

# Spatial Analysis of Air Pollution and Mortality in California

**Author list:** Michael Jerrett, PhD<sup>1</sup>, Richard T. Burnett, PhD<sup>2</sup>, Bernardo S. Beckerman, MS<sup>1</sup>, Michelle C. Turner, PhD<sup>3</sup>, Daniel Krewski, PhD<sup>3,4</sup>, George Thurston, ScD<sup>5</sup>, Randall Martin, PhD<sup>6</sup>, Aaron von Donkelaar, PhD<sup>6</sup>, Edward Hughes, PhD<sup>7</sup>, Yuanli Shi MD<sup>3</sup>, Susan M. Gapstur, PhD<sup>8</sup>, Michael J. Thun, MD<sup>8</sup>, C. Arden Pope III, PhD<sup>9</sup>

## **Affiliations:**

1. Division of Environmental Health Sciences, School of Public Health, University of California Berkeley, Berkeley, CA, USA
2. Population Studies Division, Health Canada, Ottawa, ON, Canada
3. McLaughlin Centre for Population Health Risk Assessment, Institute of Population Health, University of Ottawa, Ottawa, ON, Canada
4. Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada
5. New York University School of Medicine, Tuxedo, NY, USA
6. Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, Canada
7. Edward Hughes Consulting, Ottawa, Ontario, Canada
8. Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA
9. Department of Economics, Brigham Young University, Provo, UT, USA

Corresponding author: Michael Jerrett, PhD. 50 University Hall, School of Public Health.

University of California Berkeley, Berkeley, California 94720-7360. jerrett@berkeley.edu

### **Authors' contributions:**

M. Jerrett conceived the study, led all analyses, contributed to the development of the exposure models, drafted much of the text, and responded to comments from co-author reviewers.

B. Beckerman ran many of the statistical models that led to the exposure assessments, conducted geographic analyses, contributed text, and assisted with interpreting the results.

R.T. Burnett supplied expert statistical advice on the analyses, drafted sections of the paper, and assisted with the interpretation of the results.

E. Hughes developed the statistical programs used to interpret the random effects models, helped to interpret the results, and supplied key statistical advice on the interpretation.

D. Krewski contributed to the original grant proposal, assisted with interpretation of the results, wrote sections of the paper, and assisted with the interpretation.

C.A. Pope contributed to the statistical analyses, wrote sections of the text, and assisted with interpreting the results.

S. Gapstur is the Principle Investigator of the ACS, CPS-II cohort and commented on the final draft of the paper. She also oversaw the geocoding process for exposure assignment.

M. Thun assisted with interpretation of the statistical models and supplied expert medical epidemiological advice on the results. At the time of submission of this manuscript, Dr. Thun retired from the American Cancer Society.

G. Thurston assisted with the conception of the study, supplied key information on interpreting the pollution models, and commented on several drafts of the paper, which changed the interpretation of the results.

M. Turner contributed text and tables, helped to assemble supporting data, assisted with the statistical modeling, interpret results, and served as a liaison with the American Cancer Society for code review and data access.

R. Martin and A. von Donkelaar contributed the remote sensing models used to derive estimates of PM<sub>2.5</sub>, supplied text, edited versions of the paper, and gave advice on atmospheric chemistry issues.

Y. Shi ran the statistical models, managed the data, prepared code for review by the American Cancer Society, prepared all of the tables and associated text, and assisted with the interpretation of the results.

**Sources of support:** This work was supported in part by a contract with the California Air Resources Board. Additional funding came from the Environmental Public Health Tracking Program of the Centers for Disease Control. D. Krewski is the Natural Sciences and Engineering Research Council of Canada Chair in Risk Science at the University of Ottawa. Dr. Thurston was also supported in part by the NYU-NIEHS Center of Excellence Grant ES00260.

**Short running head:** Air Pollution and Mortality in California

**Descriptor number:** 6.1

**Total word count:** 3469

**At a Glance Commentary:**

Several cohort studies have examined whether long-term exposure to air pollution is associated with premature death. The results of these studies have been mixed, possibly due to errors introduced in the exposure assessment process.

To address this potential problem, this study assigns members of the American Cancer Society Cancer Prevention Study II Cohort residing in California more precise exposure assignments at their home address using advanced exposure models.

The study makes several important contributions to the field, including: (1) the first evidence that ozone is significantly associated with cardiovascular mortality, particularly from ischemic heart disease; (2) a strong association between nitrogen dioxide (NO<sub>2</sub>) and lung cancer; and (3) a finding that fine particulate matter, with aerodynamic diameter of 2.5 microns or less (PM<sub>2.5</sub>), and nitrogen dioxide both associate independently with premature death from all causes and cardiovascular disease. In total the findings confirm earlier evidence on PM<sub>2.5</sub> associations with mortality and expand the evidence base markedly on associations between ozone or NO<sub>2</sub> and premature death.

## Abstract

**Rationale:** Although substantial scientific evidence suggests that chronic exposure to ambient air pollution contributes to premature mortality, uncertainties exist in the size and consistency of this association. Uncertainty may arise from inaccurate exposure assessment.

**Objective:** To assess the associations of three criteria air pollutants (fine particulate matter, ozone and nitrogen dioxide) with risk of mortality in a large cohort of California adults using individualized exposure assessments.

**Methods:** For fine particulate matter and nitrogen dioxide, we used land use regression models to derive predicted individualized exposure at the home address. For ozone, we estimated exposure with an inverse distance weighting interpolation. Standard and multilevel Cox survival models were used to assess the association between air pollution and mortality.

**Measurements:** Data for 73,711 subjects that resided in California were abstracted from the American Cancer Society Cohort Cancer Prevention II Cohort, with baseline ascertainment of individual characteristics in 1982 and follow up of vital status through to 2000. Exposure data were derived from government monitors.

**Results:** Exposure to fine particulate matter, ozone and nitrogen dioxide were all positively associated with ischemic heart disease mortality. Nitrogen dioxide (a marker for traffic pollution) and fine particulate matter were also associated with mortality from all causes combined. Only nitrogen dioxide had significant positive association with lung cancer mortality.

