



RESEARCH REPORT

HEALTH
EFFECTS
INSTITUTE

Number 183
Part 3
April 2016

Development of Statistical Methods for Multipollutant Research

**Part 3. Modeling of Multipollutant Profiles and
Spatially Varying Health Effects with Applications
to Indicators of Adverse Birth Outcomes**

John Molitor, Eric Coker, Michael Jerrett, Beate Ritz,
and Arthur Li



Development of Statistical Methods for Multipollutant Research

Part 3. Modeling of Multipollutant Profiles and
Spatially Varying Health Effects with Applications
to Indicators of Adverse Birth Outcomes

John Molitor, Eric Coker, Michael Jerrett, Beate Ritz, and Arthur Li

with a Critique by the HEI Health Review Committee

Research Report 183, Part 3

Health Effects Institute

Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

Publishing history: This document was posted at www.healtheffects.org in April 2016.

Citation for this document:

Molitor J, Coker E, Jerrett M, Ritz B, Li A. 2016. Part 3. Modeling of Multipollutant Profiles and Spatially Varying Health Effects with Applications to Indicators of Adverse Birth Outcomes. In: Development of Statistical Methods for Multipollutant Research. Research Report 183. Boston, MA:Health Effects Institute.

© 2016 Health Effects Institute, Boston, Mass., U.S.A. Cameographics, Belfast, Me., Compositor. Printed by Recycled Paper Printing, Boston, Mass. Library of Congress Catalog Number for the HEI Report Series: WA 754 R432.

♻️ Cover paper: made with at least 55% recycled content, of which at least 30% is post-consumer waste; free of acid and elemental chlorine. Text paper: made with 100% post-consumer waste recycled content; acid free; no chlorine used in processing. The book is printed with soy-based inks and is of permanent archival quality.

CONTENTS

About HEI	v
About This Report	vii
Preface	ix
HEI STATEMENT	1
INVESTIGATORS' REPORT <i>by Molitor et al.</i>	3
ABSTRACT	3
INTRODUCTION	4
SPECIFIC AIMS	4
METHODS AND STUDY DESIGN	4
Continuous Exposures	5
Categorical Exposures	6
Combinations of Categorical and Continuous Exposures	6
Finding Clustering that Best Fits the Data	6
MODELING APPROACHES: DATA ANALYSIS	7
Associations Between Air Pollution Profiles and Poverty in Los Angeles County	7
Introduction	7
Methods	7
Results	9
Discussion	13
A Place-Based Approach to Examining Associations Between Multipollutant Exposure Profiles and TLBW	14
Introduction	14
Methods	15
Results: Separate Multipollutant and Contextual Clusters	17
Results: Joint Contextual–Multipollutant Clusters	23
Discussion	26
IMPLICATIONS OF FINDINGS	28
CONCLUSION	29
REFERENCES	29
HEI QUALITY ASSURANCE STATEMENT	34
APPENDIX A	35
ABOUT THE AUTHORS	46
OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH	46
ABBREVIATIONS AND OTHER TERMS	47

Research Report 183 Part 3

CRITIQUE <i>by the Health Review Committee</i>	49
INTRODUCTION	49
SIDEBAR: PROCESS FOR STATISTICAL METHODS DEVELOPMENT AND EVALUATION	50
APPROACH	50
THE HEALTH REVIEW COMMITTEE'S EVALUATION	51
SUMMARY AND CONCLUSIONS	54
ACKNOWLEDGMENTS	54
REFERENCES	54
 Related HEI Publications	 57
 HEI Board, Committees, and Staff	 59

ABOUT HEI

The Health Effects Institute is a nonprofit corporation chartered in 1980 as an independent research organization to provide high-quality, impartial, and relevant science on the effects of air pollution on health. To accomplish its mission, the institute

- Identifies the highest-priority areas for health effects research;
- Competitively funds and oversees research projects;
- Provides intensive independent review of HEI-supported studies and related research;
- Integrates HEI's research results with those of other institutions into broader evaluations; and
- Communicates the results of HEI's research and analyses to public and private decision makers.

HEI typically receives balanced funding from the U.S. Environmental Protection Agency and the worldwide motor vehicle industry. Frequently, other public and private organizations in the United States and around the world also support major projects or research programs. HEI has funded more than 330 research projects in North America, Europe, Asia, and Latin America, the results of which have informed decisions regarding carbon monoxide, air toxics, nitrogen oxides, diesel exhaust, ozone, particulate matter, and other pollutants. These results have appeared in more than 260 comprehensive reports published by HEI, as well as in more than 1000 articles in the peer-reviewed literature.

HEI's independent Board of Directors consists of leaders in science and policy who are committed to fostering the public-private partnership that is central to the organization. The Health Research Committee solicits input from HEI sponsors and other stakeholders and works with scientific staff to develop a Five-Year Strategic Plan, select research projects for funding, and oversee their conduct. The Health Review Committee, which has no role in selecting or overseeing studies, works with staff to evaluate and interpret the results of funded studies and related research.

All project results and accompanying comments by the Health Review Committee are widely disseminated through HEI's Web site (www.healtheffects.org), printed reports, newsletters and other publications, annual conferences, and presentations to legislative bodies and public agencies.

ABOUT THIS REPORT

Research Report 183, Part 3, *Modeling of Multipollutant Profiles and Spatially Varying Health Effects with Applications to Indicators of Adverse Birth Outcomes*, presents a research project funded by the Health Effects Institute and conducted by Dr. John Molitor of the College of Public Health and Human Sciences, Oregon State University, Corvallis, and his colleagues. The report contains three main sections.

The HEI Statement, prepared by staff at HEI, is a brief, nontechnical summary of the study and its findings; it also briefly describes the Health Review Committee's comments on the study.

The Investigators' Report, prepared by Dr. John Molitor and colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

The Critique is prepared by members of the Health Review Committee with the assistance of HEI staff; it places the study in a broader scientific context, points out its strengths and limitations, and discusses remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI's rigorous review process. When an HEI-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. This draft report is first examined by outside technical reviewers and a biostatistician. The report and the reviewers' comments are then evaluated by members of the Health Review Committee, an independent panel of distinguished scientists who have no involvement in selecting or overseeing HEI studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Critique reflects the information provided in the final version of the report.

PREFACE

HEI's Research Program to Develop Methods for Analyzing Multiple Air Pollutants and Health Outcomes

INTRODUCTION

Air pollution is a complex mixture of gaseous, liquid, and solid components that varies greatly in composition and concentration across the United States and around the world owing to differences in sources, weather, and topography. Air pollution also varies from day to day and by season within a region. Although it is clear that people are exposed to complex mixtures of pollutants emitted by diverse sources, the U.S. Clean Air Act — and most existing air quality guidelines and standards to protect public health — focuses on controlling a common set of pollutants individually (called criteria pollutants in the United States). Given this regulatory approach, it is perhaps not surprising that the majority of data on ambient air pollution levels and on human exposures and their health effects have focused on individual pollutants.

Since the air we breathe is a mixture, the scientific community has considered the possibility that the observed adverse health effects associated with individual pollutants may be partly attributable to the combined effects of multiple pollutants. However, the challenges of determining whether effects are additive, synergistic, or less-than-additive, and of identifying possible effect modifiers in epidemiologic studies, are substantial (Mauderly and Samet 2009). Often, a high degree of correlation exists among levels of different pollutants emitted from similar sources or generated through similar atmospheric processes; and there may be nonlinear interactions among pollutants in relation to health outcomes. These issues complicate and may even preclude the use of conventional linear regression approaches. Exposure measurement and exposure modeling errors contribute additional complications; pollutants that are

measured relatively easily (i.e., more frequently and accurately because their concentrations are well above detection levels) will tend to dominate the estimation, even if their effects are less strong than those of other pollutants.

HEI issued Request for Applications (RFA) 09-1, "Methods to Investigate the Effects of Multiple Air Pollution Constituents" in 2009 because it was clear that advancing scientific understanding would require improved statistical methods to determine how the health effects of a pollutant mixture as a whole differ from the effects of individual pollutants within the mixture.

GOALS OF THE RESEARCH PROGRAM

RFA 09-1 solicited research proposals that would address the methodologic difficulties associated with investigating the health effects of multiple pollutants through the development of innovative statistical methods. HEI primarily sought applications for research in which existing statistical approaches (including those from fields outside epidemiology) could be modified, extended, or combined, and then applied to a real-world exposure and health problem, rather than proposals for the development of purely theoretical statistical approaches. RFA 09-1 defined two specific objectives:

1. The research should support the development of innovative statistical methods for studying the combined effects of individual pollutants within complex pollutant mixtures. Analytic approaches could include improvements to existing multivariate methods and the development of strategies for their application or the proposal of new approaches. Of particular interest were multivariate

methods adapted to studying highly correlated pollutants and methods to detect the presence of interactions between two or more pollutants and to evaluate their combined effects. All methods proposed were required to include validation of the approach either by using simulation studies or by conducting a thorough sensitivity analysis with widely available data sets.

2. The research should support the development of innovative statistical methods for studying health effects of air pollution mixtures in animal models and human populations. Of particular interest were methods for characterizing mixtures emitted by specific pollutant sources or groups of sources.

The RFA welcomed proposals for methods that would explore how the effects of a pollutant mixture as a whole differ from the effects of individual pollutants within the mixture. Applicants were expected to employ methods that would be able to analyze both highly correlated pollutant concentration variables and assess the potential effects of measurement error within the chosen statistical framework.

BACKGROUND

At the time the RFA was issued, adequate statistical methods designed for analyzing the relationships among multiple pollutants and health effects were unavailable. In order to better understand the health effects of exposure to the mixture of air pollutants that people actually breathe, to delineate the contribution of individual pollutants or mixtures to adverse health effects, and to address emissions from the sources of those pollutants more cost-effectively, approaches that would go beyond the single-pollutant framework were clearly needed. A 2004 report from the National Research Council (NRC) Committee on Air Quality Management in the United States called for changing the entire air quality management system to a multipollutant approach. The report recommended that the U.S. Environmental Protection Agency (U.S. EPA) consider multiple pollutant scenarios in the National Ambient Air Quality Standards (NAAQS) review and standard setting process: "Although the committee does not believe that the science has evolved to a sufficient extent to permit the development of multipollutant NAAQS, it would be scientifically prudent to begin to review and

develop NAAQS for related pollutants in parallel and simultaneously" (NRC 2004).

The U.S. EPA responded to the NRC report by undertaking a number of activities in support of multipollutant research and a NAAQS targeted specifically to multipollutant mixtures. In late 2006, the Agency hosted the first of several workshops on multipollutant research and commenced efforts to develop a multipollutant NAAQS in 2010 (U.S. EPA 2006, 2011). In 2007, the U.S. EPA also began development of its first two-pollutant Integrated Science Assessment for nitrogen dioxide (NO₂) and sulfur dioxide (SO₂), which was finalized in December 2008.

Following the NRC recommendations, HEI also included multipollutant research as part of its research agenda, specifying in its Strategic Plan for 2005–2010 the health effects of air pollution mixtures as a priority research and review topic. Specifically, this plan called for HEI to "undertake targeted research programs on PM (particulate matter) and gases and on air toxics, two important mixtures within the broader air pollution mixture". Following the discussions about research needs at the U.S. EPA workshops, HEI issued RFA 09-1 in 2009.

At the time, some existing multipollutant modeling approaches were available to researchers in the fields of epidemiology and air pollution exposure. The process of attributing measured concentrations of multiple pollutants to the emissions from specific categories of sources, known as source apportionment, had been evolving and had become increasingly standardized during the early 2000s (Thurston et al. 2005). When statistically feasible, researchers also employed variations on linear regression, such as multivariate regression models, which simultaneously incorporated covariates for multiple pollutants. Both approaches are briefly described here.

SOURCE APPORTIONMENT

When strong correlations among pollutants in given mixtures preclude the use of multiple individual exposure variables in conventional health effects models, source apportionment is used to analyze the mixture of pollutants over time and space. It is a latent-variable method, usually applied in models that include multiple variables, at least one of which is unobserved (or latent). Factor analysis is a special type of latent-variable model used in source apportionment where the analysis assumes that multiple variables are linked together

through their association with a small number of latent variables, called factors. Source apportionment is the process of attributing emission sources to factors based on the composition of the factor: For example, a factor analysis of roadside particulate pollution data may yield a factor in which levels of copper and iron are high and vary together; in a source apportionment, this factor might be attributed to tire and brake wear given what we know about the composition of tires and brakes.

Using source apportionment to classify mixtures — based on source-specific markers in a mixture — can also link health effects with emissions from specific sources (such as facilities or activities). This approach uses the resulting quantification of components that comprise the different source mixtures in a given environment to evaluate their individual or combined contributions to health effects.

However, source-apportionment techniques are not capable of assessing the effects of interactions among the different source-apportioned mixtures, and they may not take into account the underlying biological plausibility of any given mixture to affect health. In addition, when HEI issued RFA 09-1, many researchers were using source-apportionment methods and multivariate-receptor models as “black box tools” and were not linking them sufficiently to rigorous statistical practice or demonstrating an understanding of method limitations. Moreover, the inherent uncertainty of variables generated through source apportionment, due in part to errors in the measurement of individual pollutant concentrations, was not reflected in the estimates of their associations with the health outcomes, thus rendering reproducibility and comparison among different studies difficult.

MULTIVARIATE REGRESSION MODELS

When data sets contain measurements of many constituents of air pollution obtained at different places and time points together with information about health outcomes, and when there is sufficient variability in these data, multivariate analyses of the association between constituents and health outcomes may be possible. Such analyses are aimed primarily at estimating the effects of specific constituents of interest while accounting for the potential effects of confounding. Moreover, multivariate regression models can be used to detect whether the effects of various pollutants are additive or

not. However, there are limitations to the value of simply introducing a number of pollutant variables and interaction terms simultaneously into a regression analysis and carrying out multivariate rather than univariate regressions. For example, high degrees of correlation among covariates render the results statistically unstable and difficult to interpret, and stepwise methods are inadequate in the presence of strong collinearity.

STUDIES FUNDED UNDER RFA 09-1

The three studies funded under RFA 09-1 represent a variety of statistical approaches and of data sets used to test them. The studies by Dr. Brent Coull and Dr. Eun Sug Park and their colleagues are described in Parts 1 and 2 of Research Report 183, and were published in 2015. The study by Dr. John Molitor and colleagues is described in Part 3 in this volume. The studies are described briefly below.

Statistical Learning Methods for the Effects of Multiple Air Pollution Constituents, Brent Coull, Harvard T.H. Chan School of Public Health (Principal Investigator)

Coull and colleagues developed a new analysis framework based on methods that simultaneously quantify variability in health outcomes and exposure data for multiple pollutants in order to identify the mixture profiles (groupings of pollutants and concentrations) most highly associated with the health outcomes. They developed and applied these methods using simulations, pollutant concentration and health outcomes data from the “Maintenance of Balance, Independent Living, Intellect, and Zest in the Elderly of Boston” (MOBILIZE) study cohort of senior citizens living in the Boston area, and toxicologic data from canine studies.

Development of Enhanced Statistical Methods for Assessing Health Effects Associated with an Unknown Number of Major Sources of Multiple Air Pollutants, Eun Sug Park, Texas A&M Transportation Institute (Principle Investigator)

Park and colleagues developed enhanced statistical methods to jointly assess source factors and health effects using multivariate source-characterization and source-apportionment models together with a health outcomes analysis. The investigators' approach incorporated the

uncertainty in the source apportionments into the estimation of the source-related health effects. They applied their methods to data sets for daily pollutant concentrations and acute health outcomes in Phoenix, Arizona, and Houston, Texas, and compared the results with those obtained using conventional methods of estimation.

Modeling of Multipollutant Profiles and Spatially Varying Health Effects with Applications to Indicators of Adverse Birth Outcomes, John Molitor, Oregon State University (Principal Investigator)

Molitor and colleagues developed and applied statistical methods to examine associations among geographically based patterns of air pollutant concentrations, birth outcomes, and socioeconomic status. The investigators used a large data set of pollutant concentrations (including NO₂ and PM ≤ 2.5 μm in aerodynamic diameter) and data on birth outcomes from Los Angeles County, California. They first used Bayesian statistical methods to identify clusters of specific mixtures of pollutants and pollutant concentrations frequently found together in census units, and then associated those pollutant profiles with data on socioeconomic status and health outcomes using regression methods.

REFERENCES

Mauderly JL, Samet JM. 2009. Is there evidence for synergy among air pollutants in causing health effects? *Environ Health Perspect* 117:1–6. doi:10.1289/ehp.11654.

National Research Council. 2004. *Research Priorities for Airborne Particulate Matter. IV. Continuing Research Progress*. Washington, DC:National Academy Press.

Thurston GD, Ito K, Mar T, Christensen WF, Eatough DJ, Henry RC, et al. 2005. Work-group report: Workshop on source apportionment of particulate matter health effects — Intercomparison of results and implications. *Environ Health Perspect* 113:1768–177.

U.S. Environmental Protection Agency. 2006. Workshop on Interpretation of Epidemiologic Studies of Multipollutant Exposure and Health Effects. *Federal Register*/Vol. 71, No. 225 / Wednesday, November 22, 2006. <http://docs.regulations.justia.com/entries/2006-11-22/E6-19806.pdf>.

U.S. Environmental Protection Agency. 2011. Overview of EPA Multipollutant Science and Risk Analysis Workshop: February 22–24, 2011. www.epa.gov/ncer/publications/workshop/04_07_2011/djohns_cac_110408.pdf.

ABBREVIATIONS AND OTHER TERMS

NAAQS	National Ambient Air Quality Standard
NO ₂	nitrogen dioxide
NRC	National Research Council
PM	particulate matter
RFA	request for applications
U.S. EPA	U.S. Environmental Protection Agency

HEI STATEMENT

Synopsis of Research Report 183, Part 3

Modeling of Multipollutant Profiles and Spatially Varying Health Effects

INTRODUCTION

Although it is clear that people are exposed to complex mixtures of pollutants emitted by diverse sources of air pollution, air quality standards worldwide are geared toward control of individual, or small sets of, pollutants. Consequently, most epidemiologic studies of air pollution and health to date have focused on estimating the adverse effects associated with ambient exposure in single-pollutant models. Employing multipollutant models using conventional statistical approaches frequently produces results that are difficult to interpret because air pollutant levels are often highly correlated. Therefore, advanced statistical methods are needed to investigate the health effects of air pollution mixtures.

HEI issued Request for Applications (RFA) 09-1, “Methods to Investigate the Effects of Multiple Air Pollution Constituents,” to solicit research proposals that would address these methodologic challenges through the development of innovative statistical methods. Three studies were funded under RFA 09-1 that represent a variety of statistical approaches and applications. The studies by Dr. Brent Coull and Dr. Eun Sug Park and their colleagues are described in Research Report 183, Parts 1 and 2. For the current study, Dr. John Molitor and colleagues proposed to develop and apply statistical methods to examine associations between spatial patterns of correlated air pollutants and outcomes of health and poverty.

APPROACH

The investigators built on their previous work to develop Bayesian clustering methods to identify spatial clusters of air pollution exposures — and of other covariates such as socioeconomic status — and to estimate the association of health outcomes

with those clusters. They use the term *profile* to define a set of pollutants (or more generally exposures). Their approach has three components: a prior for cluster allocation, a profile assignment submodel to assign exposure profiles to clusters, and a health effects submodel to link clusters of exposure profiles to the health outcome. The Bayesian models described by Molitor and colleagues are mostly fit using Markov chain Monte Carlo techniques. Their Bayesian framework allows a supervised (joint) estimation (meaning that they allowed the relationship between health outcomes and exposures to inform the formation of the clusters).

What This Study Adds

- Advanced statistical methods are needed to investigate health effects of air pollution mixtures. Molitor and colleagues extended their cluster methods to include continuous exposures and successfully implemented them to analyze multipollutant mixtures.
- Their approach was aimed at identifying spatial clusters of air pollution exposures — and other covariates such as socioeconomic status — and estimating health outcomes associated with those clusters. The approach is flexible, for example, the number of clusters does not need to be predefined, and uncertainty related to cluster allocation is accounted for.
- Future work is necessary to fully evaluate the methods, including simulation studies, comparison to traditional statistical methods, application in other settings, and inclusion of more pollutants.

Research Report 183, Part 3

An important feature of these clustering methods is that they are flexible. For example, the number of clusters does not need to be predefined. In addition, these methods quantify the uncertainty related to the clustering allocation and propagate it in the health analyses. To group exposure profiles into clusters, Molitor and colleagues used Dirichlet-process mixture modeling techniques and combined the resulting clusters with multilevel regression models to estimate health outcomes. Subsequently, they developed postprocessing Bayesian model-averaging techniques to find clusters that best represent the data and to assess uncertainty in the cluster allocation.

The investigators conducted analyses using three applications to demonstrate these methods on measures of poverty and adverse birth outcomes in Los Angeles County using census and birth certificate data.

A maximum of four pollutants were considered, including PM_{2.5} and NO₂.

MAIN RESULTS AND INTERPRETATION

In its independent review of the study, the HEI Review Committee concluded that the investigators extended their cluster methods to include continuous exposures and successfully implemented these methods to analyze multipollutant mixtures. Their analyses demonstrated that their approach can be applied to real-world data sets and that they produced results that were largely concordant with a priori expectations. Results indicate that the effects of pollutants, as well as socioeconomic status variables, vary spatially and that they vary in a complex interconnected manner. The Committee thought the difficult subject matter was made much more accessible through the investigators' approach to presenting their results. For example, the Committee liked the spatially-varying maps of effects, which they believe are a useful and effective tool to communicate results.

The Committee appreciated the flexibility of the clustering approach. The explicit inclusion of spatially-varying contextual factors (e.g., socioeconomic status variables) as inputs to the clusters, in a way similar to the treatment of air pollutants, was considered

unique and can potentially provide new insight into understanding vulnerable and susceptible populations.

The methods developed by Molitor and colleagues are complex. The investigators have put their models in a unified Bayesian framework as one way to allow a supervised (joint) estimation. In general, there are several important practical features of supervised modeling approaches that are worth considering. For example, there is a potential for feedback due to unbalanced data and misspecification of the models. The clusters identified are dependent on the health outcome, and changing the health outcome will generally change the definition of the clusters to some extent.

In addition, typically, they are computationally demanding.

The Committee noted that effects of the various data simplifications were not studied, such as the aggregation of exposure from the individual to the census tract or census block group level.

