.984 (0.885, 1.095) 0.49	0.921 (0.854, 0.994) 0.993 (0.943, 1.045) 1.30	0.908 (0.853, 0.967) 0.983 (0.942, 1.026) 0.36	0.883 (0.785, 0.992) 0.962 (0.889, 1.041) 8.23
+ LA * $PM_{2.5}$	**SC** **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.833 (0.629, 1.103) 1.047 (0.958, 1.145) 1.012 (0.981, 1.044)	0.950 (0.623, 1.449) 1.033 (0.903, 1.182) 1.058 (1.007, 1.112)	0.828 (0.471, 1.458) 1.063 (0.889, 1.272) 1.121 (1.050, 1.196)	0.694 (0.271, 1.777) 1.069 (0.793, 1.441) 1.024 (0.929, 1.130)	0.402 (0.139, 1.161) 1.337 (0.955, 1.872) 0.954 (0.853, 1.068)	0.844 (0.513, 1.388) 1.028 (0.877, 1.205) 0.989 (0.937, 1.044)	0.768 (0.506, 1.166) 1.055 (0.924, 1.205) 0.977 (0.934, 1.022)	0.606 (0.281, 1.307) 1.127 (0.883, 1.439) 0.950 (0.876, 1.031)
	RE **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.835 (0.629, 1.110) 1.046 (0.956, 1.145) 1.012 (0.980, 1.045) 0.94	0.961 (0.623, 1.483) 1.030 (0.897, 1.182) 1.058 (1.005, 1.115) 5.01	0.831 (0.470, 1.470) 1.062 (0.886, 1.273) 1.120 (1.049, 1.197) 2.64	0.694 (0.271, 1.780) 1.069 (0.793, 1.441) 1.025 (0.929, 1.130) 0.75	0.401 (0.139, 1.162) 1.337 (0.955, 1.872) 0.954 (0.853, 1.068) 0.48	0.841 (0.510, 1.387) 1.029 (0.877, 1.208) 0.989 (0.937, 1.045) 1.36	0.768 (0.505, 1.167) 1.055 (0.923, 1.206) 0.977 (0.934, 1.023) 0.35	0.620 (0.285, 1.349) 1.120 (0.875, 1.434) 0.950 (0.874, 1.033) 7.28

REVIEW DRAFT

Table 30: Hazard ratios of $PM_{2.5}$KRG individual-level (IQR=8.52902) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level covariates (1990), stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,135.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,432)	(n=8,327)	(n= 4,710)	(n=2,044)	(n=1,540)	(n=6,739)	(n=9,676)	(n= 2,937)
Individual covariates	**SC**	0.993 (0.968, 1.018)	1.100 (1.057, 1.144)	1.157 (1.098, 1.218)	0.963 (0.889, 1.044)	0.933 (0.850, 1.024)	0.927 (0.887, 0.969)	0.920 (0.886, 0.954)	0.902 (0.844, 0.965)
	RE	0.993 (0.967, 1.021) 4.30	1.099 (1.055, 1.145) 5.28	1.156 (1.095, 1.219) 7.54	0.963 (0.888, 1.045) 4.17	0.933 (0.850, 1.024) 0.647	0.928 (0.887, 0.970) 2.14	0.920 (0.886, 0.956) 1.93	0.904 (0.843, 0.968) 10.33
+7 1990 Ecological	**SC**	0.993 (0.966, 1.021)	1.097 (1.051, 1.145)	1.154 (1.090, 1.223)	0.986 (0.902, 1.077)	0.933 (0.842, 1.033)	0.923 (0.880, 0.969)	0.917 (0.881, 0.955)	0.905 (0.841, 0.973)
	RE	0.993 (0.965, 1.021) 0.82	1.096 (1.048, 1.147) 4.03	1.154 (1.088, 1.223) 3.65	0.986 (0.902, 1.078) 1.161	0.933 (0.842, 1.033) 0.57	0.923 (0.879, 0.969) 0.61	0.917 (0.881, 0.955) 0.34	0.905 (0.840, 0.976) 6.97
+ LA indicator	**SC** **LA Ind.** **$PM_{2.5}$**	0.966 (0.924, 1.009) 1.010 (0.975, 1.046)	1.037 (0.969, 1.109) 1.077 (1.020, 1.138)	0.977 (0.895, 1.068) 1.167 (1.087, 1.254)	0.825 (0.717, 0.950) 1.080 (0.967, 1.206)	1.010 (0.859, 1.186) 0.928 (0.817, 1.055)	0.957 (0.886, 1.034) 0.943 (0.887, 1.002)	0.933 (0.875, 0.995) 0.949 (0.902, 0.998)	0.882 (0.784, 0.992) 0.961 (0.876, 1.053)
	RE **LA Ind.** **$PM_{2.5}$**	0.966 (0.924, 1.010) 1.009 (0.974, 1.046) 0.62	1.038 (0.969, 1.113) 1.076 (1.017, 1.138) 4.11	0.978 (0.894, 1.070) 1.166 (1.084, 1.254) 3.53	0.825 (0.716, 0.951) 1.080 (0.967, 1.207) 1.01	1.010 (0.859, 1.186) 0.928 (0.817, 1.055) 0.57	0.957 (0.886, 1.034) 0.943 (0.887, 1.003) 0.58	0.933 (0.875, 0.995) 0.949 (0.901, 0.998) 0.27	0.883 (0.785, 0.995) 0.960 (0.875, 1.054) 4.69
+ LA * $PM_{2.5}$	**SC** **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.643 (0.485, 0.852) 1.189 (1.056, 1.338) 0.996 (0.960, 1.033)	0.874 (0.573, 1.332) 1.075 (0.900, 1.284) 1.071 (1.011, 1.133)	0.637 (0.356, 1.138) 1.199 (0.940, 1.528) 1.151 (1.069, 1.240)	0.319 (0.116, 0.878) 1.493 (0.980, 2.274) 1.053 (0.940, 1.180)	0.515 (0.180, 1.476) 1.333 (0.856, 2.077) 0.907 (0.794, 1.037)	0.646 (0.396, 1.054) 1.183 (0.962, 1.455) 0.930 (0.873, 0.991)	0.564 (0.372, 0.855) 1.240 (1.040, 1.477) 0.933 (0.885, 0.984)	0.398 (0.181, 0.876) 1.402 (1.008, 1.949) 0.938 (0.852, 1.032)
	RE **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.644 (0.485, 0.854) 1.188 (1.055, 1.339) 0.996 (0.960, 1.033) 0.37	0.878 (0.570, 1.352) 1.074 (0.895, 1.287) 1.069 (1.008, 1.134) 4.00	0.638 (0.355, 1.148) 1.197 (0.937, 1.530) 1.150 (1.067, 1.240) 3.04	0.319 (0.115, 0.878) 1.494 (0.980, 2.276) 1.053 (0.939, 1.181) 0.97	0.515 (0.180, 1.477) 1.333 (0.855, 2.077) 0.907 (0.794, 1.037) 0.55	0.646 (0.395, 1.055) 1.183 (0.962, 1.456) 0.930 (0.873, 0.991) 0.56	0.564 (0.372, 0.856) 1.239 (1.040, 1.477) 0.933 (0.885, 0.984) 0.23	0.402 (0.182, 0.887) 1.397 (1.003, 1.946) 0.938 (0.851, 1.032) 3.06

REVIEW DRAFT

Table 31: Hazard ratios of $PM_{2.5}$LUR individual-level (IQR= 5.35) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level covariates (1990), stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
Individual covariates	**SC**	0.993 (0.966, 1.020)	1.072 (1.027, 1.119)	1.127 (1.065, 1.193)	0.952 (0.873, 1.038)	0.991 (0.898, 1.094)	0.960 (0.916, 1.006)	0.944 (0.908, 0.982)	0.910 (0.847, 0.977)
	RE	0.996 (0.968, 1.026) 4.35	1.073 (1.026, 1.123) 6.84	1.129 (1.064, 1.198) 11.05	0.953 (0.873, 1.040) 3.61	0.991 (0.898, 1.094) 0.56	0.961 (0.915, 1.009) 4.52	0.946 (0.908, 0.985) 3.50	0.914 (0.849, 0.984) 10.64
+7 1990 Ecological	**SC**	1.005 (0.976, 1.035)	1.079 (1.031, 1.130)	1.147 (1.080, 1.218)	0.985 (0.898, 1.081)	1.017 (0.915, 1.131)	0.967 (0.920, 1.017)	0.956 (0.916, 0.997)	0.931 (0.863, 1.005)
	RE	1.006 (0.976, 1.036) 0.92	1.080 (1.030, 1.132) 4.81	1.148 (1.079, 1.221) 5.14	0.986 (0.898, 1.082) 0.95	1.018 (0.915, 1.131) 0.51	0.967 (0.919, 1.018) 1.85	0.956 (0.917, 0.998) 1.08	0.935 (0.865, 1.011) 8.76
+ LA indicator	**SC** **LA Ind.** **$PM_{2.5}$**	0.961 (0.925, 1.000) 1.019 (0.987, 1.053)	1.063 (1.002, 1.129) 1.055 (1.003, 1.110)	1.023 (0.945, 1.107) 1.138 (1.064, 1.216)	0.863 (0.760, 0.979) 1.037 (0.936, 1.149)	0.940 (0.814, 1.085) 1.040 (0.925, 1.169)	0.911 (0.851, 0.976) 1.000 (0.945, 1.057)	0.898 (0.848, 0.951) 0.993 (0.948, 1.040)	0.871 (0.784, 0.967) 0.977 (0.898, 1.063)
	RE **LA Ind.** **$PM_{2.5}$**	0.962 (0.924, 1.000) 1.020 (0.987, 1.053) 0.68	1.064 (1.000, 1.131) 1.056 (1.002, 1.112) 4.37	1.022 (0.942, 1.108) 1.138 (1.063, 1.219) 5.01	0.862 (0.760, 0.979) 1.037 (0.936, 1.149) 0.91	0.940 (0.814, 1.085) 1.040 (0.925, 1.169) 0.50	0.912 (0.851, 0.977) 1.000 (0.945, 1.058) 0.96	0.898 (0.848, 0.952) 0.993 (0.948, 1.040) 0.34	0.872 (0.784, 0.969) 0.979 (0.899, 1.067) 5.93
+ LA * $PM_{2.5}$	**SC** **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.829 (0.655, 1.050) 1.060 (0.966, 1.164) 1.011 (0.976, 1.046)	1.045 (0.730, 1.497) 1.006 (0.873, 1.160) 1.054 (0.997, 1.114)	0.835 (0.517, 1.347) 1.084 (0.898, 1.308) 1.123 (1.044, 1.209)	0.618 (0.282, 1.351) 1.141 (0.840, 1.550) 1.019 (0.913, 1.138)	0.523 (0.219, 1.251) 1.261 (0.897, 1.772) 1.006 (0.886, 1.142)	0.608 (0.401, 0.922) 1.174 (0.997, 1.383) 0.977 (0.920, 1.038)	0.710 (0.501, 1.006) 1.098 (0.957, 1.259) 0.980 (0.932, 1.031)	1.030 (0.545, 1.946) 0.934 (0.726, 1.202) 0.986 (0.901, 1.079)
	RE **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.832 (0.656, 1.054) 1.059 (0.965, 1.163) 1.011 (0.976, 1.047) 0.54	1.056 (0.732, 1.524) 1.002 (0.867, 1.158) 1.055 (0.996, 1.117) 4.37	0.836 (0.515, 1.358) 1.083 (0.895, 1.310) 1.124 (1.043, 1.211) 4.86	0.617 (0.282, 1.351) 1.142 (0.840, 1.551) 1.019 (0.913, 1.138) 0.91	0.523 (0.219, 1.252) 1.261 (0.897, 1.772) 1.005 (0.886, 1.142) 0.49	0.609 (0.401, 0.924) 1.174 (0.997, 1.383) 0.977 (0.920, 1.039) 0.75	0.711 (0.501, 1.008) 1.097 (0.957, 1.259) 0.980 (0.932, 1.031) 0.30	1.037 (0.546, 1.969) 0.932 (0.723, 1.202) 0.989 (0.902, 1.084) 6.07

REVIEW DRAFT

...Continues from previous page (Table 31).

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
+CMSA 5-indicator	SC								
	CM1	0.785 (0.683, 0.903)	0.735 (0.587, 0.919)	0.872 (0.660, 1.153)	0.775 (0.492, 1.223)	1.225 (0.786, 1.909)	0.829 (0.651, 1.057)	0.845 (0.693, 1.030)	0.878 (0.623, 1.237)
	CM2	0.988 (0.937, 1.041)	1.061 (0.977, 1.152)	1.159 (1.037, 1.295)	0.844 (0.716, 0.994)	0.980 (0.803, 1.197)	1.016 (0.925, 1.116)	0.966 (0.894, 1.044)	0.864 (0.752, 0.992)
	CM3	0.969 (0.905, 1.038)	0.939 (0.842, 1.047)	1.018 (0.880, 1.179)	0.916 (0.742, 1.131)	1.075 (0.843, 1.369)	1.024 (0.910, 1.153)	1.026 (0.931, 1.131)	1.033 (0.869, 1.227)
	CM4	1.003 (0.935, 1.076)	0.938 (0.838, 1.051)	1.115 (0.961, 1.294)	0.806 (0.647, 1.003)	0.943 (0.723, 1.230)	1.178 (1.045, 1.328)	1.115 (1.009, 1.232)	0.992 (0.826, 1.190)
	CM5	1.048 (0.997, 1.102)	1.039 (0.959, 1.124)	1.157 (1.039, 1.289)	0.927 (0.795, 1.082)	1.140 (0.949, 1.370)	1.113 (1.020, 1.215)	1.092 (1.016, 1.174)	1.050 (0.923, 1.196)
	PM2.5LUR	1.038 (1.001, 1.077)	1.078 (1.018, 1.141)	1.140 (1.056, 1.231)	1.058 (0.942, 1.188)	1.079 (0.943, 1.234)	0.998 (0.937, 1.064)	1.008 (0.956, 1.064)	1.031 (0.936, 1.135)
	RE								
	CM1	0.785 (0.682, 0.904)	0.735 (0.586, 0.922)	0.872 (0.658, 1.155)	0.775 (0.491, 1.222)	1.225 (0.785, 1.909)	0.829 (0.650, 1.058)	0.845 (0.693, 1.030)	0.878 (0.622, 1.240)
	CM2	0.987 (0.936, 1.041)	1.061 (0.975, 1.155)	1.158 (1.035, 1.296)	0.843 (0.715, 0.994)	0.980 (0.803, 1.196)	1.016 (0.925, 1.116)	0.966 (0.893, 1.044)	0.865 (0.752, 0.995)
	CM3	0.969 (0.905, 1.038)	0.938 (0.839, 1.050)	1.018 (0.878, 1.181)	0.915 (0.741, 1.131)	1.075 (0.843, 1.370)	1.024 (0.910, 1.154)	1.027 (0.931, 1.132)	1.036 (0.871, 1.234)
	CM4	1.003 (0.934, 1.076)	0.938 (0.835, 1.055)	1.115 (0.959, 1.296)	0.806 (0.647, 1.004)	0.942 (0.722, 1.230)	1.178 (1.045, 1.328)	1.115 (1.009, 1.233)	0.992 (0.825, 1.193)
	CM5	1.048 (0.997, 1.102)	1.041 (0.959, 1.129)	1.157 (1.037, 1.291)	0.927 (0.794, 1.082)	1.140 (0.949, 1.370)	1.113 (1.020, 1.215)	1.092 (1.016, 1.174)	1.050 (0.921, 1.197)
	PM2.5LUR	1.039 (1.001, 1.078)	1.079 (1.018, 1.144)	1.141 (1.056, 1.232)	1.058 (0.942, 1.189)	1.079 (0.943, 1.234)	0.999 (0.937, 1.065)	1.009 (0.956, 1.064)	1.032 (0.936, 1.137)
		0.325	3.061	2.379	0.957	0.537	0.496	0.249	3.237

REVIEW DRAFT

Table 32: Hazard ratios of $PM_{2.5}$RS individual-level (IQR= 5.39) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level (1990) covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
Individual covariates	**SC**	0.995 (0.966, 1.024)	1.082 (1.035, 1.132)	1.139 (1.074, 1.209)	0.960 (0.876, 1.053)	0.978 (0.881, 1.087)	0.963 (0.916, 1.013)	0.938 (0.900, 0.979)	0.884 (0.819, 0.955)
	RE	0.998 (0.967, 1.029) 4.38	1.082 (1.032, 1.134) 6.17	1.140 (1.071, 1.214) 10.04	0.962 (0.877, 1.055) 3.51	0.979 (0.881, 1.087) 0.55	0.964 (0.916, 1.015) 4.60	0.940 (0.900, 0.982) 3.49	0.887 (0.820, 0.959) 9.93
+7 1990 Ecological	**SC**	0.999 (0.969, 1.030)	1.077 (1.027, 1.129)	1.139 (1.069, 1.213)	0.988 (0.896, 1.090)	0.992 (0.886, 1.109)	0.968 (0.917, 1.021)	0.945 (0.904, 0.988)	0.896 (0.827, 0.972)
	RE	0.999 (0.969, 1.031) 0.88	1.076 (1.024, 1.131) 4.32	1.139 (1.067, 1.215) 4.52	0.989 (0.896, 1.091) 0.94	0.992 (0.886, 1.110) 0.50	0.968 (0.917, 1.022) 1.79	0.946 (0.904, 0.989) 0.96	0.898 (0.827, 0.976) 7.78
+ LA indicator	**SC** **LA Ind.** **$PM_{2.5}$**	0.966 (0.930, 1.004) 1.011 (0.978, 1.045)	1.069 (1.008, 1.134) 1.052 (0.998, 1.108)	1.037 (0.959, 1.121) 1.125 (1.050, 1.205)	0.864 (0.763, 0.979) 1.038 (0.933, 1.154)	0.957 (0.831, 1.102) 1.007 (0.891, 1.138)	0.911 (0.852, 0.975) 1.000 (0.943, 1.060)	0.904 (0.854, 0.956) 0.979 (0.933, 1.028)	0.888 (0.801, 0.985) 0.934 (0.855, 1.020)
	RE **LA Ind.** **$PM_{2.5}$**	0.967 (0.930, 1.005) 1.011 (0.978, 1.046) 0.66	1.070 (1.007, 1.137) 1.051 (0.995, 1.109) 3.98	1.037 (0.957, 1.122) 1.124 (1.048, 1.206) 4.35	0.864 (0.763, 0.979) 1.038 (0.934, 1.154) 0.96	0.957 (0.830, 1.102) 1.007 (0.891, 1.138) 0.50	0.912 (0.852, 0.976) 1.000 (0.943, 1.060) 0.96	0.904 (0.854, 0.957) 0.979 (0.933, 1.028) 0.33	0.889 (0.801, 0.987) 0.934 (0.855, 1.022) 5.17
+ LA * $PM_{2.5}$	**SC** **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.796 (0.678, 0.935) 1.114 (1.021, 1.216) 0.992 (0.956, 1.029)	0.948 (0.744, 1.209) 1.069 (0.937, 1.220) 1.038 (0.980, 1.100)	0.872 (0.630, 1.207) 1.101 (0.923, 1.312) 1.105 (1.024, 1.193)	0.498 (0.290, 0.856) 1.357 (1.015, 1.814) 0.988 (0.880, 1.110)	0.610 (0.332, 1.117) 1.285 (0.927, 1.783) 0.965 (0.843, 1.104)	0.672 (0.504, 0.895) 1.185 (1.015, 1.385) 0.972 (0.912, 1.036)	0.752 (0.592, 0.956) 1.108 (0.973, 1.262) 0.963 (0.913, 1.015)	0.988 (0.640, 1.523) 0.941 (0.742, 1.194) 0.943 (0.857, 1.038)
	RE **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.796 (0.677, 0.936) 1.114 (1.020, 1.217) 0.992 (0.956, 1.029) 0.41	0.955 (0.745, 1.225) 1.065 (0.930, 1.219) 1.038 (0.978, 1.102) 3.69	0.873 (0.627, 1.214) 1.100 (0.920, 1.315) 1.105 (1.022, 1.194) 4.13	0.498 (0.289, 0.856) 1.358 (1.015, 1.815) 0.989 (0.880, 1.111) 0.94	0.609 (0.332, 1.117) 1.286 (0.927, 1.783) 0.965 (0.843, 1.104) 0.50	0.671 (0.503, 0.896) 1.186 (1.014, 1.387) 0.972 (0.911, 1.036) 0.84	0.752 (0.592, 0.956) 1.108 (0.973, 1.262) 0.963 (0.913, 1.015) 0.31	0.988 (0.638, 1.531) 0.942 (0.741, 1.197) 0.944 (0.856, 1.040) 5.09

REVIEW DRAFT

Table 33: Hazard ratios of Ozone IDW (IQR=24.07) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level and ecologic covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,135.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,432)	(n=8,327)	(n=4,710)	(n=2,044)	(n=1,540)	(n=6,739)	(n=9,676)	(n= 2,937)
Individual covariates	**SC**	0.989 (0.967, 1.012)	1.062 (1.024, 1.101)	1.114 (1.062, 1.168)	0.979 (0.910, 1.053)	0.936 (0.860, 1.018)	0.927 (0.890, 0.965)	0.939 (0.908, 0.971)	0.968 (0.911, 1.029)
	RE	0.987 (0.963, 1.012) 4.40	1.062 (1.022, 1.103) 7.06	1.113 (1.059, 1.171) 10.96	0.977 (0.908, 1.052) 4.82	0.936 (0.860, 1.018) 0.66	0.925 (0.888, 0.964) 3.83	0.938 (0.906, 0.971) 3.67	0.971 (0.911, 1.034) 13.02
+7 1990 Ecol	**SC**	0.979 (0.955, 1.003)	1.051 (1.011, 1.091)	1.097 (1.043, 1.154)	0.974 (0.901, 1.052)	0.917 (0.839, 1.002)	0.919 (0.880, 0.959)	0.930 (0.897, 0.964)	0.958 (0.898, 1.022)
	RE	0.979 (0.955, 1.003) 0.83	1.051 (1.009, 1.094) 5.77	1.097 (1.041, 1.156) 7.15	0.973 (0.901, 1.052) 1.24	0.917 (0.839, 1.002) 0.57	0.918 (0.880, 0.959) 0.84	0.930 (0.897, 0.964) 0.53	0.960 (0.898, 1.025) 8.81
+ LA indicator	**SC** **LA Ind.** **Ozone**	0.982 (0.945, 1.020) 0.984 (0.958, 1.011)	1.081 (1.020, 1.145) 1.028 (0.986, 1.071)	1.051 (0.974, 1.135) 1.082 (1.024, 1.143)	0.870 (0.769, 0.984) 1.011 (0.930, 1.100)	1.006 (0.874, 1.157) 0.915 (0.830, 1.009)	0.954 (0.893, 1.021) 0.931 (0.889, 0.976)	0.922 (0.872, 0.975) 0.952 (0.916, 0.990)	0.855 (0.772, 0.947) 1.000 (0.932, 1.073)
	RE **LA Ind.** **Ozone**	0.982 (0.945, 1.021) 0.984 (0.957, 1.011) 0.71	1.081 (1.018, 1.148) 1.027 (0.984, 1.073) 5.16	1.051 (0.971, 1.137) 1.082 (1.022, 1.145) 6.75	0.870 (0.769, 0.984) 1.011 (0.930, 1.099) 1.04	1.006 (0.874, 1.158) 0.915 (0.830, 1.010) 0.57	0.955 (0.893, 1.021) 0.931 (0.888, 0.976) 0.69	0.922 (0.872, 0.976) 0.952 (0.915, 0.990) 0.29	0.856 (0.772, 0.949) 1.001 (0.932, 1.075) 4.86
+ LA * O3	**SC** **LA Ind.** **$LA*O_3$** **Ozone**	1.119 (0.885, 1.413) 0.948 (0.863, 1.041) 0.989 (0.961, 1.017)	1.495 (1.049, 2.131) 0.876 (0.759, 1.010) 1.040 (0.996, 1.087)	1.585 (0.989, 2.540) 0.846 (0.700, 1.023) 1.099 (1.037, 1.164)	1.905 (0.876, 4.144) 0.727 (0.531, 0.994) 1.036 (0.950, 1.130)	1.092 (0.465, 2.561) 0.967 (0.687, 1.360) 0.918 (0.829, 1.017)	0.804 (0.532, 1.214) 1.072 (0.909, 1.264) 0.926 (0.882, 0.972)	0.776 (0.548, 1.098) 1.072 (0.933, 1.232) 0.946 (0.909, 0.986)	0.704 (0.369, 1.340) 1.082 (0.836, 1.399) 0.994 (0.924, 1.069)
	RE **LA Ind.** **$LA*O_3$** **Ozone**	1.122 (0.886, 1.422) 0.947 (0.861, 1.041) 0.988 (0.961, 1.017) 0.74	1.483 (1.029, 2.138) 0.879 (0.758, 1.018) 1.040 (0.994, 1.089) 4.61	1.576 (0.972, 2.553) 0.848 (0.698, 1.029) 1.099 (1.036, 1.166) 6.01	1.909 (0.876, 4.157) 0.726 (0.530, 0.994) 1.035 (0.949, 1.129) 1.06	1.093 (0.466, 2.566) 0.966 (0.687, 1.359) 0.918 (0.829, 1.017) 0.57	0.805 (0.532, 1.217) 1.071 (0.908, 1.264) 0.926 (0.881, 0.972) 0.66	0.777 (0.548, 1.100) 1.072 (0.932, 1.232) 0.946 (0.909, 0.986) 0.26	0.716 (0.374, 1.371) 1.075 (0.829, 1.393) 0.995 (0.924, 1.072) 4.21