**Conclusions:** Using the first individualized exposure assignments in this important cohort, we found positive associations of fine particulate matter, ozone and nitrogen dioxide with mortality. The positive associations of nitrogen dioxide suggest that traffic pollution relates to premature death.

*Abstract length: 250 words*

## Introduction

A substantial body of evidence suggests that long-term exposure to combustion-related air pollution contributes to the development of chronic disease and can lead to premature death (1-6). Exposure to air pollution affects huge populations globally. As a result, the public health impact can be large (7, 8).

Using data from the American Cancer Society's (ACS) Cancer Prevention Study II (CPS-II), a nationwide cohort study of nearly 1.2 million adults who have been followed for mortality since 1982, several studies have been published examining associations of metropolitan-level air pollution and mortality (3, 9-11). In those studies, exposure data were derived at the metropolitan scale, relying on between-city exposure contrasts using central monitor data.

In addition, two studies using CPS-II data evaluated within-city (i.e., Los Angeles and New York) exposure contrasts in fine particulate matter with aerodynamic diameter of 2.5 microns or less ( $PM_{2.5}$ ) (2). Both studies assigned exposure to the ZIP code postal area of residence, but in the study from Los Angeles (2), the  $PM_{2.5}$ -mortality dose-response relationship was stronger than that for the full nationwide cohort and in the study from New York City, the relationship was weaker (3). The ZIP code areas, while more specific than the metropolitan area, may still have introduced error in the exposure assignment that led to the inconsistent results. Another recent study based on individualized exposures found little association between  $PM_{2.5}$  exposure and mortality in a cohort of male health professionals (12); however, in that study if home address records were missing, then workplace addresses were used for exposure assignment, possibly leading to measurement error. Conversely, an earlier study based on a large cohort of nurses reported

strong and significant associations of  $PM_{2.5}$  with mortality, using essentially the same exposure model, but with complete home address information for exposure assignment (13). Viewed together, these findings suggest that uncertainties in the characterization of the dose-response may be due partly to the errors in exposure estimates arising from the lack of specificity of the coordinates used to link addresses to the exposure estimates. A need therefore exists to investigate and understand how individualized estimates of exposure at the home address influence the observed dose-response function.

In the present analysis, individualized exposure estimates were developed and assigned to the home address for more than 73,000 California residents enrolled in CPS-II. These estimates were used to assess the association of three criteria air pollutants ( $PM_{2.5}$ , ozone and nitrogen dioxide) with risk of mortality. We also sought to understand the joint effects of the pollutants in co- and multi-pollutant models. Although CPS-II is a nationwide cohort, we limited this analysis to California because the state has a wide range of pollution exposures and a good monitoring network.

## **Methods**

The ACS CPS-II cohort was enrolled in 1982 (details are presented elsewhere (3, 14)). For purposes of this paper, vital status was ascertained through to 2000. Subjects with valid postal addresses had their residential locations geocoded. After limiting to residence in the State of California and making exclusions for missing data on key covariates, there were 73,711 subjects available for analysis.

We assigned exposure for  $PM_{2.5}$ , nitrogen dioxide ( $NO_2$ ) and ozone ( $O_3$ ). Monthly average monitoring data for  $PM_{2.5}$  were available at 112 sites between 1998 and 2002.  $NO_2$

and O<sub>3</sub> data were available over the period 1988-2002 at 138 and 262 sites, respectively. PM<sub>2.5</sub> and NO<sub>2</sub> exposures were assessed using land use regression models (LUR) that were selected from more than 70 possible land use covariates (15). The PM<sub>2.5</sub> model additionally included an advanced remote sensing model coupled with atmospheric modeling (16). LUR models were selected with the deletion/substitution/addition (DSA) algorithm (17). The DSA algorithm, which aggressively tests nearly all polynomial covariate combinations, uses v-fold cross-validation to evaluate potential models. In this instance of v-fold cross-validation, data are first partitioned into 10 roughly equal parts, i.e., folds. The model is then trained on 9 folds and cross-validated on the left out fold. This is repeated a total of 10 times so every fold is used as a cross-validation data set. The model selection method avoids the potential problems of over-fitting on all the data or on a large training set and then using a cross-validation subset (details presented elsewhere (15, 18)). For ozone, we extracted monthly averaged values from 1988-2002 and calculated the Inverse Distance Weighting (IDW) models with the decay parameter set to the inverse of the square of the distance from all sites within a 50 km radius of operational monitors during any particular month. Estimates for all pollutants were then assigned to geocoded baseline residential addresses of the CPS-II subjects, and the monthly values were averaged for the entire time period available.

We employed a comprehensive set of individual risk factor variables operationalized through 42 covariates similar to those used in previous studies of the CPS-II cohort (3, 18). Individual-level variables controlled for lifestyle, dietary, demographic, occupational, and educational factors, and ecological variables extracted from the 1990 US Census in the ZIP code of residence were used to control for potential “contextual” neighborhood

confounding (including unemployment, poverty, income inequality, and racial composition).

We assessed the association between air pollution and mortality using standard and multilevel Cox proportional hazards regression models. Control for place of residence was also applied in the five largest conurbations – defined by the four consolidated metropolitan statistical areas (CMSA) of California and the metropolitan statistical area (MSA) of San Diego – that potentially have lower mortality rates than non-metropolitan areas. This pattern is consistent with what has been termed the “non-metropolitan mortality penalty,” where non-metropolitan areas tend to have higher death rates compared to metropolitan areas (19). Because metropolitan areas also have high pollution, failure to control for residence in large urban areas has the potential to confound associations between mortality and air pollution.

We evaluated the association between air pollution and several causes of death, including cardiovascular disease (CVD), ischemic heart disease (IHD), stroke, respiratory disease, and lung cancer. We also evaluated “all other” causes of death, excluding the preceding causes, to serve as a negative “control”. Finally, we evaluated mortality from all causes combined.

## **Results**

Table 1 compares characteristics of the nationwide CPS-II cohort used in previous analyses to the subset selected for this analysis (a detailed description of exclusions and sample selection is provided elsewhere (18)). Minor differences in alcohol consumption and education are apparent, but overall the California cohort appears to have

characteristics similar to the nationwide cohort. Subjects included in this analysis were widely distributed across California, giving comprehensive coverage for much of the State's population (54 of 58 California counties were represented).