Finally, the Committee thought it would have been worthwhile to understand how the methods perform under known conditions and to compare the methods to traditional statistical methods for which the research community has already developed a deep understanding of their properties and performance.

CONCLUSIONS

Dr. Molitor and colleagues developed methods to address an important question in multipollutant research, that is, what are the combined effects of various constituents of an air pollution mixture. The Committee concluded that the multipollutant methods developed show promise, but that the full extent to which they will be useful remains to be seen. Future work is necessary to fully evaluate these methods, including simulation studies, comparison to traditional statistical methods, application in other settings, and inclusion of more pollutants. Such analyses could help to determine the degree to which these new methods will lead to a better understanding of how pollutant mixtures contribute to health effects, and ultimately, to better decisions about how to control them.

Part 3. Modeling of Multipollutant Profiles and Spatially Varying Health Effects with Applications to Indicators of Adverse Birth Outcomes

John Molitor, Eric Coker, Michael Jerrett, Beate Ritz, and Arthur Li

College of Public Health and Human Sciences, Oregon State University, Corvallis, OR (J.M., E.C.); School of Public Health, University of California–Los Angeles, CA (M.J., B.R.); Department of Information Science, City of Hope National Cancer Center, Duarte, CA (A.L.)

ABSTRACT

The highly intercorrelated nature of air pollutants makes it difficult to examine their combined effects on health. As such, epidemiological studies have traditionally focused on single-pollutant models that use regression-based techniques to examine the marginal association between a pollutant and a health outcome. These relatively simple, additive models are useful for discerning the effect of a single pollutant on a health outcome with all other pollutants held to fixed values. However, pollutants occur in complex mixtures consisting of highly correlated combinations of individual exposures. For example, evidence for synergy among pollutants in causing health effects has been recently reviewed by Mauderly and Samet (2009). Also, studies cited in the Ozone Criteria Document (U.S. Environmental Protection Agency [U.S. EPA*] 2006) confirmed that synergisms between ozone and other pollutants have been demonstrated in laboratory studies involving humans and animals. Thus, the highly correlated nature of air pollution exposures makes marginal, single-pollutant models inadequate. This issue was raised

in a report by the National Research Council (NRC 2004), which called for a multipollutant approach to air quality management.

Here we present and apply a series of statistical approaches that treat patterns of covariates as a whole unit, stochastically grouping pollutant patterns into clusters and then using these cluster assignments as random effects in a regression model. Using this approach, the effect of a multipollutant pattern, or *profile*, is determined in a manner that takes into account the uncertainty in the clustering process. The models are set in a Bayesian framework, and in general, Markov chain Monte Carlo (MCMC) techniques (Gilks et al. 1998). For interpretation purposes, a *best clustering* is derived, and the uncertainty related to this best clustering is determined by utilizing model averaging techniques, in a manner such that consistent clustering obtained by the estimation process generally yields smaller standard errors while inconsistent clustering is generally associated with larger errors.

These multivariate methods are applied to a range of different problems related to air pollution exposures, namely an association of multipollutant profiles with indicators of poverty and to an assessment of the association between measures of various air pollutants, patterns of socioeconomic status (SES), and birth outcomes. All of these studies involve an examination of regional-level exposures, at the census tract (CT) and census block group (CBG) levels, and individual-level outcomes throughout Los Angeles (LA) County. Results indicate that effects of pollutants vary spatially and vary in a complex interconnected manner that cannot be discerned using standard additive linear models. Results obtained from these studies can be used to efficiently use limited resources to inform policies in targeting areas where air pollution reductions result in maximum health benefits.

This Investigators' Report is one part of Health Effects Institute Research Report 183, Part 3, which also includes a Critique by the Health Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. John Molitor, College of Public Health and Human Sciences, Oregon State University, 157 Milam Hall, Corvallis, OR 97331-6406; e-mail: John.Molitor@oregonstate.edu.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR-83467701 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

* A list of abbreviations and other terms appears at the end of the Investigators' Report.

INTRODUCTION

Air pollution studies traditionally rely on standard regression-based methods that adjust for individual-level confounders such as indicators of SES and then examine the effect of one or more pollutants on a health outcome using a linear additive model (Kutner et al. 2004). The air pollution effect is generally measured linearly, with a regression coefficient that indicates the change in the outcome of interest (e.g., log odds of preterm birth) for every unit change in exposure to the air pollutant in question. These traditional approaches typically estimate a single global effect of air pollution, that is, the effect of a change in the exposure to an air pollutant on the health outcome, assuming that such effect is constant throughout the entire region in question. Such an analysis, while standard in the literature, is inadequate as effects of air pollution vary spatially, often affecting vulnerable subpopulations more than wealthy ones, as aspects of neighborhoods such as heavily trafficked roadways, food access, crime rates, and poorer access to health centers may make these subpopulations more vulnerable to the harmful effects of individual pollutants or to elevated levels of toxic components such as diesel exhaust. Further, pollutants occur in complex combinations, not as single-pollutant exposures. Standard regression approaches expanded to accommodate multiple pollutants with a series of interaction terms and associated regression coefficients make inference unwieldy. Teasing out the combined effect of multiple pollutants becomes cumbersome and is limited by a lack of power necessary to adequately estimate coefficients corresponding to a large set of interaction terms.

In this report, we develop and apply advanced Bayesian statistical modeling framework to examine the joint effects of multiple exposures on health. Our overall approach clusters joint patterns of air pollution exposures, denoted as an *air pollution profile*, and relates these clusters of exposures to health outcomes. The methods utilize recently developed powerful Bayesian dimension-reduction clustering techniques that characterize the pollutant patterns. The multipollutant profile approach adopts a global point of view, where inference is based on the joint pattern of pollution exposures.

SPECIFIC AIMS

The current study involved building and applying sophisticated multilevel Bayesian models to analyze air pollution profiles throughout LA County. The models were estimated using stochastic MCMC methods (Gilks et al. 1998) and were developed in a manner that made them

applicable to a wide variety of settings where investigators are interested in examining associations between patterns of covariates (e.g., air pollution, SES) and health outcomes (e.g., measures of adverse birth outcomes). Our aims are as follows:

1. Further develop and apply sophisticated Bayesian clustering methods based on well-established Dirichlet-process techniques (Dahl 2006; Escobar 1994; Neal 2000) to characterize profiles of pollutants or other patterns of covariates relevant to the study of health effects.
2. Associate multipollutant profiles found in Aim 1 to health outcomes using cluster assignments as random effects in a regression model. This regression approach allows for adjustment of relevant confounders, thus enabling researchers to examine the residual effect of pollutant patterns on health after taking into account variables related to risk factors such as smoking or SES. This approach will properly take into account uncertainty in the clustering process.
3. Utilize ways to assess a *best* clustering and then assess the uncertainty related to this best clustering by model averaging (Raftery et al. 2003) through clustering obtained via the stochastic estimation process. As a result of this model averaging, consistent clustering results in lower standard errors, and haphazard inconsistent clustering is associated with higher errors. This approach represents a compromise between highly interpretable *hard clustering* methods that obtain a single best clustering and stochastic *fuzzy clustering* approaches that produce a less interpretable, though perhaps more accurate, output consisting of potentially thousands of different clusterings.
4. Analyze real data sets assessing associations between air pollution exposure profiles and poverty, and analyze associations between multipollutant profiles with measures of adverse birth outcomes, all in LA County. The results obtained are examples of how to use the developed methodologies, but they contain substantively relevant findings suitable for publication in applied health-related journals.

METHODS AND STUDY DESIGN

As part of this project we enhanced previous work to develop an advanced Bayesian statistical modeling framework to examine the joint effects of multiple exposures on health. Our overall approach was to cluster joint patterns of air pollution exposures, denoted as an air pollution profile, and potentially relate these clusters of exposures to health outcomes. The methods developed would then

utilize recently developed powerful Bayesian dimension-reduction clustering techniques to characterize the pollutant patterns. The multipollutant profile approach adopted a joint point of view, where inference was based on the joint pattern of pollution exposures. The methodology consisted of the following key components:

1. A **profile assignment submodel**, which assigns single multipollutant profiles to clusters. We utilized advanced Dirichlet-process mixture-modeling techniques to group multivariate exposure profiles into clusters, allowing the number of clusters to vary. Further, postprocessing techniques were developed to find clusters that best represented the data, and Bayesian model averaging modeling techniques were used to assess uncertainty and estimate relevant cluster-specific parameters.
2. A **health effects submodel**, which links clusters of exposure profiles to a health outcome of interest via a regression model. Thus the exposure clusters were, in some cases, informed also by the health effects. The health effects submodel can be used to cluster covariates on multiple domains, so, for example, indoor exposures can be clustered separately from outdoor exposures, or confounders can be clustered into risk groups that are separate from the risk groups formed by the air pollution variables. In the multiple-domains model, all domains contributed to explain the health effects through easier to interpret domain-specific clusters.
3. A **prior for cluster allocation** was developed that allowed patterns of similar exposure levels to be grouped in a manner that generally resulted in spatially contiguous clusters. In our application, we utilized a spatial stick-breaking construct that works well with the flexible Dirichlet-process mixture-modeling framework proposed.

We formulated the model in a Bayesian context, and, ideally, all components of the modeling framework, including the health effects submodel, would be fitted jointly using MCMC methods (Gilks et al. 1998). The Bayesian clustering aspect of the proposed modeling framework has a number of advantages over traditional clustering approaches in that it allows the number of clusters to vary, allows comparison of different clusters of profiles, can incorporate a priori known structures as well as separate exposure domains, uncovers clusters based on their association with an outcome of interest, and fits the model as a unit, allowing an individual's health outcome to influence cluster membership. Some of these methodologies have been previously developed where the method was illustrated on an analysis of epidemiological profiles

using data from a children's health survey, and these profiles were used to predict the mental health status of the child (Molitor et al. 2010).

More technically, we first constructed an allocation submodel of the probability that an area is assigned to a particular cluster. The basic model we used to cluster profiles is a standard discrete mixture model, the kind described in Jain and Radford (2004) and Neal (2000). Mixture models have been applied to a wide range of applications such as classification and density-estimation problems (Everitt 1984; Everitt and Hand 1981; McLachlan and Basford 1988) and latent class analysis (Lazarsfeld and Henry 1968). Our mixture model incorporated a Dirichlet-process prior on the mixing distribution. (For further background information regarding mixture models with Dirichlet-process priors, see Green and Richardson 2001.) Mathematically, we denoted for each area, i , covariate profiles with the form $\mathbf{x}_i = (x_1, x_2, \dots, x_p)$, where each covariate, x_p , $p = 1, \dots, P$, within each multipollutant profile denotes a measure of exposure for pollutant p in basic experimental unit, i . (Note that the basic experimental unit may be a person in some examples, or a residential area in other applications.) Profiles are grouped into clusters, c , and an allocation variable, $z_i = c$ indicates the c^{th} cluster to which area, i , belongs, where C denotes the maximum number of clusters.

CONTINUOUS EXPOSURES

Air pollution exposures will generally be characterized using continuous distributions, such as (log)normal. In this case our basic mixture model for assignment is then

$$\begin{aligned} \Pr(\mathbf{x}_i) &= \sum_{c=1}^C \Pr(z_c = c) \prod_{p=1}^P f(x_{ip} | z_c = c) \\ &= \sum_{c=1}^C \psi_c \prod_{p=1}^P f(x_{ip} | \mu_{z_i}^p, \sigma_{z_i}^p), \end{aligned} \quad (1)$$

where $f(x_{ip} | \mu_{z_i}^p, \sigma_{z_i}^p)$ will typically denote a normal or a lognormal distribution with location and scale parameters μ_c and σ_c . Using this setup, we are assuming, as is done in latent class analysis (McHugh 1956), that exposures are conditionally independent given cluster assignment. Unconditionally, they are of course dependent, as a profile's overall covariate pattern will affect the cluster to which the profile is assigned, and thus the probability that a particular covariate takes on a certain value. However, this conditional independence assumption can be relaxed by simply specifying a multivariate normal distribution for $f(\mathbf{x}_i)$, $p = 1, \dots, P$, as, $f(\mathbf{x}_i | \boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i})$, where $\boldsymbol{\Sigma}_{z_i}$ denotes a covariance matrix.

The mixture weights corresponding to a maximum of C clusters, denoted as $\boldsymbol{\psi} = (\psi_c, c = 1, \dots, C)$. We place a stick-breaking prior (Green and Richardson 2001; Ishwaran and James 2001) on the mixture weights, $\boldsymbol{\psi}$, using the following construction. We define a series of independent random variables, V_1, V_2, \dots, V_{C-1} , each having distribution $V_c \sim \text{Beta}(1, \alpha)$. This generative process is referred to as a stick-breaking formulation since one can think of V_1 as representing the first portion broken from a stick of length 1, leaving a remainder of $(1 - V_1)$, and then a portion V_2 being broken off, leaving a remainder of $(1 - V_1)(1 - V_2)$, and so on. Then, we specify prior cluster assignment probabilities as $\psi_c = V_c \prod_{p < c} (1 - V_p)$, $1 \leq c \leq C - 1$. Since we have little a priori information regarding the specification of α , we place a noninformative uniform prior on this parameter. This parameter is important, since it determines the degree of clustering that takes place, and we want this to be driven by the data as opposed to prior beliefs.

Note that we set the maximum number of clusters, C , at 20, but allowed clusters to be empty. As such we have approximated the standard Dirichlet-process infinite cluster model with a finite one (see, for example, Ohlssen et al. 2007). The advantage of this construction is that it can be easily coded in standard Bayesian modeling software, such as WinBUGS (Spiegelhalter et al. 2003).

CATEGORICAL EXPOSURES

Exposures may be irregularly distributed, for example, exposure data may contain some extremely high or low exposure values, and we may therefore choose to categorize the exposures. In this case, we will have M_p categories for the p^{th} exposure. We denote with ψ_c the probability of assignment to the c^{th} cluster and let $\phi_c^p(x)$ denote the probability that the p^{th} covariate in cluster c is equal to x . In other words, for each cluster, c , the parameters, ϕ_c^p , $p = 1, \dots, P$ define the prototypical profile for that cluster. Our basic mixture model for assignment is

$$\begin{aligned} \mathbf{x}_i &= \sum_{c=1}^C \Pr(z_c = c) \prod_{p=1}^P f(x_p | z_c = c) \\ &= \sum_{c=1}^C \psi_c \prod_{p=1}^P \phi_{z_i}^p(x_{ip}). \end{aligned} \quad (2)$$

COMBINATIONS OF CATEGORICAL AND CONTINUOUS EXPOSURES

In some instances, it will be desirable to model certain exposures categorically and other exposures using continuous distributions. In this case we can model categorical exposures with latent underlying continuous distribution, with threshold parameters used to determine the appropriate exposure category. This latent-variable approach to modeling categorical and continuous data has been used successfully in other contexts. (See, for example, Albert and Chib 1993.)

FINDING CLUSTERING THAT BEST FITS THE DATA

One important aspect of our flexible Bayesian modeling framework is that our model implementation allows the number of air pollution exposure clusters to change from iteration to iteration of the MCMC sampler, and this added flexibility leads to a rich output that requires careful interpretation. Therefore, we have developed procedures to process the output of our method to make useful, interpretable inference.

There are two main areas of interest, namely: (1) find the partition (grouping of exposures) that is most supported by the data, and (2) assess uncertainty associated with clusters of this best partition in a manner which exploits the entire MCMC output of the sampler. Regarding the former, we wish to find the general, typical, or best way in which the stochastic algorithm groups profiles into clusters based on multidimensional exposure patterns. This problem has been addressed in the literature by many authors in the context of mixture models (see, for example, Dahl 2006). The technique involves constructing a probability matrix with cells indicating the percentage of time that two exposure profiles fall into the same cluster, and then examining which *partition* of exposure profiles obtained at each iteration of the sampler best matches this probability matrix. This best partition will often have important substantive meaning as it indicates which exposure patterns tend to group together.

Given this typical or best grouping of exposure patterns, we wish to examine the uncertainty associated with exposure clusters in this best partition in a manner that exploits the entire MCMC output of the sampler by exploiting modern Bayesian model-averaging techniques. This is important, since every data set (including *noisy* ones) will exhibit a best clustering, and it is important to examine, with proper consideration for uncertainty, the characteristics associated with the exposure clusters present in any chosen partition of the data set. Therefore, at each iteration of the sampler, we average exposure cluster parameters for all exposure profiles within a particular cluster in the best

partition defined above. This approach yields a posterior distribution for various cluster parameters that averages over all the different clustering of profiles produced by the MCMC algorithm. These posterior distributions can then be used to assess which pollutants are most important within a particular exposure cluster, and for some models, assess the level of association between the exposure cluster and the health effect in question.

MODELING APPROACHES: DATA ANALYSIS

Our approach has been demonstrated in a variety of contexts, which has led to substantively relevant analyses associating pollution profiles and other relevant information to deprivation and health-related outcomes. As such, we present work funded by this grant into investigations covering three main applications:

1. Associations between air pollution profiles and poverty;
2. Spatially varying effects of $PM_{2.5}$ on term low birth weight (TLBW) (see Appendix A); and
3. A place-based approach to examining associations between multipollutant exposure profiles and TLBW.

We chose to put the work corresponding to application 2 in Appendix A, as it represents single-pollutant modeling but lays the framework for what is done in application 3. We have chosen to disseminate all of this work to peer-reviewed scientific journals. The material presented below consists of work that has already been published, or is currently under review.

ASSOCIATIONS BETWEEN AIR POLLUTION PROFILES AND POVERTY IN LOS ANGELES COUNTY

Introduction

Growing health disparities exist in the United States between people with high SES and people who have a lower SES or among people of color. These disparities translate into higher rates for mortality, morbidity, and disability for the lower SES groups and for people of color (Brulle and Pellow 2006). Although these health disparities are frequently attributed to individual health behaviors such as smoking, individual factors account for only a fraction of the overall inequalities between these groups (Lantz et al. 2001). As such, social epidemiology research has focused on the effects of SES on many health outcomes (Kim and Durden 2007), on differential access to health and social services (Shumka and Benoit 2007), and on neighborhood or community characteristics that may promote or adversely

affect health (Do et al. 2008). Researchers and policy makers concerned about environmental justice also investigate whether disadvantaged groups experience higher environmental exposures. These studies generally report that areas with a greater proportion of residents with incomes near or below the poverty line, or who are not white, face higher single and cumulative environmental exposures (Brulle and Pellow 2006; Marshall 2008; Morello-Frosch et al. 2001, 2002, 2011a). Environmental justice researchers argue that such socially disadvantaged groups bear a greater environmental exposure burden and are more susceptible to the effects of these exposures due to factors such as psychosocial stressors, underlying health conditions, and occupational exposures (Institute of Medicine 1999). These disparities in environmental exposures are increasingly recognized as potential determinants of health inequities (Finkelstein and Jerrett 2007; Morello-Frosch and Jesdale 2006).

Methods

Here, we applied our air pollution profile approach that examines associations between several components of air pollution exposures and poverty. Consistent with the methods described previously, we addressed these problems by using, as its basic unit of inference, a profile consisting of a joint pattern of air pollution exposure values. These profiles were grouped into clusters and associated with poverty levels in LA County census tracts (CT), which are small, relatively stable geographic areas for which the U.S. Census Bureau publishes sample data; a CT usually has a population between 2500 and 8000 people. The multipollutant profiles examined consisted of regional estimates of air pollution concentrations for nitrogen dioxide (NO_2), particulate matter $\leq 2.5 \mu m$ in aerodynamic diameter ($PM_{2.5}$), diesel on-road exposure, and diesel off-road exposure. Briefly, NO_2 estimates were obtained via a land use regression (LUR). The variables used for NO_2 model selection included land use information (e.g., commercial and industrial), road network, traffic, population distribution, physical properties, and remote sensing-derived greenness and soil brightness. Methods are described in detail by Su and colleagues (2009a,b). The model was based on field measurements at 201 locations in LA. The measurements were obtained in two seasons during summer 2006 and winter 2007, each for a two-week period representing the seasonal mean. These measurements were averaged to represent the annual mean. Some monitors were stolen or vandalized, leaving 181 sites for the analysis. Sixteen measurements were chosen at random to use as cross-validation sites. The measurements from the remaining 167 sites then were

used as the dependent variables in a spatial LUR model with traffic, land use, population, and physical geography as predictors of pollution concentrations. The model was highly predictive; the R^2 between the field measured and predicted pollutant concentration was 86% with similar performance at the out-of-sample cross-validation sites predictions. To estimate PM_{2.5} exposure, we interpolated from 23 state and local district monitoring stations in the LA basin for year 2000 with a universal kriging algorithm (Jerrett et al. 2005).

On-road and off-road diesel exposures were estimated through the Assessment System for Population Exposure Nationwide, a model known as ASPEN (Version 1.1), with a Gaussian dispersion that accounted for meteorological conditions, wind speed, and atmospheric chemistry (U.S. EPA 2000).

All exposure concentrations were obtained at the CT level. The number of people living below the poverty level were obtained for each CT from the U.S. Census Bureau for the year 2000.

While the air pollution profile approach has already been described previously, the specific implementation is described below.

Exposure Profile Assignment Submodel Our basic data structure consists of, for each CT, i , a covariate profile, $\mathbf{x}_i = (x_1, x_2, \dots, x_P)$, where each covariate, x_p , $p = 1, \dots, P$, within each multipollutant profile denotes a measure of exposure for pollutant p in area i . We first construct an allocation submodel of the probability that any given area is assigned to a particular cluster. Profiles of areas are grouped into clusters, and an allocation variable, $z_i = c$, indicates the c^{th} cluster to which area, i , belongs. Our assignment submodel is then

$$f(\mathbf{x}_i) = \sum_{c=1}^C \psi_c f(\mathbf{x}_i | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c), \quad (3)$$

where $f(\mathbf{x}_i | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$, denotes a multivariate normal distribution with location parameters $\boldsymbol{\mu}_c = (\mu_c^1, \dots, \mu_c^P)$ and covariance matrix $\boldsymbol{\Sigma}_c$.