REVIEW DRAFT

Table 34: Hazard ratios of Ozone BME at ZIP-level (IQR=24.241) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level and ecologic covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 69,551.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=18,542)	(n=7,554)	(n=4,276)	(n=1,828)	(n=1,402)	(n=6,142)	(n=8,811)	(n= 2,669)
Individual covariates	**SC**	0.985 (0.961, 1.010)	1.058 (1.018, 1.099)	1.099 (1.045, 1.157)	0.967 (0.894, 1.045)	0.930 (0.851, 1.017)	0.925 (0.887, 0.966)	0.935 (0.903, 0.969)	0.960 (0.900, 1.024)
	RE	0.984 (0.958, 1.011) 4.99	1.058 (1.016, 1.102) 7.28	1.099 (1.042, 1.160) 10.59	0.966 (0.892, 1.045) 3.36	0.930 (0.851, 1.018) 0.55	0.924 (0.883, 0.966) 6.21	0.935 (0.901, 0.971) 5.17	0.964 (0.900, 1.032) 17.24
+7 1990 Ecol	**SC**	0.973 (0.948, 0.999)	1.045 (1.003, 1.090)	1.076 (1.019, 1.136)	0.962 (0.884, 1.046)	0.900 (0.817, 0.990)	0.917 (0.876, 0.960)	0.925 (0.890, 0.961)	0.945 (0.881, 1.013)
	RE	0.973 (0.947, 1.000) 1.18	1.046 (1.001, 1.092) 6.59	1.076 (1.017, 1.139) 6.94	0.961 (0.884, 1.046) 0.96	0.900 (0.818, 0.991) 0.45	0.917 (0.875, 0.960) 1.24	0.925 (0.890, 0.961) 0.52	0.948 (0.882, 1.019) 11.88
+ LA indicator	**SC** **LA Ind.** **Ozone**	0.990 (0.951, 1.030) 0.977 (0.948, 1.006)	1.096 (1.030, 1.165) 1.013 (0.967, 1.062)	1.073 (0.988, 1.164) 1.051 (0.989, 1.118)	0.875 (0.768, 0.998) 1.005 (0.915, 1.104)	1.063 (0.915, 1.235) 0.881 (0.789, 0.983)	0.959 (0.893, 1.030) 0.930 (0.883, 0.980)	0.926 (0.872, 0.982) 0.950 (0.909, 0.992)	0.856 (0.768, 0.954) 0.995 (0.921, 1.076)
	RE **LA Ind.** **Ozone**	0.991 (0.951, 1.032) 0.976 (0.947, 1.006) 1.12	1.097 (1.028, 1.170) 1.013 (0.964, 1.064) 5.88	1.073 (0.986, 1.167) 1.051 (0.987, 1.120) 6.42	0.875 (0.768, 0.998) 1.005 (0.914, 1.104) 0.76	1.063 (0.915, 1.235) 0.881 (0.789, 0.983) 0.45	0.960 (0.893, 1.031) 0.930 (0.882, 0.980) 1.06	0.926 (0.872, 0.983) 0.950 (0.909, 0.992) 0.29	0.857 (0.768, 0.957) 0.997 (0.921, 1.080) 8.38
+ LA * O3	**SC** **LA Ind.** **$LA*O_3$** **Ozone**	1.070 (0.841, 1.362) 0.968 (0.878, 1.067) 0.980 (0.950, 1.011)	1.559 (1.082, 2.248) 0.864 (0.745, 1.003) 1.030 (0.980, 1.082)	1.763 (1.083, 2.868) 0.815 (0.668, 0.993) 1.075 (1.008, 1.146)	1.748 (0.784, 3.896) 0.752 (0.543, 1.042) 1.031 (0.935, 1.137)	0.933 (0.388, 2.243) 1.054 (0.740, 1.502) 0.876 (0.780, 0.983)	0.767 (0.501, 1.173) 1.096 (0.923, 1.301) 0.922 (0.873, 0.974)	0.689 (0.481, 0.986) 1.128 (0.976, 1.303) 0.939 (0.897, 0.982)	0.528 (0.272, 1.026) 1.217 (0.932, 1.589) 0.977 (0.900, 1.061)
	RE **LA Ind.** **$LA*O_3$** **Ozone**	1.074 (0.839, 1.374) 0.967 (0.875, 1.068) 0.979 (0.948, 1.012) 1.14	1.545 (1.056, 2.258) 0.868 (0.744, 1.013) 1.029 (0.977, 1.084) 5.25	1.750 (1.064, 2.878) 0.817 (0.668, 1.001) 1.074 (1.005, 1.148) 1.14	1.750 (0.784, 3.904) 0.752 (0.543, 1.042) 1.030 (0.934, 1.137) 0.76	0.934 (0.388, 2.247) 1.054 (0.740, 1.502) 0.876 (0.780, 0.983) 0.45	0.767 (0.500, 1.177) 1.096 (0.922, 1.302) 0.922 (0.872, 0.974) 0.94	0.690 (0.481, 0.987) 1.128 (0.976, 1.303) 0.939 (0.897, 0.983) 0.23	0.539 (0.275, 1.057) 1.208 (0.921, 1.583) 0.979 (0.900, 1.065) 6.40

REVIEW DRAFT

Table 35: Hazard ratios of NO_2 IDW (IQR=17.17) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 75,364.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,162)	(n=8,237)	(n=4,667)	(n=2,012)	(n=1,518)	(n=6,636)	(n=9,534)	(n= 2,898)
Individual covariates	**SC**	0.990 (0.967, 1.014)	1.080 (1.042, 1.120)	1.124 (1.071, 1.179)	0.928 (0.861, 1.001)	0.949 (0.870, 1.035)	0.949 (0.911, 0.989)	0.933 (0.902, 0.966)	0.898 (0.844, 0.956)
	RE	0.992 (0.967, 1.017) 4.49	1.081 (1.040, 1.123) 6.46	1.123 (1.068, 1.181) 8.98	0.929 (0.861, 1.002) 3.64	0.949 (0.870, 1.035) 0.62	0.951 (0.911, 0.992) 2.98	0.934 (0.902, 0.968) 2.51	0.899 (0.843, 0.959) 11.15
+7 1990 Ecol	**SC**	1.007 (0.981, 1.034)	1.104 (1.060, 1.150)	1.166 (1.105, 1.230)	0.956 (0.879, 1.040)	0.968 (0.878, 1.067)	0.957 (0.913, 1.002)	0.943 (0.907, 0.980)	0.913 (0.852, 0.979)
	RE	1.008 (0.981, 1.035) 0.91	1.104 (1.059, 1.151) 3.63	1.165 (1.104, 1.230) 1.77	0.956 (0.879, 1.040) 1.20	0.968 (0.878, 1.067) 0.57	0.957 (0.913, 1.002) 1.10	0.943 (0.907, 0.980) 0.70	0.914 (0.852, 0.981) 8.77
+ LA indicator	**SC** **LA Ind.** **NO_2**	0.935 (0.891, 0.981) 1.044 (1.006, 1.082)	1.001 (0.929, 1.079) 1.103 (1.042, 1.168)	0.912 (0.826, 1.006) 1.225 (1.136, 1.321)	0.836 (0.715, 0.977) 1.048 (0.934, 1.177)	0.963 (0.808, 1.147) 0.987 (0.863, 1.128)	0.910 (0.836, 0.990) 1.005 (0.943, 1.071)	0.898 (0.837, 0.964) 0.997 (0.946, 1.052)	0.873 (0.767, 0.993) 0.979 (0.890, 1.078)
	RE **LA Ind.** **NO_2**	0.935 (0.891, 0.982) 1.043 (1.006, 1.083) 0.42	1.003 (0.929, 1.082) 1.102 (1.040, 1.169) 3.64	0.912 (0.826, 1.007) 1.224 (1.135, 1.320) 1.20	0.836 (0.715, 0.977) 1.048 (0.934, 1.177) 1.05	0.963 (0.808, 1.147) 0.987 (0.863, 1.129) 0.57	0.910 (0.836, 0.990) 1.005 (0.943, 1.072) 0.87	0.898 (0.837, 0.964) 0.997 (0.945, 1.052) 0.38	0.876 (0.769, 0.998) 0.978 (0.888, 1.078) 6.15
+ LA * NO2	**SC** **LA Ind.** **LA*NO_2** **NO_2**	0.752 (0.603, 0.937) 1.110 (1.001, 1.231) 1.027 (0.988, 1.069)	0.940 (0.676, 1.307) 1.030 (0.883, 1.203) 1.098 (1.032, 1.169)	0.923 (0.589, 1.446) 0.994 (0.806, 1.225) 1.226 (1.130, 1.330)	0.379 (0.170, 0.845) 1.452 (1.005, 2.098) 1.002 (0.885, 1.135)	0.719 (0.316, 1.636) 1.151 (0.782, 1.693) 0.967 (0.836, 1.118)	0.782 (0.533, 1.148) 1.075 (0.898, 1.288) 0.994 (0.928, 1.066)	0.694 (0.502, 0.961) 1.131 (0.972, 1.318) 0.979 (0.924, 1.038)	0.520 (0.281, 0.960) 1.281 (0.964, 1.704) 0.946 (0.853, 1.050)
	RE **LA Ind.** **LA*NO_2** **NO_2**	0.752 (0.603, 0.938) 1.110 (1.001, 1.231) 1.027 (0.987, 1.069) 0.36	0.947 (0.676, 1.326) 1.028 (0.877, 1.204) 1.097 (1.030, 1.170) 3.58	0.925 (0.589, 1.451) 0.993 (0.805, 1.225) 1.225 (1.129, 1.330) 1.20	0.379 (0.170, 0.846) 1.452 (1.005, 2.099) 1.002 (0.884, 1.135) 1.01	0.719 (0.315, 1.636) 1.151 (0.783, 1.694) 0.967 (0.836, 1.119) 0.57	0.780 (0.531, 1.147) 1.077 (0.898, 1.291) 0.994 (0.927, 1.066) 0.95	0.694 (0.501, 0.961) 1.132 (0.972, 1.319) 0.979 (0.924, 1.038) 0.38	0.524 (0.282, 0.972) 1.278 (0.959, 1.703) 0.946 (0.851, 1.051) 4.90

REVIEW DRAFT

Table 36: Hazard ratios of NO_2LUR (IQR= 4.12) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
Individual covariates	**SC**	0.993 (0.966, 1.020)	1.072 (1.027, 1.119)	1.127 (1.065, 1.193)	0.952 (0.873, 1.038)	0.991 (0.898, 1.094)	0.960 (0.916, 1.006)	0.944 (0.908, 0.982)	0.910 (0.847, 0.977)
	RE	0.996 (0.968, 1.026) 4.35	1.073 (1.026, 1.123) 6.84	1.129 (1.064, 1.198) 11.05	0.953 (0.873, 1.040) 3.61	0.991 (0.898, 1.094) 0.56	0.961 (0.915, 1.009) 4.52	0.946 (0.908, 0.985) 3.50	0.914 (0.849, 0.984) 10.64
+7 1990 Ecological	**SC**	1.005 (0.976, 1.035)	1.079 (1.031, 1.130)	1.147 (1.080, 1.218)	0.985 (0.898, 1.081)	1.017 (0.915, 1.131)	0.967 (0.920, 1.017)	0.956 (0.916, 0.997)	0.931 (0.863, 1.005)
	RE	1.006 (0.976, 1.036) 0.92	1.080 (1.030, 1.132) 4.81	1.148 (1.079, 1.221) 5.14	0.986 (0.898, 1.082) 0.95	1.018 (0.915, 1.131) 0.51	0.967 (0.919, 1.018) 1.85	0.956 (0.917, 0.998) 1.08	0.935 (0.865, 1.011) 8.76
+ LA indicator	**SC** **LA Ind.** **$PM_{2.5}$**	0.961 (0.925, 1.000) 1.019 (0.987, 1.053)	1.063 (1.002, 1.129) 1.055 (1.003, 1.110)	1.023 (0.945, 1.107) 1.138 (1.064, 1.216)	0.863 (0.760, 0.979) 1.037 (0.936, 1.149)	0.940 (0.814, 1.085) 1.040 (0.925, 1.169)	0.911 (0.851, 0.976) 1.000 (0.945, 1.057)	0.898 (0.848, 0.951) 0.993 (0.948, 1.040)	0.871 (0.784, 0.967) 0.977 (0.898, 1.063)
	RE **LA Ind.** **$PM_{2.5}$**	0.962 (0.924, 1.000) 1.020 (0.987, 1.053) 0.68	1.064 (1.000, 1.131) 1.056 (1.002, 1.112) 4.37	1.022 (0.942, 1.108) 1.138 (1.063, 1.219) 5.01	0.862 (0.760, 0.979) 1.037 (0.936, 1.149) 0.91	0.940 (0.814, 1.085) 1.040 (0.925, 1.169) 0.50	0.912 (0.851, 0.977) 1.000 (0.945, 1.058) 0.96	0.898 (0.848, 0.952) 0.993 (0.948, 1.040) 0.34	0.872 (0.784, 0.969) 0.979 (0.899, 1.067) 5.93
+ LA * $PM_{2.5}$	**SC** **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.829 (0.655, 1.050) 1.060 (0.966, 1.164) 1.011 (0.976, 1.046)	1.045 (0.730, 1.497) 1.006 (0.873, 1.160) 1.054 (0.997, 1.114)	0.835 (0.517, 1.347) 1.084 (0.898, 1.308) 1.123 (1.044, 1.209)	0.618 (0.282, 1.351) 1.141 (0.840, 1.550) 1.019 (0.913, 1.138)	0.523 (0.219, 1.251) 1.261 (0.897, 1.772) 1.006 (0.886, 1.142)	0.608 (0.401, 0.922) 1.174 (0.997, 1.383) 0.977 (0.920, 1.038)	0.710 (0.501, 1.006) 1.098 (0.957, 1.259) 0.980 (0.932, 1.031)	1.030 (0.545, 1.946) 0.934 (0.726, 1.202) 0.986 (0.901, 1.079)
	RE **LA Ind.** **LA*$PM_{2.5}$** **$PM_{2.5}$**	0.832 (0.656, 1.054) 1.059 (0.965, 1.163) 1.011 (0.976, 1.047) 0.54	1.056 (0.732, 1.524) 1.002 (0.867, 1.158) 1.055 (0.996, 1.117) 4.37	0.836 (0.515, 1.358) 1.083 (0.895, 1.310) 1.124 (1.043, 1.211) 4.86	0.617 (0.282, 1.351) 1.142 (0.840, 1.551) 1.019 (0.913, 1.138) 0.91	0.523 (0.219, 1.252) 1.261 (0.897, 1.772) 1.005 (0.886, 1.142) 0.49	0.609 (0.401, 0.924) 1.174 (0.997, 1.383) 0.977 (0.920, 1.039) 0.75	0.711 (0.501, 1.008) 1.097 (0.957, 1.259) 0.980 (0.932, 1.031) 0.30	1.037 (0.546, 1.969) 0.932 (0.723, 1.202) 0.989 (0.902, 1.084) 6.07

REVIEW DRAFT

...Continues from previous page (Table 36).

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
+CMSA 5-indicator	SC								
	CM1	0.799 (0.697, 0.917)	0.765 (0.614, 0.954)	0.941 (0.716, 1.238)	0.811 (0.518, 1.268)	1.251 (0.812, 1.927)	0.819 (0.646, 1.039)	0.844 (0.695, 1.025)	0.899 (0.642, 1.257)
	CM2	0.987 (0.937, 1.038)	1.070 (0.988, 1.159)	1.193 (1.070, 1.329)	0.874 (0.746, 1.023)	0.948 (0.783, 1.148)	0.993 (0.907, 1.087)	0.955 (0.886, 1.030)	0.879 (0.769, 1.006)
	CM3	0.956 (0.892, 1.025)	0.922 (0.825, 1.030)	0.995 (0.858, 1.155)	0.924 (0.745, 1.144)	1.022 (0.799, 1.308)	1.008 (0.893, 1.137)	1.015 (0.919, 1.121)	1.036 (0.869, 1.235)
	CM4	1.002 (0.934, 1.074)	0.943 (0.843, 1.056)	1.134 (0.978, 1.314)	0.825 (0.665, 1.025)	0.916 (0.704, 1.193)	1.159 (1.029, 1.305)	1.107 (1.002, 1.222)	1.003 (0.838, 1.202)
	CM5	1.025 (0.973, 1.080)	1.002 (0.922, 1.088)	1.098 (0.981, 1.228)	0.924 (0.787, 1.084)	1.064 (0.879, 1.288)	1.094 (0.999, 1.198)	1.078 (1.000, 1.162)	1.047 (0.915, 1.197)
	NO2LUR	1.047 (1.012, 1.084)	1.072 (1.016, 1.132)	1.100 (1.024, 1.181)	1.002 (0.898, 1.118)	1.169 (1.031, 1.325)	1.042 (0.982, 1.106)	1.030 (0.980, 1.083)	1.002 (0.916, 1.097)
	RE								
	CM1	0.799 (0.696, 0.917)	0.766 (0.613, 0.957)	0.941 (0.714, 1.241)	0.810 (0.518, 1.267)	1.251 (0.811, 1.927)	0.819 (0.646, 1.040)	0.844 (0.695, 1.025)	0.899 (0.642, 1.260)
	CM2	0.986 (0.937, 1.038)	1.070 (0.986, 1.162)	1.192 (1.068, 1.330)	0.873 (0.746, 1.023)	0.948 (0.783, 1.148)	0.993 (0.906, 1.087)	0.955 (0.886, 1.030)	0.880 (0.769, 1.008)
	CM3	0.956 (0.892, 1.026)	0.921 (0.821, 1.032)	0.995 (0.855, 1.158)	0.923 (0.744, 1.144)	1.022 (0.799, 1.309)	1.008 (0.893, 1.137)	1.016 (0.919, 1.122)	1.039 (0.870, 1.241)
	CM4	1.002 (0.934, 1.074)	0.943 (0.839, 1.059)	1.133 (0.975, 1.317)	0.825 (0.665, 1.025)	0.916 (0.703, 1.193)	1.159 (1.028, 1.306)	1.107 (1.002, 1.222)	1.004 (0.836, 1.205)
	CM5	1.025 (0.972, 1.080)	1.003 (0.922, 1.092)	1.097 (0.979, 1.230)	0.923 (0.786, 1.084)	1.064 (0.879, 1.289)	1.094 (0.998, 1.198)	1.078 (0.999, 1.162)	1.046 (0.913, 1.198)
	NO2LUR	1.048 (1.012, 1.085)	1.074 (1.016, 1.134)	1.101 (1.024, 1.183)	1.002 (0.898, 1.118)	1.169 (1.031, 1.325)	1.042 (0.982, 1.107)	1.031 (0.980, 1.083)	1.003 (0.916, 1.098)
		0.376	3.112	3.127	0.889	0.508	0.553	0.266	3.089

REVIEW DRAFT

Table 37: Hazard ratios of PM_{10} (IQR=17.3150) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,135.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,432)	(n=8,327)	(n=4,710)	(n=2,044)	(n=1,540)	(n=6,739)	(n=9,676)	(n= 2,937)
Individual covariates	SC	1.001 (0.978, 1.025)	1.081 (1.042, 1.121)	1.137 (1.084, 1.193)	0.982 (0.911, 1.059)	0.969 (0.889, 1.056)	0.955 (0.916, 0.995)	0.946 (0.914, 0.980)	0.929 (0.872, 0.989)
	RE	1.002 (0.977, 1.028) 4.33	1.082 (1.041, 1.124) 6.25	1.138 (1.082, 1.197) 9.42	0.982 (0.910, 1.060) 4.26	0.969 (0.889, 1.056) 0.68	0.955 (0.916, 0.996) 3.34	0.947 (0.914, 0.981) 3.27	0.930 (0.872, 0.992) 11.55
+7 1990 Ecol	SC	1.002 (0.977, 1.029)	1.079 (1.036, 1.123)	1.136 (1.077, 1.198)	1.007 (0.927, 1.093)	0.972 (0.884, 1.069)	0.957 (0.915, 1.001)	0.946 (0.911, 0.983)	0.924 (0.863, 0.990)
	RE	1.002 (0.976, 1.029) 0.84	1.079 (1.034, 1.125) 4.69	1.136 (1.075, 1.200) 5.10	1.006 (0.926, 1.094) 1.15	0.972 (0.884, 1.069) 0.59	0.957 (0.914, 1.002) 0.95	0.947 (0.911, 0.983) 0.59	0.925 (0.862, 0.993) 7.53
+ LA indicator	SC LA Ind. PM10	0.966 (0.929, 1.003) 1.013 (0.985, 1.043)	1.063 (1.003, 1.127) 1.058 (1.012, 1.106)	1.028 (0.952, 1.111) 1.126 (1.062, 1.194)	0.847 (0.748, 0.959) 1.059 (0.968, 1.159)	0.964 (0.837, 1.110) 0.983 (0.886, 1.092)	0.925 (0.865, 0.990) 0.981 (0.933, 1.031)	0.909 (0.859, 0.961) 0.976 (0.936, 1.017)	0.873 (0.787, 0.968) 0.965 (0.895, 1.040)
	RE LA Ind. PM_{10}	0.966 (0.929, 1.004) 1.013 (0.984, 1.043) 0.61	1.064 (1.002, 1.130) 1.057 (1.009, 1.108) 4.41	1.028 (0.950, 1.112) 1.126 (1.061, 1.196) 4.95	0.847 (0.748, 0.960) 1.059 (0.968, 1.159) 1.02	0.964 (0.837, 1.110) 0.984 (0.886, 1.092) 0.59	0.926 (0.865, 0.990) 0.981 (0.933, 1.031) 0.70	0.909 (0.859, 0.962) 0.976 (0.936, 1.017) 0.30	0.874 (0.788, 0.970) 0.965 (0.894, 1.041) 4.45
+ LA *PM10	SC LA Ind. $LA*PM_{10}$ PM10	0.722 (0.568, 0.918) 1.142 (1.024, 1.273) 1.005 (0.976, 1.035)	0.946 (0.656, 1.363) 1.054 (0.894, 1.244) 1.054 (1.007, 1.104)	0.711 (0.438, 1.156) 1.183 (0.951, 1.471) 1.114 (1.049, 1.183)	0.497 (0.221, 1.117) 1.274 (0.886, 1.832) 1.046 (0.954, 1.147)	0.751 (0.315, 1.790) 1.121 (0.757, 1.660) 0.976 (0.877, 1.087)	0.644 (0.423, 0.980) 1.181 (0.976, 1.428) 0.971 (0.923, 1.022)	0.606 (0.425, 0.864) 1.203 (1.025, 1.412) 0.965 (0.925, 1.008)	0.525 (0.272, 1.014) 1.262 (0.938, 1.697) 0.952 (0.881, 1.029)
	RE LA Ind. $LA*PM_{10}$ PM_{10}	0.723 (0.568, 0.921) 1.141 (1.023, 1.273) 1.005 (0.976, 1.035) 0.37	0.955 (0.655, 1.390) 1.050 (0.886, 1.245) 1.054 (1.005, 1.106) 4.28	0.714 (0.436, 1.169) 1.180 (0.946, 1.474) 1.115 (1.048, 1.186) 4.29	0.497 (0.221, 1.118) 1.274 (0.886, 1.833) 1.046 (0.954, 1.147) 1.00	0.752 (0.315, 1.791) 1.120 (0.756, 1.660) 0.977 (0.877, 1.087) 0.59	0.644 (0.422, 0.981) 1.181 (0.976, 1.429) 0.971 (0.923, 1.023) 0.66	0.607 (0.425, 0.865) 1.203 (1.025, 1.412) 0.965 (0.925, 1.008) 0.25	0.530 (0.273, 1.028) 1.257 (0.932, 1.694) 0.953 (0.881, 1.030) 3.18

REVIEW DRAFT

Table 38: Hazard ratios of Sulfate (IQR=2.6705) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual level covariates and Ecologic 1990, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,135.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,432)	(n=8,327)	(n=4,710)	(n=2,044)	(n=1,540)	(n=6,739)	(n=9,676)	(n= 2,937)
Individual covariates	**SC**	0.982 (0.957, 1.009)	1.068 (1.025, 1.113)	1.105 (1.046, 1.166)	0.886 (0.817, 0.966)	0.926 (0.842, 1.019)	0.972 (0.929, 1.017)	0.935 (0.901, 0.972)	0.858 (0.801, 0.920)
	RE	0.984 (0.957, 1.012) 4.16	1.068 (1.023, 1.116) 7.26	1.103 (1.042, 1.169) 12.23	0.889 (0.817, 0.966) 2.02	0.926 (0.842, 1.019) 0.66	0.972 (0.928, 1.019) 3.93	0.936 (0.900, 0.974) 2.83	0.859 (0.800, 0.922) 5.91
+7 1990 Ecol	**SC**	1.002 (0.973, 1.032)	1.091 (1.042, 1.143)	1.145 (1.007, 1.217)	0.918 (0.836, 1.009)	0.957 (0.859, 1.066)	0.992 (0.943, 1.044)	0.950 (0.910, 0.991)	0.859 (0.795, 0.929)
	RE	1.002 (0.973, 1.033) 0.84	1.092 (1.041, 1.146) 4.87	1.144 (1.074, 1.219) 5.45	0.918 (0.836, 1.009) 0.98	0.957 (0.859, 1.067) 0.60	0.992 (0.942, 1.045) 1.34	0.950 (0.910, 0.992) 0.79	0.860 (0.795, 0.931) 3.42
+ LA indicator	**SC** **LA Ind.** **Sulfate**	0.957 (0.917, 0.999) 1.024 (0.987, 1.062)	1.051 (0.984, 1.123) 1.064 (1.004, 1.127)	1.001 (0.917, 1.092) 1.144 (1.061, 1.234)	0.895 (0.779, 1.028) 0.968 (0.864, 1.085)	0.975 (0.834, 1.141) 0.969 (0.849, 1.106)	0.881 (0.818, 0.950) 1.056 (0.992, 1.124)	0.894 (0.839, 0.951) 1.004 (0.953, 1.058)	0.926 (0.824, 1.040) 0.893 (0.810, 0.983)
	RE **LA Ind.** **Sulfate**	0.957 (0.917, 1.000) 1.024 (0.987, 1.062) 0.60	1.052 (0.982, 1.127) 1.063 (1.001, 1.129) 4.77	1.001 (0.916, 1.094) 1.143 (1.058, 1.236) 5.46	0.895 (0.779, 1.028) 0.968 (0.864, 1.086) 1.02	0.975 (0.834, 1.141) 0.969 (0.849, 1.106) 0.60	0.882 (0.818, 0.950) 1.056 (0.991, 1.124) 0.59	0.894 (0.839, 0.952) 1.004 (0.952, 1.058) 0.33	0.927 (0.824, 1.042) 0.893 (0.810, 0.985) 2.92
+ LA *SO_4	**SC** **LA Ind.** **LA*SO_4** **Sulfate**	0.727 (0.537, 0.984) 1.184 (0.984, 1.423) 1.016 (0.979, 1.055)	0.852 (0.538, 1.349) 1.137 (0.860, 1.505) 1.057 (0.996, 1.121)	0.739 (0.398, 1.374) 1.204 (0.826, 1.754) 1.134 (1.049, 1.226)	0.330 (0.113, 0.962) 1.841 (0.965, 3.511) 0.946 (0.842, 1.064)	0.836 (0.281, 2.482) 1.099 (0.565, 2.139) 0.965 (0.842, 1.105)	0.671 (0.395, 1.140) 1.182 (0.857, 1.632) 1.048 (0.983, 1.118)	0.757 (0.487, 1.177) 1.107 (0.846, 1.449) 1.000 (0.947, 1.055)	1.023 (0.463, 2.260) 0.940 (0.579, 1.526) 0.895 (0.811, 0.988)
	RE **LA Ind.** **LA*SO_4** **Sulfate**	0.726 (0.535, 0.986) 1.185 (0.984, 1.427) 1.016 (0.979, 1.055) 0.55	0.870 (0.543, 1.394) 1.124 (0.843, 1.498) 1.057 (0.994, 1.124) 4.48	0.749 (0.399, 1.406) 1.195 (0.815, 1.753) 1.134 (1.047, 1.228) 4.99	0.330 (0.113, 0.962) 1.842 (0.965, 3.518) 0.946 (0.842, 1.064) 1.01	0.835 (0.281, 2.481) 1.100 (0.565, 2.142) 0.965 (0.842, 1.105) 0.61	0.669 (0.393, 1.138) 1.185 (0.857, 1.638) 1.048 (0.983, 1.118) 0.66	0.755 (0.485, 1.176) 1.109 (0.847, 1.452) 0.999 (0.947, 1.055) 0.35	1.019 (0.459, 2.262) 0.942 (0.579, 1.534) 0.895 (0.810, 0.990) 2.86