Table 2A shows the mean, variance and percentiles of each pollutant as estimated by the different models employed here. All models display considerable variation in the exposures assigned to the home address. Most pollutants show moderate to high positive correlations (Table 2B). The exception is between interpolated ozone and NO<sub>2</sub> estimates, which displays a weak negative correlation.

Estimates of adjusted relative risks (RR) and 95% confidence intervals (CI) are reported in Table 3. All RR estimates are given over the interquartile range of each specific pollutant. We assessed residual spatial autocorrelation in the health effect estimates with a multilevel Cox model (3). Because the multilevel clustering and autocorrelation analysis had minimal impact on the risk estimates, only results for the standard Cox models are reported.

For PM<sub>2.5</sub> we observe significantly elevated RR for mortality from all causes (RR = 1.032, 95% CI: 1.002, 1.068), CVD (RR = 1.064, 95% CI: 1.016, 1.114), and IHD (RR = 1.111, 95% CI: 1.045, 1.181). Deaths from stroke, respiratory causes and lung cancer had positive relative risks with less precision and confidence intervals that included unity. No association is present with other causes.

NO<sub>2</sub> is significantly and positively associated with all cause (RR = 1.031, 95% CI: 1.008, 1.056), CVD (RR = 1.048, 95% CI: 1.010, 1.087), IHD (RR = 1.066, 95% CI: 1.015, 1.119), stroke (RR = 1.078, 95% CI: 1.016, 1.145), and lung cancer (RR = 1.078, 95% CI: 1.016,

1.145) mortality. Respiratory deaths and those from all other causes were not associated with NO<sub>2</sub>.

While there was no association between O<sub>3</sub> and all-cause mortality, there was a positive association with CVD mortality (RR =1.045, 95% CI: 0.986, 1.108) and a significantly elevated risk for IHD death (RR = 1.104, 95% CI: 1.021, 1.194). O<sub>3</sub> had a positive association with stroke and respiratory deaths that lacked precision and a marginally significant negative association with deaths from lung cancer. There was no association with other causes.

We compared the risk estimates obtained from single pollutant models with risk estimates from two-pollutant and multipollutant models (see Table 4). In models that included PM<sub>2.5</sub> and NO<sub>2</sub>, the PM<sub>2.5</sub> associations with mortality from all causes were reduced to about half the size of those in the single pollutant models and the estimates became insignificant. When ozone and PM<sub>2.5</sub> were included in the same all-cause mortality model, the effects from PM<sub>2.5</sub> remained significantly elevated and became slightly larger. A similar pattern was observed with CVD and IHD, where the effects of PM<sub>2.5</sub> were attenuated with NO<sub>2</sub> but remained unchanged in the presence of the ozone estimates (See Figure 1).

The NO<sub>2</sub> associations with CVD and IHD were attenuated when PM<sub>2.5</sub> was included in the model, but they became slightly larger when O<sub>3</sub> was included. O<sub>3</sub> continued to show elevated risks for CVD and IHD in the two-pollutant models with either NO<sub>2</sub> or PM<sub>2.5</sub> included. For respiratory deaths, PM<sub>2.5</sub> continued to have elevated but insignificant risk estimates, while neither of the other two pollutants were associated with respiratory mortality. For lung cancer, NO<sub>2</sub> consistently displayed significantly elevated risks in two-

pollutant models. When combined with ozone, PM<sub>2.5</sub> associations with lung cancer increased, but remained insignificant.

In multipollutant models containing all three pollutants, NO<sub>2</sub> had the strongest associations with all-cause mortality, CVD and with lung cancer, while PM<sub>2.5</sub> tended to have stronger effects on deaths from IHD. Intercorrelations among the various pollutants, however, likely contribute to bias in individual pollutant risk estimates in such simultaneous pollutant models, so these results must be interpreted with caution. In multipollutant models, PM<sub>2.5</sub> continued to produce elevated risks for all cause, CVD, IHD and respiratory mortality, but none of these estimates was statistically significant. O<sub>3</sub> had elevated risks on CVD, and it remained a significant predictor of IHD deaths even with the other pollutants in the model.

There was little evidence of associations with the other causes of death in the two pollutant or multipollutant models.

In Figure 1, we illustrate results from cumulative risk index (CRI) models for CVD and IHD mortality that show the extent to which one pollutant confounds the others (see the Online Supplement for details of the CRI methods). Comparisons of CRI based on combinations of pollutants estimated jointly and independently can also provide a means of understanding the joint impacts of the atmospheric mixture on survival. For example with CVD mortality, the combined hazard ratio (HR) of NO<sub>2</sub> and O<sub>3</sub> assuming independence is  $1.048 \times 1.045 = 1.095$ . However, the combined HR based on the two-pollutant survival model is 1.121, suggesting a synergy of effect among the pollutants. A similar pattern of synergy is also observed for IHD mortality.

Such a comparative assessment is illustrated in Figure 1 for three pollutants, NO<sub>2</sub>, O<sub>3</sub>, and PM<sub>2.5</sub> and two causes of death, CVD and IHD. The hazard ratios evaluated at their respective IQRs for the three pollutants are presented singly, based on the three possible two-pollutant models, and the single three-pollutant model. There is some modest increase in the CRI for models containing PM<sub>2.5</sub> and either NO<sub>2</sub> or O<sub>3</sub> compared to each of the models with just a single pollutant. The model with NO<sub>2</sub> and O<sub>3</sub>, however, is larger than either of the other two-pollutant models and has a similar CRI to the three-pollutant model, suggesting that a combination of NO<sub>2</sub> and O<sub>3</sub> is sufficient to characterize the toxicity of the pollutant mixture in this study, at least with respect to the three pollutants considered.

The CRI suggests that there is little marginal contribution to CVD and IHD mortality from the addition of PM<sub>2.5</sub> in the presence of the mixture represented by NO<sub>2</sub> and O<sub>3</sub>. We also caution that in this interpretation the confidence intervals clearly overlap each of the CRIs we have calculated. This limits our ability to infer the set of minimally sufficient pollutants required to fully capture the toxicity of the atmosphere in California.