Because it is possible that clusters will be empty, we cannot assign noninformative, *flat*, priors to cluster parameters. Therefore, we adopt an empirical Bayes approach and assign a prior for the mean of each pollutant across clusters as $\mu_c^p \sim N(v^p, \phi^p)$, where each v^p is set to the observed empirical average, \bar{x}^p , but each ϕ^p is set equal to the square of the empirical range squared as suggested in Richardson

and Green (1997). Similarly, we assign a Wishart prior for the precision matrices as $\boldsymbol{\Sigma}_c^{-1} \sim \text{Wish}(\rho R, \rho)$, where $\rho = P$. Since under this formulation the mean of the Wishart distribution is $E(\boldsymbol{\Sigma}_c^{-1}) = R^{-1}$, we set R to the empirical variance, namely, $R = \hat{\boldsymbol{\Sigma}}$. Note in our model formulation cluster hyperparameters are assumed to come from distributions centered on empirical averages. Thus cluster-specific parameters are used to represent subgroups that deviate from a single empirically derived population.

Poverty Submodel This submodel uses the allocation variables defined for the exposure profile submodel above, namely, for each CT, i , $z_i = c$, $c = 1, \dots, C$, indicates the cluster to which individual i belongs. However, in this submodel, the c^{th} cluster is assigned a random-effect parameter that measures the cluster's influence on the outcome (on the logistic scale) denoted as θ_c . Since it is possible for a particular θ_c to be associated with an empty cluster, these parameters must be assigned a proper prior. Therefore we assign to each θ_c a proper t density function with 7 degrees of freedom and scale 2.5 as a prior, as discussed in Gelman and colleagues (2008), which corresponds to the baseline case of one-half of a success and one-half of a failure for a single binomial trial with probability $p = \text{logit}^{-1}(\theta_c)$. Our response model, which links the clusters with the number of individuals living below the poverty line, y_i for CT, i , is simply $y_i \sim \text{Bin}(n_i, p_i)$, with

$$\text{logit}(p_i) = \theta_{z_i} + \varepsilon_i, \quad (4)$$

where $\varepsilon_i \sim N(0, \sigma^2)$ represents unexplained CT-level variation in the outcome not explained by air pollution exposures, n_i indicates the number of individuals in CT, i , and p_i indicates the probability that a randomly chosen individual in CT, i , is living below the poverty line.

At each iteration of the sampler, we define, $V^0 = \text{Var}(\theta_{z_i})$, and $V^\varepsilon = \text{Var}(\varepsilon_i)$ across all regions, so we can then obtain a posterior distribution for the overall amount of variation in poverty explained by air pollution clusters versus unexplained, residual error, defined as

$$\rho = V^0 / (V^0 + V^\varepsilon). \quad (5)$$

Note that the posterior distribution for ρ is *not* based on the best clustering referred to earlier, but rather represents the ratio obtained by *model averaging* through the entire MCMC output, thus properly taking into account uncertainty regarding cluster assignment and the number of clusters used.

At each iteration of our algorithm, CTs are grouped into a relatively small number of clusters with a set of parameters associated with each cluster. However, as cluster membership and the number of clusters used changes from iteration to iteration, the iterative process will create, for each CT, a unique posterior distribution for each parameter of interest, such as air pollution risk and modeled exposure parameters (NO_2 , $\text{PM}_{2.5}$, on-road and off-road diesel). This feature of using shared cluster parameters estimated at each iteration of the model-fitting process to form unique posterior distributions at the “individual” level is well known as Bayesian partitions models. For an overview of these models see Denison and Holmes (2001).

Results

Since we are interested in the joint distribution of exposures, we examine the best clustering obtained using the profile-based Bayesian modeling approach with the number of individuals living below the poverty line per CT as the outcome. The best clustering is displayed in Figure 1, with mean values and posterior credible intervals (CIs) listed in Table 1 and graphically displayed in Figure 2. In Table 1, clusters are sorted according to poverty risk. Table 1 includes the value of $\rho = V^0/(V^0 + V^e)$ that indicates the proportion of variance explained by the air pollution clusters relative to the residual error. Clusters

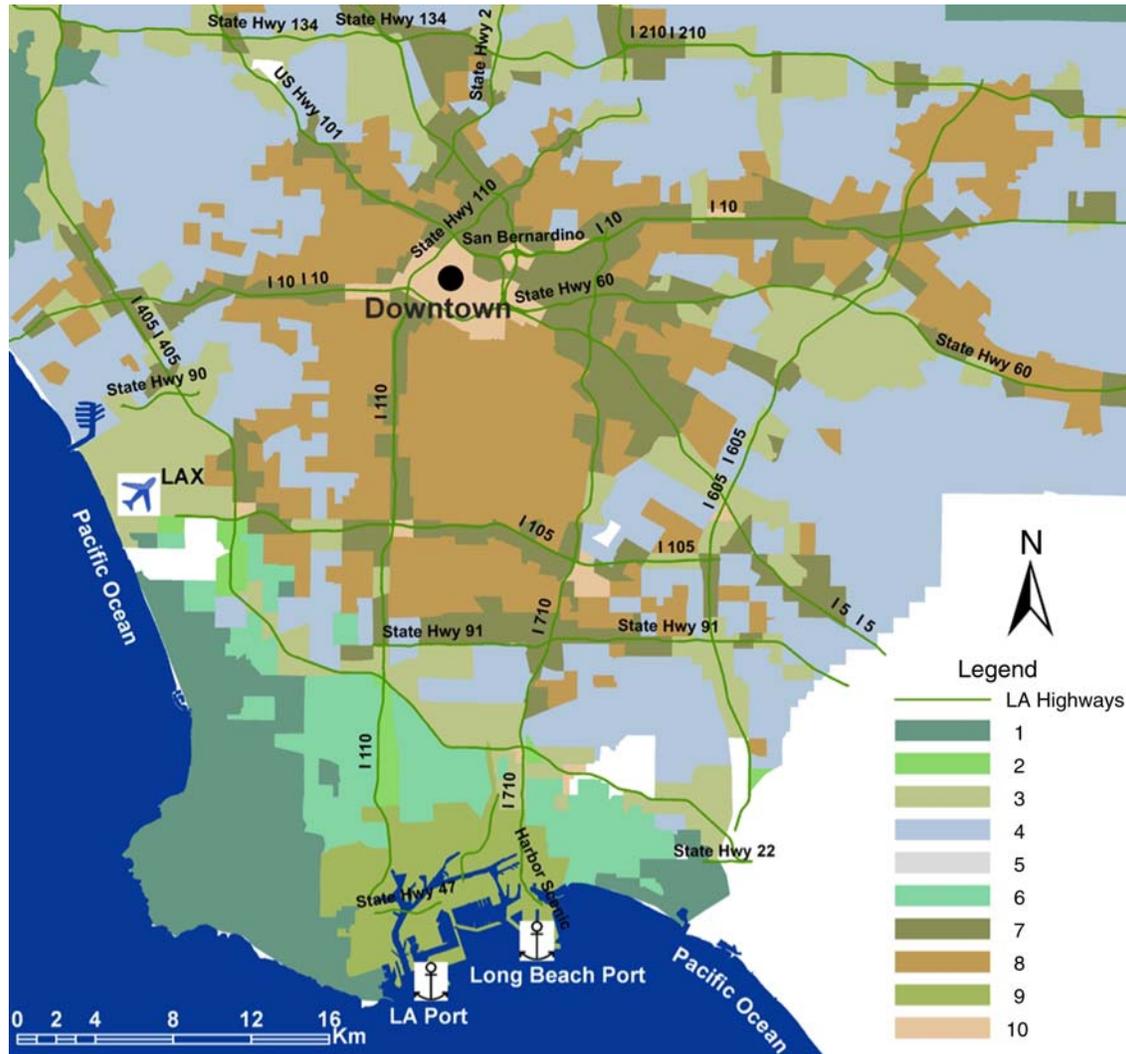


Figure 1. Best clusters as defined in Methods.

Table 1. Modeled Values for Air Pollution Clusters and Poverty Risk ^a

Cluster	n	Air Pollution Clusters ^b				Poverty Risk
		NO ₂ ppb (95% CI)	PM _{2.5} µg/m ³ (95% CI)	Road Diesel µg/m ³ (95% CI)	Off-Road Diesel µg/m ³ (95% CI)	
1	192	15.50 (14.94 to 16.13)	17.00 (16.63 to 17.39)	0.45 (0.43 to 0.49)	1.12 (1.04 to 1.20)	0.05 (0.04 to 0.06)
2	12	22.08 (19.70 to 24.19)	19.33 (18.16 to 20.34)	1.37 (0.96 to 1.81)	1.82 (1.22 to 3.06)	0.08 (0.05 to 0.13)
3	203	22.18 (21.40 to 22.90)	20.18 (19.89 to 20.45)	0.96 (0.90 to 1.01)	1.09 (1.04 to 1.16)	0.10 (0.10 to 0.11)
4	543	21.82 (21.55 to 22.09)	21.22 (21.10 to 21.33)	0.60 (0.59 to 0.62)	1.08 (1.06 to 1.10)	0.11 (0.10 to 0.11)
5	72	16.77 (15.56 to 18.02)	12.02 (10.92 to 13.27)	0.33 (0.29 to 0.38)	0.62 (0.53 to 0.73)	0.13 (0.12 to 0.16)
6	178	19.95 (19.38 to 20.61)	18.46 (18.16 to 18.77)	0.59 (0.56 to 0.65)	1.54 (1.41 to 1.67)	0.16 (0.15 to 0.18)
7	285	26.69 (26.22 to 27.14)	21.68 (21.55 to 21.81)	1.21 (1.15 to 1.26)	1.29 (1.26 to 1.33)	0.23 (0.22 to 0.25)
8	479	24.18 (23.90 to 24.47)	21.70 (21.63 to 21.77)	0.72 (0.70 to 0.74)	1.42 (1.39 to 1.46)	0.25 (0.24 to 0.26)
9	38	20.64 (19.42 to 21.84)	16.67 (16.07 to 17.30)	0.90 (0.74 to 1.08)	7.91 (6.53 to 9.27)	0.28 (0.24 to 0.32)
10	36	32.60 (30.42 to 34.81)	21.94 (21.58 to 22.31)	2.49 (2.10 to 2.89)	1.80 (1.60 to 2.03)	0.34 (0.29 to 0.38)
Overall mean		22.33	20.25	0.77	1.36	0.17

^a Percentage of poverty explained by air pollution clusters: $\rho = 0.79$ (0.47 to 0.97). **Bold** rows indicate clusters with statistically significant poverty risks.

^b 95% CI = 95% credible interval.

with statistically significant associations with poverty are displayed in Figure 3.

In Table 1, the value of $\rho = 0.79$ (0.47–0.97) reveals that variation in air pollution exposures throughout LA County coincide with variation in poverty levels. If we examine the clusters significantly associated with poverty in Figure 3, we see that populations living in the port neighborhoods of LA and Long Beach mainly suffer from nonroad diesel impacts, probably from goods movement vessels (cluster 9). Further, the roadways (cluster 7) exhibit higher than average concentrations of NO₂, PM_{2.5}, and road diesel, while the high-traffic area of downtown LA (cluster 10) exhibits higher than average concentrations of all pollutants. These results reveal that people who live in the port neighborhoods of LA and Long Beach, the main artery near those neighborhoods, the LA downtown core area, and the central areas not only suffer from poverty but also face significant pollution impacts from multiple air pollutants.

In general, the results depicted in Table 1 and Figure 2 reveal that areas with higher concentrations of air pollution exposures are associated with higher levels of poverty. However, the association between air pollutants and poverty is not entirely linear. For example, cluster 9 (LA and Long Beach ports) has a higher poverty risk than cluster 7 (roads), 0.28 (0.24–0.32) versus 0.23 (0.22–0.25). While the marginal CIs for cluster risks just barely overlap, the joint

probability that the risk for cluster 9 is greater than cluster 7 is significant as $\Pr(p_9 > p_7) = 0.99$. As one might expect, cluster 9 with its higher poverty risk has much higher concentrations of off-road diesel emissions, 7.91 µg/m³ (6.53–9.27), compared with cluster 7, 1.29 (1.26–1.33), with its lower poverty risk. However cluster 9 has lower concentrations of NO₂ (ppb), PM_{2.5} (µg/m³), and on-road diesel (µg/m³), with exposure values and CIs of 20.64 (19.42–21.84), 16.67 (16.07–17.30), and 0.90 (0.74–1.08) compared with exposure values for cluster 7 of 26.69 (26.22–27.14), 21.68 (21.55–21.81), and 1.21 (1.15–1.26). Thus, what is different between clusters 9 and 7 cannot be summarized by an additive effect of all pollutants, but is instead related to a contrast between off-road diesel and other pollutants. Similar remarks can be made for other clusters. For example, the relatively higher poverty risk cluster 10 (downtown) has higher concentrations of NO₂, on-road diesel emissions, and off-road diesel emissions, 32.60 (30.42–34.18), 2.49 (2.10–2.89), and 1.80 (1.60–2.03), compared with lower emission concentrations corresponding to the relatively lower poverty risk cluster 8 (central LA, off-roads), 24.18 (23.90, 24.47), 0.72 (0.70, 0.74), and 1.42 (1.39–1.46). The concentrations of PM_{2.5}, however, are nearly the same and not statistically different, with values of 21.94 (21.58–22.31) for cluster 10 versus 21.70 (21.63–21.77) for cluster 8, despite differences in poverty

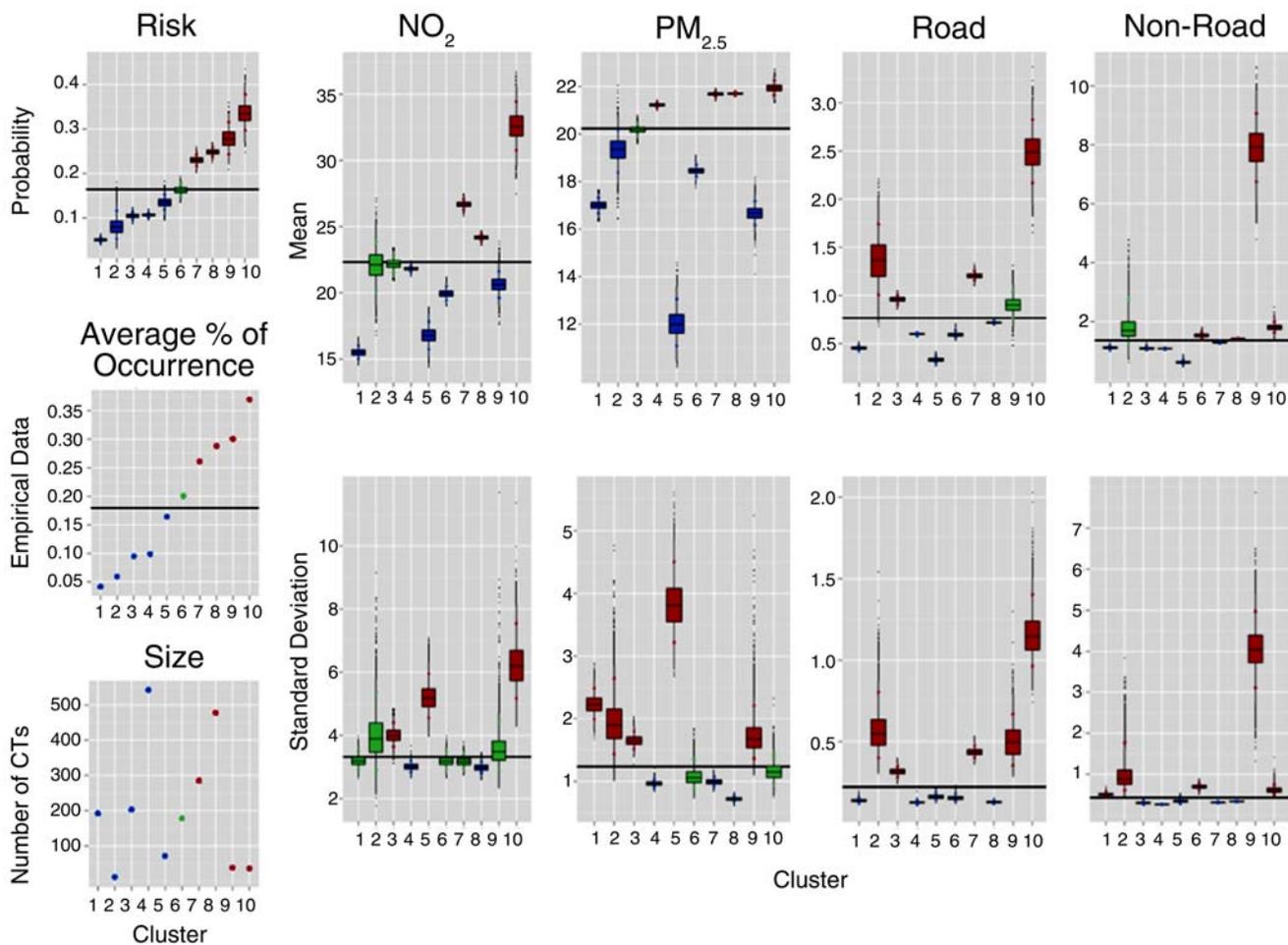


Figure 2. Best cluster summaries ($N_{clusters} = 10$). These boxplots are color coded to indicate profile cluster distributions considered high exposures (red), average exposures (green), and low exposures (blue). Black lines indicate overall averages. **Left vertical panels:** (top) poverty risk associated with each cluster, (center) average poverty risk associated with each cluster, (bottom) number of CTs in each cluster. **Right vertical panels:** (top four) cluster means for each pollutant, (bottom four) cluster standard deviations for each pollutant.

levels. Therefore, while it might be generally true that increased poverty is associated with increased air pollution exposures, the nature of these associations in LA County is complex and nonlinear.

Discussion

There has been increased interest in the air pollution literature in the examination of the combined effect of air pollution and SES. In this section of the report, we examined the joint effects of air pollution mixtures to help identify vulnerable populations in LA County. The results showed a general relationship between elevated levels of air pollution exposures and poverty. The results also revealed that the relationship is complex in that poverty levels do not increase linearly with increased levels of exposure, as is assumed when such relationships are examined using linear additive regression models. The approach employed here examined the combined effects of several air pollutants on poverty, revealing vulnerable populations were not always subject to elevated levels of different exposures uniformly, but rather different combinations of exposure levels were associated with different subgroups of poverty populations.

The approach used here clusters exposure profiles into risk groups that were then associated with poverty. The flexible MCMC-based parameter estimation techniques allowed the assignment of exposure profiles to risk groups and the number of risk groups to vary throughout the run of the sampler. The results displayed exploratory best clustering of profiles along with more robust results obtained from the model averaging through the clustering patterns obtained from the sampler. The approach identifies cumulative environmental hazard inequalities within a region at the CT level. It further extends the framework that identified cumulative environmental risks at the regional level (Su et al. 2009c).

Not surprisingly, the results often display a decidedly nonlinear pattern, as some clusters display extremely high values for one pollutant but average or below average values for other pollutants. Unlike a conventional linear model approach, the clustering approach applied here allows one to examine the manner in which pollutants vary together.

The empirical results we observed here are broadly consistent with the literature on environmental hazard inequalities in the United States (see Morello-Frosch et al. [2011a] for a recent review). Substantial evidence now suggests that numerous environmental hazards, including air pollution, are worse in poor neighborhoods and places with high proportions of racial or ethnic minority groups. In an international context, the findings from this study fit

within a fairly consistent pattern, suggesting that air pollution and other environmental risks remain unequally distributed with an inverse social gradient. Even in economically advanced countries with many income and social equalization programs, universal access to health care, and some of the highest life expectancies, air pollution and other environmental risks remain unequally distributed with an inverse social gradient. Unequal distributions of traffic pollution by race and SES have been documented, albeit with more mixed results than in the United States. Pearce and colleagues (2006) used atmospheric dispersion modeling to demonstrate a relationship between traffic pollution and disadvantaged social groups in New Zealand, finding higher levels of air pollution in areas of relatively high poverty. In England, Brainard and colleagues (2002) found that carbon monoxide (CO) and NO₂, both markers of traffic pollution, related strongly to racial and ethnic minority status and to poverty. In Sweden, Chaix and colleagues (2006) investigated the distribution of NO₂ in relation to young children. They reported higher concentrations of NO₂ for children living in poorer housing and neighborhoods. A Canadian study based on a LUR prediction of NO₂ in Toronto reported that lower SES was related to higher air pollution exposures, but there were exceptions that contrasted with the U.S. literature (Jerrett and Finkelstein 2005). For example, in Toronto racial minority groups tended to be less exposed to pollutants than were other groups, probably due to the city's role as a gateway city for highly-educated immigrants. Dwelling values also took an unexpected positive sign, which may have been partly explained by the dense urban structure of the downtown area and the relatively high traffic and land rents in this district. Similar diversions from the pattern were reported in an Italian study (Cesaroni et al. 2010). These subtle differences highlight the need to examine the specific intricacies of place, but also to employ methods used in this paper, which may elucidate more subtle patterns and relationships.

In this section, we have demonstrated the profiling method in exploring the joint distribution of poverty and several important air pollutants. Such an investigation is important, since, at present, current public health protections do not take into account these cumulative exposures and susceptibilities, which may be significant contributing factors to observed health inequalities that follow social gradients.

A PLACE-BASED APPROACH TO EXAMINING ASSOCIATIONS BETWEEN MULTIPOLLUTANT EXPOSURE PROFILES AND TLBW

Introduction

Increased prevalence of TLBW has been linked with increasing exposures to various outdoor air pollutants including NO₂, nitric oxide (NO), and PM_{2.5} (Geer 2014; Ritz and Wilhelm 2008). Yet most of this evidence has relied heavily on single-pollutant exposure modeling that does not account for multipollutant mixtures (Ritz and Wilhelm 2008). A few studies (Brauer and Tamburic 2009; Gouveia et al. 2004; Wilhelm et al. 2011a) have examined exposures to multipollutant mixtures in air as they relate to adverse birth outcomes, however, these studies are limited in their ability to infer which pollutant or combination of pollutants are most hazardous regarding associations with birth outcomes (Ritz and Wilhelm 2008), nor have they investigated the spatial patterning of multipollutant-related birth outcome risks.