REVIEW DRAFT

Table 39: Hazard ratios of Road Buffers (EXP100 or Mjr50) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
Individual covariates	SC	1.021 (0.985, 1.059)	1.034 (0.978, 1.093)	1.045 (0.971, 1.126)	1.001 (0.894, 1.120)	0.922 (0.803, 1.058)	0.976 (0.914, 1.042)	1.008 (0.955, 1.063)	1.081 (0.984, 1.188)
	RE	1.021 (0.984, 1.059) 4.33	1.036 (0.979, 1.095) 7.88	1.047 (0.972, 1.128) 14.29	1.001 (0.894, 1.120) 4.47	0.922 (0.803, 1.058) 0.57	0.974 (0.912, 1.040) 5.41	1.006 (0.953, 1.062) 5.04	1.082 (0.984, 1.189) 15.45
+7 1990 Ecol	SC	1.014 (0.978, 1.052)	1.029 (0.973, 1.088)	1.036 (0.962, 1.116)	0.985 (0.880, 1.104)	0.914 (0.797, 1.050)	0.971 (0.910, 1.037)	1.001 (0.949, 1.056)	1.069 (0.973, 1.175)
	RE	1.014 (0.978, 1.052) 0.88	1.030 (0.974, 1.090) 5.97	1.038 (0.963, 1.118) 8.93	0.985 (0.880, 1.104) 0.96	0.914 (0.796, 1.050) 0.50	0.970 (0.909, 1.036) 2.28	1.001 (0.948, 1.056) 1.90	1.070 (0.974, 1.177) 12.07
+ LA indicator	SC LA Ind. RDbuf	0.972 (0.938, 1.007) 1.013 (0.977, 1.050)	1.096 (1.039, 1.157) 1.035 (0.978, 1.094)	1.096 (1.020, 1.178) 1.042 (0.967, 1.123)	0.878 (0.783, 0.986) 0.979 (0.874, 1.097)	0.956 (0.840, 1.089) 0.913 (0.795, 1.048)	0.910 (0.855, 0.968) 0.966 (0.905, 1.032)	0.895 (0.850, 0.943) 0.995 (0.943, 1.050)	0.863 (0.785, 0.949) 1.060 (0.965, 1.165)
	RE LA Ind. RDbuf	0.972 (0.938, 1.008) 1.013 (0.977, 1.050) 0.67	1.097 (1.037, 1.160) 1.036 (0.979, 1.095) 4.82	1.095 (1.017, 1.180) 1.043 (0.968, 1.124) 7.46	0.878 (0.782, 0.986) 0.979 (0.874, 1.097) 0.88	0.956 (0.840, 1.089) 0.913 (0.795, 1.048) 0.49	0.910 (0.855, 0.969) 0.966 (0.905, 1.031) 1.04	0.895 (0.850, 0.943) 0.995 (0.943, 1.050) 0.35	0.865 (0.786, 0.953) 1.061 (0.965, 1.167) 6.83
+ LA * Mjr50	SC LA Ind. LA*RDbuf RDbuf	0.964 (0.928, 1.002) 1.052 (0.962, 1.149) 1.002 (0.962, 1.044)	1.080 (1.018, 1.145) 1.095 (0.959, 1.249) 1.014 (0.951, 1.080)	1.065 (0.985, 1.152) 1.183 (0.995, 1.407) 1.003 (0.921, 1.092)	0.885 (0.782, 1.003) 0.951 (0.709, 1.275) 0.988 (0.872, 1.120)	0.985 (0.858, 1.131) 0.801 (0.558, 1.149) 0.952 (0.817, 1.108)	0.902 (0.844, 0.964) 1.061 (0.902, 1.249) 0.955 (0.887, 1.027)	0.890 (0.842, 0.941) 1.033 (0.903, 1.183) 0.989 (0.931, 1.050)	0.865 (0.781, 0.959) 0.982 (0.771, 1.252) 1.064 (0.958, 1.181)
	RE LA Ind. LA*RDbuf RDbuf	0.965 (0.928, 1.002) 1.052 (0.962, 1.149) 1.002 (0.962, 1.044) 0.65	1.080 (1.016, 1.147) 1.096 (0.961, 1.251) 1.014 (0.951, 1.081) 4.90	1.064 (0.982, 1.153) 1.184 (0.995, 1.408) 1.004 (0.922, 1.093) 7.39	0.885 (0.782, 1.003) 0.951 (0.709, 1.275) 0.988 (0.872, 1.120) 0.88	0.985 (0.858, 1.131) 0.801 (0.558, 1.149) 0.952 (0.817, 1.108) 0.50	0.902 (0.844, 0.965) 1.061 (0.902, 1.249) 0.955 (0.887, 1.027) 1.03	0.891 (0.842, 0.942) 1.033 (0.903, 1.183) 0.989 (0.931, 1.050) 0.35	0.868 (0.782, 0.963) 0.981 (0.769, 1.250) 1.065 (0.959, 1.183) 6.86

REVIEW DRAFT

Table 40: Hazard ratios of Road Buffer Mjr50 for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and Ecologic covariates level, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
Individual covariates	SC	1.012 (0.976, 1.051)	1.025 (0.968, 1.086)	1.048 (0.972, 1.130)	0.977 (0.870, 1.097)	0.902 (0.782, 1.040)	0.964 (0.902, 1.032)	1.003 (0.950, 1.060)	1.092 (0.992, 1.202)
	RE	1.012 (0.975, 1.050) 4.36	1.027 (0.970, 1.088) 7.85	1.050 (0.973, 1.133) 14.32	0.977 (0.870, 1.097) 4.47	0.902 (0.782, 1.040) 0.57	0.963 (0.900, 1.030) 5.47	1.002 (0.948, 1.059) 5.07	1.092 (0.992, 1.203) 15.40
+7 1990 Ecol	SC	1.007 (0.970, 1.045)	1.021 (0.965, 1.082)	1.041 (0.965, 1.123)	0.963 (0.857, 1.082)	0.896 (0.777, 1.034)	0.961 (0.898, 1.028)	0.998 (0.944, 1.054)	1.080 (0.981, 1.190)
	RE	1.007 (0.970, 1.045) 0.88	1.023 (0.966, 1.084) 5.93	1.043 (0.966, 1.125) 8.95	0.963 (0.857, 1.082) 0.96	0.896 (0.777, 1.034) 0.50	0.960 (0.897, 1.027) 2.32	0.997 (0.944, 1.054) 1.91	1.081 (0.982, 1.191) 12.02
+ LA indicator	SC LA Ind. Mjr50	0.972 (0.938, 1.006) 1.005 (0.968, 1.043)	1.096 (1.038, 1.156) 1.027 (0.970, 1.088)	1.096 (1.020, 1.178) 1.047 (0.970, 1.129)	0.877 (0.782, 0.985) 0.957 (0.852, 1.075)	0.956 (0.839, 1.088) 0.895 (0.776, 1.032)	0.910 (0.855, 0.968) 0.956 (0.894, 1.023)	0.895 (0.850, 0.942) 0.992 (0.938, 1.048)	0.863 (0.785, 0.949) 1.071 (0.973, 1.180)
	RE LA Ind. Mjr50	0.972 (0.938, 1.007) 1.005 (0.968, 1.043) 0.66	1.096 (1.036, 1.159) 1.028 (0.971, 1.089) 4.80	1.095 (1.017, 1.180) 1.048 (0.971, 1.131) 7.48	0.877 (0.782, 0.985) 0.957 (0.852, 1.075) 0.87	0.956 (0.839, 1.089) 0.895 (0.776, 1.032) 0.49	0.910 (0.855, 0.968) 0.956 (0.893, 1.023) 1.05	0.895 (0.850, 0.943) 0.992 (0.938, 1.048) 0.35	0.866 (0.786, 0.954) 1.072 (0.974, 1.181) 6.82
+ LA * Mjr50	SC LA Ind. LA*Mjr50 Mjr50	0.965 (0.929, 1.002) 1.047 (0.955, 1.147) 0.996 (0.955, 1.038)	1.082 (1.021, 1.147) 1.080 (0.943, 1.238) 1.010 (0.946, 1.078)	1.068 (0.989, 1.155) 1.173 (0.982, 1.400) 1.010 (0.926, 1.101)	0.887 (0.784, 1.004) 0.928 (0.684, 1.259) 0.970 (0.853, 1.102)	0.985 (0.859, 1.130) 0.782 (0.536, 1.140) 0.936 (0.800, 1.096)	0.904 (0.846, 0.965) 1.050 (0.887, 1.243) 0.947 (0.878, 1.021)	0.890 (0.842, 0.941) 1.039 (0.904, 1.194) 0.984 (0.926, 1.046)	0.860 (0.776, 0.952) 1.027 (0.803, 1.314) 1.066 (0.958, 1.186)
	RE LA Ind. LA*Mjr50 Mjr50	0.965 (0.929, 1.003) 1.047 (0.955, 1.147) 0.996 (0.955, 1.038) 0.65	1.083 (1.020, 1.149) 1.082 (0.944, 1.240) 1.010 (0.947, 1.079) 4.84	1.068 (0.986, 1.157) 1.172 (0.981, 1.401) 1.011 (0.927, 1.103) 7.36	0.887 (0.784, 1.004) 0.928 (0.684, 1.259) 0.970 (0.853, 1.102) 0.87	0.985 (0.859, 1.130) 0.782 (0.536, 1.140) 0.937 (0.800, 1.096) 0.49	0.904 (0.846, 0.966) 1.050 (0.887, 1.243) 0.947 (0.878, 1.021) 1.04	0.890 (0.842, 0.941) 1.039 (0.904, 1.194) 0.984 (0.926, 1.047) 0.35	0.862 (0.778, 0.956) 1.026 (0.802, 1.312) 1.067 (0.959, 1.188) 6.78

REVIEW DRAFT

Table 41: Hazard ratios of Road Buffer EXP100 for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
Individual covariates	SC	1.051 (0.938, 1.178)	1.035 (0.865, 1.237)	0.902 (0.701, 1.161)	1.195 (0.852, 1.676)	1.148 (0.778, 1.693)	1.055 (0.864, 1.288)	1.027 (0.868, 1.216)	0.967 (0.706, 1.325)
	RE	1.050 (0.936, 1.177) 4.38	1.034 (0.864, 1.237) 7.72	0.902 (0.700, 1.162) 14.13	1.194 (0.851, 1.676) 4.36	1.148 (0.778, 1.692) 0.56	1.053 (0.862, 1.287) 5.15	1.027 (0.867, 1.216) 5.07	0.971 (0.708, 1.330) 15.26
+7 1990 Ecol	SC	1.037 (0.925, 1.162)	1.022 (0.854, 1.222)	0.886 (0.688, 1.140)	1.170 (0.834, 1.643)	1.139 (0.772, 1.679)	1.042 (0.853, 1.272)	1.014 (0.856, 1.200)	0.952 (0.695, 1.305)
	RE	1.037 (0.925, 1.162) 0.88	1.021 (0.853, 1.222) 5.81	0.886 (0.688, 1.142) 8.69	1.170 (0.834, 1.643) 0.96	1.139 (0.772, 1.679) 0.50	1.041 (0.852, 1.272) 2.07	1.014 (0.856, 1.201) 1.90	0.956 (0.698, 1.310) 11.73
+ LA indicator	SC LA Ind. EXP100	0.971 (0.938, 1.006) 1.036 (0.925, 1.162)	1.094 (1.037, 1.155) 1.024 (0.856, 1.224)	1.093 (1.017, 1.175) 0.887 (0.689, 1.142)	0.879 (0.783, 0.987) 1.170 (0.833, 1.642)	0.960 (0.843, 1.093) 1.137 (0.771, 1.678)	0.911 (0.857, 0.970) 1.041 (0.852, 1.271)	0.895 (0.850, 0.943) 1.012 (0.855, 1.199)	0.860 (0.782, 0.946) 0.949 (0.693, 1.302)
	RE LA Ind. EXP100	0.972 (0.938, 1.007) 1.036 (0.925, 1.162) 0.66	1.094 (1.035, 1.157) 1.023 (0.855, 1.224) 4.70	1.092 (1.014, 1.176) 0.888 (0.689, 1.143) 7.27	0.879 (0.783, 0.987) 1.170 (0.833, 1.643) 0.88	0.960 (0.843, 1.093) 1.137 (0.771, 1.677) 0.49	0.912 (0.857, 0.970) 1.040 (0.852, 1.271) 0.94	0.895 (0.850, 0.943) 1.012 (0.855, 1.199) 0.35	0.862 (0.783, 0.950) 0.952 (0.694, 1.304) 6.35
+ LA * EXP100	SC LA Ind. LA*EXP100 EXP100	0.972 (0.938, 1.007) 0.975 (0.749, 1.270) 1.043 (0.914, 1.189)	1.092 (1.034, 1.152) 1.149 (0.778, 1.695) 0.984 (0.795, 1.218)	1.088 (1.012, 1.170) 1.334 (0.779, 2.283) 0.813 (0.597, 1.107)	0.880 (0.784, 0.989) 0.943 (0.422, 2.106) 1.186 (0.806, 1.746)	0.967 (0.848, 1.101) 0.629 (0.214, 1.848) 1.240 (0.812, 1.893)	0.912 (0.857, 0.971) 0.952 (0.592, 1.530) 1.053 (0.838, 1.322)	0.897 (0.852, 0.945) 0.845 (0.558, 1.280) 1.051 (0.869, 1.271)	0.866 (0.787, 0.952) 0.588 (0.245, 1.413) 1.050 (0.745, 1.480)
	RE LA Ind. LA*EXP100 EXP100	0.972 (0.938, 1.008) 0.976 (0.750, 1.271) 1.043 (0.914, 1.189) 0.65	1.092 (1.032, 1.155) 1.154 (0.781, 1.704) 0.982 (0.793, 1.216) 4.75	1.088 (1.009, 1.172) 1.337 (0.780, 2.290) 0.813 (0.597, 1.108) 7.31	0.880 (0.783, 0.989) 0.943 (0.422, 2.107) 1.186 (0.806, 1.746) 0.88	0.967 (0.848, 1.102) 0.629 (0.214, 1.848) 1.239 (0.811, 1.893) 0.49	0.912 (0.857, 0.971) 0.952 (0.593, 1.530) 1.052 (0.837, 1.322) 0.94	0.897 (0.852, 0.946) 0.845 (0.558, 1.280) 1.051 (0.869, 1.271) 0.35	0.868 (0.788, 0.956) 0.587 (0.244, 1.411) 1.053 (0.747, 1.484) 6.39

REVIEW DRAFT

Table 42: Hazard ratios of $PM_{2.5}$ KRG with Road Buffer (EXP100 and Mjr50) for selected causes of death in the ACS cohort with follow-up from 1982 to 2000, adjusting for individual and ecologic level covariates, stratifying the baseline hazard function by age (1-year groupings), gender and race using the Standard Cox Model and the Random Effects model, 1 cluster level (ZIP), (95% confidence intervals given in parenthesis), ZIP variance (10^{-3}) beneath, n = 76,105.

Covariate		Cause of Death							
		All Cause	CVD	IHD	Respiratory	Lung Cancer	All Cancers	Other	Others no Cancers
		(n=20,358)	(n=8,300)	(n=4,693)	(n=2,034)	(n=1,535)	(n=6,713)	(n=9,639)	(n= 2,926)
42 Cov. +7 1990 Ecol	SC								
	EXP100	1.036 (0.924, 1.161)	1.025 (0.857, 1.226)	0.888 (0.690, 1.144)	1.173 (0.836, 1.647)	1.140 (0.772, 1.682)	1.040 (0.852, 1.271)	1.010 (0.853, 1.196)	0.941 (0.686, 1.290)
	Mjr50	1.006 (0.969, 1.044)	1.027 (0.970, 1.088)	1.051 (0.974, 1.134)	0.961 (0.855, 1.080)	0.892 (0.773, 1.029)	0.956 (0.893, 1.023)	0.992 (0.939, 1.049)	1.075 (0.976, 1.184)
	PM2.5KR G	0.993 (0.965, 1.022)	1.097 (1.049, 1.146)	1.149 (1.083, 1.219)	0.987 (0.901, 1.082)	0.925 (0.833, 1.028)	0.921 (0.876, 0.968)	0.918 (0.880, 0.957)	0.912 (0.845, 0.983)
	RE								
	EXP100	1.036 (0.924, 1.161)	1.024 (0.856, 1.225)	0.888 (0.690, 1.144)	1.173 (0.836, 1.647)	1.140 (0.772, 1.682)	1.040 (0.851, 1.271)	1.010 (0.853, 1.196)	0.944 (0.688, 1.294)
	Mjr50	1.006 (0.969, 1.044)	1.028 (0.970, 1.089)	1.052 (0.975, 1.135)	0.961 (0.855, 1.080)	0.892 (0.773, 1.029)	0.956 (0.893, 1.022)	0.992 (0.939, 1.049)	1.076 (0.977, 1.185)
	$PM_{2.5}$KR G	0.993 (0.965, 1.023)	1.096 (1.047, 1.148)	1.149 (1.082, 1.220)	0.987 (0.901, 1.082)	0.926 (0.833, 1.028)	0.921 (0.876, 0.968)	0.918 (0.880, 0.957)	0.913 (0.845, 0.987)
		0.86	3.62	3.17	0.95	0.47	0.86	0.40	8.92

REVIEW DRAFT

Table 43: Long-term and sub-acute cardiovascular mortality log-hazard ratios per 1 ppb of nitrogen dioxide or ozone using a follow-up period of January 1989 to December 2000, adjusted for individual and ecological risk factors using the Standard Cox survival model by definition of exposure time: long-term, 12 month moving average, or joint exposure (long-term and sub-acute). Standard errors are given in parenthesis.

Pollutant	Long-term Exposure (β β)	12 Month Moving Average (γ γ)	Joint Exposure	
			Long-term (ϕ ϕ)	Sub-Acute (λ λ)
Nitrogen Dioxide (per 1 ppb)	0.00571*** (0.00153)	0.00547*** (0.00159)	0.00517** (0.00160)	-0.01167 (0.00876)
Ozone (per 1 ppb)	0.00379* (0.00157)	0.00347* (0.00156)	0.00385* (0.00159)	-0.00263 (0.00489)

*: 0.01<p<0.05
**: 0.001<p<0.01
***: p<0.001

REVIEW DRAFT

Additional figures referenced in results

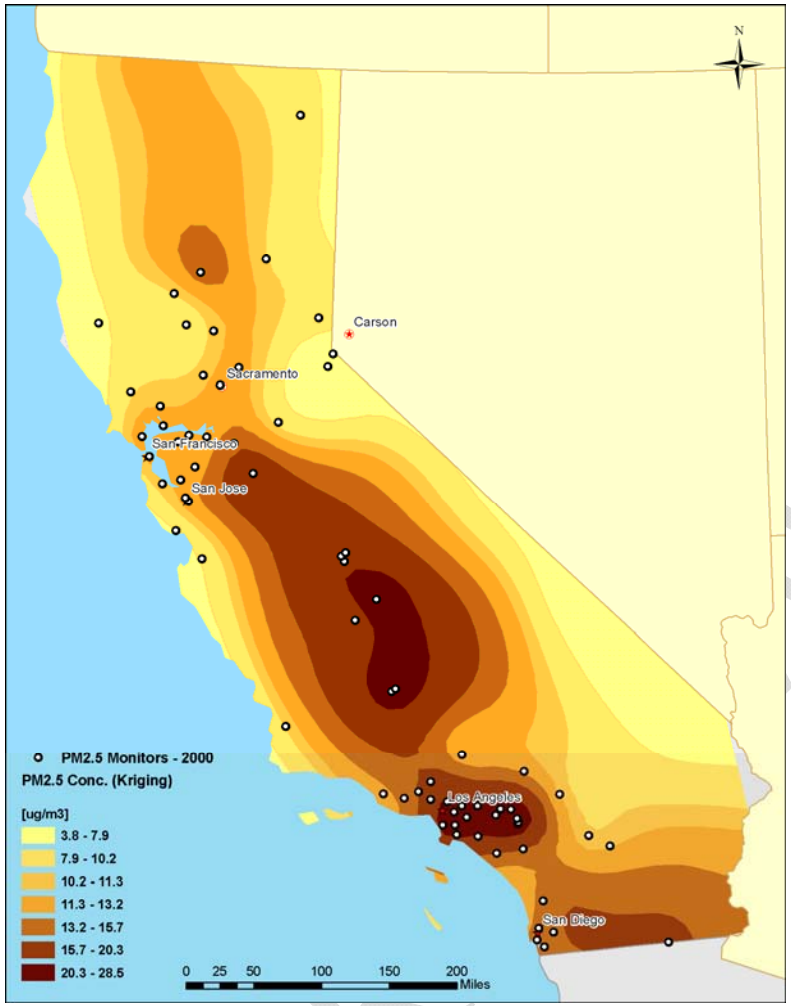


Figure 17: $PM_{2.5}$ map from the kriging interpolation model for 2000

REVIEW DRAFT

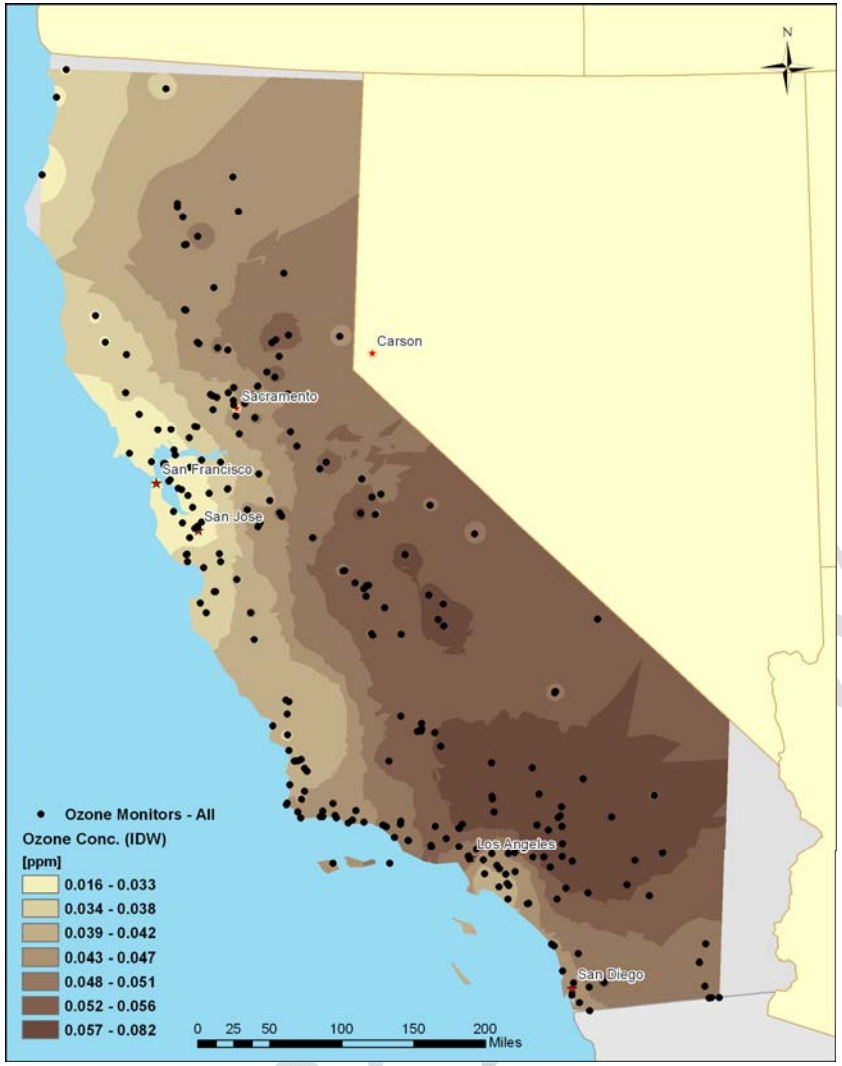


Figure 18: Ozone map from the inverse distance weighting model interpolation model for 1988-2002 Average