## **Discussion**

We sought to estimate the effects of three criteria air pollutants on premature death in California. This study was motivated by earlier research from Los Angeles that showed PM<sub>2.5</sub> exerted a large significant effect on all-cause mortality and mortality from CVD. Other studies, including those based on data from the ACS CPS-II, showed heterogeneous health effect estimates that potentially resulted from lack of precision in the exposure assessment. To address this problem, we developed detailed exposure assessment models that included auxiliary information and assigned resulting estimates of exposure to the

baseline residential address of more than 73,000 subjects with valid data from the ACS CPS-II cohort.

Several important results deserve mention. First, findings of associations of  $PM_{2.5}$  with all-cause and cardiovascular mortality are consistent with those reported from our previous analyses of the full, nationwide CPS-II cohort (see Krewski et al. (3)). Table 5 shows that results for all-cause, CVD and IHD mortality from the current study are similar, though slightly weaker than from the study of the nationwide cohort. The difference in exposure metrics had little impact on the risk estimates for  $PM_{2.5}$ . We also fit models specifically for Los Angeles to compare to earlier results (2). While the sample size is different here due to limitations in the geocoding, the results show that effects in Los Angeles continue to be higher than those in either the national study or the rest of the state. We also examine the dose-response function for non-linearity because levels in Los Angeles are generally higher than in many other parts of the state, but we found no evidence of non-linearity in the dose response function based on visual inspection of spline plots and formal measures of model fit (AIC and BIC results not shown). This suggests that either the population of LA is more susceptible to air pollution, the air pollution there is more toxic, or both.

Second, the strongest associations with mortality appear to be for exposures that are effectively markers for traffic-related air pollution. The largest predictors of  $NO_2$  in the LUR model were measures of roadway length near the monitors, although we cannot rule out other contributions to modeled concentrations such as heating and industrial sources, particularly given the generally higher concentrations of  $NO_2$  during the winter when home heating contributes to emissions of  $NO_2$  precursors (20). This exposure measure

demonstrated significant associations with all-cause, CVD, IHD, and lung cancer mortality. In multipollutant models, these associations remained elevated, but became insignificant in some models, probably due to multicollinearity among the pollutants. We also examined direct measures of proximity to roadways in earlier studies (18), and found these markers of traffic had positive coefficients, but the findings were null, suggesting that the improved exposure estimates with the LUR model may have reduced exposure measurement error.

Our results are broadly consistent with several studies from Europe where NO<sub>2</sub> exposure was positively associated with mortality (21, 22). In an American study of male truck drivers, NO<sub>2</sub> was found to be independently associated with all-cause and cause-specific mortality, even after controlling for occupational exposures (23). In a comprehensive review by the Health Effects Institute, effects of traffic-related pollution on mortality were identified as suggestive, but insufficient to establish a causal association (24). When viewed in the context of the emerging literature, our results strengthen the evidence base on the effects of traffic-related air pollution on mortality.

Although acute exposure to ozone has been related to mortality (25), here we observed a significant positive association between long-term ozone exposure and cardiovascular disease mortality, notably for IHD. The strength of association for ozone was similar to that of PM<sub>2.5</sub> and NO<sub>2</sub>. The association of ozone with IHD were mildly confounded by PM<sub>2.5</sub>; however, the two exposures had moderately high correlation and given the extensive auxiliary information in the PM<sub>2.5</sub> model, the PM<sub>2.5</sub> estimates may have dominated by virtue of lower exposure measurement error (26). Ozone nevertheless continued to exhibit a significant association with IHD, even with PM<sub>2.5</sub> in the model.

Interestingly, positive RR estimates for ozone became larger when NO<sub>2</sub> was included in the model (see Figure 1). We hypothesize that this results from the negative correlation between the two pollutants due to the atmospheric chemistry, such that in areas where ozone is high, NO<sub>2</sub> tends to be low and vice versa (27, 28). If both pollutants represent harmful constituents of the complex mixture of ambient air pollutions, each would contaminate the comparison for calculating “clean” atmospheres when assessing the risk of the other pollutant. In such instances, the comparison groups with lower pollution levels may also have higher mortality resulting in part from higher levels of the other pollutant that occupies the opposite spatial pattern. We found a negative, significant association between ozone and lung cancer, which became insignificant when NO<sub>2</sub> was included in the model. These findings together suggest the importance of having both ozone and NO<sub>2</sub> in models that attempt to predict health effects from either pollutant. We did observe a weak negative correlation between the two pollutants; however, subsequent analyses showed that in four of the five major urban regions of California, NO<sub>2</sub> and moderately high negative correlations with ozone (see Online Supplement for details), which supports the possibility of the positive confounding we have observed here and of the hypothesis that both pollutants need to be in the model for correct inference on either.

Unlike previous analyses (14) we did not see a significant association between respiratory disease and ozone. In the present analysis, however, the number of respiratory deaths was much smaller than in the earlier national study. The point estimate here was elevated and of similar size to that reported in an earlier analysis of the nationwide cohort (3); consequently, the lack of significant association may have resulted from the lower event numbers. In contrast to earlier results, PM<sub>2.5</sub> did have a positive association with

respiratory mortality, which tended to get stronger with the inclusion of co-pollutants, particularly ozone. In the correlational analyses done by major urban regions (see Appendix), we observed significant negative correlations between ozone and PM<sub>2.5</sub> suggesting again the potential for positive confounding.

Several strengths and limitations merit mention. For NO<sub>2</sub> and PM<sub>2.5</sub>, we used advanced exposure assessment models informed by auxiliary information that had good predictive capacity. These models, however, were based on government monitoring data, and the placement of the government monitoring sites might be less representative of all exposure domains because they are chosen to represent background conditions. For the most part, near road environments are not well represented in this network, limiting the ability to predict small-area variations near roadways. Our estimates of ozone exposure likely do not capture the small area variation that can occur in open space areas and other areas away from roadways (27). Nonetheless, by assigning exposures that vary among individuals within cities, this study extends the applicability of the risk estimates to support studies that have an interest in assessing the health impacts of air pollutants within cities, which is being increasingly done to justify the health benefits of urban planning and climate mitigation interventions (29, 30).