A large body of evidence indicates the existence of large within-city variations in air pollution concentrations and recent research further demonstrates that pollutant concentrations and their chemical components and sources are correlated spatially within urban regions (Geer 2014; Hasheminassab et al. 2014; Houston et al. 2014; Levy et al. 2013). Moreover, air pollution in complex urban environment such as LA has been characterized by highly localized pollutant concentration gradients and complex chemical mixtures (particularly related with distances from major roadways). The variety of factors that determine such complex gradients and pollutant mixtures in urban regions include local traffic volumes and congestion, the types of fuel and engines, operating conditions of emitting sources, background ambient air pollution levels, local meteorology, chemical reactions between pollutants, types of land use, and local topographies (Austin et al. 2012; Boehmer et al. 2013; Cho et al. 2009; Zhang and Batterman 2013). The chemical species itself is also an important determinant regarding pollutant spatial dispersal and thus variations in exposure levels within urban communities. For instance, particulate matter (PM) pollution is often observed to have a more homogenous distribution over an urban area compared to nitrogen oxides (NO_x) that display a much larger degree of heterogeneity in an urban environment (Geer 2014). Thus, it can be expected that NO_x exposure gradients are more likely to be locally influenced by distance to major roadways, while PM exposure gradients are less likely to be influenced by distance from roadways compared to NO_x. However, different PM fraction sizes are likely to be dispersed differentially based on

distances from major roadways (Buonocore et al. 2009; Kuhn et al. 2005; Zwack et al. 2011).

Despite such marked differences in terms of spatial heterogeneity among PM concentrations, different PM fraction sizes, and NO_x species, there remains a large degree of correlation among the various air pollutants (Levy et al. 2013). Correlations among various outdoor air pollutants are problematic within the context of standard regression techniques when relating chronic exposures with respect to health outcomes (Mauderly et al. 2010), particularly because health effects estimates may become unstable under such circumstances. This has created a major epidemiologic impediment in our knowledge of multipollutant health effects and thus hampers our ability to disentangle whether and which pollutants act individually or together (either in synergism or antagonism) to contribute to health risks (Mauderly and Samet 2009).

Not only are outdoor air pollutants highly collinear with the potential for interaction on health effects, but the relative toxicity of individual pollutants such as fine PM may or may not differ by distance to major roadways based on differences in the physicochemical characteristics of the particulates (Buonocore et al. 2009; Cho et al. 2009; Kuhn et al. 2005; Wagner et al. 2012). Recent evidence suggests that birth outcomes may be particularly sensitive to various components and sources of PM pollution (Bell et al. 2010; Laurent et al. 2014; Wilhelm et al. 2011b). Further, the potential toxicity of various multipollutant mixtures may well be differential across an urban environment (Geer 2014).

It is also well recognized that disadvantaged socioeconomic and demographic groups in urban settings are more likely to live in closer spatial proximity to sources of major traffic-related air pollutants and off-road air pollutants (i.e., industrial), and are thus exposed to higher concentrations of outdoor air pollutants and complex urban air pollution mixtures (Geer 2014; Su et al. 2012). However, the general positive relationship that has been observed between increased sociodemographic disadvantage and higher traffic and off-road sources of outdoor air pollution exposures may be nonlinear and spatially correlated, as recent research data from LA County in this report would suggest. Also, air pollution and birth outcomes research studies have generally been limited in terms of accounting for spatial clustering of multiple neighborhood-level vulnerabilities (i.e., race-ethnicity, poverty, and housing conditions, etc.) that could confound or interact with localized multipollutant exposure-response relationships (Geer 2014; Morello-Frosch et al. 2011b; Ponce 2005). Since health research data indicates that neighborhood contextual factors can influence the spatial patterning of

adverse birth outcomes (English et al. 2003; Messer et al. 2006; Metcalfe et al. 2011), we included contextual census block group (CBG)-level variables as clustering covariates. A CBG is the smallest geographical area for which the U.S. Census Bureau publishes sample data; it usually has a population between 600 and 3000 people.

The work in this section addresses spatial patterns of multipollutant exposure profile clusters and their relative association with TLBW. This place-based approach is intended to highlight the distinct spatial patterning of multipollutant effects and how these exposures occur within the context of clustered indicators of neighborhood-level contextual factors such as SES, housing characteristics, and neighborhood race and ethnicity composition, which may confound or exacerbate relationships between multipollutant exposures and adverse health effects.

Methods

Institutional Review Board Approval Human subjects research was approved through the University of California, Los Angeles' Office of the Human Subjects Protection Program, the California Committee for the Protection of Human Subjects, and the University of Southern California's Office for the Protection of Research Subjects.

Study Population and Birth Outcomes Electronic birth certificates from the California Department of Public Health were collected for data on birth weight and individual-level covariates. Data was available for births between 1/1/1995 and 12/31/2006 in LA County ($N = 1,518,676$). We restricted the birth certificate data set to the years 2000–2006 ($N = 804,726$) to better align with our available air pollution estimates (described below). Individual data from the birth records included length of gestation, maternal age, race and ethnicity, education, total number of previous maternal births, and residential address, along with information on the infant and birth (abnormalities, birth season, gestational age at birth, birth weight, and baby's sex). The data set was restricted to singleton births without apparent abnormalities, while births with extreme gestational days (less than 140 days or greater than 320 days) or birth weight ($< 500\text{g}$ or $> 5000\text{g}$) were excluded from the analysis as such extreme values are likely attributable to recording errors. We defined TLBW as full term (≥ 259 gestation days) infants with a birth weight $< 2500\text{g}$. Geocoding of residential addresses is explained in Goldberg and colleagues (2008).

Exposure Estimation Two separate LUR models estimated individual-level exposures for $\text{PM}_{2.5}$, NO, and NO_2 (Beckerman et al. 2013ab; Su et al. 2009b). LUR exposure predictions for NO and NO_2 were based on traffic volumes,

truck routes and road networks, land use data, satellite-derived vegetation greenness and soil brightness, truck route slope gradients, and air monitoring data collected during 2-week time periods in the summer of 2006 and the winter of 2007 (Su et al. 2009b). The $\text{PM}_{2.5}$ exposure estimates were based on a LUR model that utilized long-term governmental monitoring data of $\text{PM}_{2.5}$ measurements and a combination of remote sensing data and atmospheric modeling (Beckerman et al. 2013b). A machine learning deletion/substitution technique (Beckerman et al. 2013a) assessed as many as 70 covariates to develop the final $\text{PM}_{2.5}$ LUR model, such as land use data (i.e., agricultural, barren, all developed land, high-density development, green space, water, and wetland), long-term traffic counts (1990–2001), and road networks from the year 2000 (Beckerman et al. 2013a; Jerrett et al. 2013). All of the available data from the LUR model estimates were then averaged over CBGs in order to develop geographically bounded (or place-based) air pollution exposure profiles. Data aggregation at the CBG-level of the individual estimates was performed since we were interested in assessing *between neighborhood* multipollutant exposure-related TLBW risks.

Bayesian Profile Regression We implemented the profile regression using the PReMiuM package in R (Liverani et al. 2014), which implements the approach described previously but uses an infinite cluster version of the Dirichlet-process model. Since we were interested in obtaining clustering that best fits the data for subregions within the LA County area, we utilized a feature of the PReMiuM package that excludes the outcome variable from the profile regression model (Liverani et al. 2014). We relied on *hard* clustering (Fang et al. 2011) in the sense that the best clustering assignments derived from the Bayesian averaging process were used as input variables for the multilevel random-effects model. Briefly, for each CBG, j , a covariate profile is defined as $\mathbf{x}_j = (x_1, x_2, \dots, x_p)$, where every covariate, x_p , $p = 1, \dots, P$, within each profile signifies a level of exposure for covariate p in region j . The primary model for cluster profiles was defined by a multivariate normal mixture model.

We performed three separate profile regressions in order to develop a set of three separate exposure profile clusters to be fit in the TLBW risk model. The first clustering procedure developed multipollutant-only exposure profile clusters, while the second developed contextual-only exposure profile clusters. Lastly we fit the pollutants and contextual variables jointly. The covariate pollutants for our multipollutant-only exposure profile regression included average CBG concentrations for NO_2 , NO, and $\text{PM}_{2.5}$. The contextual-only exposure profile regression included CBG-level race–ethnicity (% nonHispanic white, % nonHispanic black, and % Hispanic), median household income, and percentage of

homes older than 1950. Even though our multivariate model appropriately adjusts for individual-level maternal race–ethnicity, CBG-level racial and ethnic composition is included as a contextual covariate because some birth outcomes research indicates that area-level racial and ethnic composition may act as a contextual risk factor for TLBW — separate from an individual’s race–ethnicity (Debbink and Bader 2011). Furthermore, while our multivariate model also adjusts for individual-level SES (i.e., maternal education), we similarly used a contextual SES variable in the clustering procedure, under the same rationale that area-level SES acts as a contextual risk factor for TLBW (Grady 2006, 2011). Furthermore, the percentage of older housing for CBGs is included since disparities in housing quality, housing characteristics, or both may act as an important contextual risk factor in TLBW risk (Ghosh et al. 2013; Grady 2011). Finally, our last profile regression clustering procedure fit each of the aforementioned pollutants and contextual variables jointly, which we denote as *contextual–multipollutant exposure profile clusters*. Each of these exposure profile clusters were then analyzed as random effects in two separate multilevel risk models described in turn. Given computational challenges, we *pre-clustered* exposure profiles and then used the R-INLA (integrated nested Laplace approximations in R) package to implement the Bayesian multilevel random-effects model described in Equation 6. R-INLA estimates Bayesian posterior marginal distributions (Rue et al. 2009, 2014) without relying on computationally intensive MCMC techniques, which is important given the dimensionality of our data set. Thus, a fully joint Bayesian modeling approach was not taken in this application.

Multilevel Risk Model Our multilevel logistic regression model was set in a Bayesian framework with the multipollutant exposure profile clusters used as a random-effect variable in the regression equation, along with spatially structured and independent error terms fit as random effects. The model specification is summarized below:

$$\begin{aligned} \text{logit}(p_i) = & \alpha + V\eta' \\ & + \gamma_{k[i]}^{\text{pollutant-cluster}} \\ & + \gamma_{l[i]}^{\text{contextual-cluster}} \\ & + S_j \\ & + \varepsilon_j, \end{aligned} \quad (6)$$

where p_i denotes the logit probability of TLBW ($y_i = 1$) for individual i , $V\eta'$ represents the individual-level and CBG-level covariate fixed effects,

$$\gamma_k^{\text{pollutant-cluster}} \sim N\left(0, \sigma_{\text{pollutant-cluster}}^2\right), k = 1, \dots, 15$$

represents the random effects for the multipollutant clusters, and

$$\gamma_l^{\text{contextual-cluster}} \sim N\left(0, \sigma_{\text{contextual-cluster}}^2\right), l = 1, \dots, 15$$

denotes random effects for contextual clusters. We use the notation $k[i]$ to denote the pollutant profile group k to which individual i belongs and $l[i]$ to denote the contextual group l to which individual i belongs. Thus, each multipollutant random error term represents the variation in TLBW prevalence in the multipollutant profile clusters and likewise each contextual random error term represents the variation in TLBW in the contextual cluster.

Here S_j and ε_j denote spatial and independent residual error terms, respectively, with the restriction $\sum_j S_j = 0$ imposed for indefinability reasons. While the independent error term is defined in the standard way as $\varepsilon_j \sim N(0, \sigma^2)$, the spatial error term is defined as

$$S_j |_{k \neq j} \sim N\left(\frac{\sum_{k \neq j} w_{jk} S_k}{\sum_{k \neq j} w_{jk}}, \frac{\sigma^2}{\sum_{k \neq j} w_{jk}}\right), \quad (7)$$

where the weights $w_{j,k}$ are elements of the zero-one neighborhood adjacency matrix defined to be equal to one when CT’s i, k are adjacent and zero otherwise. This approach is an implementation of the Besag-York-Mollé model (Besag et al. 1991) and has been successfully employed in a variety of exposure–health association studies. (See, for example, Molitor et al. 2007.) Our second multilevel model employed a random effect for the joint contextual–multipollutant profile clusters using the same basic model delineated above, but we have one set of random effects, γ^{joint} .

Individual-level covariates adjusted for in our statistical model were maternal factors including age (< 20 years, 20–24 years, 25–29 years, 30–34 years, ≥ 35 years), race–ethnicity (nonHispanic white, nonHispanic black, Hispanic, Asian, and Other race), highest education level attained (< 9 years, 9–12 years, 13–15 years, and ≥ 16 years), parity, along with infant factors such as gestational days, gestational days squared, and sex (male or female).

Given the large size of our data, we preclustered exposure profiles and then used the R-INLA package to implement the Bayesian multilevel random-effects model described earlier in equation 6. R-INLA estimates Bayesian posterior marginal distributions (Rue et al. 2009, 2014)

without relying on computationally intensive MCMC techniques, which is important given the dimensionality of our data set.

Assessing Uncertainty in Multipollutant Random Effects

Since our analysis is set in a Bayesian framework, we are able to assess the uncertainty with respect to relative TLBW random-effects estimates attributable to multipollutant profiles. To assess uncertainty we calculated the posterior probabilities that a profile’s random effect was greater than zero. The specific posterior probabilities were subsequently mapped in ArcGIS to explore the spatial distribution of these random effects on TLBW risks.

Results: Separate Multipollutant and Contextual Clusters

The study population included birth weight data on 804,726 full term births between 2000–2006, with an overall TLBW prevalence of 2.07% (95% CI: 2.04–2.11, $n = 16,694$). Our study also included complete information on air pollution estimates for 899,554 individuals, with estimated individual average and interquartile range (IQR) for NO_2 , NO, and $\text{PM}_{2.5}$ of 22.49 ppb (IQR: 19.68–25.30), 21.84 ppb (IQR: 16.05–26.11), and 16.94 $\mu\text{g}/\text{m}^3$ (IQR: 15.96–18.18), respectively.

There was extensive evidence for correlation between average CBG-level data for the pollutants and contextual variables considered in our study (Figure 4). Correlation

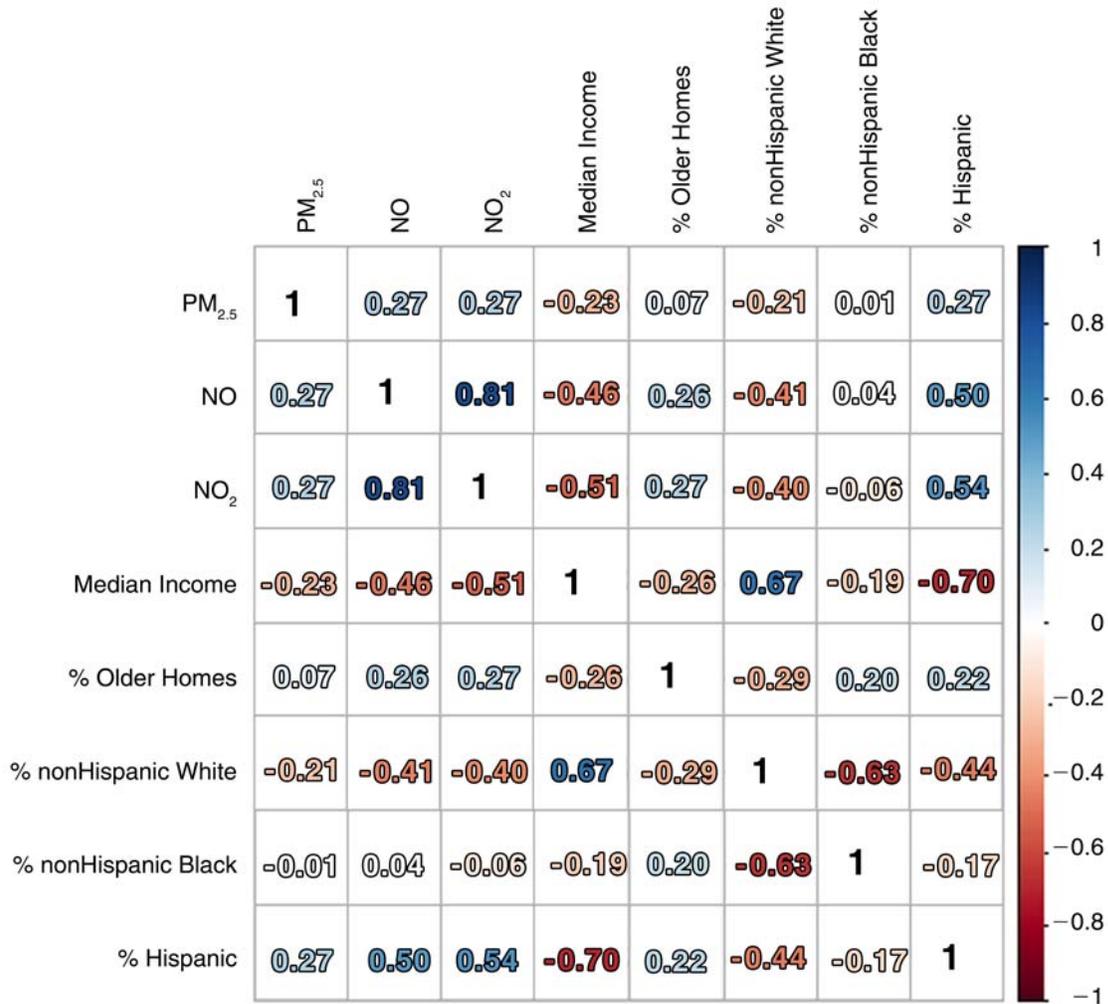


Figure 4. Correlation matrix of CBG-level air pollutants and contextual variables.

between mean CBG-level NO₂ and NO was highly positive, while positive correlations between PM_{2.5} and NO₂ and between PM_{2.5} and NO were much weaker. Correlation between CBG median income and NO₂ and NO were moderately negative, while the negative correlation between PM_{2.5} and median income was weak. The correlation between percentage of homes older than 1950 with median income was negative, while correlations between study air pollutants and older housing were all positive; however, positive correlations were much stronger between the NO_x species and older housing compared to PM_{2.5} and older housing. Percent Hispanic had the strongest positive correlations for all study pollutants, whereas percent nonHispanic white was negatively correlated with study pollutants. Percent nonHispanic black and percent Hispanic were similarly positively correlated with percentage of older homes

per CBG, whereas percent nonHispanic white was negatively correlated with percentage of older homes.

Formation of Multipollutant Clusters The multipollutant profile regression identified 15 distinct multipollutant exposure profile clusters from the 6280 CBGs for which there were complete air pollution data. Summary statistics for mean and IQR CBG-level concentrations for each pollutant are provided in Table 2, with graphical summaries of each pollutant further summarized in Figure 5A. According to Figure 5A, the high NO₂ concentration clusters (red boxplots) were cluster numbers 3, 7, 10, 11, and 15, whereas high NO concentration clusters were clusters 3, 7, 10, 11, 13, 14, and 15. The high PM_{2.5} concentration clusters were clusters 7, 8, 9, 10, 11, and 15. Just four of the 15 pollutant profile clusters exhibited high exposure concentrations for

Table 2. Summary Statistics of CBG-level Pollutant Concentrations for Multipollutant Exposure Profile Clusters

Pollutant Cluster	CBG (n)	PM _{2.5} µg/m ³ (IQR)	NO ₂ ppb (IQR)	NO ppb (IQR)
1	283	13.99 (13.01 to 15.00)	19.98 (18.53 to 21.53)	11.63 (9.80 to 13.53)
2	224	14.29 (12.78 to 14.93)	21.67 (18.99 to 24.53)	19.95 (15.58 to 24.45)
3	550	14.95 (14.65 to 15.26)	26.46 (24.67 to 28.05)	25.52 (20.79 to 28.72)
4	166	11.74 (10.29 to 13.21)	13.98 (12.96 to 13.40)	8.47 (6.91 to 10.27)
5	684	16.03 (15.43 to 16.53)	20.76 (19.72 to 21.83)	19.99 (18.66 to 21.27)
6	1289	17.06 (16.56 to 17.53)	21.96 (19.88 to 24.20)	17.27 (14.28 to 20.32)
7	579	17.56 (17.02 to 18.03)	28.55 (26.96 to 36.27)	34.35 (30.09 to 37.70)
8	358	19.29 (18.88 to 19.65)	16.74 (15.42 to 18.23)	13.17 (11.39 to 15.31)
9	223	17.49 (17.10 to 17.93)	14.81 (13.16 to 16.92)	12.90 (10.14 to 15.43)
10	1667	18.22 (17.87 to 18.53)	23.48 (21.99 to 24.97)	24.71 (22.07 to 27.37)
11	58	17.32 (16.99 to 17.48)	38.82 (35.73 to 40.38)	58.50 (52.65 to 65.57)
12	29	10.36 (9.33 to 11.01)	11.38 (6.52 to 15.29)	13.18 (7.92 to 18.60)
13	58	16.71 (16.24 to 16.99)	14.98 (14.18 to 15.84)	28.15 (22.00 to 33.17)
14	21	17.37 (16.60 to 18.19)	15.76 (14.99 to 16.72)	75.67 (64.46 to 94.39)
15	91	22.95 (22.12 to 23.91)	28.75 (27.10 to 30.37)	25.61 (23.71 to 27.85)
Overall	(N = 6280)	16.94 (15.96 to 18.18)	22.49 (19.68 to 25.30)	21.84 (16.05 to 26.11)

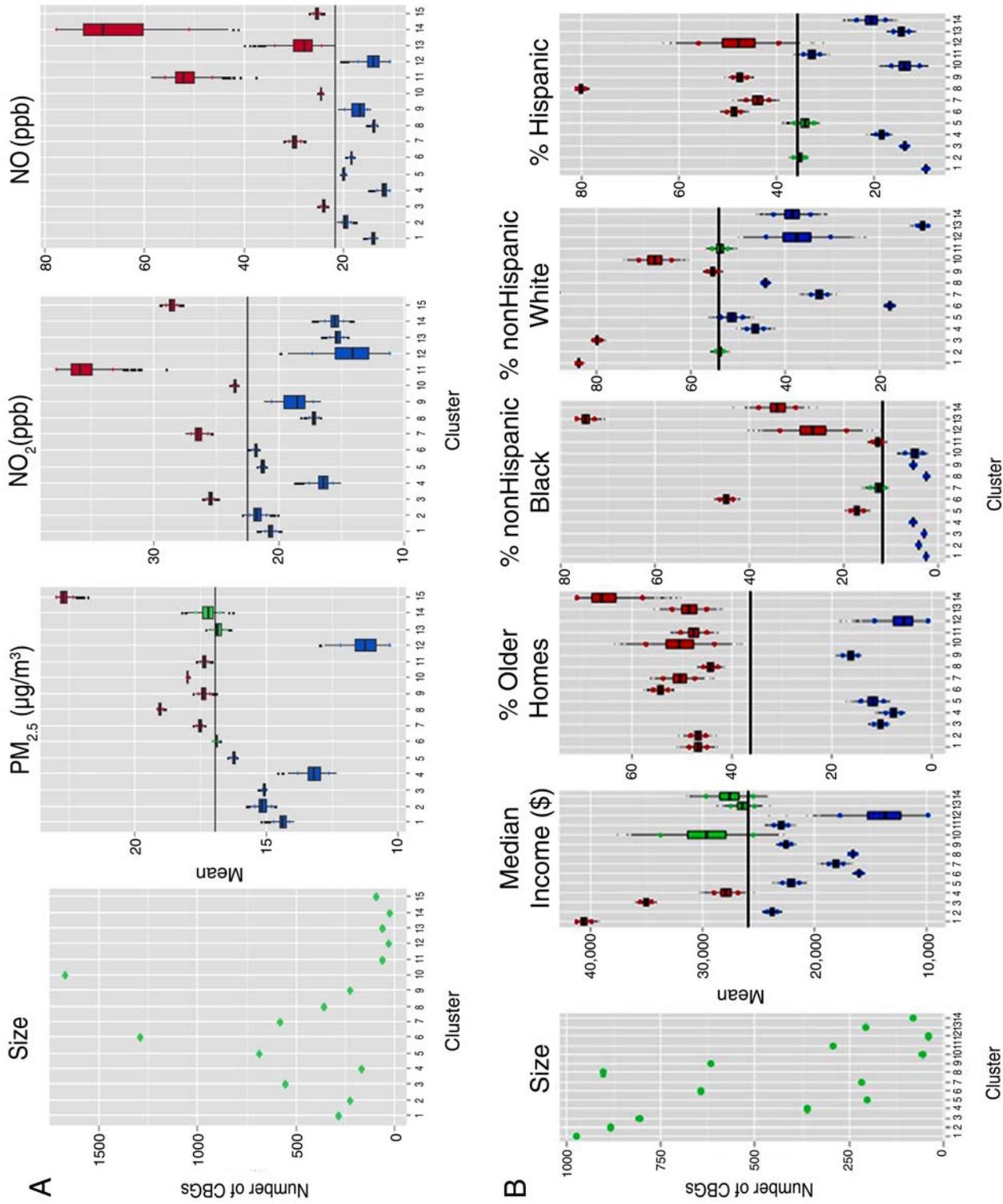


Figure 5. Exposure profiles and number of CBGs for multipollutant and contextual clusters. These boxplots are color coded to indicate profile cluster distributions considered high exposures (red), average exposures (green), and low exposures (blue). Black lines indicate overall averages. **A:** pollutant concentrations for multipollutant clusters ($N_{clusters} = 15$) and **B:** median income (\$), % older homes (older than 1950), % nonHispanic black, % nonHispanic white, and % Hispanic for contextual clusters ($N_{clusters} = 14$).

each pollutant considered, which were clusters 7, 10, 11, and 15, whereas profile clusters 1, 2, 4, 5, and 12 exhibited low concentrations of exposure for all pollutants (blue boxplots). In addition, the spatial distributions of each multipollutant profile cluster is mapped in Figure 6A. The

clusters with elevated concentrations for all pollutants were located within Central LA and South-Central LA (clusters 7, 10, and 11) and portions of East LA County (cluster 15).