REVIEW DRAFT

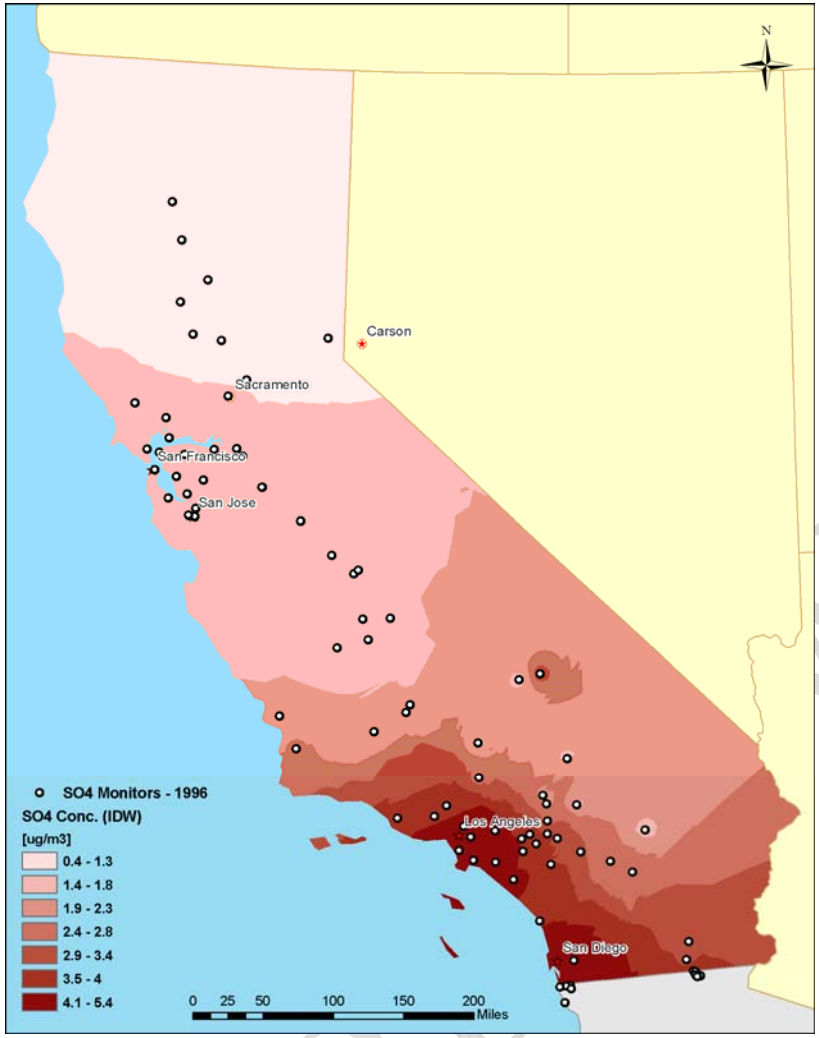


Figure 19: Sulfate map from the inverse distance weighting model interpolation model for 1996

REVIEW DRAFT

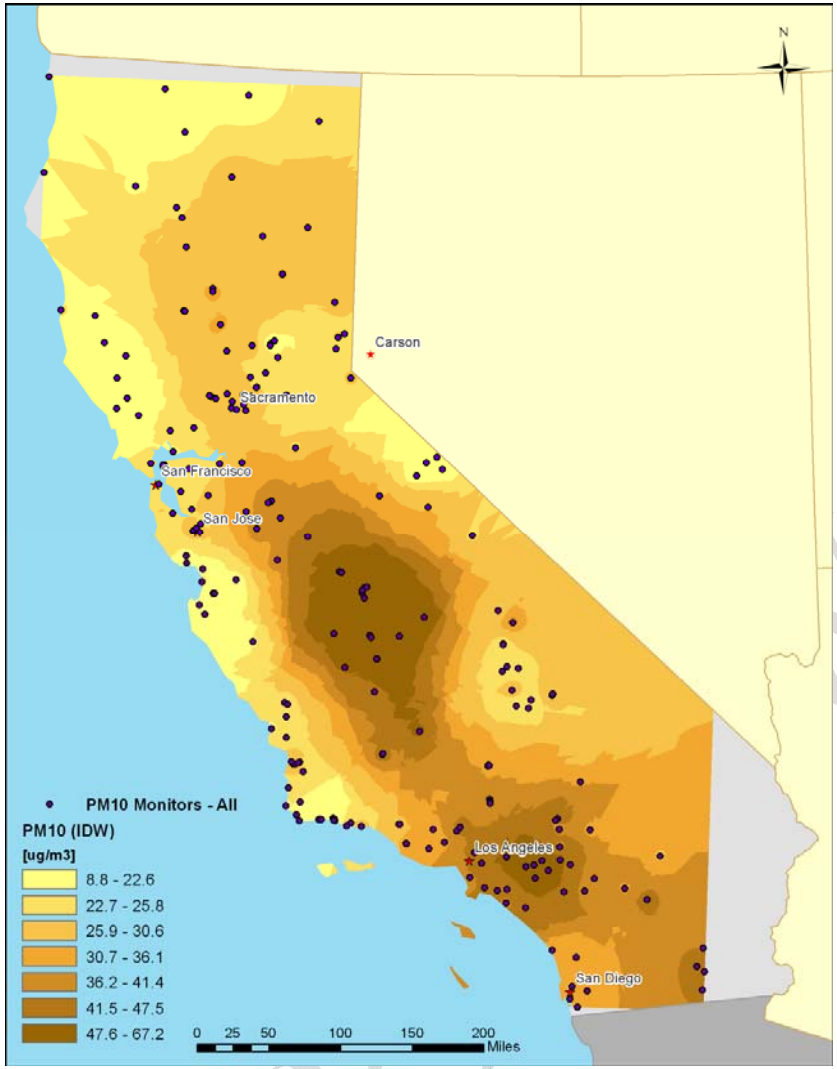


Figure 20: PM_{10} map from the inverse distance weighting interpolation model for 1988-2000 Average

REVIEW DRAFT

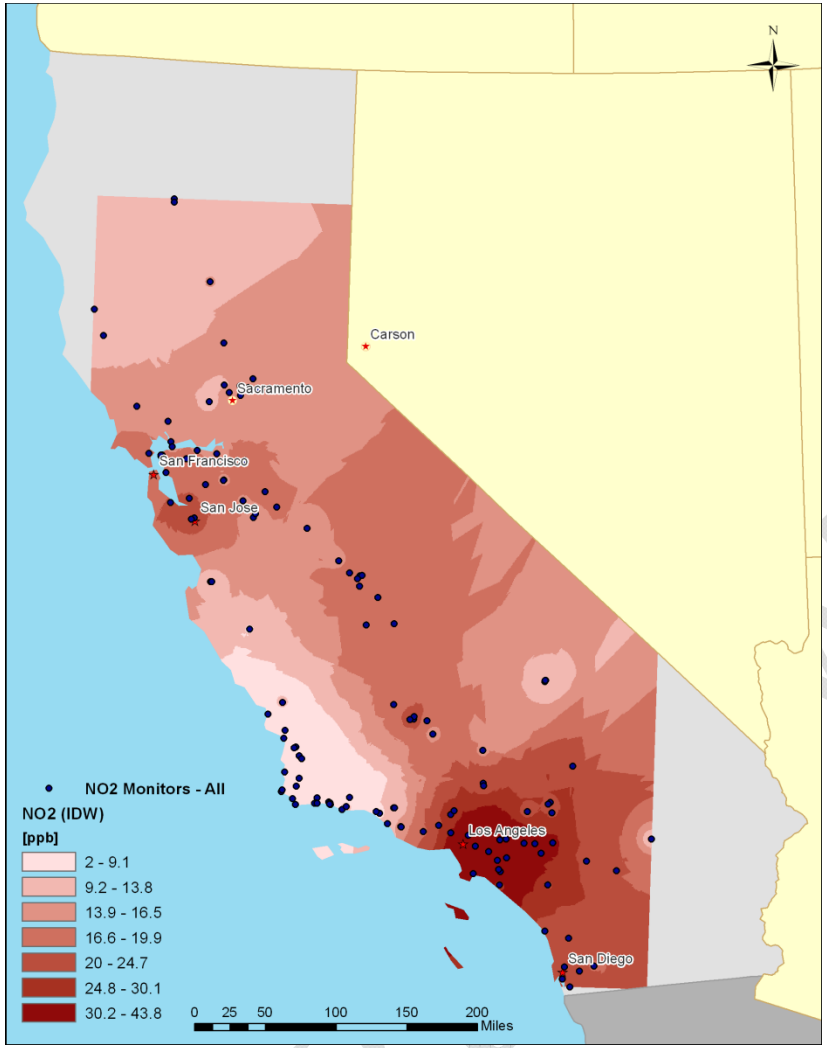


Figure 21: NO_2 map from the inverse distance weighting interpolation model for 1988-2000 Average

REVIEW DRAFT

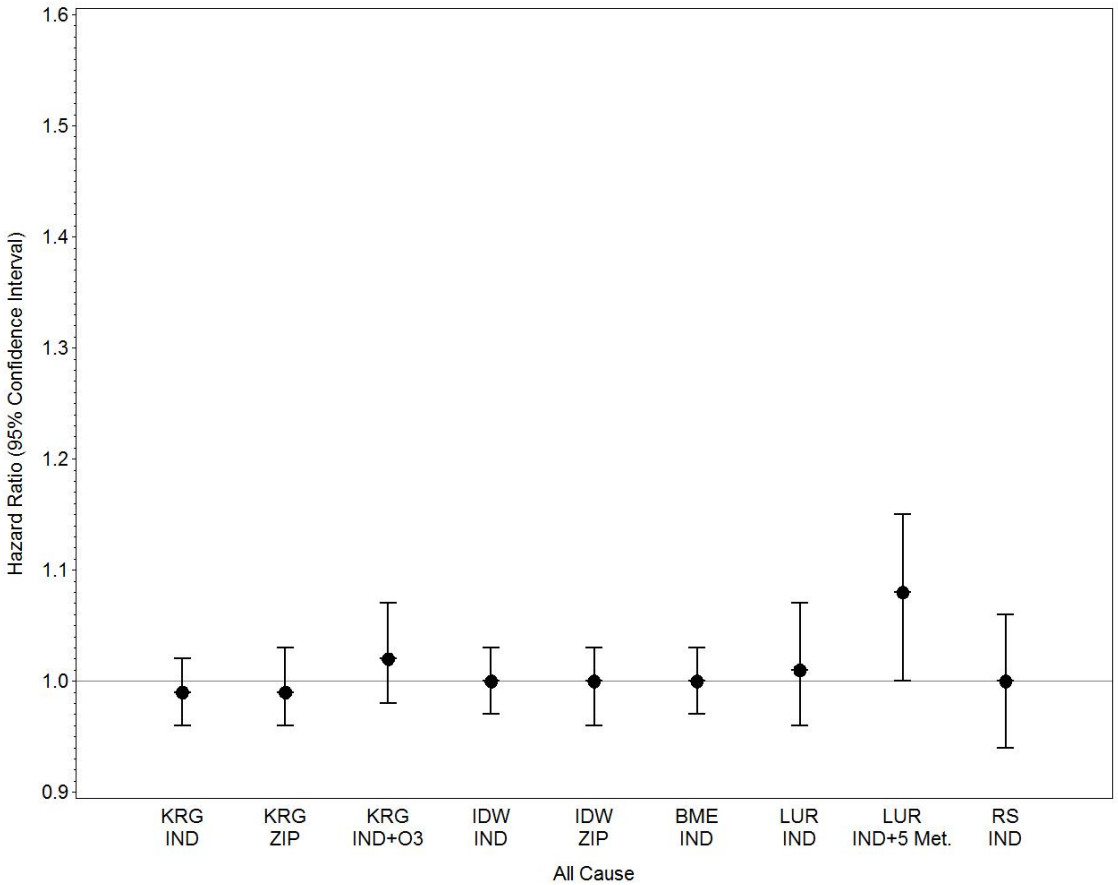


Figure 22: Hazard ratios and 95% confidence intervals for the association between different $PM_{2.5}$ indicators (each 10 ug/m3) at both the individual and ZIP code-level and all cause mortality, follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level covariates (1990), stratifying the baseline hazard function by age (1-year groupings), gender and race using the Random Effects model, 1 cluster level (ZIP)

DRAFT

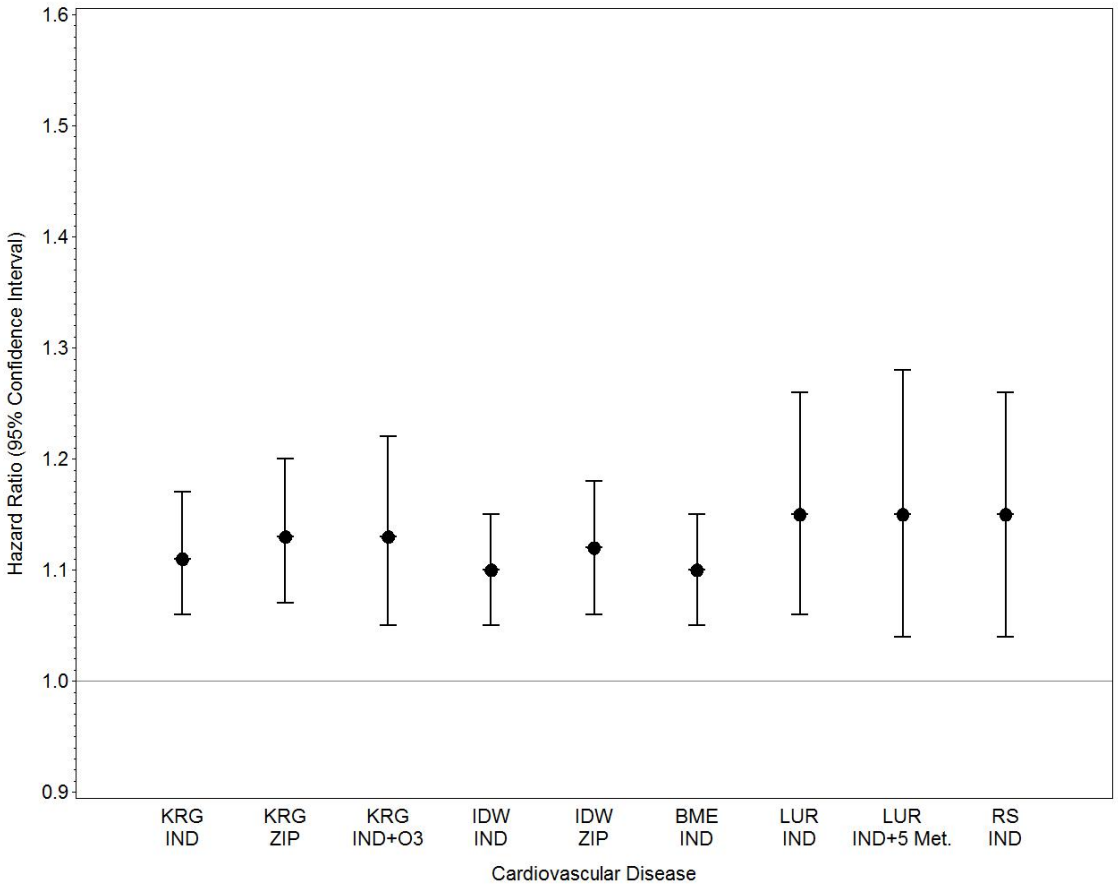


Figure 23: Hazard ratios and 95% confidence intervals for the association between different $PM_{2.5}$ indicators (each 10 ug/m3) at both the individual and ZIP code-level and cardiovascular mortality, follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level covariates (1990), stratifying the baseline hazard function by age (1-year groupings), gender and race using the Random Effects model, 1 cluster level (ZIP)

DRAFT

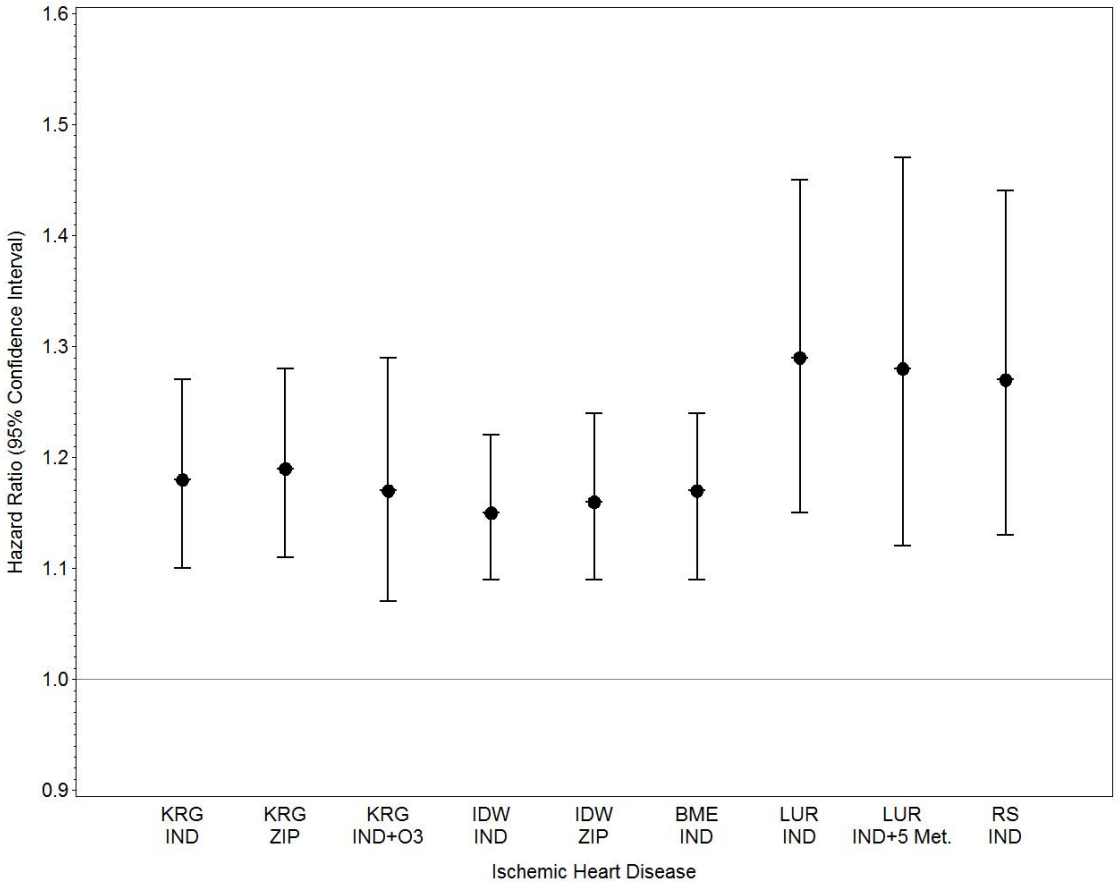


Figure 24: Hazard ratios and 95% confidence intervals for the association between different $PM_{2.5}$ indicators (each 10 ug/m3) at both the individual and ZIP code-level and ischemic heart disease mortality, follow-up from 1982 to 2000, adjusting for individual level covariates and ecologic level covariates (1990), stratifying the baseline hazard function by age (1-year groupings), gender and race using the Random Effects model, 1 cluster level (ZIP)

DRAFT

DISCUSSION

In these analyses we sought to estimate the effects of $PM_{2.5}$ and other air pollutants on premature death in California. This study was motivated by earlier research from Los Angeles that showed $PM_{2.5}$ exerted a large, significant effect on all cause mortality and mortality from CVD and by a lack of statewide dose-response functions for benefits estimates. In the earlier analyses, effects for all causes, CVD, and IHD outcomes were larger than those observed in our national level studies using the ACS CPS II [5]. But in a more recent follow up[3], the effects tended to increase for CVD and IHD in the national study and were of similar size to those observed in LA. The effects on all cause mortality were still about twice the size in LA compared to the recent national study, although they were more uncertain due to the smaller sample size. Consequently, uncertainty exists on the effects that would be observed in a statewide model for California.

Below we summarize the key findings from the present investigation. We then offer narrative interpretation.

Key Findings

1. CVD deaths, especially those from IHD, are consistently and robustly associated with measures of fine particulate and traffic-related air pollution. The effects on CVD and IHD in California are virtually identical to those of the national ACS study (see Table 41).

Table 44: Comparison of Relative Risk Estimates from the California and National American Cancer Society Cohorts for $PM_{2.5}$ using a 10 µg/m^3 Exposure Increment

	California*		National Level**	
	Hazard Ratio	95% CI	Hazard Ratio	95% CI
All Cause	1.08	(1.00, 1.15)	1.08	(1.04, 1.11)
CVD	1.15	(1.04, 1.28)	1.17	(1.11, 1.24)
IHD	1.28	(1.12, 1.47)	1.29	(1.18, 1.40)

* California study uses residential address with a Land Use Regression estimate of exposure

** National level study uses metropolitan area of residence with the average of all $PM_{2.5}$ monitors within the metropolitan area as the exposure estimate; source for the National estimate for all-cause and IHD from Krewski et al. 2009 [3] Table 9; CVD estimate produced for this report for comparison with the California using the same model and sample as in the Krewski report (i.e., two level random effects, with no spatial autocorrelation – referred to as MSA and DIFF in Table 9). Note numbers slightly differ from the Krewski report due to rounding.

Models for both risk estimates control for individual risk factors (e.g., smoking), contextual risk factors (e.g., unemployment in area of residence) and are stratified by age, race and sex. Results for the California cohort are also additionally adjusted for place of residence in five major urban conurbations. Follow up period for both studies was from 1982-2000

REVIEW DRAFT

2. All-cause mortality is significantly associated with fine particulate matter exposure, but the results are sensitive to statistical model specification and to the exposure model used to generate the estimates. When we applied control for residence in the largest urban conurbations, and we employed the LUR model, we found significantly elevated effects on all cause mortality. The effect sizes are the same as those in the national study (see Table 41 for a comparison). We observed effects that were of similar size, but of borderline significance when using other exposure models.

3. The strongest and most consistent effects are observed when there is finer-scale spatial resolution in the exposure predictions. All-cause mortality is associated with $PM_{2.5}$ in Los Angeles where there is a dense monitoring network with reasonably good spatial resolution and in models using the LUR estimate that serve as markers of relatively local variation in pollution (see Figure 25 for a comparison of the risks from statewide LUR models of $PM_{2.5}$ and NO_2 for various causes of death).

DRAFT

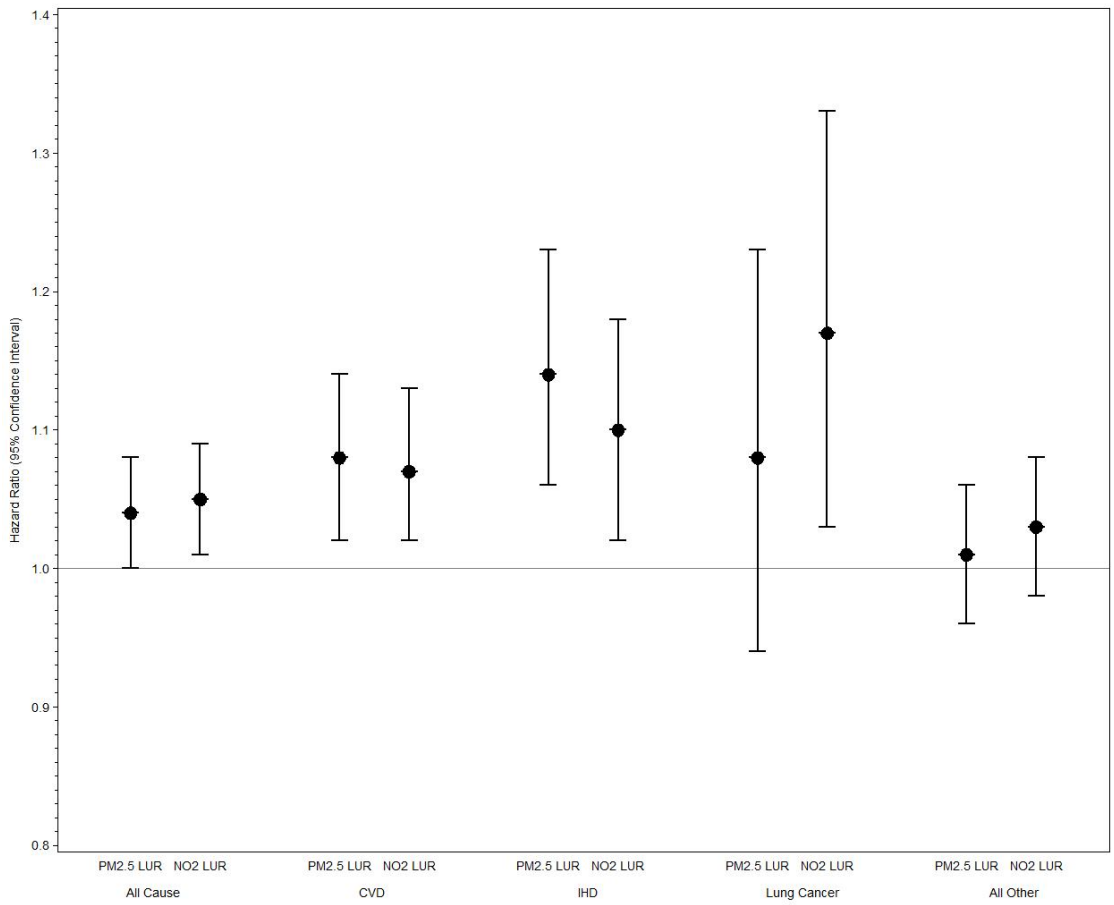


Figure 25: Summary of key results for $PM_{2.5}$ and NO_2 with all-cause and cause specific death. Estimates derived from single pollutant models and calibrated to the inter-quartile range of exposure for each pollutant

4. The strongest evidence of mortality effects is with exposure models that are markers of traffic-related air pollution. The NO_2 LUR estimate has significant associations with all cause, CVD, IHD, and lung cancer deaths. Exposure estimates based on roadway proximity had elevated, but insignificant risks, suggesting weaker effects than with the NO_2 model, probably due to increased exposure measurement error.