Regarding limitations, there were no follow up surveys conducted in the full CPS-II, and key lifestyle characteristics may have changed during the follow up (e.g., smoking rates declined precipitously across California between 1982 and 2000) (31). If the declines in smoking rates were spatially associated with the air pollution levels, these would have the capacity to confound our air pollution risk estimates. We also lacked information on mobility during the follow up and on key microenvironments such as in-transit exposures,

which contribute substantially to inter-individual variability in air pollution exposures (32).

In conclusion, our results suggest that several components of the combustion-related air pollution mixture are significantly associated with increased all-cause and cause-specific mortality. Associations with CVD deaths in general and IHD in particular stand out as most consistent in our analyses. The strong associations of NO<sub>2</sub> with all-cause, CVD and lung cancer mortality are suggestive of traffic-related pollution as a cause of premature death. The potential for positive confounding between ozone and NO<sub>2</sub> requires increased attention in future research. Given the indications that ozone may relate significantly to CVD mortality, future research may lead to refined ozone exposure assessment with lower measurement error. In sum, the associations observed here reduce key uncertainties on the relationship between air pollution and mortality and confirm that air pollution is a significant risk factor for mortality.

### **Acknowledgements**

This work was supported in part by a contract with the California Air Resources Board. Additional funding came from the Environmental Public Health Tracking Program of the Centers for Disease Control. D. Krewski is the Natural Sciences and Engineering Research Council of Canada Chair in Risk Science at the University of Ottawa. Dr. Thurston was also supported in part by the NYU-NIEHS Center of Excellence Grant ES00260.

## References

1. Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, Holguin F, Hong Y, Luepker RV, Mittleman MA, et al. Particulate Matter Air Pollution and Cardiovascular Disease. *Circulation* 2010; 121: 2331-2378.
2. Jerrett M, Burnett RT, Ma R, Pope CA, 3rd, Krewski D, Newbold KB, Thurston G, Shi Y, Finkelstein N, Calle EE, et al. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 2005; 16: 727-736.
3. Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA, 3rd, Thurston G, Calle EE, et al. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. *Res Rep Health Eff Inst* 2009: 5-114; discussion 115-136.
4. Pope CA, III, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D, Godleski JJ. Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. *Circulation* 2004; 109: 71-77.
5. Pope CA, III, Burnett RT, Turner MC, Cohen A, Krewski D, Jerrett M, Gapstur SM, Thun MJ. Lung Cancer and Cardiovascular Disease Mortality Associated with Ambient Air Pollution and Cigarette Smoke: Shape of the Exposure-Response Relationships. *Environ Health Perspect* 2011; 119.
6. Chen H, Goldberg MS, Villeneuve PJ. A systematic review of the relation between long-term exposure to ambient air pollution and chronic diseases. *Rev Environ Health* 2008; 23: 243-297.
7. Pope CA, III, Dockery DW. Health effects of fine particulate air pollution: lines that connect. *J Air Waste Manag Assoc* 2006; 56: 709-742.
8. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, Amann M, Anderson HR, Andrews KG, Aryee M, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380: 2224-2260.
9. Krewski D, Burnett RT, Goldberg MS, Hoover K, Siemiatycki J, Abrahamowicz M, White WH. 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. Cambridge, MA: Health Effects Institute; 2000.
10. Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama-J Am Med Assoc* 2002; 287: 1132-1141.
11. Pope CA, 3rd, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW, Jr. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am J Respir Crit Care Med* 1995; 151: 669-674.
12. Puett RC, Hart JE, Schwartz J, Hu FB, Liese AD, Laden F. Are particulate matter exposures associated with risk of type 2 diabetes? *Environ Health Perspect* 2011; 119: 384-389.

13. Puett RC, Hart JE, Yanosky JD, Paciorek C, Schwartz J, Suh H, Speizer FE, Laden F. Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *Environ Health Perspect* 2009; 117: 1697-1701.
14. Jerrett M, Burnett RT, Pope CA, 3rd, Ito K, Thurston G, Krewski D, Shi Y, Calle E, Thun M. Long-term ozone exposure and mortality. *N Engl J Med* 2009; 360: 1085-1095.
15. Beckerman BS, Jerrett M, Martin RV, van Donkelaar A, Ross Z, Burnett RT. Application of the Deletion/Substitution/Addition Algorithm to Selecting Land Use Regression Models for Interpolating Air Pollution Measurements. *Atmos Environ*; doi: 10.1016/j.atmosenv.2013.04.024.
16. van Donkelaar A, Martin RV, Brauer M, Kahn R, Levy R, Verduzco C, Villeneuve PJ. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ Health Perspect* 2010; 118: 847-855.
17. Sinisi SE, van der Laan MJ. Deletion/substitution/addition algorithm in learning with applications in genomics. *Statistical Applications in Genetics and Molecular Biology* 2004; 3: Article18.
18. Jerrett M, Burnett RT, Pope III A, Krewski D, Thurston G, Christakos G, Hughes E, Ross Z, Shi Y, Thun M, et al. Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort: Final Report. California Air Resources Board; 2011.
19. Cosby AG, Neaves TT, Cossman RE, Cossman JS, James WL, Feierabend N, Mirvis DM, Jones CA, Farrigan T. Preliminary evidence for an emerging nonmetropolitan mortality penalty in the United States. *Am J Public Health* 2008; 98: 1470-1472.
20. Spengler J, Schwab M, Ryan PB, Colome S, Wilson AL, Billick I, Becker E. Personal exposure to nitrogen dioxide in the Los Angeles Basin. *Air & waste : journal of the Air & Waste Management Association* 1994; 44: 39-47.
21. Brunekreef B. Health effects of air pollution observed in cohort studies in Europe. *Journal of exposure science & environmental epidemiology* 2007; 17 Suppl 2: S61-65.
22. Cesaroni G, Badaloni C, Gariazzo C, Stafoggia M, Sozzi R, Davoli M, Forastiere F. Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. *Environ Health Perspect* 2013; 121: 324-331.
23. Hart JE, Garshick E, Dockery DW, Smith TJ, Ryan L, Laden F. Long-term ambient multipollutant exposures and mortality. *Am J Respir Crit Care Med* 2011; 183: 73-78.
24. HEI. Health Effects Institute: HEI Panel on the Health Effects of Traffic-Related Air Pollution. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure and Health Effects: Special Report 17. May, 2009; 2009.
25. Henrotin JB, Zeller M, Lorgis L, Cottin Y, Giroud M, Bejot Y. Evidence of the role of short-term exposure to ozone on ischaemic cerebral and cardiac events: the Dijon Vascular Project (DIVA). *Heart* 2010; 96: 1990-1996.
26. Zidek JV, Wong H, Le ND, Burnett R. Causality, measurement error and multicollinearity in epidemiology. *Environmetrics* 1996; 7: 441-451.
27. Beckerman B, Jerrett M, Brook JR, Verma DK, Arain MA, Finkelstein MM. Correlation of nitrogen dioxide with other traffic pollutants near a major expressway. *Atmos Environ* 2008; 42: 275-290.
28. McConnell R, Berhane K, Yao L, Lurmann FW, Avol E, Peters JM. Predicting residential ozone deficits from nearby traffic. *Sci Total Environ* 2006; 363: 166-174.