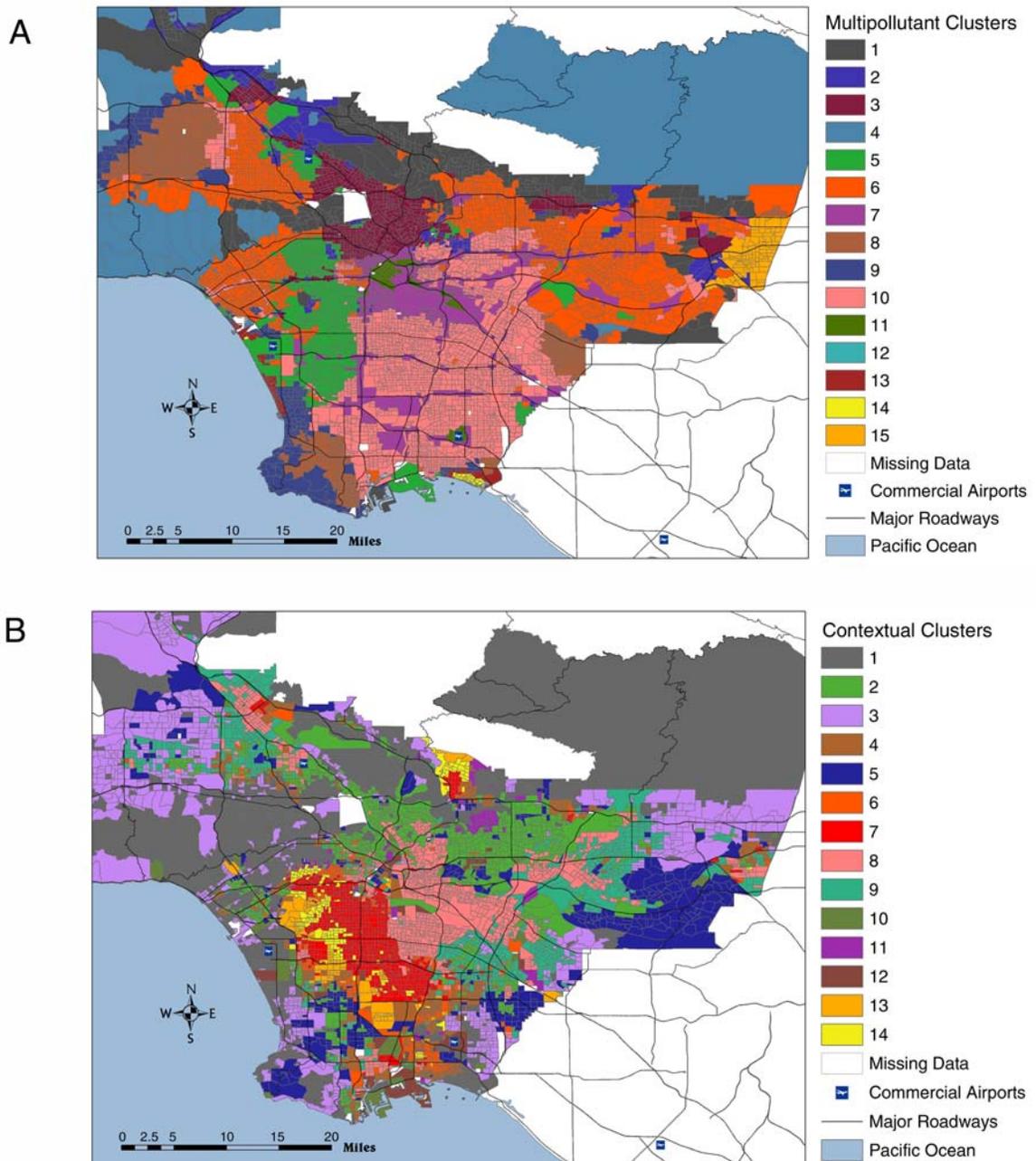
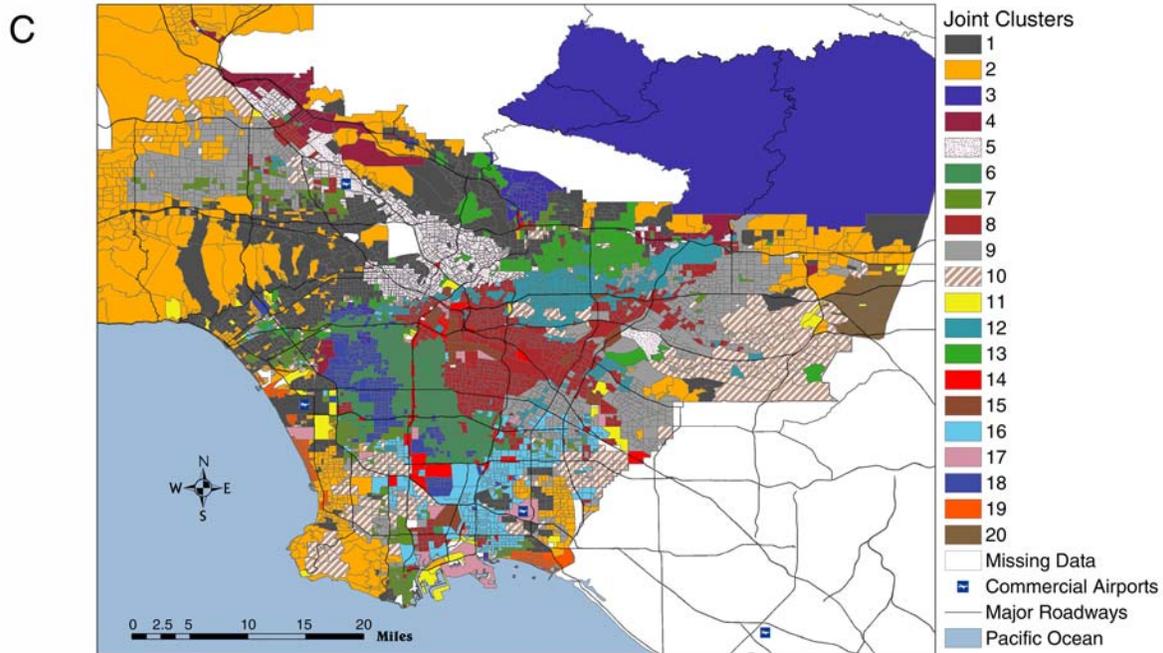


Figure 6. Spatial distributions of exposure profile clusters. A: multipollutant exposure profile clusters; B: contextual exposure profile clusters; C: joint contextual–multipollutant exposure profile clusters. (Figure continues next page.)



(Figure 6 continued from previous page.)

Formation of Contextual Clusters The contextual profile regression further identified 14 distinct exposure profile clusters. According to Figure 5B, the low income contextual clusters (blue boxplots) were cluster numbers 2, 5, 6, 7, 8, 9, 11, and 12, whereas high income clusters (red boxplots) were contextual clusters 1, 3, and 4. The high percentages of older housing stock were clusters 1, 2, 6, 7, 8, 10, 11, 13, and 14, whereas contextual clusters 3, 4, 5, 9, and 12 had low percentages of older housing stock. The clusters with elevated percentage nonHispanic black were clusters 5, 6, 11, 12, 13, and 14. For nonHispanic white we find elevated percentages within clusters 1, 3, 9, and 10. The clusters with elevated percentage Hispanic include clusters 6, 7, 8, 9, and 12. The spatial distributions of each contextual profile cluster is mapped in Figure 6B.

Regression Results

Fixed Effects Results The fixed effects for contextual factors and individual-level variables are summarized in Table 3. Individual-level covariates associated with the

odds of TLBW were sex of infant, maternal parity, age, race–ethnicity, education level, gestational days, and gestation squared. A decrease in median household income and an increase in the percentage of homes older than 1950 were each independently associated with increased odds of TLBW (data not shown).

Multipollutant Cluster Random Effects For each multipollutant cluster, the prevalence of TLBW along with the random-effect estimates and their respective probabilities for an effect above zero are presented in Table 4. According to the random effects portion of our multilevel hierarchical model, after adjusting for the individual and contextual covariates indicated above, mothers residing in cluster 7 had the highest probability (96.8%) of having a random effect above zero. Mothers from clusters 10, 11, and 15 had an 80.8%, 85.4%, and 89.4% probability of having random effects above zero, respectively. All other cluster-specific effects were characterized by probabilities below 80% for an effect above zero (Table 4).

Table 3. Fixed Effects Odds Ratios of TLBW for Contextual CBG-Level and Individual-Level Covariates ($N = 804,726$)

Individual-Level Covariates	Odds Ratio	2.5% Quantile	97.5% Quantile
Female	1.45	1.40	1.49
Parity	0.59	0.57	0.61
Maternal age			
< 20 yr	1		
20–24 yr	0.97	0.92	1.03
25–29 yr	0.91	0.85	0.96
30–34 yr	0.91	0.86	0.97
≥ 35 yr	1.07	1.00	1.14
Maternal education			
0–8 yr	1		
9–12 yr	0.90	0.86	0.95
13–15 yr	0.75	0.71	0.80
≥ 16 yr	0.67	0.62	0.71
Race and ethnicity			
nonHispanic white	1		
Hispanic	1.08	1.02	1.15
nonHispanic black	2.16	2.01	2.32
Asian	1.41	1.31	1.52
Other	1.82	1.68	1.97
Gestation (days)	0.32	0.30	0.33
Gestation-squared	1.0019	1.0018	1.002

Table 4. Prevalence of TLBW for Multipollutant Clusters and Model Results for Multipollutant Exposure Profile Cluster Random Effects ($N = 804,726$)

Cluster	Births (n)	TLBW (n)	% TLBW (95% CI) ^a	Cluster Effect Size (95% CI) ^a	Probability (Cluster Effect > 0)
1	23,946	381	1.59 (1.44 to 1.76)	−0.117 (−0.245 to 0.009)	0.033
2	35,297	623	1.77 (1.63 to 1.91)	−0.127 (−0.242 to −0.012)	0.015
3	70,534	1,400	1.98 (1.88 to 2.09)	−0.017 (−0.119 to 0.084)	0.368
4	18,364	282	1.54 (1.37 to 1.72)	−0.030 (−0.173 to 0.111)	0.336
5	73,794	1,796	2.43 (2.33 to 2.55)	0.037 (−0.063 to 0.136)	0.766
6	154,954	3,017	1.95 (1.88 to 2.02)	0.015 (−0.077 to 0.107)	0.625
7	89,022	2,100	2.35 (2.26 to 2.46)	0.092 (−0.005 to 0.188)	0.968
8	42,117	750	1.78 (1.66 to 1.91)	−0.040 (−0.153 to 0.072)	0.240
9	20,628	339	1.64 (1.48 to 1.83)	−0.027 (−0.159 to 0.103)	0.340
10	246,348	5,384	2.19 (2.13 to 2.24)	0.040 (−0.050 to 0.131)	0.808
11	3,950	96	2.43 (1.99 to 2.96)	0.099 (−0.086 to 0.281)	0.854
12	3,202	67	2.09 (1.65 to 2.65)	0.032 (−0.174 to 0.236)	0.621
13	4,296	65	1.51 (1.19 to 1.93)	0.003 (−0.199 to 0.201)	0.512
14	1,056	13	1.23 (0.72 to 2.11)	−0.039 (−0.292 to 0.211)	0.380
15	17,218	381	2.21 (2.00 to 2.44)	0.089 (−0.050 to 0.226)	0.894

^a 95% CI = 95% credible interval.

The spatial distribution of multipollutant profile cluster probabilities for random effects above zero are mapped in Figure 7A. The map indicates that multipollutant profile clusters with the highest probabilities (depicted as red CBGs) for deviations above the overall TLBW log odds are concentrated within LA County's urban core of Central LA, South-Central LA, and parts of East LA. The map clearly depicts the influence that major roadways are likely to have on the log odds of TLBW infants. With respect to exposures to multipollutant mixtures, the certainty of elevated TLBW risk is confined to CBGs within close proximity to major roadways.

Contextual Cluster Random Effects According to the random-effects results, after adjusting for the individual and contextual covariates indicated above, mothers residing in contextual cluster 7 had the highest probability (96.1%) of having a random effect above zero, while all other cluster-specific effects were characterized by probabilities below 80% for an effect above zero. The spatial distribution of contextual profile cluster probabilities are mapped in Figure 7B. This map indicates that contextual profile clusters with the highest probabilities for deviations above zero are predominantly located within the urban core of LA County, including Central and South Central LA.

Results: Joint Contextual–Multipollutant Clusters

As demonstrated earlier, there were considerably strong correlations between contextual variables and air pollution concentrations. Thus our final clustering procedure included both the contextual variables (% of homes older than 1950, median income, and % nonHispanic black, % nonHispanic white, and % Hispanic) and the three pollutant variables ($PM_{2.5}$, NO_2 , and NO) clustered together. This final clustering procedure resulted in a total of 20 contextual–pollutant clusters, which are summarized graphically in Figure 8. These joint contextual–multipollutant profile clusters were analyzed in a separate multilevel model fit as random effects as described earlier, and results are shown in Table 5. The cluster map of these joint clusters (Figure 6C) and associated probabilities (Figure 7C) further displays the spatial distribution of the random effects for the joint contextual–pollutant clusters. This map indicates that when contextual and multipollutant variables are clustered jointly, the spatial pattern of TLBW risk changes substantively. CBGs with TLBW-related cluster-specific risks are similarly localized within Central LA, South Central LA, and East LA, and the high risk clusters expand into neighborhoods surrounding the ports of LA and Long Beach.

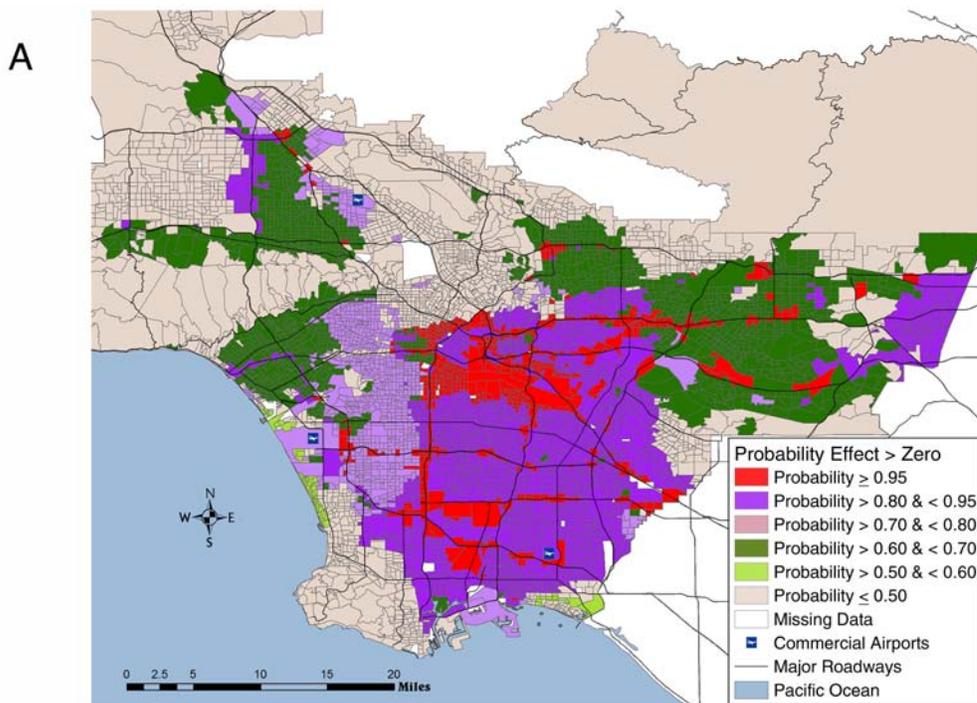
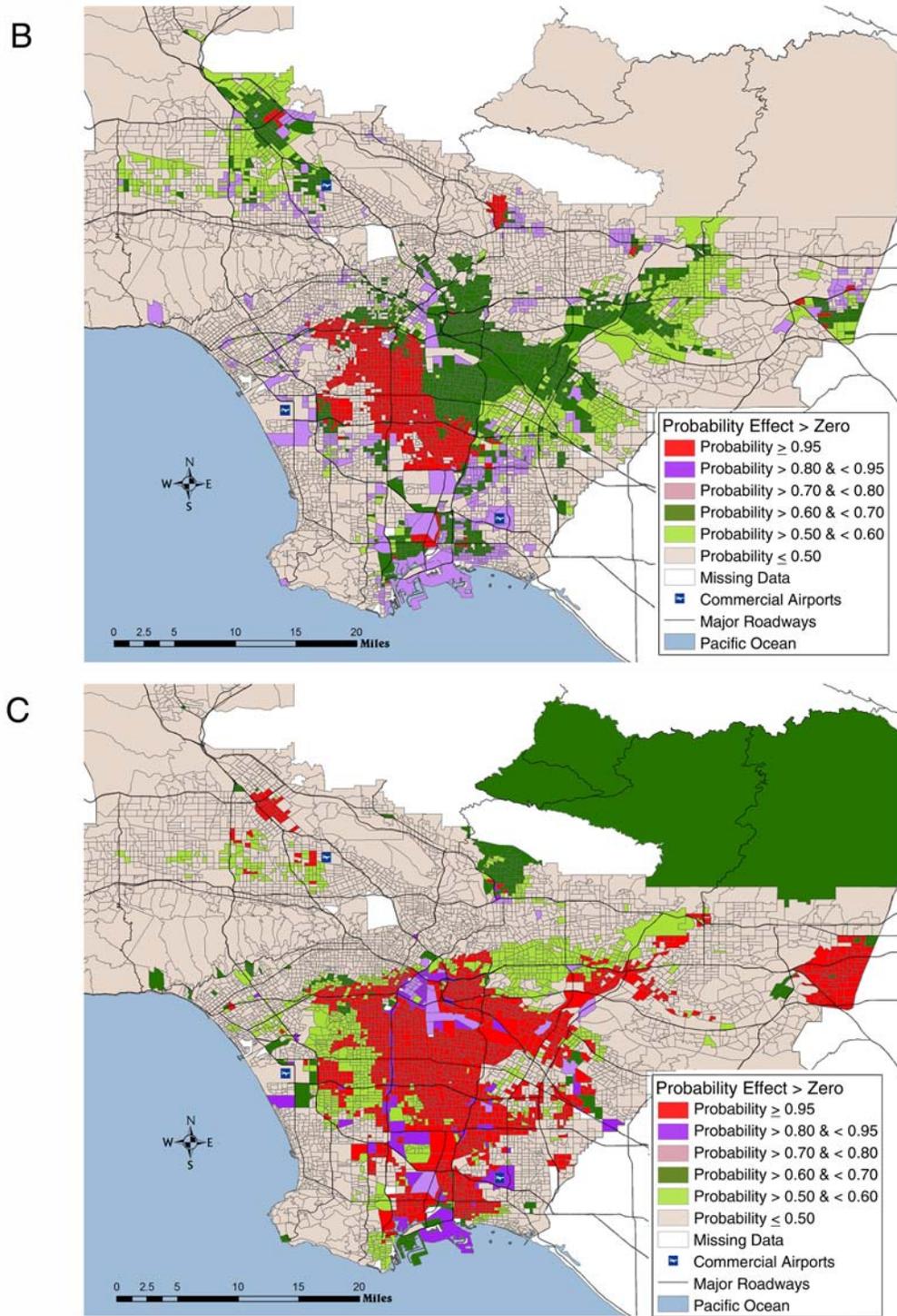


Figure 7. Spatial distribution of TLBW probabilities that random effect deviates from zero. **A:** multipollutant random effects posterior probabilities; **B:** contextual random effects posterior probabilities; **C:** joint contextual–multipollutant random effects posterior probabilities. (*Figure continues next page.*)



(Figure 7 continued from previous page.)