5. With regard to other causes of death, there is no evidence of an air pollution effect. In fact for some regional $PM_{2.5}$ exposure there is some evidence of negative association, but when residence in the five largest urban conurbations was accounted for in the model, the effects become positive, but insignificant.

6. Other pollutants − namely PM_{10}, sulfate derived from PM_{10} filters, NO_2, and ozone estimates from interpolation models − all show consistent associations with CVD that are similar in size to those observed for $PM_{2.5}$. In general, the interpolation estimates of these pollutants were highly correlated with each other and with $PM_{2.5}$. Therefore caution must be exercised in interpreting effects from any single pollutant when the exposure estimate relies solely on interpolation.

Narrative

The overall results of these analyses indicate that there is a measurable and significant link between air pollution and cardiovascular disease deaths in California. These results are remarkably robust to controlling for individual and spatial/contextual variables and the use of various approaches to assign pollution exposures. These air pollution-cardiovascular disease death associations are also largely consistent with similar findings from other cohorts in the U.S. and elsewhere. All-cause mortality was associated with $PM_{2.5}$ air pollution exposure in the Los Angeles area and in the state of California as a whole in some model specifications. Consistent with other findings [74], NO_2 was significantly related to all-cause mortality in statewide models.

We found evidence of different effects in Los Angeles with several exposure models for $PM_{2.5}$. With these models there were significantly elevated causes of death for all causes and for respiratory effects in LA that were greater than those observed in statewide analyses or the national study. Many other causes of death also seemed to have elevated risks in LA, which despite having a lower overall death rate, tended to display a steeper dose-response relationship between $PM_{2.5}$ and death. LA has the densest network of monitors in the state. It also has topographic and meteorological conditions that increase the variation in $PM_{2.5}$ compared to other parts of the state [75]. We attribute the larger effects there to the combination of better monitoring data and the relatively high variation in $PM_{2.5}$, both of which may contribute to more precise risk estimates because of lower measurement error and higher inherent variation.

REVIEW DRAFT

We conducted several sensitivity analyses in an effort to understand more fully why the statewide effects for all cause and lung cancer deaths were more sensitive to model specification than in either the national model or the LA-specific model. These included the following:

1. Threshold models to assess whether there was a level below which no effect was observed. We hypothesized that because LA had higher average $PM_{2.5}$ levels than much of the state, the larger effects there may represent a threshold. We found no significant evidence of a threshold effect in the dose-response function.
2. Controlling for demographics by excluding people of Hispanic ethnicity who may have lower rates for some types of cancer and also tend to live in more highly polluted regions, but this had no appreciable effect on the results.
3. Excluding baseline cancer subjects to ensure this did not bias the results, but this did not alter the results.
4. Including more individual covariates for predicting cancer than in our standard model with 20 variables; again the results remained roughly the same.
5. Examining long-term residents only (those in the same resident for 10 years or more at baseline). We observed larger effects in this group, but not significantly different than unity for all causes of death in models that did not control for residence in a large conurbation.
6. Including variables to control for the percent of the county that is urban or rural, which had little influence on the results.
7. Clustering on two levels – county and ZIP code areas – with spatial autocorrelation at each level, and again the results were stable and the similar to other models.
8. Including indicator variables for all major combined statistical areas and the primary statistical area of San Diego to represent the potential influence of the urban versus rural differences in death rates and pollution. These analyses revealed $PM_{2.5}$ effects on all cause mortality when we employed the LUR model, which incorporated remote sensing predictions and land use information likely to influence local variations in the pollutant.

Results must be interpreted with caveats that, in several sensitivity analyses, the exposure models may have lacked the data support necessary to derive exposure estimates capable of detecting effects outside of LA. As we have discussed elsewhere [5], there may be a more toxic mixture of air pollution in LA due to the presence of major port facilities and a relatively higher contribution there from traffic and diesel exhaust, but we are unable to test this effect directly because the monitoring data needed to predict markers of diesel, such as elemental carbon, are unavailable.

In all exposure models, we observed elevated risks for IHD and CVD. Statistical model specification, whether using standard Cox models or multilevel models clustered on the ZIP code of residence, did not substantively influence the results. In addition, the inclusion of updated ecologic covariates from 1990 did not appreciably affect the risks, and if anything had a small positive confounding effect.

We did see some evidence of significant negative associations between cancer and $PM_{2.5}$, but these became null in the more fully specified models with an indicator for LA or with the indicator for the five largest urban conurbations. Models controlling for residence in the five

REVIEW DRAFT

major urban conurbations probably yield the most valid results because they account for the non-metropolitan mortality penalty discussed earlier. Within the indicator variables, there is evidence of a negative association between other causes, including some of the cancers, and air pollution, which is likely indicative of negative confounding due to the non-metropolitan elevation in mortality and the fact that many of the non-metropolitan areas also have lower levels of pollution. Failure to control for this effect probably introduces a negative bias into the air pollution coefficients.

In comparing results obtained with the different $PM_{2.5}$ exposure models (Figures 22-24), the largest HRs per each 10 $\mu g/m^3$ increase in $PM_{2.5}$ were observed for $PM_{2.5}$ RS and $PM_{2.5}$ LUR. In contrast, somewhat lower HRs were observed for $PM_{2.5}$ KRG and $PM_{2.5}$ IDW. Results were comparable for $PM_{2.5}$ KRG and $PM_{2.5}$ IDW at both the individual and ZIP code-levels, although slightly lower for the individual-level assignments. These findings suggest that in models with auxiliary information beyond the monitoring sites, this additional information may reduce exposure measurement error and therefore provide more valid estimates of exposure. The BME model contained auxiliary information, but also had a large smoothing influence from the relatively sparse monitoring network, which tended to increase correlations with other interpolation pollution estimates that did not use auxiliary information. Therefore with the BME, it is not surprising that risk estimates tended to resemble those from the interpolation models. We caution that we have not undertaken formal analyses of measurement error. Thus, our conclusion that more spatially resolved exposure models produce higher risks is based solely on statistical theory about the impact of measurement error and the empirical findings rather than formal analyses testing various models against a gold standard.

Many of the other pollutants tested also displayed relationships that were similar in magnitude and sign to those of $PM_{2.5}$. Ozone, NO_2 (interpolated), PM_{10} and sulfate all had significant associations with CVD and IHD. When tested in co-pollutant models, ozone effects were confounded to the null when $PM_{2.5}$ was in the model. $PM_{2.5}$ estimates were relatively unaffected by the inclusion of ozone.

We observed significantly elevated effects on CVD and IHD deaths using our land use regression (LUR) model for NO_2 (largely influenced by the density of roads). We also found significant NO_2 effects on all-cause mortality and lung cancer deaths. When this NO_2 estimate was included in models with $PM_{2.5}$, both pollutants had significantly elevated effects on mortality from CVD and IHD. The effects of NO_2 on all causes and on deaths from lung cancer remained significantly elevated. The trend toward elevated risks from the road buffers, which measured proximity to highways and major roads, also supports the traffic effect. Although these effects were not significantly elevated, they were consistently positive. As shown in the Appendix, roadways in different areas of the state carry different traffic volumes in different counties. Combined with the other limitation of road buffers for measuring actual pollution exposures (because they do not take into account the continuous field of exposure), we interpret the elevated but non-significant effects as suggestive of a traffic influence on mortality. In other studies where road buffers have been explicitly compared to LUR results, the continuous NO_2 exposure fields from the LUR have been found to be more consistently predictive of mortality than road buffers [74], and that conclusion is supported by our findings here.

REVIEW DRAFT

CONCLUSION

Taken together, the results from this investigation indicate consistent and robust effects of $PM_{2.5}$ − and other pollutants commonly found in the combustion-source mixture with $PM_{2.5}$ − on deaths from CVD and IHD. We also found significant associations between $PM_{2.5}$ and all causes of death, although these findings were sensitive to model specification. In Los Angeles, where the monitoring network is capable of detecting intraurban variations in $PM_{2.5}$, we observed large effects on death from all causes, CVD, IHD, and respiratory disease. These results are consistent with past ACS analyses and with findings from other national or international studies reviewed in this report. Our strongest results are from a land use regression estimate of NO_2, which represents traffic sources, indicating significantly elevated effects on deaths from all causes, CVD, IHD, and lung cancer. **We conclude that combustion-source air pollution, especially from traffic, is significantly associated with premature death in this large cohort of Californians.**

REVIEW DRAFT

Appendices

Appendix A: ZP4 Output Fields

Address (delimited)
Address (final)
Address (original)
Address leftovers
Building/firm
City (final)
City (original)
City (preferred)
City state ZIP (final)
City state ZIP (preferred)
County code
County name
Date certified
DefaultMatch
DPV CMRA
DPV confirm
DPV fastnotes
DPV footnotes
DPV violation
Error codes (combined)
Error message
Error numbers (detailed)
Field5
Field6
Field7
Field8
ID
LACS flag
LACS indicator
LACS return
State (final)
State (original)
Unique ZIP
Version
ZIP (final)
ZIP (five-digit)
ZIP (four-digit add-on)

REVIEW DRAFT

Appendix B: Deriving Estimates of Traffic for the Land Use Regression Models

Summary

Consistent traffic is one of the best predictors of air pollutant levels in land use regression and other deterministic models, particularly for nitrogen dioxide and particulate matter. However, traffic is also one of the weakest data sources in terms of completeness and accuracy. Traffic data for highways is available from the Highway Performance Monitoring System. For large cities, traffic data is available for major roads from Municipal Planning Organizations. Both of these data sources tend to suffer from spatial inaccuracy and fail to include smaller roads that represent a high proportion of overall road density. As a result, existing traffic data sources are inadequate for a statewide study in California.

In lieu of traffic data, several studies have used road density as a proxy. This variable appears to perform reasonably well as an air pollutant predictor but fails to capture the enormous variation between types of roads and between very different areas of the state of California. In order to improve upon simple road density measures and provide a traffic-related variable for use in a statewide study we used more than 56,000 traffic counts stratified by road type and location to generate traffic weights for 2.5 million road links in California. The traffic weights were then used to calculate traffic-weighted road density for air monitoring locations for use in land use regression and related air pollutant models.

Traffic Count Data Used to Generate Traffic Weights

Traffic point-count data was purchased from TeleAtlas. The data provides the count location, traffic volume and year for the most recent counts and often several historical counts for the same location. Although the data includes historical counts going back to the mid-1970s, and the highest density of counts are available for the years 1990-2001. The traffic count data includes a unique ID that enabled a one-to-one link with the TeleAtlas road network data.

Temporal Trends in Traffic Volumes

Between 1990 and 2000, the years with high numbers of counts, we see a slight increase in average traffic volume.

REVIEW DRAFT

Counts as far back as 1976 are included in the TeleAtlas dataset but the volume of data really picks up in 1990 (and even more in 1995) and drops off precipitously after 2001 (Figure 1). In general, the mean traffic volume increases slightly from 1990 to 2001 (Figures 2-4), for the most part this increase is driven by counts on roads with a Functional Class Code of A1 (FCCs are defined in table at right).[8] A2 (Figure 4), A3 and A4 (not shown) have very little, or no, consistent change from year to year. When counts are limited to those since 1995 the increase is still visible though somewhat less pronounced. The slope when volume is regressed against year is about 1400 (1800 with all counts since 1990) meaning that for each year counts increase by about 1400. This increase, however, is primarily limited to FCC A1 and given that the overall mean of A1 counts is about 56,000 the increase is negligible and was not adjusted for. In addition, based on the number of yearly counts available we restricted the dataset included in the analysis to 1990-2001.

FCC Category	Description
A1	Primary interstate
A2	Primary US and state highways
A3	Secondary state and county highways
A4	Local, neighborhood, rural road, city street
Possibly Excluded Categories With Limited Counts	
A5	Vehicular trail, road (4WD) vehicle
A6	Various (access ramp, cul-de-sac, traffic circle)
A7	Various (other thoroughfare, alley, driveway)

Many locations have more than one traffic count.

The data included count information for 56,000 different locations statewide, but each location may include data on as many as five historical counts and we evaluate the potential of using this additional information to strengthen estimates. For example, 43% of the locations have more than one count, 8% have six counts (see table at left). Figure 5 shows the within-location averages against the most recent count and you can see that there is a strong linear relationship with some scatter. In the end the very strong correlation between the most recent count and historical counts provided confidence that using the most current information would provide representative values for a location.

Number of Counts	Frequency	Proportion
1	32327	0.57
2	10354	0.18
3	4021	0.07
4	2671	0.05
5	3053	0.05
6	4257	0.08
Total	56683	1.00

Assigning Traffic Values to Unmeasured Road Links

Statewide averages/medians would be too coarse of a measure

We determined that assigning traffic weights to road links based on statewide medians would be too coarse and not take advantage of the information available in the traffic count data and instead would use "targeted" medians and several approaches were evaluated.

Traffic counts were intersected with county and Census-defined urbanized area layers. Mean, median and counts of traffic volume data was then calculated by county by urban versus non-

[8]Note that the TeleAtlas FCC code has been aggregated to two digits from the original three. The two digit codes represent the broad road-type categories while the extra digit distinguishes between, for example, separated vs un-separated roads.

REVIEW DRAFT

urban and by specific urbanized areas. The county-level means and medians are depicted in Figure 6-10.

County-level not recommended due to distinct differences in within county traffic volumes.

Based on this analysis it was determined that county counts would be less informative than counts related to urbanization or population. This was due to the fact that, in many cases, there can be different levels of urbanization in different areas of the same county that can lead to different traffic volume patters. For example, in Los Angeles County median traffic volume in urbanized areas on FCC A1 is 109,500 while this same number for non-urbanized areas in LA County is just 38,500. Similarly for San Diego median volume on A1 in urbanized areas is 75,000 compared with 32,500 in the same county in non-urbanized areas. In addition, several counties do not have adequate counts within particular FCC categories to calculate stable medians.

There are distinct differences in traffic volume between urbanized and non-urbanized areas.

Box plots of traffic volumes by FCC broken out by urban versus non-urbanized areas suggest that an urban/non-urban split could be a convenient way to aggregate the data (Figure 11). In all FCC categories with enough data urbanized areas have distinctly higher traffic volumes than non-urbanized areas.

Similar to statewide and county-levels, urban versus non-urban statewide, however, may be too coarse a split. We see that within urbanized areas the means and medians differ substantially with a median of 109,500 in urban areas of Los Angeles County, 79,500 in urban areas of Alameda County, 33,000 in urban Fresno.

Population levels within an urbanized area could be a convenient way to cut the data but fails to account for proximity to major urban areas such as Los Angeles

The possibility of breaking all data down into population-urbanization categories to account for differences between urbanized areas was tested. Box and other types of plots suggest that this is a reasonable approach. Figure 12 shows a preliminary categorization of urbanized areas based on population with small, medium, large, very large and Los Angeles with its own category. Figure 13 shows that these categories do a reasonable job of categorization particularly for the highways (A1).

But figure 13 also shows that size alone may not be enough to distinguish very well between small, medium or large urbanized areas, particularly in non-FCCA1. Figures 14 and 15 further illustrate the bias that would result from assignment based on size alone. Using medians based on all traffic counts by urbanized areas this plot shows that assignment based on population mis-specifies some of the urbanized areas. For example, the second and third-ranked cities (after Los Angeles) in terms of median traffic are Mission Viejo and Thousand Oaks. These are relatively small cities in terms of population but their proximity to Los Angeles affects their traffic counts. Instead, a slight alteration of the categorization based on population potential (a distance

REVIEW DRAFT

weighted population calculation) improves categorization but still leaves a series of cities with nearly the same median traffic levels in each of the population categories.

Final Traffic Weight Assignment Method

In the end it was determined that five different broad categories could be used to assign traffic weights:

Urban – These are FCC codes in a specific urban area with at least 15 traffic count points. So, for example, San Francisco FCC A1 has greater than 15 counts and, therefore, all FCC1 in the San Francisco urbanized area will be assigned SF UA-specific medians. All links in the specific urban area with this FCC, regardless of county would be assigned this value.

Urban County – These are FCC codes in a specific urban area that do not have at least 15 traffic count points but the county itself has at least 15 traffic count points in urbanized areas generally. In this case we first calculated county-specific urban area-specific medians (i.e., we would have Los Angeles County—Los Angeles Urban Area; Los Angeles County – Lancaster Urbanized Area etc) and then took the mean of these values. Note, first, that only the portion of an urbanized area within a specific county was included so that we have Orange County – Los Angeles Urban; Ventura County -- Los Angeles Urban etc because the Los Angeles Urbanized area extends in to several counties. Note that we chose to calculate urbanized area-specific medians first and then average these values as opposed to taking the overall median of traffic counts in urbanized areas in a county. We did this so that one urbanized area with a significant number of counts did not swamp all the rest of the urbanized areas.

Urban County Size – These are links in urban areas that don't have the required 15 counts for an FCC AND the county also doesn't have the required number of urban counts. For these instances the state's 58 counties were broken into five size categories Very Low (0-100,000 people; 23 counties), Low (100,001-250,000 13 counties), Medium (250,001-1 million; 14 counties), High (1 million-2.5 million; 5 counties) and Very High (>2.5 million; 3 counties) (see map in Figure 16). Then the median traffic counts using all traffic counts in urban areas in each population size category was calculated.

Rural[9] County – This is basically the same as Urban County but for non-urban areas. Links within a county with enough rural counts on a particular FCC are assigned a county-specific (FCC-specific) median.

Rural County Size – Again, basically the same as the Urban County Size – for links in rural areas in counties that don't have 15+ counts for an FCC, we needed to aggregate counties of a particular population size and use these numbers.

[9] *Note that rural refers to areas not in Census-defined urban areas.*

REVIEW DRAFT

Specific Steps in Detail

see Figure 17 for traffic assignment decision making tree

1. Assign all California traffic counts to a Census-defined urbanized area and county based on location.[10]

2. Calculate two-digit abbreviated FCC code (A1-A6).

3. Limit counts to those 1990-2001 and limit to abbreviated FCC codes A1-A4. Total of 53917 counts.

4. For counts in urbanized areas aggregate most recent traffic count values ("VOLUME" in the TeleAtlas traffic data) by urbanized area and by FCC. Calculate median and the number of counts.

5. For counts in urbanized areas aggregate by county code by urbanized area by FCC. Calculate median and the number of counts.

6. For counts in urbanized areas aggregate by county population category by FCC. Calculate median and the number of counts. (This table is only 20 rows).

7. Merge together three tables representing counts in urbanized areas and create a final traffic assignment field ("trafFIN") for each urbanized-county-FCC combination where urban-area specific values are prioritized (if they have enough counts), followed by within-county averages and, finally, county-size-based estimates.

8. For counts in non-urbanized areas aggregate by county by FCC. Calculate median and the number of counts.

9. For counts in non-urbanized areas aggregate by county population category by FCC.

10. Merge together two tables representing counts in non-urbanized areas and create a final traffic assignment field ("trafFIN") for each county-FCC combination where county-level values are prioritized followed by county-size categories.

11. Append the urban and non-urban tables together and generate a "trafID" which is a concatenation of the urbanized area Census code, county fips code and FCC code separated by "-". Non urbanized areas get a ua code of "00000"

12. Link this data to the 2.5 million TeleAtlas links (A1-A4). (Reminder to Bernie that the original file sent to me has 2000 duplicated links).

[10] *There are 55 urbanized areas and 58 counties in California.*

REVIEW DRAFT

Results

There are a total of 143 possible UA-County combinations (including 58 values that begin with 00000 representing non-urban locations). Thirty-six of the 55 urbanized areas in California exist in just one county, thirteen in two, three in three, one in four and two urbanized areas (LA, Bay Area) touch five different counties. Multiplied by 4 (number of FCC categories) this gives 572 different possible "types" of traffic counts distributed through the 2.5 million records. In total, 500 of these possible combinations actually exist in the road data (for example, rural A2 links in Alameda county do not exist).

Links in urbanized areas make up 34% of road length in California. Among links in urbanized areas 90% of road length was assigned traffic estimates based on urbanized area-specific values (7% assigned based on county-specific and 3% assigned based on urban-county-size calculations). For links in rural areas assignments were split almost evenly between county-specific calculations (52%) and county-size calculations (48%).

Issues and Concerns

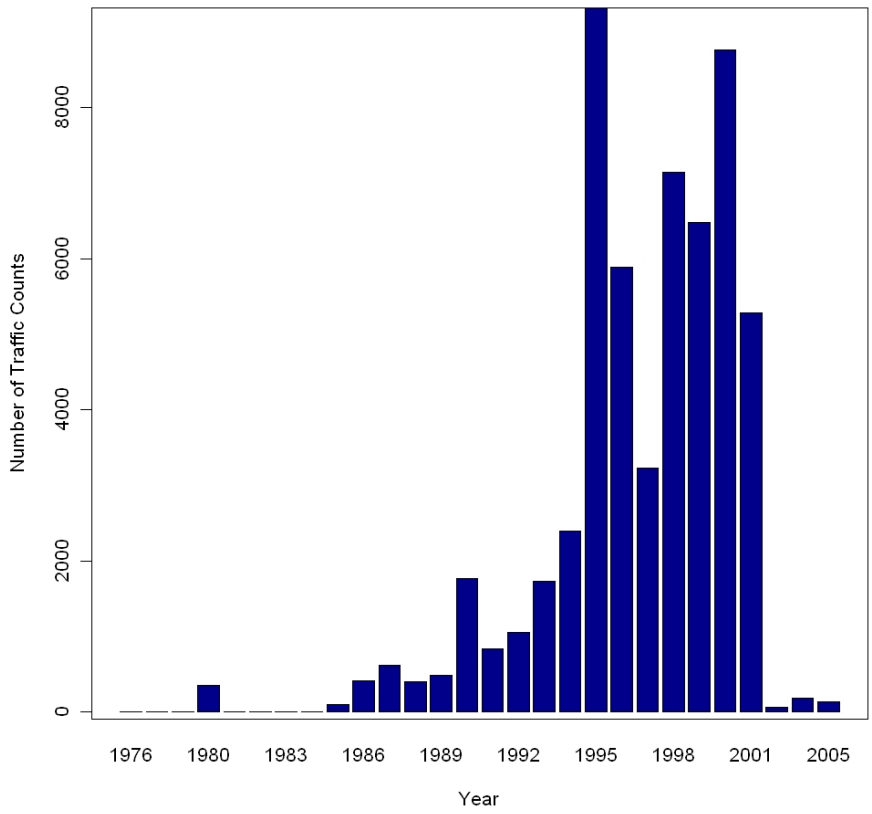
The most obvious overall concern is that assigning the medians to all links with a specific FCC (in an urban area) causes a significant amount of variation in the results. This is an important and basically unavoidable limitation of the method. We do not have enough data to accurately represent all the variation on California roadways. The method discussed above is **a traffic-weighted road density** and cannot be considered a true traffic layer.

In a few cases, urbanized areas can be diverse. For example, the San Francisco urbanized area covers both San Francisco and Alameda County (and three others). In terms of land use, transportation and other factors, these areas are quite different and yet they are treated as a single entity under this method.

Counties were weighted by population size, not density. In general this was considered the preferably choice but there may be places where population density cut-point could yield different estimates. San Bernardino or Riverside County might be examples. These counties have relatively high populations but they are concentrated in the West. As a result the more rural areas may get traffic assignments that are higher than they should be. But if we used population density both the urbanized and more rural areas closer to LA (where the majority of the population lives) might get values lower than they should be.

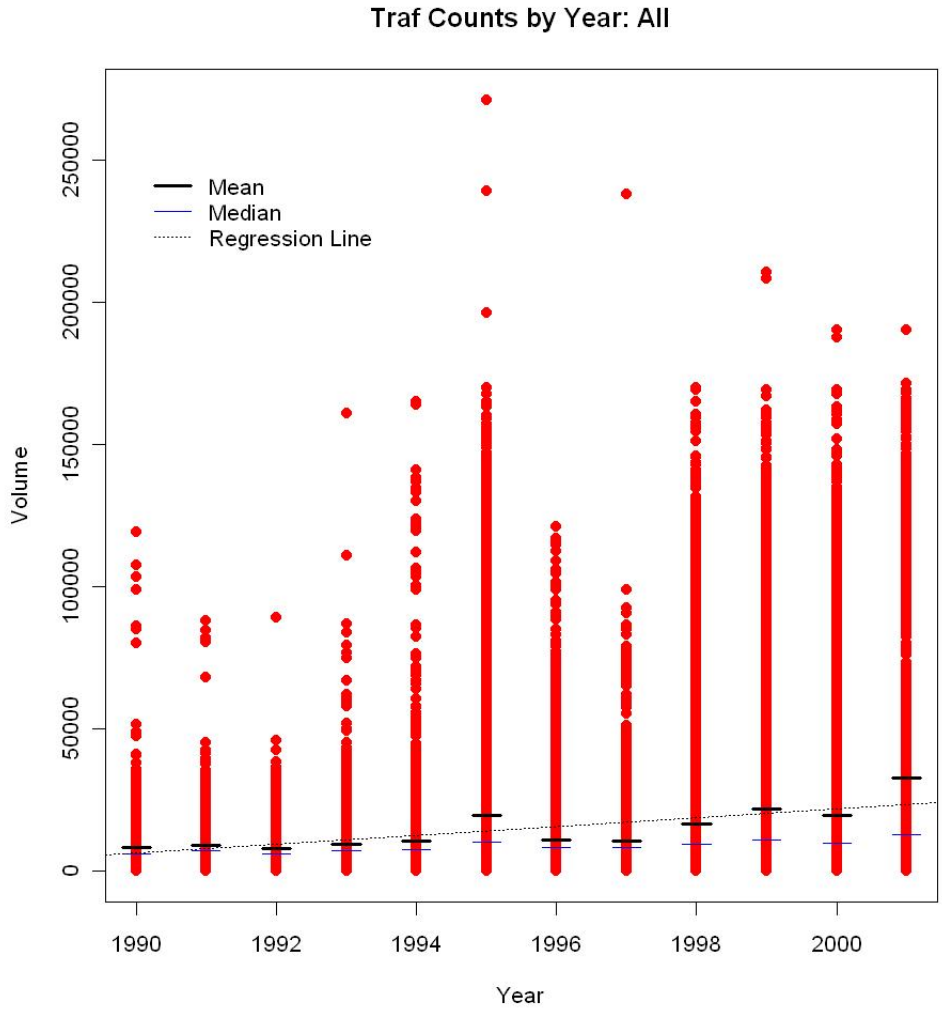
The general assumption in using a layer like this is that the variation we miss is random normal error. We will inevitably underestimate some roads and overestimate others. With tens of thousands of ACS participants and with relatively larger buffers there will not be a significant amount of bias. This is likely a fair assumption but bias will still exist in calculations based on this layer.