29. Woodcock J, Edwards P, Tonne C, Armstrong BG, Ashiru O, Banister D, Beevers S, Chalabi Z, Chowdhury Z, Cohen A, et al. Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. *Lancet* 2009; 374: 1930-1943.
30. Rojas-Rueda D, de Nazelle A, Tainio M, Nieuwenhuijsen MJ. The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *Brit Med J* 2011; 343.
31. California Department of Public Health CTCP. April 2011 [cited 2012 September 11]. Available from: <http://www.cdph.ca.gov/Pages/NR11-031SmokingChart.aspx>.
32. de Nazelle A, Nieuwenhuijsen MJ, Anto JM, Brauer M, Briggs D, Braun-Fahrlander C, Cavill N, Cooper AR, Desqueyroux H, Fruin S, et al. Improving health through policies that promote active travel: A review of evidence to support integrated health impact assessment. *Environ Int* 2011; 37: 766-777.

## List of Tables

Table 1: Participant characteristics in the Nationwide Study compared to the California Cohort as Percentages or Means with Standard Deviations

<b>Variable</b>	<b>Nationwide</b>	<b>California</b>
Participants (n)	485,426	73,711
<i>Participants died from (%)</i>		
All Causes	26.4	26.8
CPD	13.1	13.6
CVD	10.9	10.9
IHD	6.1	6.2
Respiratory	2.2	2.7
Lung Cancer	2.0	2.0
All other causes	11.3	11.2
<i>Demographics</i>		
Mean (SD) Age (yrs)	56.6 (10.5)	57.4 (10.6)
Female (%)	56.6	56.2
White (%)	94.2	91.6
Education (%)		
< High School	12.1	8.7
High School	31.3	22.9
>High School	56.6	68.4
<i>Alcohol consumption (%)</i>		
Beer	22.9	24.1
No Beer	9.5	10.9
Missing Beer	67.6	65.0
Liquor	27.6	35.1
No Liquor	8.7	8.9
Missing Liquor	63.7	56.0
Wine	23.1	37.3
No Wine	8.9	7.7
Missing Wine	68.0	55.0
<i>Smoking Status</i>		
Current Smoker (%)	21.6	19.4
Cigarettes per day	22.1 (12.4)	21.5 (12.6)
Years of smoking	33.5 (11.0)	34.1 (11.4)
Former Smoker (%)	25.9	28.9
Cigarettes per day	21.4 (14.7)	20.8 (14.7)
Years of smoking	22.2 (12.6)	22.1 (12.7)
<i>Age when started smoking (%)</i>		
< 18 yrs (current smoker)	8.9	7.7
< 18 yrs (former smoker)	10.0	10.3
Hours per day exposed to smoking	3.2 (4.4)	2.7 (4.1)

Table 2A: Distribution of air pollutants at individual level. See footnote for years represented by air pollution exposure<sup>1</sup>

Air Pollution	# Subjects	Mean	Variance	Percentiles								
				0	5	10	25	50	75	90	95	100
PM <sub>2.5</sub> LUR ( $\mu\text{g}/\text{m}^3$ )	73,711	14.09	12.42	4.25	8.29	9.45	11.60	14.03	16.90	18.42	19.36	25.09
NO <sub>2</sub> LUR (ppb)	73,711	12.27	8.54	3.04	7.93	8.81	10.21	12.12	14.33	16.22	17.09	21.94
Ozone IDW (ppb)	73,711	50.35	212.18	17.11	28.81	31.13	36.83	50.80	61.00	68.56	74.18	89.33

Table 2B: Pearson correlations (x100) between air pollutants (California overall).

	PM <sub>2.5</sub> LUR	NO <sub>2</sub> LUR
PM <sub>2.5</sub> LUR	--	--
NO <sub>2</sub> LUR	55.10	--
Ozone IDW	55.81	-0.71

<sup>1</sup> Years represented by air pollution exposure estimates

Pollutant	PM <sub>2.5</sub> LUR	NO <sub>2</sub> LUR	Ozone IDW
Years	1998-2002	1988-2002	1988-2002

Table 3: ACS cohort with follow-up from 1982 to 2000, adjusting for 42 individual level covariates, five CMSA city indicators, seven 1990 ecologic covariates stratifying the baseline hazard function by age (1-year groupings), gender, and race using the standard Cox survival model (95% confidence intervals given in parenthesis). Relative risks are shown for the interquartile range of exposure in each pollutant (i.e., 5.3037  $\mu\text{g}/\text{m}^3$  for  $\text{PM}_{2.5}$ ; 4.1167 ppb  $\text{NO}_2$ ; and 24.1782 ppb for  $\text{O}_3$ ).