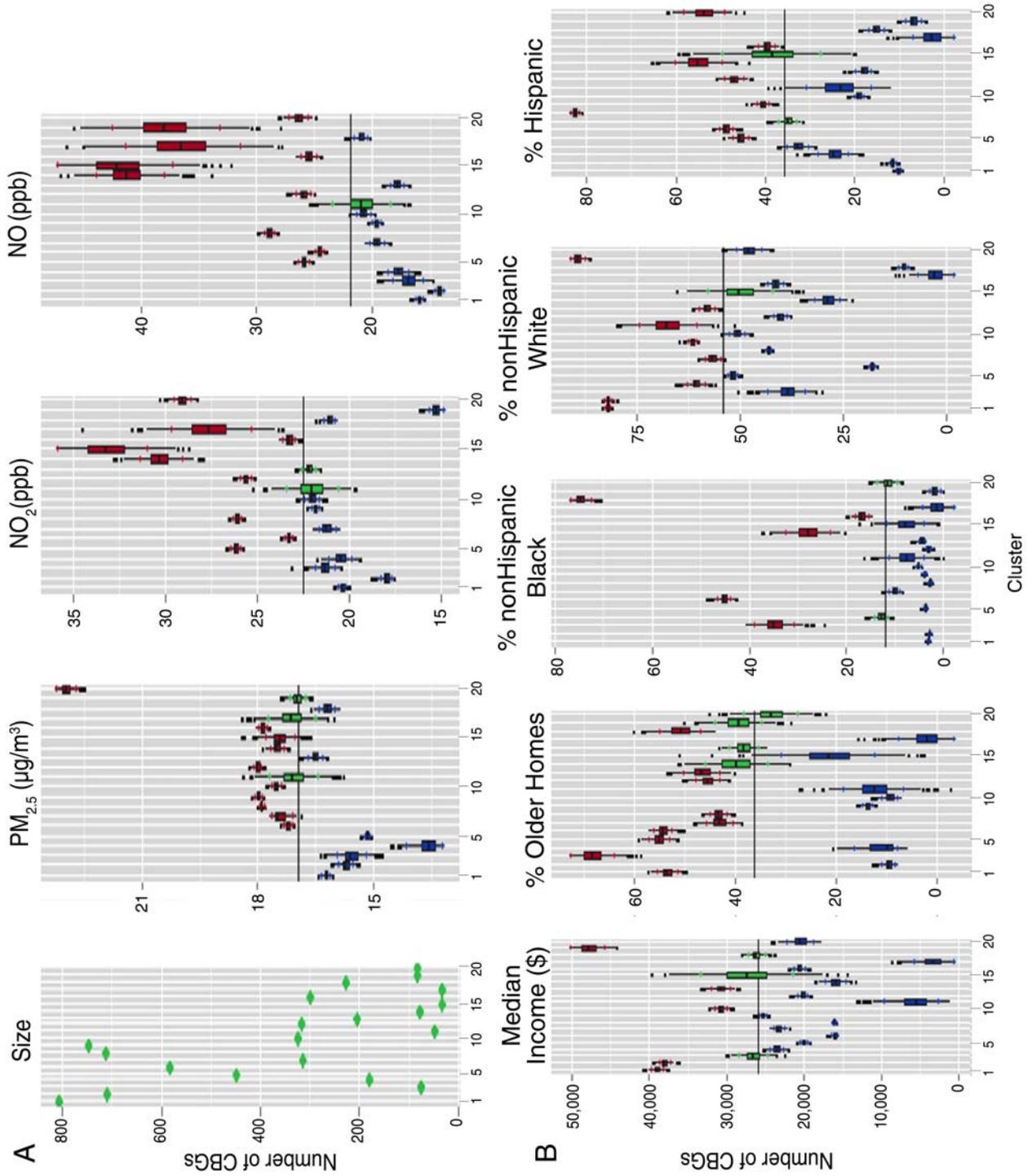


Figure 8. Distribution of pollutants and contextual variables for joint contextual-multipollutant exposure profile clusters ($N_{clusters} = 20$). These boxplots are color coded to indicate profile cluster distributions considered high exposures (red), average exposures (green), and low exposures (blue). Black lines indicate overall averages. **A:** pollutant concentrations for multipollutant clusters and **B:** median income (\$), % older homes (older than 1950), % nonHispanic black, % nonHispanic white, and % Hispanic for contextual clusters.

Table 5. Prevalence of TLBW for Multipollutant Clusters and Model Results for Joint Contextual-Multipollutant Exposure Profile Cluster Random Effects ($N = 804,726$)

Cluster	Births (n)	TLBW (n)	% TLBW (95% CI) ^a	Cluster Effect Size (95% CI) ^a	Probability (Cluster Effect > 0)
1	51,623	791	1.53 (1.43 to 1.64)	-0.124 (-0.221 to 0.028)	0.006
2	63,817	943	1.48 (1.39 to 1.57)	-0.180 (-0.275 to 0.085)	0.001
3	7,109	175	2.46 (2.13 to 2.85)	0.026 (-0.123 to 0.174)	0.635
4	35,726	649	1.82 (1.68 to 1.96)	-0.155 (-0.261 to -0.049)	0.002
5	59,269	1,176	1.98 (1.87 to 2.10)	-0.042 (-0.133 to 0.049)	0.181
6	107,507	2,807	2.61 (2.52 to 2.71)	0.121 (0.041 to 0.202)	0.998
7	40,293	840	2.08 (1.95 to 2.23)	0.011 (-0.085 to 0.107)	0.588
8	132,335	2,815	2.13 (2.05 to 2.21)	0.076 (-0.003 to 0.156)	0.970
9	106,813	2,053	1.92 (1.84 to 2.01)	-0.026 (-0.107 to 0.056)	0.268
10	37,864	765	2.02 (1.88 to 2.17)	-0.024 (-0.122 to 0.073)	0.310
11	1,038	23	2.22 (1.48 to 3.31)	0.043 (-0.188 to 0.270)	0.642
12	43,896	926	2.11 (1.98 to 2.25)	0.005 (-0.090 to 0.099)	0.537
13	16,051	287	1.79 (1.59 to 2.01)	-0.108 (-0.235 to 0.017)	0.044
14	7,989	205	2.57 (2.24 to 2.94)	0.080 (-0.061 to 0.219)	0.868
15	1,186	30	2.53 (1.77 to 3.60)	0.061 (-0.163 to 0.282)	0.705
16	50,933	1,195	2.35 (2.22 to 2.48)	0.088 (-0.005 to 0.180)	0.968
17	601	23	3.83 (2.56 to 5.69)	0.109 (-0.129 to 0.345)	0.816
18	18,101	523	2.89 (2.66 to 3.14)	0.004 (-0.107 to 0.180)	0.525
19	5,740	86	1.50 (1.21 to 1.85)	-0.060 (-0.238 to 0.114)	0.249
20	16,835	382	2.27 (2.05 to 2.51)	0.106 (-0.021 to 0.232)	0.949

^a 95% CI = 95% credible interval.

Discussion

Our clustering procedure and subsequent multilevel logistic regression analysis provides concentration estimates for pollutant mixtures and TLBW risk estimates for place-based air pollution mixtures within LA County. We found that CBGs in Central LA, South-Central LA, and CBGs in East LA that are adjacent to major highways exhibited the highest concentrations in terms of area-level exposure to multiple pollutants. Furthermore, we found that CBGs in Central LA, South-Central LA, and East LA consistently had the most hazardous air pollution mixtures as they relate to TLBW. We also identified exposure profiles of clustered contextual variables that suggested that populations vulnerable to TLBW risks attributable to neighborhood factors were concentrated in the center of LA and South-Central LA County. Importantly, we further observed that indicators of neighborhood disadvantage and hazardous

air pollution mixtures tended to be linked. Moreover, evidence from our study indicates that the urban environment of LA County is a complex one in terms of spatial disparities for multipollutant exposures and concurrent adverse neighborhood conditions, which are likely to be drivers of racial-ethnic and socioeconomic disparities in TLBW risks.

Explanations as to why our multilevel-multipollutant analysis observed distinct spatial patterning of mixed air pollution-birth outcome risks could be attributable to spatially clustered factors that influence highly localized pollutant mixtures, such as distance to major roadways combined with local traffic volumes, traffic congestion, and the types of traffic-emitting sources of air pollution. Evidence from our study would suggest that indeed distance to major roadways is likely to enhance TLBW risk with respect to elevated pollutant concentration mixtures.

Firstly, our data analysis demonstrates that the highest risk multipollutant clusters (clusters 7 and 11) were characterized by the smallest median maternal residential distances to major roadways when compared to all other cluster-specific median distances to major roadways. Hence, these results present evidence that proximity to major sources of traffic-related air pollution is likely to at least partially explain the spatial patterning of multipollutant TLBW risks within LA County.

Our study further illustrates the importance of examining the spatial patterning of multipollutant effects. For example, the CBGs that intersect major interstate highways, such as I-110, I-710, and I-10, were characterized by multipollutant mixtures with the highest effect sizes as well as the highest probabilities for TLBW. Moreover, there was apparent spatial clustering of higher risk multipollutant profiles among the interchanges where these major interstate highways converge at the center of LA County. This is suggestive of highly localized traffic and emission patterns that are distinctive to these major highways and their intersections.

An important implication from our findings is that exposure to pollutant mixtures with respect to $PM_{2.5}$ — particularly its physical and chemical properties — is worthy of further examination in terms of gradients in health effects. Pirani and colleagues (2015) recently conducted a time-series analysis to study the variation in respiratory mortality across exposure profile clusters using a Bayesian profile regression approach similar to that applied in our study. Their study found that days characterized by high concentrations of secondary particulates (e.g., nitrates and sulfates) imparted the highest risk of respiratory mortality in comparison with all other $PM_{2.5}$ constituent mixtures.

In a multivariate analysis we also found that the age of housing stock was significantly associated with TLBW, in that a larger proportion of older homes (older than 1950) was related to a higher prevalence of TLBW (data not shown). While the adverse effects estimated for the age of housing stock are relatively modest in comparison to the other contextual variable considered in our study (race–ethnicity, median household income), this finding should be explored further in future studies. Our finding could be explained by previous research indicating that housing characteristics are associated with elevated exposure to both indoor and outdoor air pollution sources (Ghosh et al. 2013; Heroux et al. 2010; Houston et al. 2004; Meng et al. 2005; Spengler et al. 1994) and TLBW (Ghosh et al. 2013; Grady 2011). In the LA area, older homes have been associated with higher PM (Houston et al. 2014) and NO_2 exposures (Spengler et al. 1994). Older homes are also associated with a higher prevalence of gas stoves in the

home (Eisner and Blanc 2003) and may be more likely to lack exhaust ventilation in homes that use gas stoves (Coker et al. 2015). Additionally, older homes are more likely have exposure to lead (i.e., lead contaminated drinking water has been associated with lead piping, solder, and piping materials [Brown and Margolis 2012]), and lead exposure is implicated as an environmental risk factor for adverse birth outcomes such as low birth weight (Andrews et al. 1994; Centers for Disease Control and Prevention 2014; González-Cossío et al. 1997; Jelliffe-Pawlowski et al. 2006). Importantly, additional research data suggests that lower income and nonwhite households are more likely to live in older housing stock (Adamkiewicz et al. 2011; Houston et al. 2004), which was the case in our study as well.

Importantly, our findings are consistent with previous studies that found variation in effects on birth weight between various air pollutants or sources of air pollution (e.g., traffic-related versus natural background sources). In LA County, Laurent and colleagues (2014) examined the relation between various components and sources of fine PM air pollution; they found statistically significant differences with respect to different PM-related components and sources for effects on TLBW. Likewise, Wilhelm and colleagues (2011b) observed differences in the exposure–response relationship between $PM_{2.5}$ and TLBW based on the source of $PM_{2.5}$ exposure (e.g., gasoline versus geologic sources) within LA County. Each of these studies suggest that the pollutant, or combination of pollutants (i.e., constituents of PM pollution), related to traffic exposure were most hazardous in terms of TLBW risks. Findings from our study would suggest the same.

It is also clear from our results that multipollutant health risks, in this case TLBW, are related to racial and ethnic disparities in exposure since populations in the highest risk CBGs were predominantly people of color, socioeconomically deprived, and tended to live in older housing stock. While we did adjust for individual-level race–ethnicity and SES in our model, we still found that TLBW occurred predominantly in minority neighborhoods. The implications of this finding is that disparities in multipollutant exposures in particular may be an important driver of disparities in environmentally-driven adverse birth effects. Another notable finding in our study is that while some high SES neighborhoods were found to have high percentages of older housing stock, these neighborhoods tended to have the lowest concentration pollution mixtures. Moreover, higher SES neighborhoods with older housing are likely to have better maintained homes with air conditioning, which may lend itself to homes with fewer leaks that would allow the penetration of outdoor air pollutants indoors.

Our approach of jointly fitting contextual variables with multiple pollutants proved to alter the spatial patterning of TLBW cluster-related risks substantively. For instance, while CBGs surrounding the ports tended to have average multipollutant-only TLBW risk, the areas surrounding the ports exhibited a significantly increased risk of TLBW once contextual factors were clustered with the pollutants. This result would suggest that neighborhoods surrounding the ports are particularly vulnerable to air pollution exposure due to neighborhood sociodemographic and housing factors. Additionally, once contextual factors were clustered with pollutants a similar pattern of varying TLBW risk was seen in areas of South Central LA.

There are several noteworthy limitations and strengths in our findings. Much like other air pollution and birth weight studies, we relied on imperfect exposure estimation. A major limitation lies in the lack of additional pollutants to include in the multipollutant model, aside from the three pollutants considered in our study. There are several other hazardous air pollutants — particularly traffic-related pollutants — that would potentially better explain the spatial patterning of the pollutant–TLBW risks. Furthermore, our study lacked speciation data for PM exposures, hence we are limited in our ability to attribute a particular source of air pollution most likely to be implicated in causing excess risk of TLBW. The spatial patterning of our results does provide strong evidence to implicate traffic pollution from major roadways. A major strength of our study lies in the large sample size and the fact that we relied on population-wide data, rather than a nonrandom selection of a study population based simply on proximity to major sources of air pollution. Moreover, our study design allowed us to examine exposure to multiple pollutants that are highly correlated with each other. In addition, given that we were able to cluster correlated pollutants to examine health effects, we were able to reduce the dimensions of statistical comparison, compared to typical fixed-effects models that attempt to examine joint effects. Also, we adjusted for individual covariates and area-level poverty factors associated with adverse effects on birth weight (English et al. 2003), and we further accounted for some residual confounding (at the CT level [Williams et al. 2007]) that is likely to be present (e.g., for example due to maternal smoking during pregnancy).

IMPLICATIONS OF FINDINGS

Our results go beyond the findings one would obtain from analyzing air pollution associations with health outcomes using standard, additive regression models. These kinds of standard approaches have value, but fail to provide

the kind of subtleties that researchers and policymakers desire when one is interested in how combinations of pollutants associate with health and in determining how effects of single pollutants and pollutant mixtures vary spatially throughout a large geographical area.

Part of the problem with using standard regression methods to analyze high- or even moderate-dimensional data sets is the fact that many of the regression methodologies were developed in a low-dimensional setting (Fisher 1922). Main effects and interaction terms are included in the model in an additive manner, which works well when there are only a small number of covariates to consider. However, in epidemiologic studies, the number of air pollutant variables and SES covariates can be large, and the additive nature of the standard regression approach becomes unwieldy, as hundreds or even thousands of terms are needed to capture the complex manner in which the covariates interact with each other in determining the overall effect. Here we have expanded on well-established Dirchlet-process clustering methods (Dunson 2009; Neal 2000) to create a set of methodologies that enables one to obtain meaningful estimates of how pollution mixtures and SES covariates jointly affect health outcomes, at both the region-wide and the local level. The local nature of the associations is important, as it allows policy makers to target areas where reduction of air pollution exposures is most beneficial in terms of reduction of adverse health outcomes.

Traditional epidemiologic approaches to assessing associations between air pollution exposures and health outcomes are generally univariate in nature. Results may indicate that an association between, say, NO₂ and low birth weight exists in the study area in question. However, as demonstrated in this report, the effects of air pollutants vary across a large geographical region. If limited funds are available to affect changes regarding exposure levels, then policy makers may decide to focus on areas where such changes are most effective in affecting health outcomes. Further, since exposures exist in complex mixtures, it is desirable to identify typical combinations of exposures and where they exist, along with the likelihood that such typical mixtures are highly associated with adverse health outcomes. The Discussion sections of this report detail how our methods benefited in making inferences on these kinds of spatially-dependent associations between exposures and health outcomes using real data sets. While great progress had been made, our work is not complete. We would like to explore how policies aimed at reduction of a single pollutant will affect concentrations of other pollutants, and how these new pollution profiles affect health outcomes. Also, much work can be done exploring how grouping of regions to pollution clusters changes over time. We will use the work done here as a platform for further exploration in addressing these important issues.

CONCLUSION

The methodology utilized and advanced in this report, profile regression, like standard regression, consists of a flexible modeling framework. Profile regression provides extensions to standard regression that can be used to address many problems dealing with multiple correlated exposures and their health effects. Much of the estimation framework has already been detailed in previous work (Molitor et al. 2010), which describes the relevant iterative parameter estimation procedures and compares the profile regression method to other approaches via extensive simulation studies.

However, an important goal of this project is to create methods that will prove to be useful in generating real, substantively relevant results. Three applications (one in Appendix A) are provided that give model-based solutions to different real-world multipollutant problems. These three sections form the basis of three papers that are of both methodological and substantive value. The substantive value of these papers is derived from the fact that the data sets used did not previously exist in their current form, and were created, with great effort from both methodological and substantive collaborators on this project, for the purpose of conducting the multipollutant analyses. They thus represent new results and do not just reproduce findings from previous work.

Since our work is data driven, we did not hand pick data sets that would make our approaches look favorable when compared with other methods. Instead, we forced ourselves to analyze and address the questions laid out in the beginning of the project, even if this created unexpected computational challenges. This was particularly true for the place-based application, where a unified fully Bayesian approach was not feasible. Note however, that a fully Bayesian approach was used for the first application, studying air pollution profiles and poverty.

Regarding the applications, one should not infer that our methods can only analyze the exact kinds of data presented. In particular, a small number of pollutants were analyzed, due to data availability. Nevertheless, our approach is likely capable of handling dozens of exposures, where both exposures and outcomes can be presented as either continuous or categorical. However, our approach would likely struggle to handle data sets that contain millions of records. Scaling up our approaches to elegantly handle such big data is an ongoing area of our research and is a big ongoing area of research in the field of Bayesian methods and statistics in general.

It should be noted that while general statements can be made regarding the efficiency of certain parameter estimation procedures, one must be wary of making statements

suggesting that a particular model is correct. In this report, we followed the edict of the late Prof. George Box who famously stated, “Essentially, all models are wrong, but some are useful” (Box and Draper 1987), and we built models that are relevant to the problem at hand. Nevertheless, the models detailed in these model-based sections can be viewed as templates that can be taken, perhaps with some modification, by other researchers and applied to their work.

REFERENCES

- Adamkiewicz G, Zota AR, Fabian MP, Chahine T, Julien R, Spengler JD, et al. 2011. Moving environmental justice indoors: understanding structural influences on residential exposure patterns in low-income communities. *Am J Public Health* 101:S238–S245; doi:10.2105/AJPH.2011.300119.
- Albert JH, Chib S. 1993. Bayesian analysis of binary and polychotomous response data. *J Am Stat Assoc* 88:669–679.
- Andrews KW, Savitz DA, Hertz-Picciotto I. 1994. Prenatal lead exposure in relation to gestational age and birth weight: a review of epidemiologic studies. *Am J Ind Med* 26:13–32; doi:10.1002/ajim.4700260103.
- Austin E, Coull B, Thomas D, Koutrakis P. 2012. A framework for identifying distinct multipollutant profiles in air pollution data. *Environ Int* 45:112–121; doi:10.1016/j.envint.2012.04.003.
- Beckerman BS, Jerrett M, Martin RV, van Donkelaar A, Ross Z, Burnett RT. 2013a. Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California. *Atmos Environ* 77:172–177; doi:10.1016/j.atmosenv.2013.04.024.
- Beckerman BS, Jerrett M, Serre M, Martin RV, Lee S-J, van Donkelaar A, et al. 2013b. A hybrid approach to estimating national scale spatiotemporal variability of PM_{2.5} in the contiguous United States. *Environ Sci Technol* 47(13): 7233–7234; doi:10.1021/es400039u.
- Bell ML, Belanger K, Ebisu K, Gent JF, Lee HJ, Koutrakis P, et al. 2010. Prenatal exposure to fine particulate matter and birth weight: variations by particulate constituents and sources. *Epidemiology* 21:884–891; doi:10.1097/EDE.0b013e3181f2f405.