REVIEW DRAFT

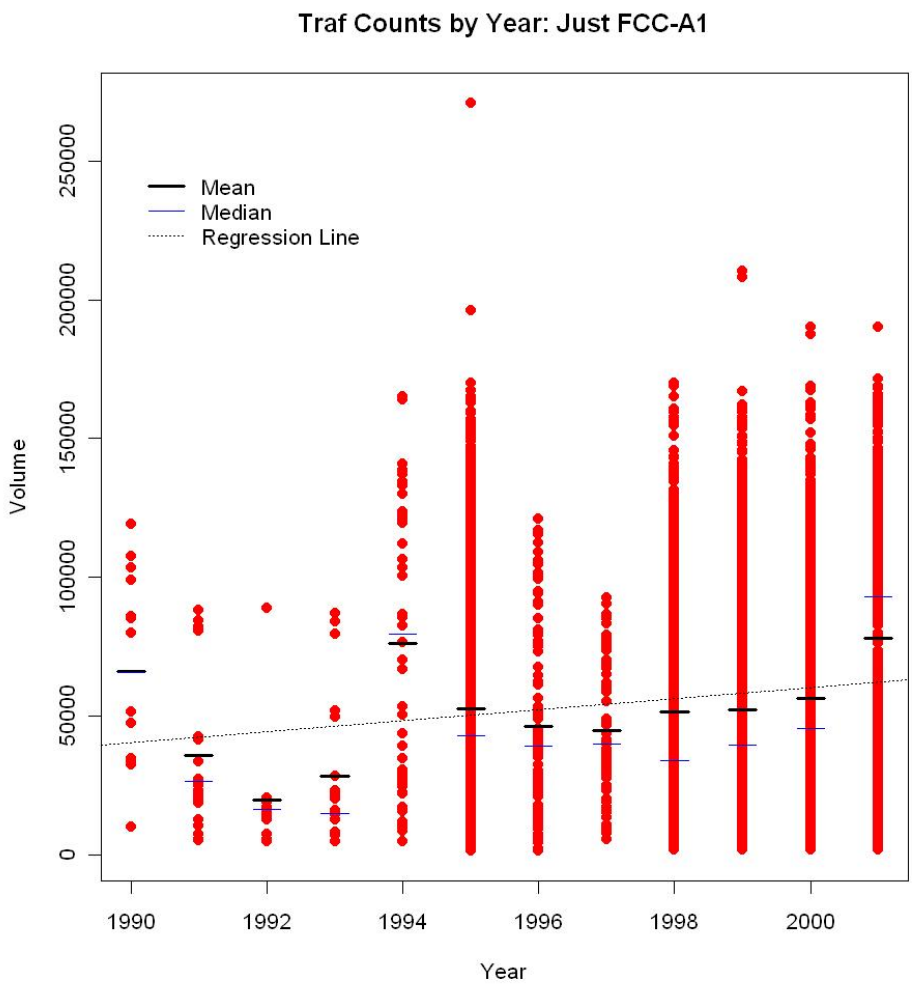


Appendix B - Figure 1: Histogram of number of available traffic counts year

REVIEW DRAFT

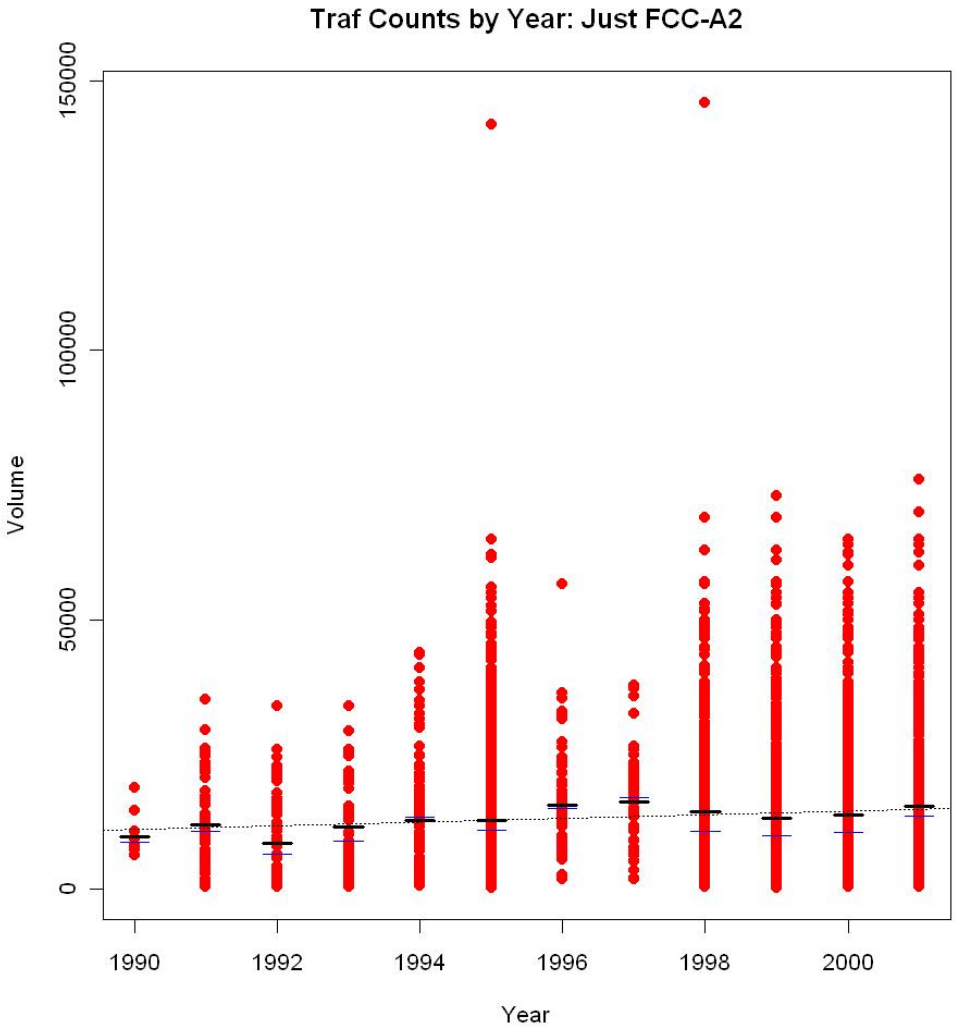


Appendix B - Figure 2: Scatterplot of traffic volumes by year with overlaid statistical measures for all road types

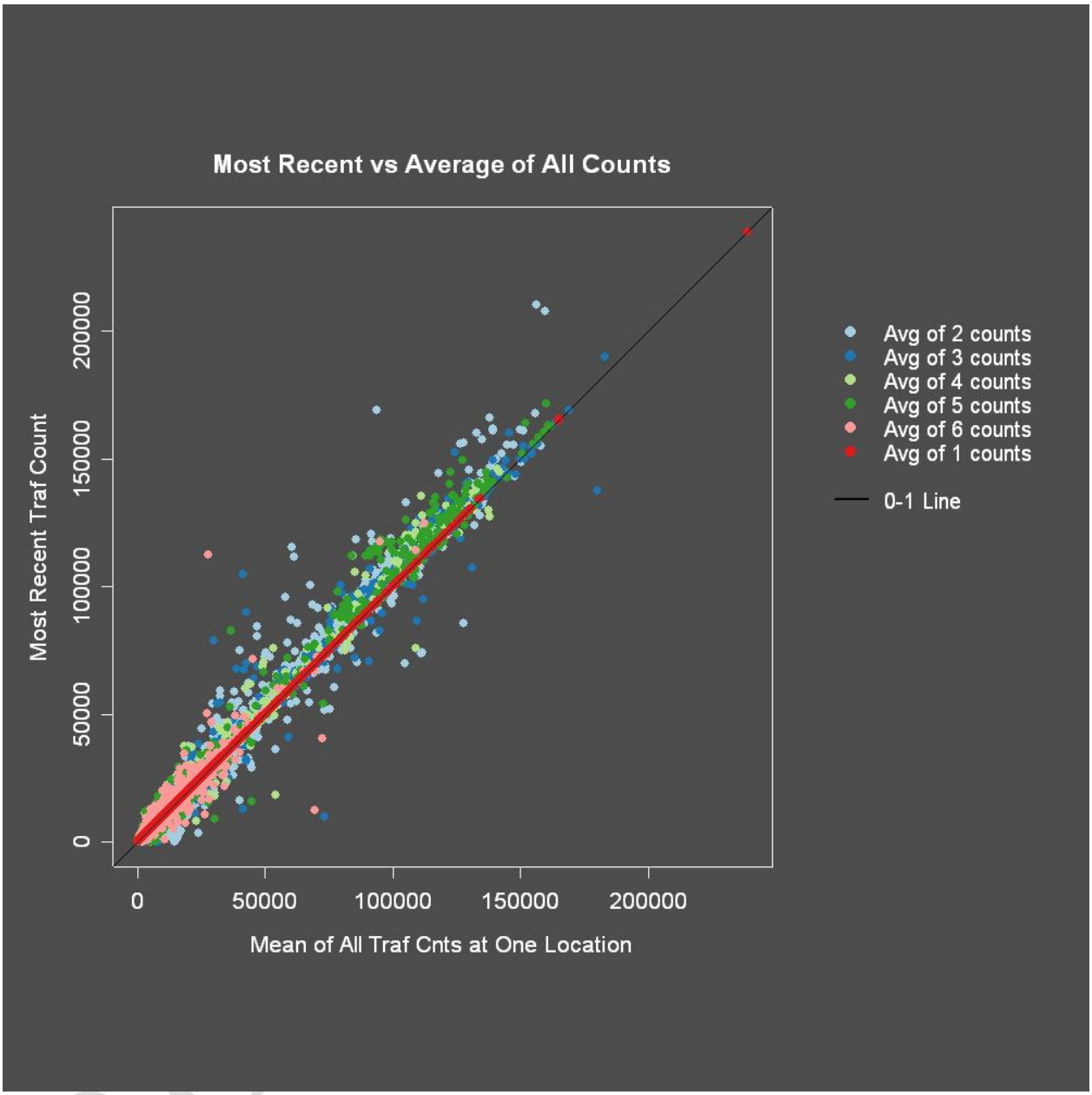


Appendix B - Figure 3: Scatterplot of traffic volumes by year with overlaid statistical measures for primary interstate highways

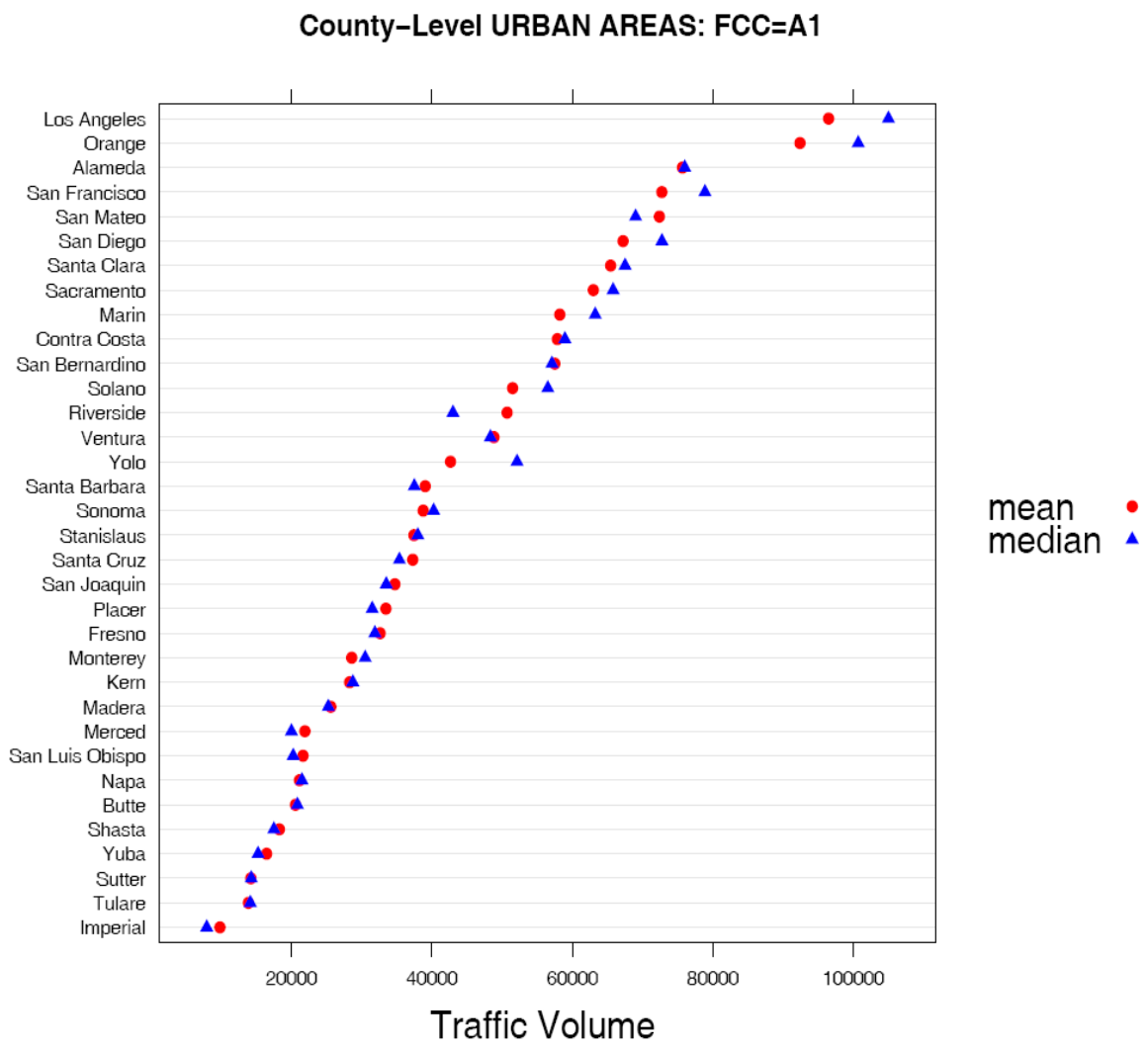
REVIEW DRAFT



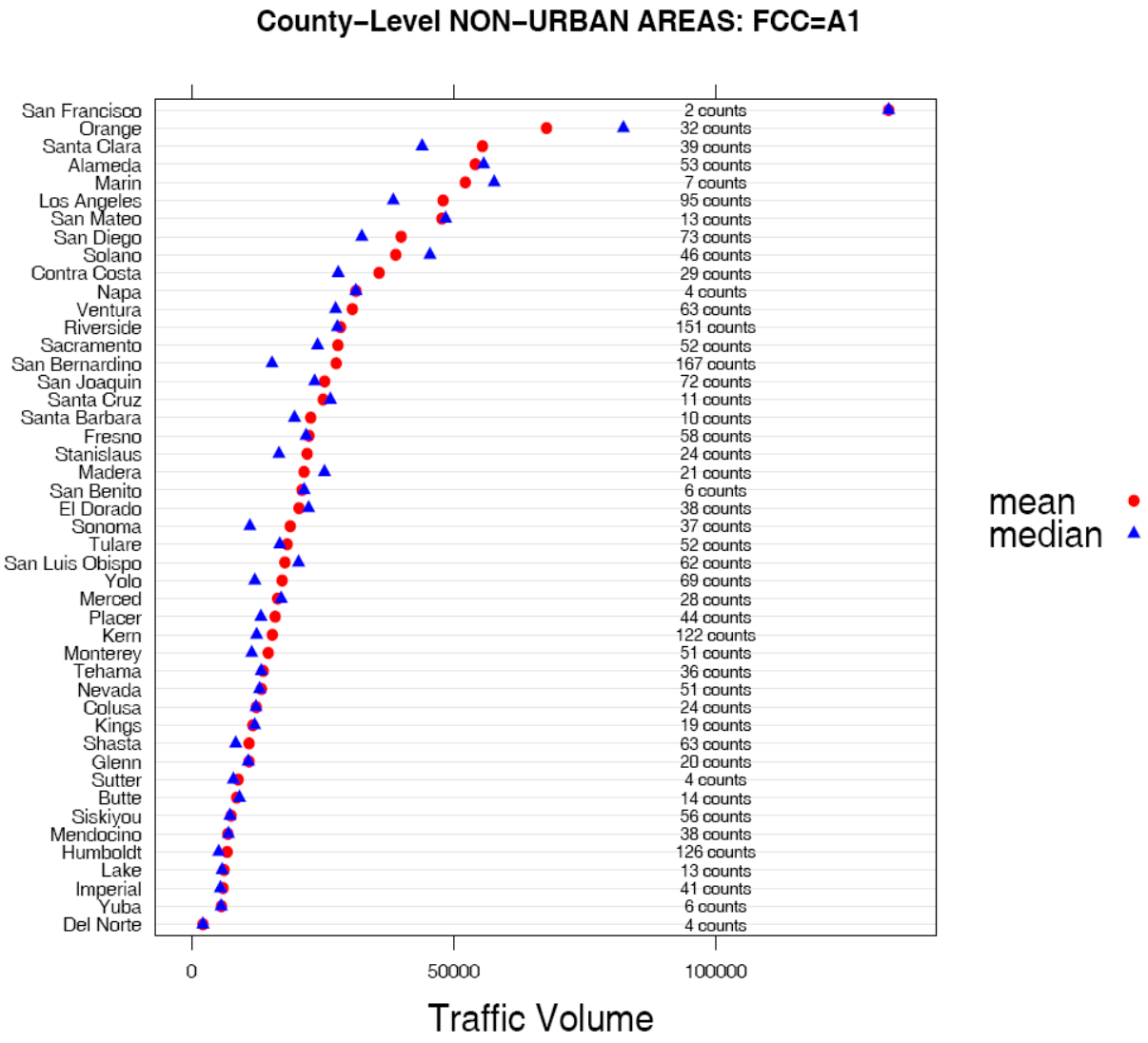
Appendix B - Figure 4: Scatterplot of traffic volumes by year with overlaid statistical measures for Primary US and states highways



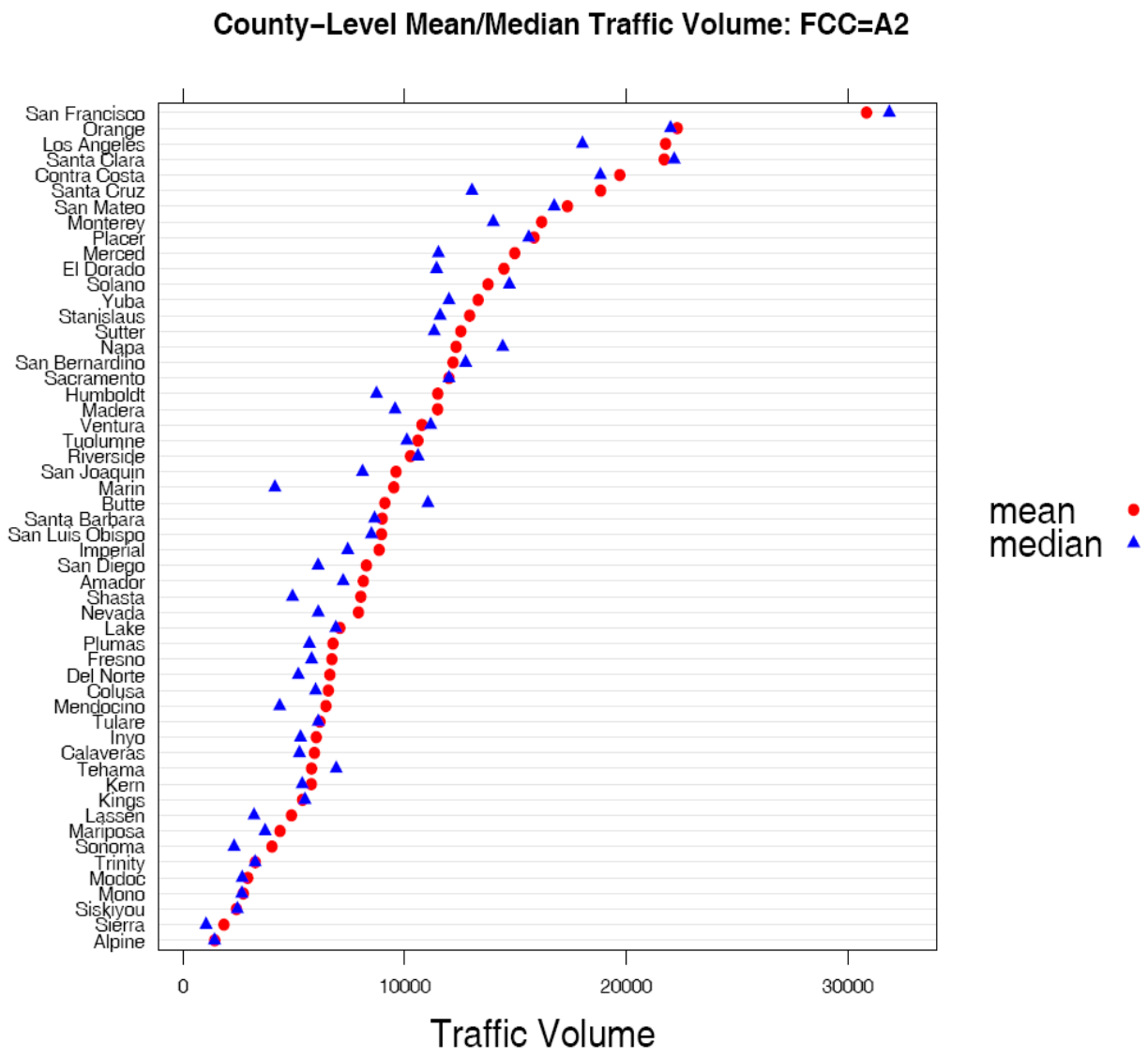
Appendix B - Figure 5: Within location averages versus the most recent counts



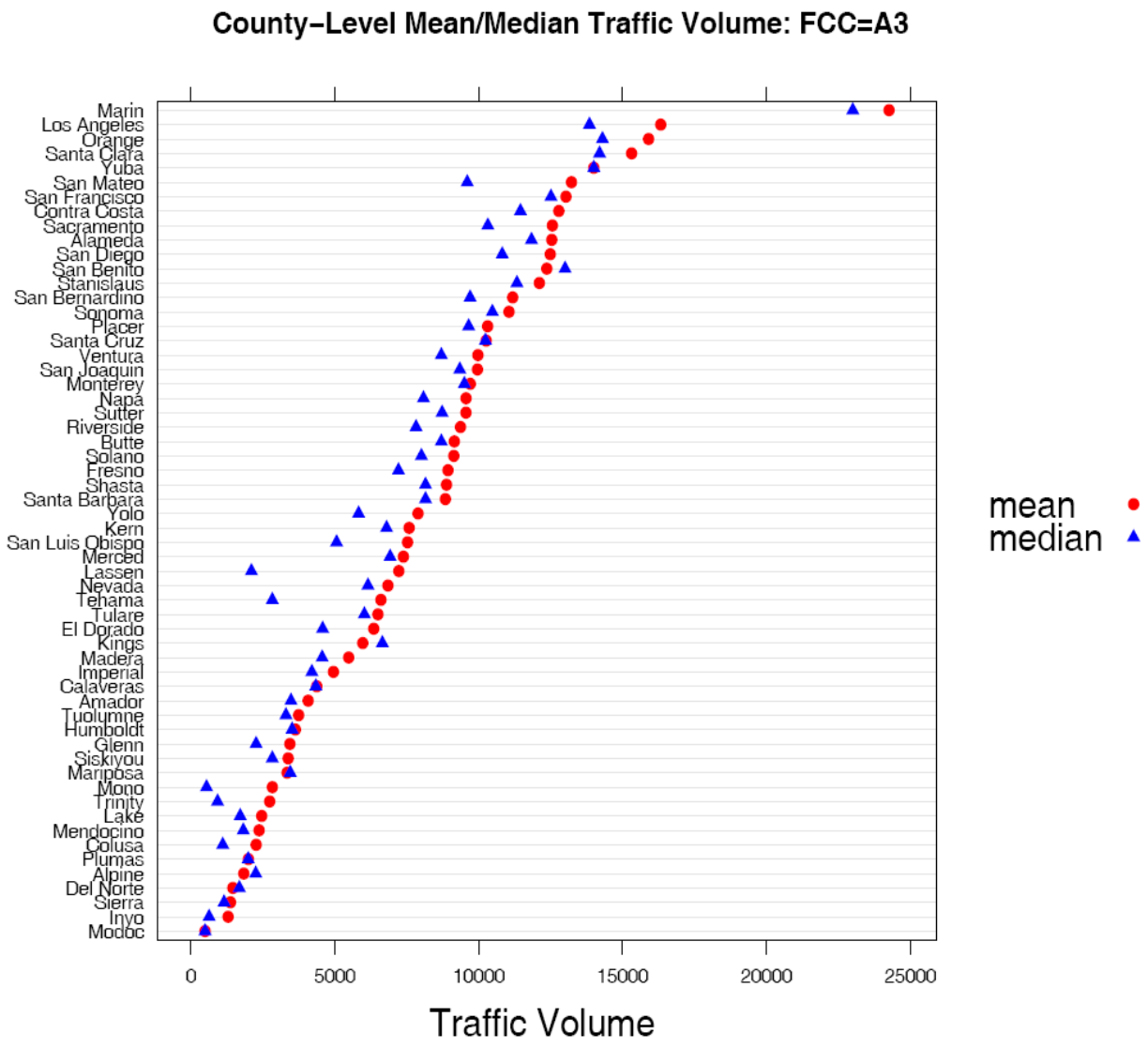
Appendix B - Figure 6: County level urban mean/median traffic volumes for primary interstates