Air Pollutant	Cause of Death						
	All Causes	Cardiovascular	Ischemic Heart	Stroke	Respiratory	Lung Cancer	All Others
	(n=19,733)	(n=8,046)	(n=4,540)	(n=3,068)	(n=1,973)	(n=1,481)	(n=8,233)
<b>PM<sub>2.5</sub> LUR</b>	1.032 (1.002, 1.062)	1.064 (1.016, 1.114)	1.111 (1.045, 1.181)	1.065 (0.988, 1.148)	1.046 (0.953, 1.148)	1.062 (0.954, 1.183)	0.994 (0.950, 1.040)
<b>NO<sub>2</sub> LUR</b>	1.031 (1.008, 1.056)	1.048 (1.010, 1.087)	1.066 (1.015, 1.119)	1.078 (1.016, 1.145)	0.999 (0.927, 1.077)	1.111 (1.020, 1.210)	1.009 (0.973, 1.046)
<b>Ozone IDW</b>	0.998 (0.960, 1.036)	1.045 (0.986, 1.109)	1.104 (1.021, 1.194)	1.011 (0.919, 1.112)	1.017 (0.902, 1.147)	0.861 (0.747, 0.992)	0.967 (0.911, 1.027)

Table 4: Two-pollutant and multipollutant model results from ACS cohort with follow-up from 1982 to 2000, adjusting for 42 individual level covariates, five CMSA city indicators, seven 1990 ecologic covariates stratifying the baseline hazard function by age (1-year groupings), gender, and race using the standard Cox survival model (95% confidence intervals given in parenthesis). Relative risks are shown for the interquartile range of exposure in each pollutant (i.e., 5.3037  $\mu\text{g}/\text{m}^3$  for  $\text{PM}_{2.5}$ ; 4.1167 ppb  $\text{NO}_2$ ; and 24.1782 ppb for  $\text{O}_3$ ).

Air Pollutant	Cause of Death						
	All Causes	Cardiovascular	Ischemic Heart	Stroke	Respiratory	Lung Cancer	All Others
	(n=19,733)	(n=8,046)	(n=4,540)	(n=3,068)	(n=1,973)	(n=1,481)	(n=8,233)
<b>PM<sub>2.5</sub> LUR</b>	1.015 (0.980, 1.050)	1.043 (0.989, 1.101)	1.090 (1.015, 1.170)	1.019 (0.934, 1.112)	1.064 (0.954, 1.185)	0.985 (0.867, 1.119)	0.984 (0.933, 1.038)
<b>NO<sub>2</sub> LUR</b>	1.025 (0.997, 1.054)	1.030 (0.987, 1.075)	1.029 (0.972, 1.090)	1.070 (0.998, 1.147)	0.973 (0.891, 1.063)	1.118 (1.010, 1.236)	1.016 (0.973, 1.060)
<b>PM<sub>2.5</sub> LUR</b>	1.035 (1.004, 1.067)	1.057 (1.008, 1.109)	1.093 (1.027, 1.165)	1.067 (0.987, 1.153)	1.045 (0.949, 1.151)	1.103 (0.985, 1.234)	1.002 (0.955, 1.050)
<b>Ozone IDW</b>	0.985 (0.947, 1.025)	1.025 (0.964, 1.089)	1.070 (0.987, 1.161)	0.988 (0.894, 1.091)	1.001 (0.883, 1.134)	0.832 (0.719, 0.964)	0.966 (0.908, 1.029)
<b>NO<sub>2</sub> LUR</b>	1.032 (1.008, 1.057)	1.055 (1.016, 1.095)	1.082 (1.029, 1.137)	1.082 (1.019, 1.150)	1.001 (0.928, 1.080)	1.097 (1.006, 1.196)	1.006 (0.970, 1.043)
<b>Ozone IDW</b>	1.006 (0.968, 1.046)	1.062 (1.000, 1.127)	1.132 (1.045, 1.227)	1.034 (0.938, 1.140)	1.017 (0.901, 1.149)	0.882 (0.764, 1.019)	0.968 (0.912, 1.029)
<b>PM<sub>2.5</sub> LUR</b>	1.015 (0.977, 1.055)	1.024 (0.965, 1.086)	1.048 (0.969, 1.133)	1.008 (0.915, 1.110)	1.070 (0.949, 1.207)	1.040 (0.902, 1.198)	0.995 (0.938, 1.056)
<b>NO<sub>2</sub> LUR</b>	1.025 (0.995, 1.056)	1.044 (0.996, 1.093)	1.059 (0.995, 1.126)	1.079 (1.000, 1.163)	0.969 (0.881, 1.066)	1.078 (0.967, 1.201)	1.008 (0.963, 1.056)
<b>Ozone IDW</b>	0.999 (0.957, 1.042)	1.050 (0.982, 1.122)	1.106 (1.012, 1.209)	1.031 (0.925, 1.149)	0.984 (0.860, 1.126)	0.866 (0.739, 1.015)	0.971 (0.908, 1.038)

Table 5: Comparison of Relative Risk Estimates from the California and National American Cancer Society Cohorts for PM<sub>2.5</sub> using a 10 µg/m<sup>3</sup> Exposure Increment

	California*		National Level†		Los Angeles Only*	
	Hazard Ratio	95% CI	Hazard Ratio	95% CI	Hazard Ratio	95% CI
<b>All-cause</b>	1.060	(1.003, 1.120)	1.065	(1.035, 1.096)	1.104	(0.968, 1.260)
<b>CVD</b>	1.122	(1.030, 1.223)	1.141	(1.086, 1.198)	1.124	(0.918, 1.375)
<b>IHD</b>	1.217	(1.085, 1.365)	1.248	(1.160, 1.342)	1.385	(1.058, 1.814)