- Besag J, York JC, Mollié A. 1991. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann Stat Math* 43:1–59.
- Boehmer TK, Foster SL, Henry JR, Woghiren-Akinnifesi EL, Yip FY, Centers for Disease Control and Prevention. 2013. Residential proximity to major highways — United States, 2010. *MMWR Surveill Summ* 62 Suppl 3:46–50.
- Box GEP, Draper NR. 1987. *Empirical Model-Building and Response Surfaces*. 1st edition. New York:Wiley.
- Brainard JS, Jones AP, Bateman IJ, Lovett AA, Fallon PJ. 2002. Modelling environmental equity: access to air quality in Birmingham, England. *Environ Plan A* 34:695–716; doi:10.1068/a34184.
- Brauer M, Tamburic L. 2009. Multi-pollutant analysis of reproductive outcomes and air pollution using the CMAQ model. *Epidemiology* 20:S72–S73; doi:10.1097/01.ede.0000362926.16562.96.
- Brown MJ, Margolis S. 2012. Lead in drinking water and human blood lead levels in the United States. *MMWR Supplements* 61(04):1–9.
- Brulle RJ, Pellow DN. 2006. Environmental justice: human health and environmental inequalities. *Annu Rev Public Health* 27:103–124; doi:10.1146/annurev.publhealth.27.021405.102124.
- Buonocore JJ, Lee HJ, Levy JI. 2009. The influence of traffic on air quality in an urban neighborhood: a community–university partnership. *Am J Public Health* 99:S629–S635; doi:10.2105/AJPH.2008.149138.
- Centers for Disease Control and Prevention. 2014. Health effects low birthweight and the environment - CDC Tracking Network. Available: <http://ephtracking.cdc.gov/showRbLBWGrowthRetardationEnv.action> [accessed 20 April 2015].
- Cesaroni G, Badaloni C, Romano V, Donato E, Perucci CA, Forastiere F. 2010. Socioeconomic position and health status of people who live near busy roads: the Rome Longitudinal Study (RoLS). *Environ Health* 9:41; doi:10.1186/1476-069X-9-41.
- Chaix B, Gustafsson S, Jerrett M, Kristersson H, Lithman T, Boalt A, et al. 2006. Children’s exposure to nitrogen dioxide in Sweden: investigating environmental injustice in an egalitarian country. *J Epidemiol Community Health* 60:234–241; doi:10.1136/jech.2005.038190.
- Cho S-H, Tong H, McGee J, Baldauf R, Krantz Q, Gilmour M. 2009. Comparative toxicity of size-fractionated airborne particulate matter collected at different distances from an urban highway. *Environ Health Perspect*; 117(11):1682–1689; doi:10.1289/ehp.0900730.
- Coker ES, Smit E, Harding AK, Molitor J, Kile ML. 2015. A cross sectional analysis of behaviors related to operating gas stoves and pneumonia in U.S. children under the age of 5. *BMC Public Health* 15:77; doi:10.1186/s12889-015-1425-y.
- Dahl D. 2006. Model-based clustering for expression data via a Dirichlet process mixture model. In *Bayesian Inference for Gene Expression and Proteomics* (Do K-A, Muller P, Vannucci M, eds), Cambridge University Press.
- Debbink MP, Bader MDM. 2011. Racial residential segregation and low birth weight in Michigan’s metropolitan areas. *Am J Public Health* 101:1714–1720; doi:10.2105/AJPH.2011.300152.
- Denison D, Holmes C. 2001. Bayesian partitioning for estimating disease risk. *Biometrics* 57:143–149.
- Do DP, Finch BK, Basurto-Davila R, Bird C, Escarce J, Lurie N. 2008. Does place explain racial health disparities? Quantifying the contribution of residential context to the black/white health gap in the United States. *Soc Sci Med* 67:1258–1268; doi:10.1016/j.socscimed.2008.06.018.
- Dunson DB. 2009. Nonparametric Bayes local partition models for random effects. *Biometrika* 96:249–262; doi:10.1093/biomet/asp021.
- Eisner MD, Blanc PD. 2003. Gas stove use and respiratory health among adults with asthma in NHANES III. *Occup Environ Med* 60:759–764.
- English PB, Kharrazi M, Davies S, Scalf R, Waller L, Neutra R. 2003. Changes in the spatial pattern of low birth weight in a southern California county: the role of individual and neighborhood level factors. *Soc Sci Med* 56:2073–2088; doi:10.1016/S0277-9536(02)00202-2.
- Escobar MD. 1994. Estimating normal means with a Dirichlet process prior. *J Am Stat Assoc* 89:268–277.
- Everitt B. 1984. *An Introduction to Latent Variable Models*. London; New York:Chapman and Hall.
- Everitt B, Hand DJ. 1981. *Finite Mixture Distributions*. London; New York:Chapman and Hall.
- Fang H, Johnson C, Stopp C, Espy KA. 2011. A new look at quantifying tobacco exposure during pregnancy using fuzzy clustering. *Neurotoxicol Teratol* 33:155–165; doi:10.1016/j.ntt.2010.08.003.

- Finkelstein MM, Jerrett M. 2007. A study of the relationships between Parkinson's disease and markers of traffic-derived and environmental manganese air pollution in two Canadian cities. *Environ Res* 104:420–432; doi:10.1016/j.envres.2007.03.002.
- Fisher RA. 1922. The goodness of fit of regression formulae, and the distribution of regression coefficients. *J R Stat Soc* 85:597; doi:10.2307/2341124.
- Geer LA. 2014. Identifying exposure disparities in air pollution epidemiology specific to adverse birth outcomes. *Environ Res Lett* 9:101001; doi:10.1088/1748-9326/9/10/101001.
- Gelman A, Jakulin A, Pittau MG, Su Y-S. 2008. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2(4):1360–1383; doi:10.1214/08-AOAS191.
- Ghosh JKC, Wilhelm M, Ritz B. 2013. Effects of residential indoor air quality and household ventilation on preterm birth and term low birth weight in Los Angeles County, California. *Am J Public Health* 103:686–694; doi:10.2105/AJPH.2012.300987.
- Gilks WR, Richardson S, Spiegelhalter DJ. 1998. *Markov chain Monte Carlo in practice*. Boca Raton, FL:Chapman & Hall.
- Goldberg DW, Wilson JP, Knoblock CA, Ritz B, Cockburn MG. 2008. An effective and efficient approach for manually improving geocoded data. *Int J Health Geogr* 7:60; doi:10.1186/1476-072X-7-60.
- González-Cossío T, Peterson KE, Sanín LH, Fishbein E, Palazuelos E, Aro A, et al. 1997. Decrease in birth weight in relation to maternal bone-lead burden. *Pediatrics* 100:856–862.
- Gouveia N, Bremner SA, Novaes HMD. 2004. Association between ambient air pollution and birth weight in São Paulo, Brazil. *J Epidemiol Community Health* 58:11–17.
- Grady SC. 2006. Racial disparities in low birthweight and the contribution of residential segregation: a multilevel analysis. *Soc Sci Med* 63:3013–3029; doi:10.1016/j.socscimed.2006.08.017.
- Grady SC. 2011. Housing Quality and Racial Disparities in Low Birth Weight: A GIS Assessment. In: *Geospatial Analysis of Environmental Health* (Maantay JA and McLafferty S, eds). Dordrecht, Netherlands:Springer Netherlands, 303–318.
- Green PJ, Richardson S. 2001. Modelling heterogeneity with and without the Dirichlet process. *Scand J Stat* 28:355–375.
- Hasheminassab S, Daher N, Saffari A, Wang D, Ostro BD, Sioutas C. 2014. Spatial and temporal variability of sources of ambient fine particulate matter (PM_{2.5}) in California. *Atmos Chem Phys* 14:12085–12097; doi:10.5194/acp-14-12085-2014.
- Heroux M-E, Clark N, Ryswyk KV, Mallick R, Gilbert NL, Harrison I, et al. 2010. Predictors of indoor air concentrations in smoking and non-smoking residences. *Int J Environ Res Public Health* 7:3080–3099; doi:10.3390/ijerph7083080.
- Houston D, Li W, Wu J. 2014. Disparities in exposure to automobile and truck traffic and vehicle emissions near the Los Angeles-Long Beach port complex. *Am J Public Health* 104:156–164; doi:10.2105/AJPH.2012.301120.
- Houston D, Wu J, Ong P, Winer A. 2004. Structural disparities of urban traffic in southern California: implications for vehicle-related air pollution exposure in minority and high-poverty neighborhoods. *J Urban Aff* 26:565–592; doi:10.1111/j.0735-2166.2004.00215.x.
- Institute of Medicine. 1999. *Toward Environmental Justice: Research, Education, and Health Policy Needs*. Washington DC:National Academies Press.
- Ishwaran H, James LF. 2001. Gibbs sampling methods for stick-breaking priors. *J Amer Stat Assoc* 96:161–173.
- Jain S, Radford N. 2004. A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *J Comput Graph Stat* 13:158–182.
- Jelliffe-Pawłowski LL, Miles SQ, Courtney JG, Materna B, Charlton V. 2006. Effect of magnitude and timing of maternal pregnancy blood lead (Pb) levels on birth outcomes. *J Perinatol Off J Calif Perinat Assoc* 26:154–162; doi:10.1038/sj.jp.7211453.
- Jerrett M, Burnett RT, Beckerman BS, Turner MC, Krewski D, Thurston G, et al. 2013. Spatial analysis of air pollution and mortality in California. *Am J Respir Crit Care Med* 188:593–599; doi:10.1164/rccm.201303-0609OC.
- Jerrett M, Burnett RT, Ma R, Pope CA 3rd, Krewski D, Newbold KB, et al. 2005. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 16:727–36.
- Jerrett M, Finkelstein M. 2005. Geographies of risk in studies linking chronic air pollution exposure to health

- outcomes. *J Toxicol Environ Health A* 68:1207–1242. doi:10.1080/15287390590936085.
- Kim J, Durden E. 2007. Socioeconomic status and age trajectories of health. *Soc Sci Med* 65:2489–2502; doi:10.1016/j.socscimed.2007.07.022.
- Kuhn T, Biswas S, Fine PM, Geller M, Sioutas C. 2005. Physical and chemical characteristics and volatility of PM in the proximity of a light-duty vehicle freeway. *Aerosol Sci Technol* 39:347–357; doi:10.1080/027868290930024.
- Kutner M, Nachtsheim C, Neter J, Li W. 2004. *Applied Linear Statistical Models*. 5th edition. Boston, MA:McGraw-Hill/Irwin.
- Lantz PM, Lynch JW, House JS, Lepkowski JM, Mero RP, Musick MA, et al. 2001. Socioeconomic disparities in health change in a longitudinal study of U.S. adults: the role of health-risk behaviors. *Soc Sci Med* 53:29–40.
- Laurent O, Hu J, Li L, Cockburn M, Escobedo L, Kleeman MJ, et al. 2014. Sources and contents of air pollution affecting term low birth weight in Los Angeles County, California, 2001–2008. *Environ Res* 134:488–495; doi:10.1016/j.envres.2014.05.003.
- Lazarsfeld PF, Henry NW. 1968. *Latent Structure Analysis*. New York:Houghton Mifflin.
- Levy I, Mihele C, Lu G, Narayan J, Brook JR. 2013. Evaluating multipollutant exposure and urban air quality: pollutant interrelationships, neighborhood variability, and nitrogen dioxide as a proxy pollutant. *Environ Health Perspect* 122(1):65–72; doi:10.1289/ehp.1306518.
- Liverani S, Hastie D, Azizi L, Papatomas M, Richardson S. 2014. PReMiuM: An R package for profile regression mixture models using Dirichlet processes. arXiv:1303.2836 [stat.CO] 3:1–29.
- Marshall JD. 2008. Environmental inequality: air pollution exposures in California’s south coast air basin. *Atmos Environ* 42:5499–5503; doi:10.1016/j.atmosenv.2008.02.005.
- Mauderly JL, Burnett RT, Castillejos M, Ozkaynak H, Samet JM, Stieb DM, et al. 2010. Is the air pollution health research community prepared to support a multipollutant air quality management framework? *Inhal Toxicol* 22 Suppl 1:1–19; doi:10.3109/08958371003793846.
- Mauderly JL, Samet JM. 2009. Is there evidence for synergy among air pollutants in causing health effects? *Environ Health Perspect* 117:1–6; doi:10.1289/ehp.11654.
- McHugh TB. 1956. Efficient estimation and local identification in latent class analysis. *Psychometrika* 21:331–347.
- McLachlan GJ, Basford KE. 1988. *Mixture Models: Inference and Applications to Clustering*. New York, NY:John Wiley & Sons.
- Meng QY, Turpin BJ, Korn L, Weisel CP, Morandi M, Colome S, et al. 2005. Influence of ambient (outdoor) sources on residential indoor and personal PM_{2.5} concentrations: analyses of RIOPA data. *J Expo Anal Environ Epidemiol* 15:17–28; doi:10.1038/sj.jea.7500378.
- Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, et al. 2006. The development of a standardized neighborhood deprivation index. *J Urban Health* 83:1041–1062; doi:10.1007/s11524-006-9094-x.
- Metcalf A, Lail P, Ghali WA, Sauve RS. 2011. The association between neighbourhoods and adverse birth outcomes: a systematic review and meta-analysis of multi-level studies: neighbourhoods and adverse birth outcomes. *Paediatr Perinat Epidemiol* 25:236–245; doi:10.1111/j.1365-3016.2011.01192.x.
- Molitor J, Jerrett M, Chang CC, Molitor NT, Gauderman J, Berhane K, et al. 2007. Assessing uncertainty in spatial exposure models for air pollution health effects assessment. *Environ Health Perspect* 115:1147–53.
- Molitor J, Papatomas M, Jerrett M, Richardson S. 2010. Bayesian profile regression with an application to the National Survey of Children’s Health. *Biostat Oxf Engl* 11:484–498; doi:10.1093/biostatistics/kxq013.
- Molitor J, Su JG, Molitor N-T, Rubio VG, Richardson S, Hastie D, et al. 2011. Identifying vulnerable populations through an examination of the association between multipollutant profiles and poverty. *Environ Sci Technol* 45:7754–7760; doi:10.1021/es104017x.
- Morello-Frosch R, Jesdale BM. 2006. Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in U.S. metropolitan areas. *Environ Health Perspect* 114:386–393; doi:10.1289/ehp.8500.
- Morello-Frosch R, Pastor M, Porras C, Sadd J. 2002. Environmental justice and regional inequality in southern California: implications for future research. *Environ Health Perspect* 110(Suppl 2):149–154.
- Morello-Frosch R, Pastor M, Sadd J. 2001. Environmental justice and southern California’s “riskscape” the distribution of air toxics exposures and health risks among diverse communities. *Urban Aff Rev* 36:551–578; doi:10.1177/10780870122184993.

- Morello-Frosch R, Zuk M, Jerrett M, Shamasunder B, Kyle AD. 2011a. Understanding the cumulative impacts of inequalities in environmental health: implications for policy. *Health Aff Proj Hope* 30:879–887; doi:10.1377/hlthaff.2011.0153.
- Morello-Frosch R, Zuk M, Jerrett M, Shamasunder B, Kyle AD. 2011b. Understanding the cumulative impacts of inequalities in environmental health: implications for policy. *Health Aff (Millwood)* 30:879–887; doi:10.1377/hlthaff.2011.0153.
- National Research Council. 2004. Committee on Air Quality Management in the United States. *Air Quality Management in the United States*. Washington DC: National Academies Press.
- Neal R. 2000. Markov chain sampling methods for Dirichlet process mixture models. *J Comput Graph Stat* 9(2):249–265.
- Ohlssen DI, Sharples LD, Spiegelhalter DJ. 2007. Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons. *Stat Med* 26:2088–2112.
- Pearce J, Kingham S, Zawar-Reza P. 2006. Every breath you take? Environmental justice and air pollution in Christchurch, New Zealand. *Environ Plan A* 38:919–938; doi:10.1068/a37446.
- Pirani M, Best N, Blangiardo M, Liverani S, Atkinson RW, Fuller GW. 2015. Analysing the health effects of simultaneous exposure to physical and chemical properties of airborne particles. *Environ Int* 79:56–64; doi:10.1016/j.envint.2015.02.010.
- Ponce NA. 2005. Preterm birth: the interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods. *Am J Epidemiol* 162:140–148; doi:10.1093/aje/kwi173.
- Raftery AE, Zheng Y, We N-, Clyde M, Hoeting J, Madigan D. 2003. Long-run performance of Bayesian model averaging. *J Am Stat Assoc* 98:931–938.
- Richardson S, Green P. 1997. On Bayesian analysis of mixtures with an unknown number of components. *J R Stat Soc Ser B* 59:731–792.
- Ritz B, Wilhelm M. 2008. Ambient air pollution and adverse birth outcomes: methodologic issues in an emerging field. *Basic Clin Pharmacol Toxicol* 102:182–190; doi:10.1111/j.1742-7843.2007.00161.x.
- Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J R Stat Soc B* 71(2):319–392; doi:10.1111/j.1467-9868.2008.00700x.
- Rue H, Martino S, Lindgren F, Simpson D, Riebler A, Krainski E. 2014. Functions which allow to perform full Bayesian analysis of latent Gaussian models using integrated nested Laplace approximation. Available: <http://inla.googlecode.com/hg-history/default/rinla/DESCRIPTION>.
- Shumka L, Benoit C. 2007. Social suffering and gaps in alternative health care for vulnerable women workers. In *Inequalities and Disparities in Health Care and Health: Concerns of Patients, Providers and Insurers*, Vol. 25 of *Research in the Sociology of Health Care*. Bingley, W Yorkshire, England: Emerald Group Publishing Limited, 253–275.
- Spengler J, Schwab M, Ryan PB, Colome S, Wilson A, Billick I, et al. 1994. Personal exposure to nitrogen dioxide in the Los Angeles Basin. *Air Waste* 44:39–47; doi:10.1080/1073161X.1994.10467236.
- Spiegelhalter D, Thomas A, Best N. 2003. WinBUGS Version 1.4 User Manual. Cambridge, UK: MRC Biostatistics Unit.
- Su JG, Jerrett M, Beckerman B. 2009a. A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. *Sci Total Env* 407:3890–8; doi:10.1016/j.scitotenv.2009.01.061.
- Su JG, Jerrett M, Beckerman B, Wilhelm M, Ghosh JK, Ritz B. 2009b. Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. *Environ Res* 109:657–670; doi:10.1016/j.envres.2009.06.001.
- Su JG, Jerrett M, Morello-Frosch R, Jesdale BM, Kyle AD. 2012. Inequalities in cumulative environmental burdens among three urbanized counties in California. *Environ Int* 40:79–87; doi:10.1016/j.envint.2011.11.003.
- Su JG, Morello-Frosch R, Jesdale BM, Kyle AD, Shamasunder B, Jerrett M. 2009c. An index for assessing demographic inequalities in cumulative environmental hazards with application to Los Angeles, California. *Environ Sci Technol* 43:7626–7634; doi:10.1021/es901041p.
- U.S. Environmental Protection Agency. 2000. User's guide for the assessment system for population exposure nationwide (ASPEN, version 1.1) model. EPA-454/R-00-017.

U.S. Environmental Protection Agency. 2006. Air Quality Criteria for Ozone and Related Photochemical Oxidants (final). Research Triangle Park, NC. EPA 600/R-05/004aF.

Wagner JG, Morishita M, Keeler GJ, Harkema JR. 2012. Divergent effects of urban particulate air pollution on allergic airway responses in experimental asthma: a comparison of field exposure studies. *Environ Health* 11:45; doi:10.1186/1476-069X-11-45.

Wilhelm M, Ghosh JK, Su J, Cockburn M, Jerrett M, Ritz B. 2011a. Traffic-related air toxics and preterm birth: a population-based case-control study in Los Angeles County, California. *Environ Health* 10:89; doi:10.1186/1476-069X-10-89.

Wilhelm M, Ghosh JK, Su J, Cockburn M, Jerrett M, Ritz B. 2011b. Traffic-related air toxics and term low birth weight in Los Angeles County, California. *Environ Health Perspect* 120:132–138; doi:10.1289/ehp.1103408.

Williams BL, Pennock-Román M, Suen HK, Magumbol MS, Ozdenerol E. 2007. Assessing the impact of the local environment on birth outcomes: a case for HLM. *J Expo Sci Environ Epidemiol* 17:445–457; doi:10.1038/sj.jes.7500537.

Zhang K, Batterman S. 2013. Air pollution and health risks due to vehicle traffic. *Sci Total Environ* 450-451:307–316; doi:10.1016/j.scitotenv.2013.01.074.

Zwack LM, Paciorek CJ, Spengler JD, Levy JL. 2011. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. *Environ Health Perspect* 119:852–859; doi:10.1289/ehp.1002519.

HEI QUALITY ASSURANCE STATEMENT

The conduct of this research project was subject to independent quality assurance (QA) oversight by Abt Associates. The audit team consisted of Dr. Sue Greco, who has over 10 years of experience in both human health risk assessment and fine PM exposure assessment and Mr. Kevin He, who has a background in environmental science and programming. The QA oversight consisted of two on-site audits and a review of the Investigators' Final Report. The dates of the QA oversight activities are listed below with the phase of the study examined.

June 23–24, 2014. Initial audit conducted on-site at the Oregon State University College of Public Health and Human Sciences in Corvallis, Oregon.

The purpose of the initial QA audit was to review the standard operating procedures and data management practices used in the study and determine if they were followed consistently by all members of the research team. The audit team met in person with the principal investigator and key team members at Oregon State University, and had a teleconference with other key team members at other institutions in the LA area. The auditors noted that the study shifted away from examining the *Relationships of Indoor, Outdoor, and Personal Air* (RIOPA) data and toward the larger LA County birth outcomes data and requested that the researchers confirm that the Institutional Review Board status of the study and Institutional Review Board training of the team members was consistent with Oregon State University requirements.

June 29, 2015. Final audit conducted on-site at the Oregon State University College of Public Health and Human Sciences.

The purpose of the final QA audit was to check the reproducibility of conclusions, tables, and figures depicted in the final report titled, "Modeling of Multipollutant Profiles and Spatially Varying Health Effects with applications to indicators of adverse birth outcomes using data from the UCLA Environment and Pregnancy Outcome Study." The auditors preselected nine tables and figures from the May 2015 version of the Investigators' Final Report for review. The researchers described how the final report outputs were generated, starting from the raw data, processing, and program scripts. The auditors identified four minor discrepancies in the regeneration exercise, none of which influenced the conclusions in the final report, and all of which were remedied in a subsequent version of the final report. The audit demonstrated that the study was conducted by an experienced team with a high concern for data quality. The final report appears to be an accurate representation of the modeling exercises.