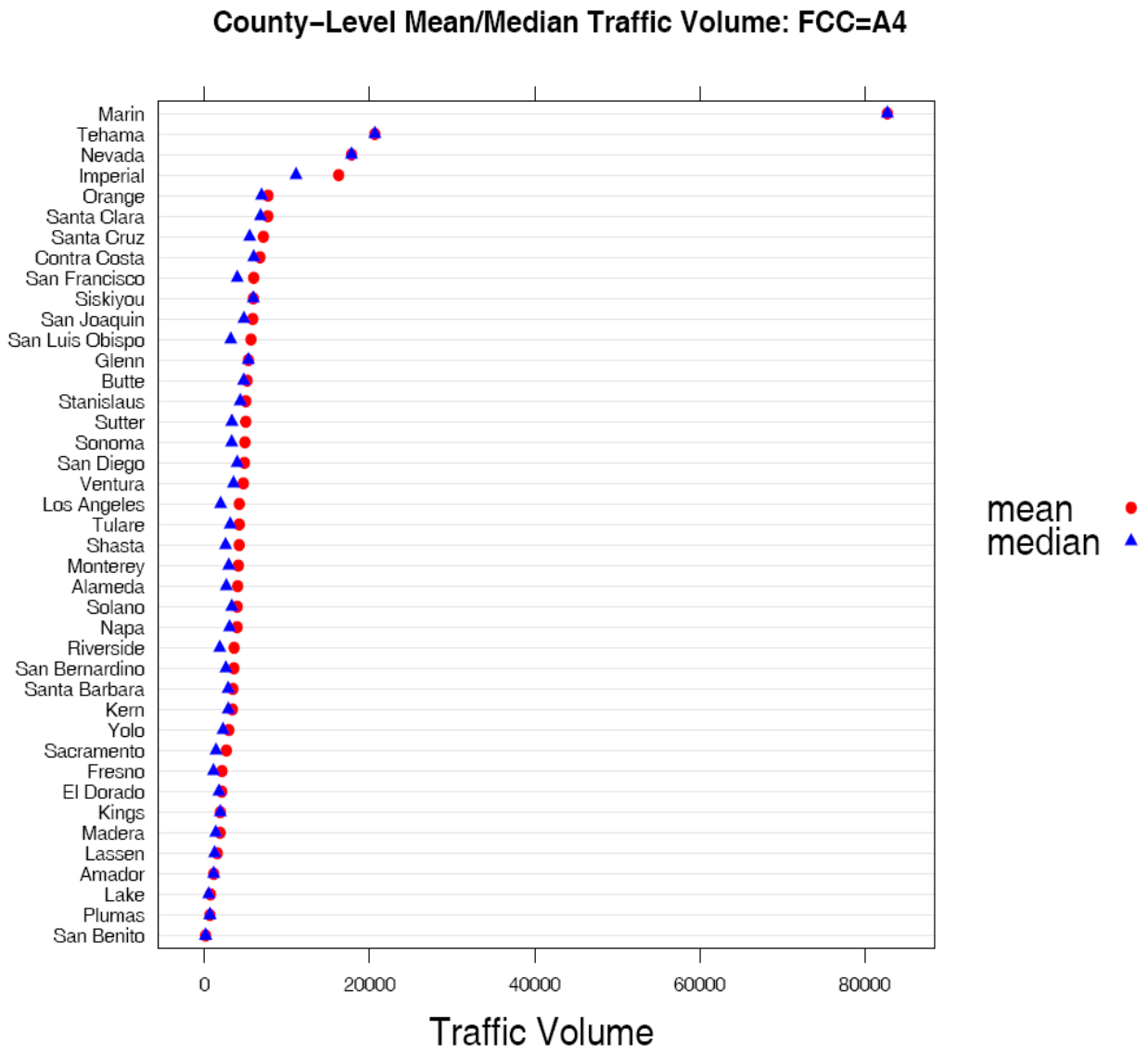
Appendix B - Figure 7County level non-urban mean/median traffic volumes for primary interstates



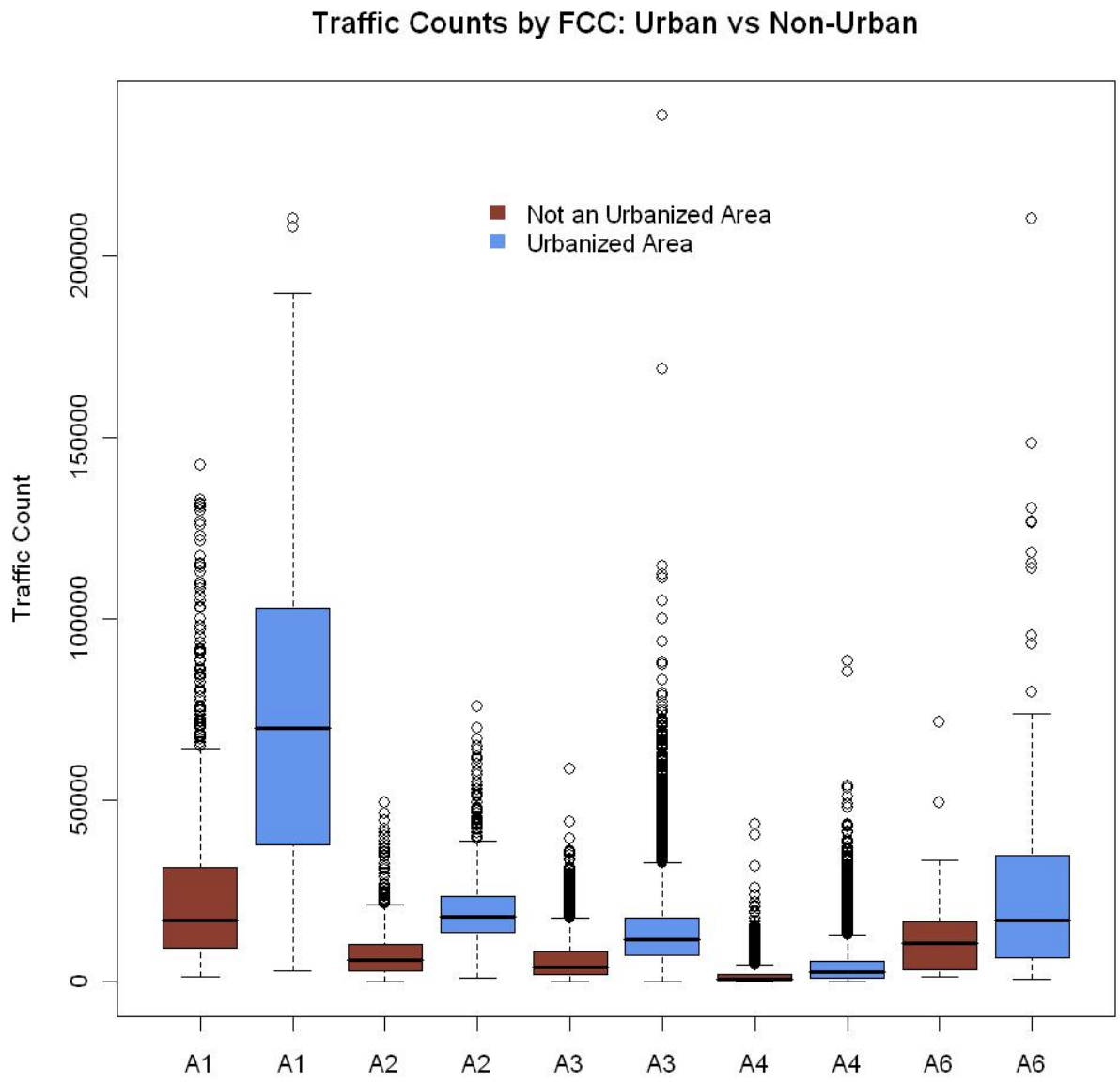
Appendix B - Figure 8: County level mean/median traffic volumes for primary US and states highways



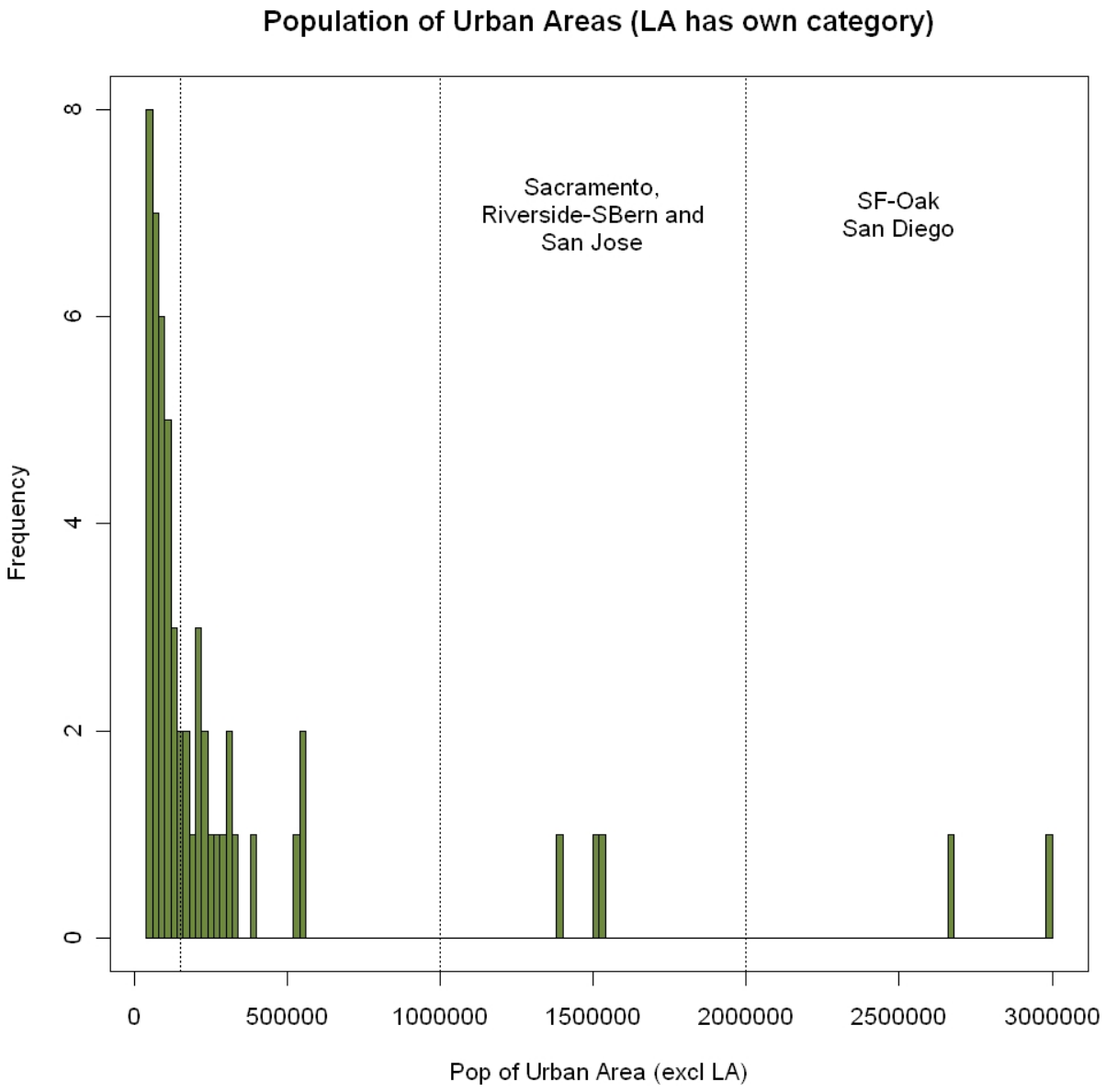
Appendix B - Figure 9: County level mean/median traffic volumes for secondary state and county highways



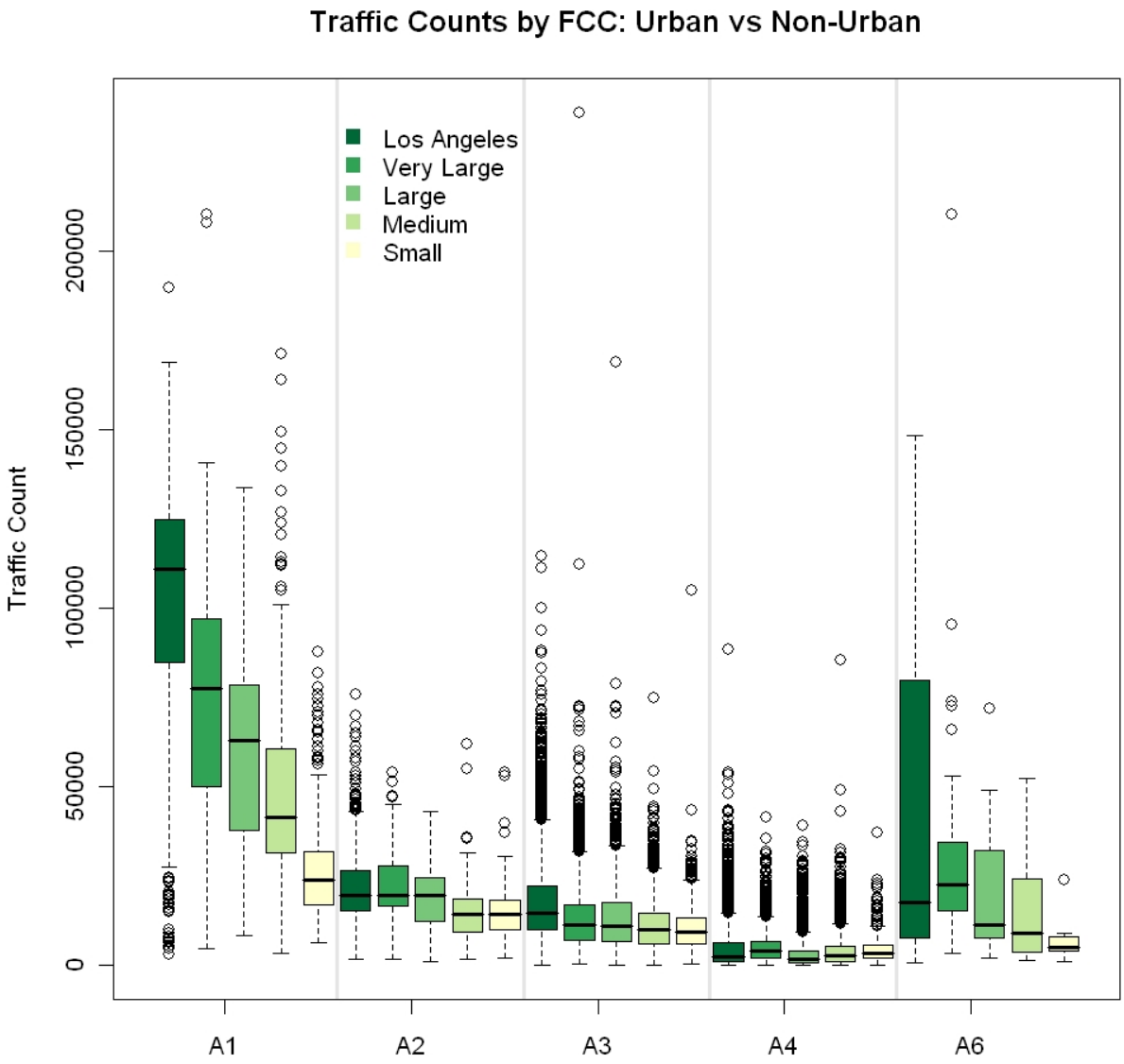
Appendix B - Figure 10: County level mean/median traffic volumes for local, neighborhood, rural road, and citystreets



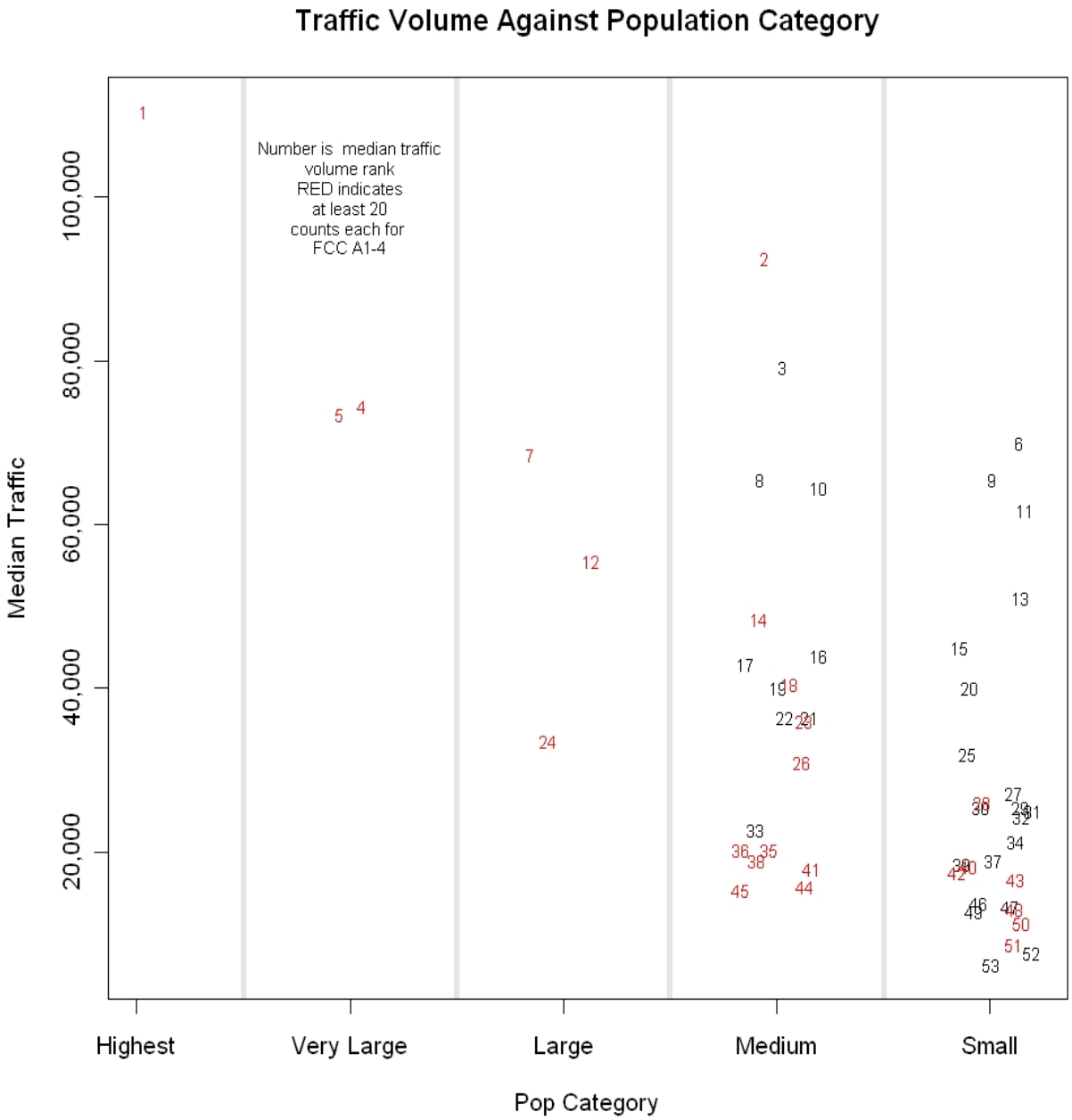
Appendix B - Figure 11: Boxplots of traffic count by road type comparing urban to non-urbanized areas



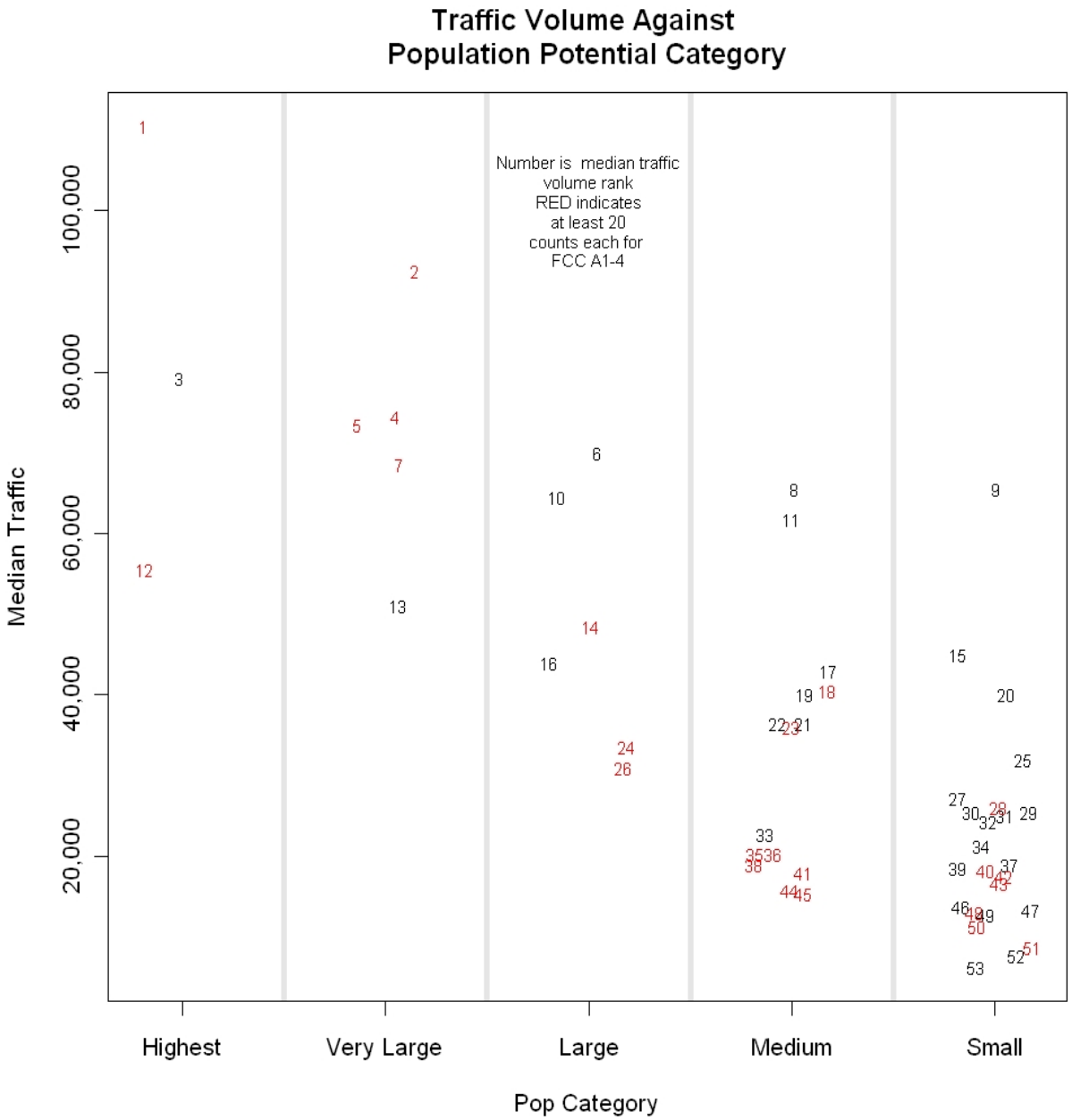
Appendix B - Figure 12: Categorization of urbanized areas based on population with small, medium, large, very large - excluding Los Angeles



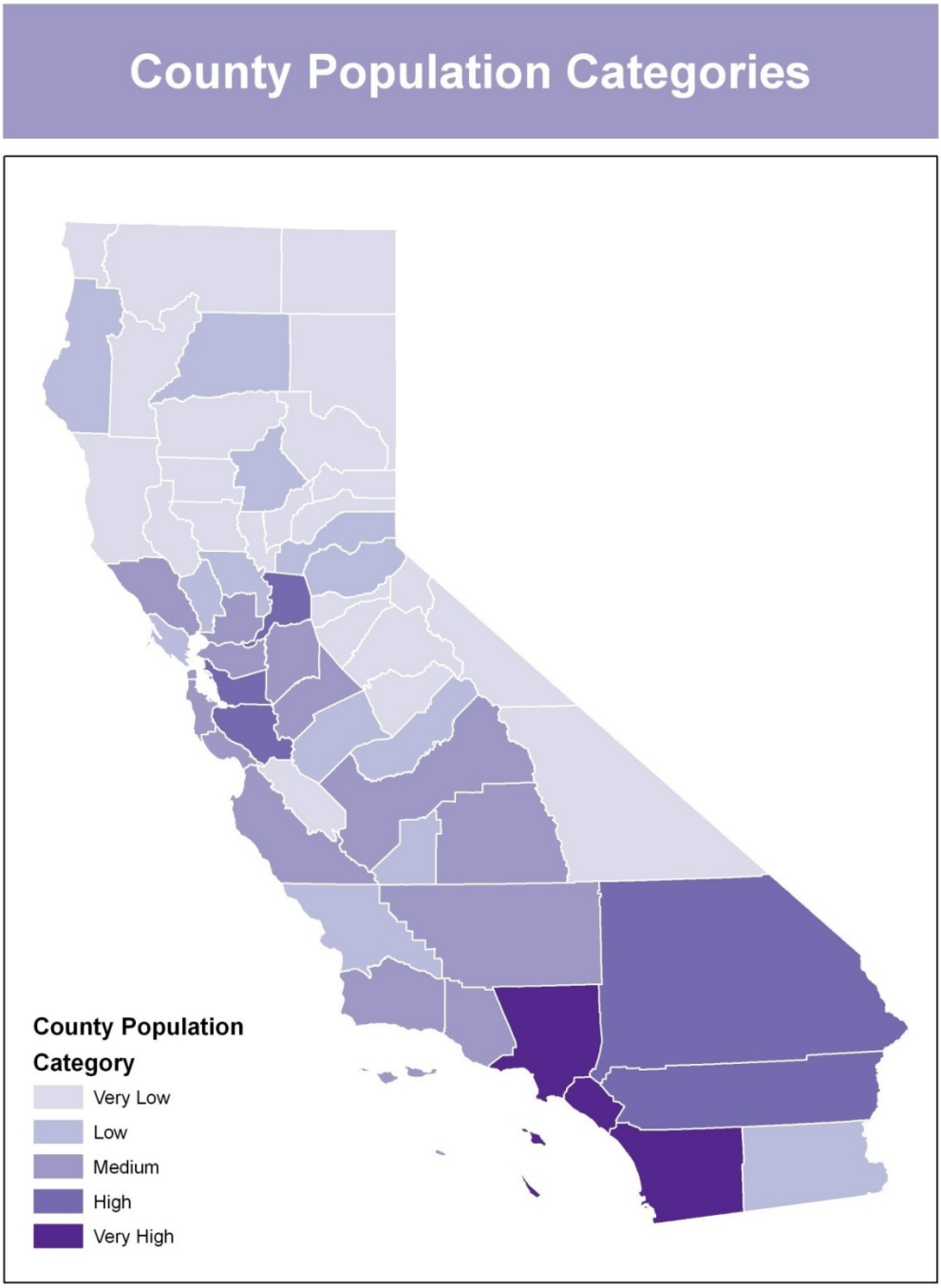
Appendix B - Figure 13: Boxplots of traffic counts by categorized urban area - Los Angeles is a singular category



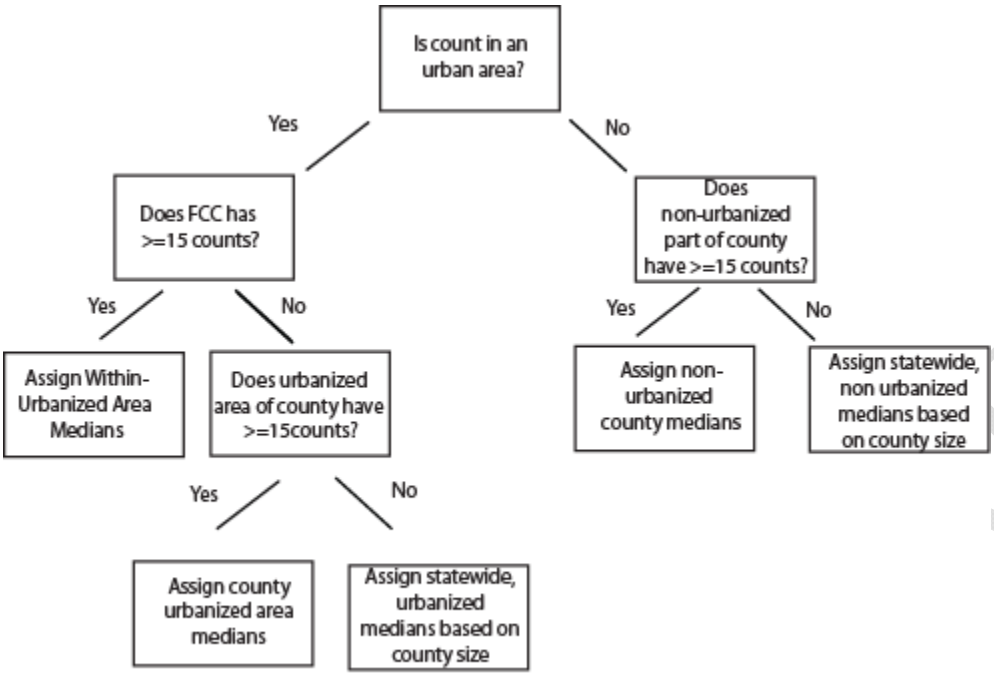
Appendix B - Figure 14: Traffic volume against population category



Appendix B - Figure 15: Traffic volume against population potential category



Appendix B - Figure 16: The plot shows how the counties were broken down in terms of population size.