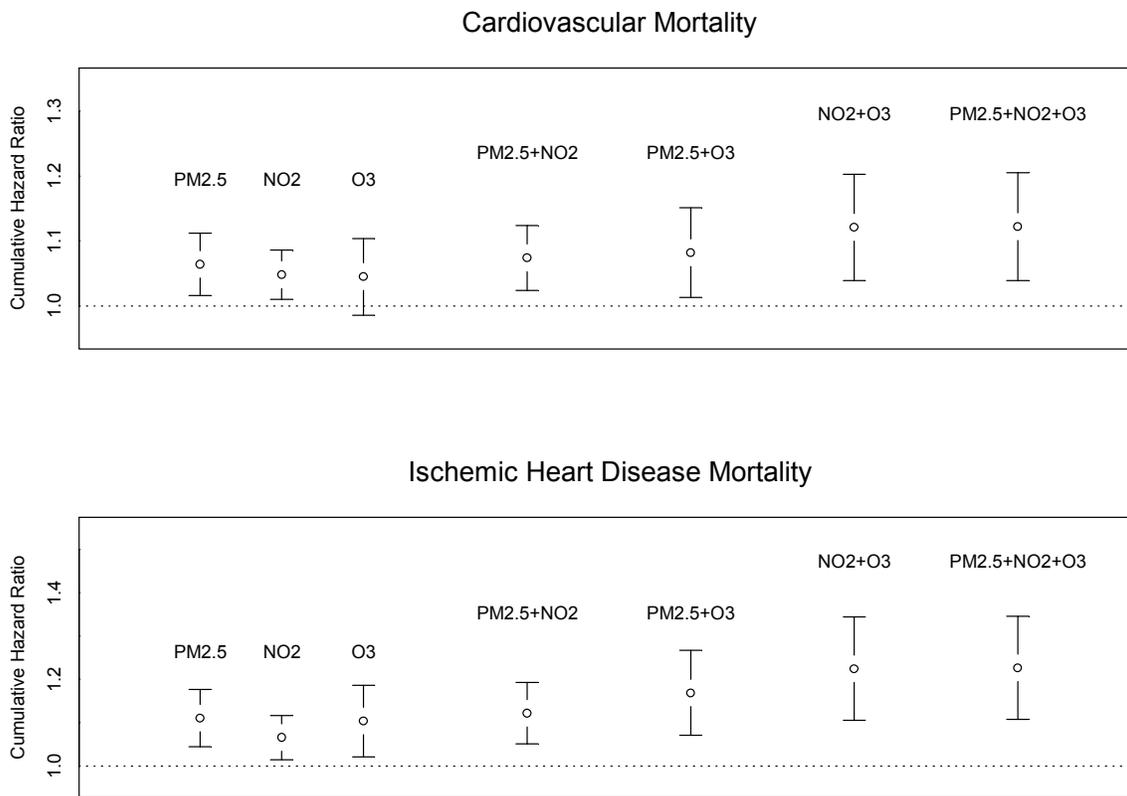
\* California and Los Angeles use residential address with a Land Use Regression estimate of exposure results using standard Cox model.

†National level study uses metropolitan area of residence with the average of all PM<sub>2.5</sub> monitors within the metropolitan area as the exposure estimate; results using two level random effects, assuming no spatial autocorrelation.

Models for both risk estimates control for an identical set of 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 all studies was from 1982-2000.

### Figure Legend

Figure 1: Summary Plot of Individual and Multipollutant Cumulative Hazard Ratios



# Spatial Analysis of Air Pollution and Mortality in California

ONLINE SUPPLEMENTAL DATA

### Supplemental 1: Calculation and Interpretation of Multiple Pollutant Survival Models used to Derive the Cumulative Risk Index

Let  $x' = (x_1, \dots, x_p)$  represent concentrations of P air pollutants (i.e., NO<sub>2</sub>, O<sub>3</sub>, and PM<sub>2.5</sub>). We denote the hazard ratio (HR) based on the combination of the P pollutants evaluated at  $x$  as the Cumulative Risk Index (CRI) and define it as

$$CRI = \exp\left\{\sum_{p=1}^P \hat{\beta}_p x_p\right\} \equiv \exp(\hat{\beta}'x) = \prod_{p=1}^P JHR_p$$

where  $\hat{\beta}' = (\hat{\beta}_1, \dots, \hat{\beta}_p)$  are the estimates of the log-hazard ratio for the P pollutants estimated in a survival model consisting of all P pollutants together, with  $JHR_p = \exp(\hat{\beta}_p x_p)$  denoting the joint hazard ratio for the  $p^{th}$  pollutant in a multi-pollutant survival model. Further denote  $Cov(\hat{\beta})$  as the variance-covariance matrix of  $\hat{\beta}$ . The 95% confidence interval of CRI is defined by:  $\exp\{\hat{\beta}'x \pm 1.96 \times \hat{\beta}Cov(\hat{\beta})\hat{\beta}'\}$ .

The interquartile range (IQR) of an exposure distribution is often used as an exposure contrast to evaluate the hazard ratio, thus representing a summary measure of the impact of a pollutant on population survival. The joint impact of multiple pollutants can be determined using the CRI with  $x$  denoting a vector of the IQRs of all P pollutants. Comparisons of CRIs based on selected subsets of the P pollutants are used as a guide to interpreting the joint effects of these pollutant subsets.

If the pollutants are uncorrelated we note that  $\prod_{p=1}^P JHR_p = \prod_{p=1}^P IHR_p$ , where  $IHR_p$  denotes the hazard ratio of the  $p^{th}$  pollutant from a survival model that only contains this pollutant, and thus each pollutant effect is estimated independently of the other pollutants. If all  $p$  pollutants are positively correlated then  $\prod_{p=1}^P JHR_p < \prod_{p=1}^P HR_p$ . That is, information in one pollutant will also be contained in the other pollutants and one does not necessary require all  $p$  pollutants to fully represent the cumulative risk impact on survival of the pollution mixture.

**Supplemental 2:** Correlations among pollutants in different geographic locals, including the entire state of California, the four largest combined statistical metropolitan areas plus the San Diego metropolitan statistical area, the entire state minus these five large urban areas, and each of the five urban areas separately. Descriptive text included in the main body of the paper.

<b>Geographic Place</b>	<b>NO<sub>2</sub>-O<sub>3</sub></b>	<b>PM<sub>2.5</sub>-O<sub>3</sub></b>	<b>NO<sub>2</sub>-PM<sub>2.5</sub></b>
Entire State of California	-0.01	0.56	0.55
Five Largest Urban Regions Combined (Listed below)	-0.04	0.56	0.49
Rest of CA minus the Five Largest Urban Regions	0.26	0.73	0.65
Fresno-Madera, CA	0.30	0.48	0.74
Los Angeles	-0.25	0.12	0.47
Sacramento-Arden-Arcade-Truckee	-0.48	-0.55	0.88
San Diego	-0.11	-0.47	0.36
San Jose-San Francisco-Oakland	-0.31	-0.12	0.67