Appendix B - Figure 17: Traffic Assignment Decision Making Tree

REVIEW DRAFT

Appendix C: Mathematical Foundations of Bayesian Maximum Entropy Estimation

BME is a non-linear, in general, estimator and integrates (*i*) composite space-time metrics (space/time analysis rather than purely spatial or purely temporal – time series analysis), (*ii*) data fluctuation, (*iii*) data uncertainty (i.e., inaccurate modeled data, extrapolation, stochastic empirical laws, missing records, desegregation or downscaling, measurement errors, and accuracy difference between remote sensing data and ground measurements), and (*iv*) secondary information correlated with a primary variable in a mathematically rigorous and unified manner. It derives spatial regression techniques involving ordinary, simple, and intrinsic kriging as its limiting cases under restrictive assumptions on spatial or temporal correlation structures, types of data, and mean trend models considered. In addition it copes with rationally non-stationary mapping situations and incorporates higher order statistical moments and the non-gaussian law. The BME approach has proven effective in numerous application contexts such as environmental exposure [76-78], adverse health effect[79, 80], risk assessment [81, 82], and urban geography [83, 84].

The random fluctuations found from the space/time distribution of air pollution are mainly due to the uncertainty in measurements, LUR estimates, and their space/time variability. The Space/Time Random Field (S/TRF), $X(\boldsymbol{p})$ is defined as an ensemble of possible realizations, χ_{map} at mapping points over space and time $\boldsymbol{p}_{\text{map}}=[\boldsymbol{p}_{\text{data}},\boldsymbol{p}_{\text{k}}]$ consisting of data points $\boldsymbol{p}_{\text{data}}$ and estimation points $\boldsymbol{p}_{\text{k}}$. The S/TRF efficiently addresses the randomness in terms of multivariate probability density function (pdf),$f_X(\chi_{\text{map}})$:

$$f_X(\chi_{\text{map}})\mathrm{d}\chi_{\text{map}} = Prob[\chi_1<X(\boldsymbol{p}_1)<\chi_1+\mathrm{d}\chi_1,\ldots,\chi_{\text{m}}<X(\boldsymbol{p}_{\text{m}})<\chi_{\text{m}}+\mathrm{d}\chi_{\text{m}}], \quad \text{(C1)}$$

where *Prob*[.] denotes the probability operator. Any equations derived through the BME framework are then based on the S/TRF representation.

Let us define the S/TRF of air pollution as $X(\boldsymbol{p})$, that is often regarded as a spatially non-homogeneous and temporally non-stationary field. We assume that the non-homogeneous/non-stationary S/TRF $X(\boldsymbol{p})$ can be modeled as the sum of a space/time mean trend $m_Z(\boldsymbol{p})$and a homogeneous/stationary residual $X'(\boldsymbol{p})$ as follow:

$$X(\boldsymbol{p})= m_Z(\boldsymbol{p}) + X'(\boldsymbol{p}), \quad \text{(C2)}$$

where the residual $X'(\boldsymbol{p})$ is a homogeneous/stationary random field in the sense that the expected value of $X'(\boldsymbol{p})$ is constant (fluctuated around zero), and the covariance of $X'(\boldsymbol{p})$ can generally be expressed solely as a function of spatial and temporal lags. The mean trend used for this study becomes the LUR results at data points, and leads to the residual with which to implement BME estimation.

(i) At the first stage, BME generates the prior pdf$f_G$providing an initial probability distributionacross space and time based on mean vector $\boldsymbol{m}_{X\text{map}}$ and covariance matrix $\boldsymbol{c}_{X\text{map}}$ of air pollution. In particular the covariance describes space/time correlation and air pollution

REVIEW DRAFT

dependencies between pairs of points. BME maximizes the entropy function [85, p.105] given $\boldsymbol{m}_{X\text{map}}$ and $\boldsymbol{c}_{X\text{map}}$, which leads to the Gaussian pdf (prior pdf) provided by:

$$f_G(\boldsymbol{\chi}_{\text{map}}) = N(\boldsymbol{\chi}_{\text{map}}; \boldsymbol{m}_{X\text{map}}, \boldsymbol{c}_{X\text{map}}), \tag{C3}$$

To obtain the $\boldsymbol{c}_{X\text{map}}$, first we were based on $X'(\boldsymbol{p})$ in Eq. (C2) to calculate experimental covariance (i.e., circles in Figs 11-13) for a given spatial lag r and temporal lag τ using the following estimator:

$$\hat{c}_X(r,\tau) \approx \frac{1}{N(r,\tau)} \sum_{i=1}^{N(r,\tau)} X_{head,i} X_{tail,i} - m_X^{\,2} \tag{C4}$$

where $N(r,\tau)$ is the number of pairs of points with values (X_{head},X_{tail}) separated by a distance of r and a time of τ. In practice we use a tolerance dr and dτ, i.e., such that

$$r - \mathrm{d}r \le \|\boldsymbol{s}_{\text{head}} - \boldsymbol{s}_{\text{tail}}\| \le r + \mathrm{d}r \text{ and } \quad \tau - \mathrm{d}\tau \le \|t_{\text{head}} - \boldsymbol{s}_{\text{tail}}\| \le \tau + \mathrm{d}\tau. \tag{C5}$$

Second, we find a permissible covariance model (i.e., $\boldsymbol{c}_{X\text{map}}$ in Eq C3) that fits the experimental covariance.

(ii) At the second stage, the site-specific knowledge available is organized into deterministic hard (Eq. C6) and probabilistic soft data (Eq. C7) and expressed in terms of suitable operators.

$$Prob[\, X(\boldsymbol{p}_{\text{hard}}) = \chi_{\text{hard}}] = 1, \tag{C6}$$

$$Prob[X(\boldsymbol{p}_{\text{soft}}) < \boldsymbol{u}] = \int_{-\infty}^{\boldsymbol{u}} d\boldsymbol{\chi}_{\text{soft}} f_S(\boldsymbol{\chi}_{\text{soft}}), \tag{C7}$$

Where χ_{hard}, $f_S(.)$, and $X(\boldsymbol{p}_{\text{soft}})$ are respectively an error-free measurement at point $\boldsymbol{p}_{\text{hard}}$, a conditionalpdf, and air pollution field at soft data locations $\boldsymbol{p}_{\text{soft}}$.

(iii) At the final stage, the initial solution $f_G$of stage *(i)* is enriched by assimilating the site-specific knowledge of stage *(ii)* through Bayesian statistics. This final solution provides the posterior pdf$f_K(\chi_k)$ (Eq. C6)for the contaminant level at each estimation point $\boldsymbol{p}_k$(i.e., 180 individual months in 1988-2002 for NO_2 and PM_{10}, each of 60 months in 1998-2002 for $PM_{2.5}$, and 10 km-gridded locations over the state of California for all of the pollutants).

$$f_K(\chi_k) = f_G[\chi_k | S] = A^{-1} \int \mathrm{d}\boldsymbol{\chi}_{\text{soft}} f_S(\boldsymbol{\chi}_{\text{soft}}) f_G(\boldsymbol{\chi}_{\text{map}}), \tag{C8}$$

REVIEW DRAFT

where A is a normalization coefficient. Since we do not consider soft data in this study, Eq. C8 is reduced to:

$$f_{\mathcal{K}}(\chi_k) = Z^{-1} f_{\mathcal{G}}(\boldsymbol{\chi}_{hard}, \chi_k), \tag{C9}$$

where Z is a normalization coefficient.

REVIEW DRAFT

REFERENCES

1. Pope, C.A., et al., *Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults.* Am J Resp Critical Care, 1995. **151**: p. 669-674.
2. Krewski, D., et al., *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, Part I: Replication and Validation*, in *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: A Special Report of the Institute's Particle Epidemiology Reanalysis Project.* 2000, Health Effects Institute: Cambridge, MA. p. 39-128.
3. Krewski, D., et al., *Extended Analysis of the American Cancer Society Study of Particulate Air Pollution and Mortality*. 2009, Health Effects Institute.
4. Industrial Economics, Incorporated, . *Expanded expert judgment assessment of the concentration-response relationship between PM2.5 exposure and mortality. Final Report*. 2006 [cited 2008 January 3]; Available from: http://www.epa.gov/ttn/ecas/regdata/Uncertainty/pm_ee_report.pdf.
5. Jerrett, M., et al., *Spatial analysis of air pollution and mortality in Los Angeles.* Epidemiol, 2005. **16**: p. 727-736.
6. Pope, C.A., et al., *Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution.* JAMA, 2002. **287**: p. 1132-1141.
7. Ostro, B., et al., *Revisiting Long-Term Exposure to PM2.5 Constituents in the California Teachers Study.* Environ Health Perspect, In Press.
8. Ostro, B., et al., *Long-Term Exposure to Constituents of Fine Particulate Air Pollution and Mortality: Results from the California Teachers Study.* Environ Health Perspect, 2009. **118**(3).
9. Brunekreef, B. and G. Hoek, *A Critique of "Fine Particulate Air Pollution and Total Mortality Among Elderly Californians, 1973-2002" by James E. Enstrom, PhD.* Inhalation Toxicology, 2006. **18**(7): p. 507-508.
10. Enstrom, J.E., *Fine Particulate Air Pollution and Total Mortality Among Elderly Californians, 1973-2002.* Inhalation Toxicology, 2005. **17**(14): p. 803-816.
11. Dockery, D.W., et al., *An association between air pollution and mortality in six US cities.* New Engl J Med, 1993. **329**: p. 1753-1759.
12. Gamble, J.F., *PM2.5 and mortality in long-term prospective cohort studies: Cause-effect or statistical associations?* Environ Health Perspect, 1998. **106**: p. 535-549.
13. Lipfert, F.W. and R.E. Wyzga, *Air pollution and mortality: issues and uncertainties.* J Air Waste Manage, 1995. **45**: p. 949-966.
14. Krewski, D., et al., *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, Part II: Sensitivity Analysis : A Special Report of the Institute's Particle Epidemiology Reanalysis Project. Health Effects Institute, Cambridge, MA.* 2000: Cambridge, MA. p. 129-240.
15. Krewski, D., et al., *Overview of the reanalysis of the Harvard Six Cities study and American Cancer Society study of particulate air pollution and mortality. J Toxicol Environ Health A 66:1507-1552.* J Toxicol Environ Health A, 2003. **66**: p. 1507-1552.
16. Curtis, S. and A. Taket, *Health and Societies: Changing Perspectives. Edward Arnold, London.* 1996, London: Edward Arnold L4 - Curtis S, Taket A. 1996. Health and Societies: Changing Perspectives.

REVIEW DRAFT

17. Duncan, C. and K. Jones, *Health-related behaviour in context: A multi-level modelling approach.* Soc SciMed, 1996. **42**: p. 817-830.
18. Macintyre, S. and A. Ellaway, *Ecological approaches: rediscovering the role of the physical and social environments.*, in *Social Epidemiology*, L. Berkman and I. Kawachi, Editors. 2000, Oxford University Press: Oxford.
19. Eyles, J., *Health, environmental assessment and population health: Tools for a complex process.* Can J Public Health, 1999. **90**: p. S31-S34.
20. Macintyre, S. and A. Ellaway, *Social and local variations in the use of urban neighbourhoods: a case study in Glasgow.* Health Place, 1998. **4**: p. 91-94.
21. Macintyre, S., S. Maciver, and A. Sooman, *Area, class and health: should we be focusing on places or people?* . J Soc Policy, 1993. **22**: p. 213-234.
22. Jerrett, M., et al., *A GIS-environmental justice analysis of particulate air pollution in Hamilton, Canada.* EnvironPlann A, 2001. **33**: p. 955-973.
23. Jerrett, M., et al., *Environmental equity in Canada: An empirical investigation into the income distribution of pollution in Ontario.* Environment and Planning, 1997. **A29**: p. 1777-1800.
24. Fotheringham, A.S., C. Brunsdon, and M. Charlton, *Quantitative geography: perspectives on spatial data analysis. ISBN 0761959475. Thousand Oaks: Sage Publications, London.* 2000.
25. Amrhein, C. and H. Reynolds, *Using the Getis statistic to explore aggregation effects in metropolitan Toronto census data. The Canadian Geographer*, 1997. **41**: p. 137-149.
26. Steel, G. and D. Holt, *Rules for random aggregation.* Environment and Planning A, 1996. **28**: p. 957-978.
27. Bailey, T.C. and A.C. Gatrell, *Interactive Spatial Data Analysis*. 1995, New York, NY: John Wiley and Sons.
28. Cutter, S., D. Holm, and L. Clark, *The role of geographic scale in monitoring environmental justice.* Risk Anal, 1996. **16**: p. 1-17.
29. Greenberg, M., *Proving environmental inequity in siting locally unwanted land uses.* Risk Issues in Health and Safety, 1993. **4**: p. 235-252.
30. McMaster, R., H. Leitner, and E. Sheppard, *GIS-based environmental equity and risk assessment: methodological problems and prospects.* Cartography and Geographic Info Systems, 1997. **24**: p. 172-189.
31. Brauer, M., et al., *Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems.* Epidemiol, 2003. **14**: p. 228-239.
32. Briggs, D., et al., *A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments.* Sci Total Environ, 2000. **253**: p. 151-167.
33. Brunekreef, B. and S. Holgate, *Air pollution and health.* Lancet, 2002. **360**: p. 1233-1242.
34. Zhu, Y., et al., *Study on ultrafine particles and other vehicular pollutants near a major highway with heavy-duty diesel traffic.* Atmos Environ, 2002. **36**: p. 4323-4335.
35. Jerrett, M., et al., *Spatial analysis of the air pollution-mortality relationship in the context of ecologic confounders.* J Toxicol Env Health, 2003. **66**: p. 1735-1777.
36. O'Neil, M., et al., *Health, wealth and air pollution.* Environ Health Persp, 2003. **111**: p. 1861-1870.

REVIEW DRAFT

37. Künzli, N., et al., *Ambient air pollution and atherosclerosis in Los Angeles.* Environmental Health Perspectives, 2005. **113**(2): p. 201-206.
38. Jerrett, M., et al., *Spatial analysis of air pollution and mortality in Los Angeles.* 2005. **16**(6): p. 727-736.
39. Nafstad, P., et al., *Lung cancer and air pollution: a 27 year follow up of 16 209 Norwegian men.* Thorax, 2003. **58**(12): p. 1071-1076.
40. Hoek, G., et al., *Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study.* Lancet, 2002. **360**(9341): p. 1203-1209.
41. Brauer, M., et al., *Estimating long-term average particulate air pollution concentrations: Application of traffic indicators and geographic information systems.* Epidemiology, 2003. **14**(2): p. 228-239.
42. Briggs, D., et al., *Changing geographic relationships between air pollution and cardio-respiratory mortality in Great Britain, 1962-1996.* Epidemiology, 2000. **11**(4): p. S103-S103.
43. Ross, Z., et al., *A land use regression model for predicting fine particulate matter concentrations in the New York City region.* Atmos Environ, 2006. **41**: p. 2255-2269.
44. Jerrett, M., et al., *A review and evaluation of intraurban air pollution exposure models.* Journal of Exposure Analysis and Environmental Epidemiology, 2005b. **15**(2): p. 185-204.
45. Bari, A., et al., *Measurement of gaseous HONO, HNO3, SO2, HCL, NH3, particulate sulfate and PM2.5 in New York, NY.* Atmospheric Environment, 2003. **37**: p. 2825-2835.
46. Higgens, M., et al., *Commentary: Health Review Committee. In Krewski D, Burnett R, Goldberg M, Hoover K, Siemiantycki J, Jerrett M, Abrahamowicz M, White W. 2000. Reanalysis of the Harvard Six-Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, Phase II: Sensitivity Analysis.* 2000, Health Effects Institute: Cambridge, MA.
47. Ma, R., D. Krewski, and R.T. Burnett, *Random effects Cox models: A Poisson modelling approach.* Biometrika, 2003. **90**: p. 157-169.
48. Pope, C.A., et al., *Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease.* Circulation, 2004. **109**: p. 71-77.
49. Krewski, D., et al., *Mortality and long-term exposure to ambient air pollution: ongoing analyses based on the American Cancer Society Cohort.* J Toxicol Environ Health A, 2005. **68**: p. 1093-1109.
50. Goddard, M.J., D.J. Murdoch, and D. Krewski, *Temporal aspects of risk characterization.* Inhal Toxicol, 1995. **7**: p. 1005-1018.
51. Zeger, S., F. Dominici, and J. Samet, *Harvesting-resistant estimates of air pollution effects on mortality.* Epidemiology, 1999. **10**: p. 171-175.
52. Schwartz, J., *Harvesting and long term exposure effects in the relation between air pollution and mortality.* Am J Epidemiol, 2000. **151**: p. 440-448.
53. Brunekreef, B., *Air pollution and life expectancy: is there a relation?* Occup Environ Med, 1997. **54**: p. 781-784.
54. Goldberg, M.S., et al., *Identifying subgroups of the general population that may be susceptible to short-term increases in particulate air pollution: A time series study in Montreal, Quebec. Health Effects Institute, Number 97, Cambridge, MA.* 2000, Hansen KA: Cambridge, MA.

REVIEW DRAFT

55. Goldberg, M.S., et al., *The association between daily mortality and short-term effects of ambient air particulate pollution in Montreal, Quebec: 1. Nonaccidental mortality.* Environ Res, 2001. **86**: p. 12-25.
56. Goldberg, M.S., et al., *The association between daily mortality and short-term effects of ambient air particle pollution in Montreal, Quebec: 2. Cause-specific mortality.* Environ Res, 2001. **86**: p. 26-36.
57. Villeneuve, P.J., et al., *Fine particulate air pollution and all-cause mortality with the Harvard Six-cities Study: Variations in risk by period of exposure.* Ann Epidemiol, 2002. **12**: p. 568-576.
58. Laden, F., et al., *Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities study.* Am J Resp Crit Care, 2006. **173**: p. 667-672.
59. Eftim, S.E., et al., *Fine Particulate Matter and Mortality: A Comparison of the Six Cities and American Cancer Society Cohorts With a Medicare Cohort.* Epidemiology, 2008. **19**(2): p. 209-216.
60. Beelen, R., et al., *Long-Term Effects of Traffic-Related Air Pollution on Mortality in a Dutch Cohort (NLCS-AIR Study).* Environ Health Perspect, 2007. **116**(2).
61. Sinisi, S. and M. van der Laan, *Deletion/Substitution/Addition Algorithm in Learning with Applications in Genomics.* Statistical Applications in Genetics and Molecular Biology, 2004. **3**(1): p. Article 18.
62. van Donkelaar, A., et al., *Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application.* Environ Health Perspect, 2010. **118**(6).
63. Calle, E.E. and D.D. Terrell, *Utility of the National Death Index for ascertainment of mortality among Cancer Prevention Study II participants.* Am J Epidemiol, 1993. **137**: p. 235-241.
64. Cox, D., *Regression models and life tables (with Discussion).* J Royal Stat Soc B, 1972. **55**: p. 187-220.
65. Cox, D. and D. Oakes, *Analysis of Survival Data*. 1984, New York: Chapman and Hall.
66. Crowder, M., *On consistency and inconsistency of estimating equations.* Econometrics Theory, 1986. **3**: p. 305-330.
67. Crowder, M., *On linear and quadratic estimating function.* Biometrika, 1987. **74**: p. 591-597.
68. Chen, H., et al., *Land use regression models for traffic pollution with temporally discordant monitoring data.* 2006(in preparation).
69. Chen, L. and Q. Shao, *Normal approximation under local dependence.* Ann Probability, 2004. **32**: p. 1985-2028.
70. Jørgensen, B., et al., *A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia.* Statist Med, 1996. **15**: p. 823-836.
71. Jørgensen, B. and S. Knudsen, *Parameter orthogonality and bias adjustment for estimating functions.* Scand J Statist, 2004. **31**: p. 93-114.
72. Jerrett, M., et al., *Long-Term Ozone Exposure and Mortality.* New England Journal of Medicine, 2009. **360**(11): p. 1085-1095.
73. Cosby, A.G., et al., *Preliminary Evidence for an Emerging Nonmetropolitan Mortality Penalty in the United States.* Am J Public Health, 2008. **98**(8): p. 1470-1472.
74. Jerrett, M., et al., *A Cohort Study of Traffic-Related Air Pollution and Mortality in Toronto, Ontario, Canada.* Environ Health Perspect, 2009. **117**(5).

REVIEW DRAFT

75. Moore, D.K., et al., *A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA.* J Environ Monit, 2007. **9**: p. 246-252.
76. Bogaert, P., et al., *Spatiotemporal modeling of ozone distribution in the state of California.* Atmospheric Research, 2009. **43**: p. 2471-2480.
77. Christakos, G. and M.L. Serre, *A spatiotemporal study of exposure-health effect associations.* Journal of Exposure Analysis and Environmental Epidemiology, 2000. **10**(2): p. 168-187.
78. Christakos, G., M.L. Serre, and J.L. Kovitz, *BME representation of particulate matter distributions in the state of California on the basis of uncertain measurements.* J. Geophys. Res., 2001. **106**(D9): p. 9717-9731.
79. Law, D.C.G., et al., *Spatial analysis and mapping of sexually transmitted diseases to optimise intervention and prevention strategies.* Sexually Transmitted Infections, 2004. **80**(4): p. 294-299.
80. Lee, S.-J., K.B. Yeatts, and M.L. Serre, *A Bayesian Maximum Entropy approach to address the change of support problem in the spatial analysis of childhood asthma prevalence across North Carolina.* Spatial and Spatio-temporal Epidemiology, 2009. **1**(1): p. 49-60.
81. Choi, K.-M., H.-L. Yu, and M. Wilson, *Spatiotemporal statistical analysis of influenza mortality risk in the State of California during the period 1997–2001.* Stochastic Environmental Research and Risk Assessment, 2008. **22**(0): p. 15-25.
82. Serre, M.L., et al., *An Application of the Holistochastic Human Exposure Methodology to Naturally Occurring Arsenic in Bangladesh Drinking Water.* Risk Analysis, 2003. **23**(3): p. 515-528.
83. Brazel, A., et al., *Determinants of changes in the regional urban heat island in metropolitan Phoenix (Arizona, USA) between 1990 and 2004.* Clim Res, 2007. **33**(2): p. 171-182.
84. Lee, S.-J., E. Wentz, and P. Gober, *Space–time forecasting using soft geostatistics: a case study in forecasting municipal water demand for Phoenix, Arizona.* Stochastic Environmental Research and Risk Assessment, 2010. **24**(2): p. 283-295.
85. Christakos, G., *Modern spatiotemporal geostatistics*. 2000, New York, NY: Oxford University Press.

REVIEW DRAFT