Sue Greco

Modeling spatial effects of PM_{2.5} on term low birth weight in Los Angeles County

Eric Coker^{a,*}, Jokay Ghosh^b, Michael Jerrett^c, Virgilio Gomez-Rubio^d,
Bernardo Beckerman^c, Myles Cockburn^e, Silvia Liverani^f, Jason Su^c, Arthur Li^g,
Molly L Kile^a, Beate Ritz^{b,1}, John Molitor^{a,1}

^a College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA

^b School of Public Health, University of California, Los Angeles, Los Angeles, CA, USA

^c School of Public Health, University of California, Berkeley, Berkeley, CA, USA

^d Department of Mathematics, Universidad De Castilla-La Mancha, Albacete, Spain

^e Preventive Medicine and Spatial Sciences, University of Southern California, Los Angeles, CA, USA

^f Department of Mathematics, Brunel University, London, United Kingdom

^g Department of Information Science, City of Hope National Cancer Center, Duarte, CA, USA

ARTICLE INFO

Article history:

Received 15 April 2015

Received in revised form

27 June 2015

Accepted 30 June 2015

Keywords:

Air pollution

PM_{2.5}

Term low birth weight

Spatial effects

Multilevel modeling

ABSTRACT

Air pollution epidemiological studies suggest that elevated exposure to fine particulate matter (PM_{2.5}) is associated with higher prevalence of term low birth weight (TLBW). Previous studies have generally assumed the exposure–response of PM_{2.5} on TLBW to be the same throughout a large geographical area. Health effects related to PM_{2.5} exposures, however, may not be uniformly distributed spatially, creating a need for studies that explicitly investigate the spatial distribution of the exposure–response relationship between individual-level exposure to PM_{2.5} and TLBW. Here, we examine the overall and spatially varying exposure–response relationship between PM_{2.5} and TLBW throughout urban Los Angeles (LA) County, California. We estimated PM_{2.5} from a combination of land use regression (LUR), aerosol optical depth from remote sensing, and atmospheric modeling techniques. Exposures were assigned to LA County individual pregnancies identified from electronic birth certificates between the years 1995–2006 ($N=1,359,284$) provided by the California Department of Public Health. We used a single pollutant multivariate logistic regression model, with multilevel spatially structured and unstructured random effects set in a Bayesian framework to estimate global and spatially varying pollutant effects on TLBW at the census tract level. Overall, increased PM_{2.5} level was associated with higher prevalence of TLBW county-wide. The spatial random effects model, however, demonstrated that the exposure–response for PM_{2.5} and TLBW was not uniform across urban LA County. Rather, the magnitude and certainty of the exposure–response estimates for PM_{2.5} on log odds of TLBW were greatest in the urban core of Central and Southern LA County census tracts. These results suggest that the effects may be spatially patterned, and that simply estimating global pollutant effects obscures disparities suggested by spatial patterns of effects. Studies that incorporate spatial multilevel modeling with random coefficients allow us to identify areas where air pollutant effects on adverse birth outcomes may be most severe and policies to further reduce air pollution might be most effective.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Extensive evidence indicates that prenatal exposure to outdoor air pollution is associated with risk of term low birth weight (Brauer et al., 2008; Fleischer et al., 2014; Ghosh et al., 2013, 2012;

Hyder et al., 2014; Padula et al., 2012; Parker et al., 2011; Ponce, 2005; Proietti et al., 2013; Ritz and Wilhelm, 2008; Shah and Balkhair, 2011; Stieb et al., 2012; Wilhelm et al., 2011; Wu et al., 2011). While TLBW contributes to racial–ethnic and socioeconomic health disparities in the United States, air pollution is thought to be an important place-based factor in the complex geography of and susceptibility to TLBW (Jerrett and Finkelstein, 2005; Morello-Frosch and Shenassa, 2006). It is reasonable to consider, however, that air pollution exposure–response effects on adverse birth outcomes, such as TLBW, vary spatially within an urban setting.

* Corresponding author.

E-mail address: cokerer@onid.orst.edu (E. Coker).

¹ Equal contribution.

First and foremost, air pollutant mixtures or components of PM air pollution can be autocorrelated spatially within urban environments – depending on local-scale air pollution sources, the intensity of emissions, and meteorology (among other factors) (Hajat et al., 2013; Molitor et al., 2011; Su et al., 2012). As a result, the intrinsic toxicity of PM_{2.5} mixtures is likely to be spatially dependent. For instance, Laurent et al. (2014) found that various components and sources of fine PM air pollution, which exhibit strong spatially varying characteristics, produced statistically significant gradients in PM-related TLBW risk in LA County. Similarly, Wilhelm et al., (2011), found that the exposure–response between PM_{2.5} and TLBW varied by PM_{2.5} source type (e.g. gasoline versus geologic sources) within LA County. Furthermore, Pedersen et al. (2015) studied eight European birth cohorts and found that the exposure–response between PM_{2.5} was dependent on its chemical composition, with OR estimates for sulfur PM_{2.5} of 1.24, compared to 1.08 for overall PM_{2.5}. Such local-scale intra-urban differences in particulate air pollution exposure and health effects patterns may therefore lead to inequalities with regard to PM-related adverse birth outcome risks (Baxter et al., 2007). Further, a wide range of contextual neighborhood factors and individual factors that are spatially correlated, from socioeconomic status (SES), demographics (i.e. racial segregation), exposure to violence (Messer et al., 2006), access to healthy food (Lane et al., 2008) or green space (Hystad et al., 2014), housing characteristic, and psychosocial, may contribute to variations in susceptibilities to air pollution that are not fully accounted for in standard regression models relying on fixed covariate effects (Morello-Frosch and Shenassa, 2006). Few studies, however, have been conducted to examine whether there is a spatial patterning – or a “risk-scape” (Morello-Frosch and Shenassa, 2006) – for PM-related birth outcomes. While previous health research has evaluated the spatial dependency of PM-related chronic health effects such as cardiovascular disease and asthma (Boehm Vock et al., 2014; Choi et al., 2009; Fuentes et al., 2006; Jerrett et al., 2005; Krewski et al., 2009; McConnell et al., 2010; Samoli et al., 2004; Shankardass et al., 2009), no studies have modeled the spatial dependency of individual-level PM_{2.5} exposure–response relationships on birth outcomes.

Several recent studies examined the spatial variation in PM_{2.5} effects on TLBW between different countries or between US states. A large collaborative multi-site international study found a substantial degree of heterogeneity in estimates for entire pregnancy exposure–response between study sites, despite the use of similar exposure assessments and statistical models in the studies (Dadvand et al., 2013; Parker et al., 2011). Hao et al. (2015) found substantial differences between states in the U.S. in terms of the magnitude and direction of effects of PM_{2.5} on TLBW. Another multi-state U.S. study also found that the size of exposure–response estimates for PM_{2.5} and TLBW depended upon study site; with odds ratios ranging from between 0.942 (95% CI: 0.817, 1.09) in Utah to as high as 1.72 (95% CI: 1.55, 1.93) in New York state (per 10-unit increase in PM_{2.5} exposure) (Harris et al., 2014). Additionally, Williams et al. (2007) demonstrated, through implementation of a multilevel linear random coefficient model, that adverse effects on average birth weight in a population varied by census tract due to hazardous air pollution emitting industrial sites. The observed statistically significant differences in effect size between census tracts remained significant even after adjusting for the number of hazardous sites per census tract, individual level confounders, and contextually relevant census tract level confounding factors (Williams et al., 2007).

Despite the recent evidence suggesting that air pollution-related adverse effects on birth weight may vary spatially, no studies have explicitly examined spatial variation in effects within a dense metropolitan region such as LA county, which we are targeting in our paper. Our guiding hypothesis is that modeling of the spatially

varying coefficients will show differences in TLBW according to LA County census tracts and thus provide evidence for localized PM_{2.5} exposure–response. Specifically, the magnitude of effect will be higher in some census tracts when compared to the global mean exposure–response for all of urban LA County. Our approach goes beyond the commonly employed estimation of an overall average PM_{2.5} effect on birth weight and will allow us to describe a spatially-patterned deviation from the average effects, thus pinpointing potential ‘hotspots’ within LA County where the magnitude and probability of PM_{2.5} effects are likely to be strongest.

In our paper we utilize an existing land use regression (LUR) PM_{2.5} exposure model within a multi-level Bayesian framework; implemented with spatially-dependent random coefficients. This information may be useful from a policy perspective to create targeted public health interventions for LA County.

2. Methods

2.1. Study population and birth outcomes

Data on infant birth weight were derived from electronic birth certificates provided by the California Department of Public Health, for LA County births between 1/1/1995 and 12/31/2006 ($N=1,522,084$). The birth records provided information on maternal characteristics such as age, race/ethnicity, education, total number of previous maternal births, and residential address, as well as characteristics of the infant (abnormalities, birth season, gestational age at birth, birth weight and baby sex). Human subjects research was approved through the University of California, Los Angeles' Office of the Human Subjects Protection Program, the California Committee for the Protection of Human Subjects, and the University of Southern California Office for the Protection of Research Subjects. Similar to previous studies, we restricted the dataset to singleton births with no recorded abnormalities (Ghosh et al., 2013, 2012; Wilhelm et al., 2011). Additionally, we excluded births with extreme gestational days (less than 140 days or greater than 320 days), births that were not full term (<259 gestation days), and births with birth weight less than 500 g or greater than 5000 g due to concerns about recording errors. For our final analyses, we further excluded births without complete information on the full set of study covariates ($n=19,017$). Finally, since we are interested in estimating within-city spatial variation in PM_{2.5} effects, the spatial analysis further excluded rural sub-region of LA County, thus leaving a final study population of $N=1,356,304$. A detailed description of methods for geocoding residential addresses are described elsewhere (Goldberg et al., 2008).

2.2. PM_{2.5} exposure assessment

A PM_{2.5} LUR model developed previously by (Jerrett et al., 2013) was used to estimate individual exposures to PM_{2.5} at each mother residential address. Such estimates are intended to best represent spatially resolved long-term exposure to annual levels of PM_{2.5} between 1995–2006, rather than pregnancy period-specific exposure. This PM_{2.5} LUR model has been used previously to examine chronic long-term exposure to PM_{2.5} and related health effects over time, in a large cohort study of California adults (Jerrett et al., 2013). This LUR method has been described in previous publications and the reader is referred to Beckerman et al. (2013) and Jerrett et al. (2013) for greater detail. Briefly, the predicted concentrations of PM_{2.5} were based on covariate data from the following sources: (1) daily observations of PM_{2.5} air monitoring collected between 1998–2002 at government monitoring sites throughout California, which was supplemented with remotely-sensed PM_{2.5} data covering the time period between 2001 and

2006 (Beckerman et al., 2013); (2) data on traffic and road networks from 1990 to 2001; (3) land use data from the year 2001; (4) population density data from the 2000 US Census; and (5) numerical output from remote sensing modeling coupled with atmospheric modeling (Van Donkelaar et al., 2010). A deletion/substitution/addition algorithm was then implemented to develop the final model covariates with a cross-validated R^2 value of 0.65.

2.3. Covariates

Since this study is a methodological extension of previous work for the LA County area (Ghosh et al., 2012; Wilhelm et al., 2011), we applied similar covariates as in the previous studies to evaluate $PM_{2.5}$ in relation to risk of TLBW. Individual-level covariates were maternal age at delivery (<20 years, 20–24 years, 25–29 years, 30–34 years, ≥ 35 years), maternal race (non-Hispanic White, non-Hispanic Black, Hispanic, Asian, and Other race), maternal years of education (<9 years, 9–12 years, 13–15 years, and ≥ 16 years), parity, gestational days, gestation days-squared (Ghosh et al., 2012; Wilhelm et al., 2011) and sex of the infant.

2.4. Statistical analysis

2.4.1. Standard analysis

While our main objective was to evaluate the spatial dependency of $PM_{2.5}$ effects on TLBW, we initially examined “global” (or L.A. County-wide) associations between $PM_{2.5}$ and TLBW using crude-unadjusted and multivariate adjusted logistic regressions techniques. The intent of implementing a global fixed effects model is to replicate exposure–response relationships between increasing $PM_{2.5}$ exposure and increasing prevalence of TLBW as demonstrated from previous research. The crude and multivariate models were implemented as a generalized linear model (*glm*) using the binomial family with the logistic function in the R statistical computing environment (R-version 3.1.2) (see Supplemental Materials for code describing the specific models employed in R (Everitt and Hothorn, 2010)). For consistency, the multivariate model utilized same fixed effects covariates as for the multilevel model described below.

2.4.2. Multilevel spatial modeling

The focus of the present study was to expand on previous work by implementing a multilevel spatial logistic regression model that would assess whether exposure–response relationships vary within L.A. County. Along with the fixed effects on the covariates, we simultaneously included a random effect coefficient for the census tract-level effect of $PM_{2.5}$ on log-odds of TLBW. The random air pollution effect coefficient is composed of a global intercept plus independent and spatial residual error terms via the Besag–York–Molly (BYM) model (Besag et al., 1991). Because this model includes both spatial and independently structured error terms, the data determined the extent of spatial smoothing employed, without requiring strong assumptions regarding residual spatial dependency. Further, this approach yields both a countywide global mean effect as well as census tract-level random coefficients indicating sub-regional (or census tract) effects of $PM_{2.5}$ on TLBW.

The variance structure of the spatial component of the BYM model requires specification of a spatial zero-one weight matrix of dimension J by J , where J is the number of census tracts. Each element i, j of the weight matrix is one if census tract i and j are adjacent to each other, and zero otherwise. The ‘*spdep*’ package (*spdep* package version 0.5-77 obtained September 30 2014) in R (Bivand et al., 2013; Bivand and Piras, 2015) was used to construct this neighborhood weight matrix and we assigned neighbors based on queens adjacency, which is defined as any neighboring census tract with a shared edge or vertex for a given area (i.e. census tract).

In fitting the model, we took advantage of the computational efficiency of Integrated Nested Laplace Approximations (INLA, version 0.0-1420281647) estimation techniques as implemented in the well-established R-INLA package (Rue et al., 2015), which has been used in several recent studies of large dimensions (Bennett et al., 2014; Castelló et al., 2013; Lee et al., 2013; Lee and Mitchell, 2014). The INLA approach avoids the computational burden related to typical Markov Chain Monte Carlo techniques (Gilks et al., 1998) often used to fit Bayesian spatial models and allows accurate approximations to posterior marginal distributions of the model parameters (Grilli et al., 2014).

In the implementation of our model using R-INLA, the sub-regional-level air pollution effects consist of an overall fixed effect (that represents the overall mean effect) plus spatial and independent random residual effects as defined in the BYM model. (Rue et al., 2014, 2009; Martino and Rue, 2009). Hence, each Sub-Regional air pollution effect is then obtained as the sum of the overall fixed effect plus spatial and non-spatial census tract-level residual terms via the linear combination feature in R-INLA. This allows us to obtain a posterior distribution for each Sub-Regional-level air pollution effect, β_j , and to examine the spatial distribution of these effects throughout L.A. County.

The full model specification is presented in Eqs. (1) and (2) below. Our first-level logistic-regression model is,

$$\text{logit}(y_i) = V\eta' + \beta_{z_i}x_i \tag{1}$$

where y_i denotes the logit probability of TLBW for individual i , $V\eta'$ represents individual-level covariates V and associated fixed effect coefficients η' , β_{z_i} represents sub-regional random effects of exposure, and x_i denotes individual-level $PM_{2.5}$ exposure. Note that $z_i = j$ indicates the census tract j to which individual i belongs, so if, say, individual 3 is in census tract number 12, then $z_3 = 12$, and $\beta_{z_3} = \beta_{12}$. There are therefore $\beta_j, j = 1, \dots, J$ effects of $PM_{2.5}$ on log-odds of TLBW corresponding to each census tract, j .

We model the effects of $PM_{2.5}$ on TLBW for each census tract, j , as

$$\beta_j = \gamma_0 + S_j + \epsilon_j^\beta \tag{2}$$

where γ_0 is the overall region-wide $PM_{2.5}$ effect, and S_j and ϵ_j^β denote spatial and independent residual error terms, respectively, with the restriction $\sum_j S_j = 0$ imposed for indefinability reasons. While the independent error term is defined in the standard way as $\epsilon_j^\beta \sim N(0, \sigma_\beta^2)$, the spatial error term is defined as,

$$S_{j|k \neq j} \sim N\left(\frac{\sum_{k \neq j} w_{jk} S_k}{\sum_{k \neq j} w_{jk}}, \frac{\tau^2}{\sum_{k \neq j} w_{jk}}\right),$$

where the weights $w_{j,k}$ are elements of the zero-one neighborhood adjacency matrix defined to be equal to one when census tracts i, k are adjacent and zero otherwise. This approach has been successfully employed in a variety of exposure/health association studies. (see, for example, Molitor et al. (2007).)

2.5. Mapping

Estimates of the posterior quantities correspond to the adjusted random air pollution effects from the multilevel model were imported into ArcGIS 10.1 (ESRI, Redlands, CA) and merged with census tract boundary shapefiles to create exposure–response census tract-level ‘effect maps’. In addition to mapping the multilevel adjusted census tract mean effects, the R-INLA package includes the ‘*inla.pmargin*’ function that computes probabilities from the posterior distribution of the marginal random effects as obtained from the linear combinations described above. This enabled us to map the marginal probabilities that a given census tract random effect coefficient lies above zero, $P(\beta_j > 0)$. Similarly,

we mapped the probability that a given census tract random effect coefficient is above the adjusted global mean effect, $P(\beta_j > \bar{\beta})$. Computation of these probabilities help illustrate where associations between PM_{2.5} and TLBW are most likely to occur (see Supplemental Materials for requisite R-INLA code needed to obtain posterior probabilities). Thus, our 'effect maps' depict probabilities that the PM_{2.5} census tract-specific exposure–response (β_j) lies above zero (or an OR above 1) and the probability that a census tract-specific air pollution effect deviates from the overall average ($P(\beta_j > \bar{\beta})$).

3. Results

3.1. Descriptive analyses

Between 1995–2006 the overall prevalence of TLBW was 2.1% and the average PM_{2.5} exposure was 17.04 μg/m³ (interquartile range=16.25, 18.21). The spatial distribution of PM_{2.5} concentrations indicated that exposures were highest within the urban core of LA County, specifically the southern, eastern, and northwest portions of urban LA (Supplemental Materials, Fig. S1). Risk factors that were associated with TLBW included maternal age, race, level of education, parity, gestation length (days), gestation squared, sex of the infant (Table 1), and were adjusted for in the following models.

3.2. PM_{2.5} regression analyses

3.2.1. Standard logistic model

The final statistical analyses included 1,356,304 births from 2,033 LA County census tracts. In unadjusted fixed effects logistic regression, the odds of TLBW was 23.2% higher (OR=1.23 [95%CI: 1.16, 1.30]) per 10 μg/m³ increase of PM_{2.5}. After adjusting for maternal age, race-ethnicity, education, parity, and infant gestation and sex, a 10 μg/m³ increase in PM_{2.5} exposure remained associated with statistically significant increased odds of TLBW (OR=1.17; 95%CI=1.10–1.24)(Table 2). The fully adjusted model results along with the model covariates are provided in detail

Table 1 Demographic characteristics overall and by TLBW and crude odds ratios for TLBW (N=1,359,284).

Parameter	Overall (N=1,359,284)		TLBW Cases (N=27,714)		Non cases (N=1,331,570)		Crude TLBW OR (95% CI)
	n	% or mean (95% CI)	n	% (95% CI)	n	% (95% CI)	
Gestational age (days)		Mean=278.91 (278.92, 278.89)					
Sex of infant							
Male	688,568	50.66 (50.57, 50.74)	11,890	42.90 (42.32, 43.49)	676,678	50.82 (50.73, 50.90)	1.00
Female	670,716	49.34 (49.26, 49.43)	15,824	57.10 (56.51, 57.68)	654,892	49.18 (49.10, 49.27)	1.38 (1.34, 1.41)
Maternal age							
< 20 years	143,265	10.54 (10.49, 10.59)	4090	14.76 (14.34, 15.18)	139,175	10.45 (10.40, 10.50)	1.00
20–24 years	318,122	23.40 (23.33, 23.47)	6959	25.11 (24.60, 25.62)	311,163	23.37 (23.30, 23.44)	0.76 (0.73, 0.79)
25–29 years	364,301	26.80 (26.73, 26.86)	6581	23.75 (23.25, 24.25)	357,720	26.86 (26.79, 26.94)	0.63 (0.60, 0.65)
30–34 years	322,341	23.71 (23.64, 23.79)	5674	20.47 (20.00, 20.95)	316,667	23.78 (23.71, 23.85)	0.61 (0.59, 0.64)
≥ 35 years	211,255	15.54 (15.48, 15.60)	4410	15.91 (15.48, 16.35)	206,845	15.55 (15.47, 15.60)	0.73 (0.69, 0.76)
Race–Ethnicity							
White	249,759	18.37 (18.31, 18.44)	3605	13.01 (12.61, 13.41)	246,154	18.49 (18.42, 18.55)	1.00
Hispanic	852,886	62.75 (62.66, 62.83)	16,260	58.67 (58.09, 59.25)	836,626	62.83 (62.75, 62.91)	1.33 (1.28, 1.38)
Black	107,237	7.89 (7.84, 7.93)	4175	15.06 (14.65, 15.49)	103,062	7.74 (7.69, 7.79)	2.77 (2.64, 2.89)
Asian	94,764	6.97 (6.93, 7.01)	2097	7.57 (7.26, 7.88)	92,667	6.96 (6.92, 7.00)	1.55 (1.46, 1.63)
Other	54,638	4.02 (3.99, 4.05)	1577	5.69 (5.42, 5.97)	53,061	3.98 (3.95, 4.02)	2.03 (1.91, 2.15)
Maternal education							
0–8 years	206,487	15.19 (15.13, 15.25)	4194	15.13 (14.71, 15.56)	202,293	15.19 (15.13, 15.25)	1.00
9–12 years	666,565	49.04 (48.95, 49.12)	14,867	53.64 (53.06, 54.23)	651,698	48.94 (48.86, 49.03)	1.10 (1.06, 1.14)
13–15 years	232,319	17.09 (17.03, 17.15)	4453	16.07 (15.64, 16.51)	227,866	17.11 (17.05, 17.18)	0.94 (0.90, 0.98)
≥ 16 years	253,913	18.68 (18.61, 18.75)	4200	15.15 (14.73, 15.58)	249,713	18.75 (18.69, 18.82)	0.81 (0.78, 0.85)
Parity							
0	522,598	38.45 (38.36, 38.53)	13,257	47.84 (47.25, 48.43)	509,341	38.25 (38.17, 38.33)	1.00
≥ 1	836,686	61.55 (61.47, 61.64)	14,457	52.16 (51.57, 52.75)	822,229	61.75 (61.67, 61.83)	0.68 (0.66, 0.69)

Table 2 Association between PM_{2.5} exposure and TLBW using standard and multilevel spatial regression methods (N=1,356,304).

Exposure	Standard model ^a OR (95% CI) ^b	Spatial multilevel model ^a OR (95% CI) ^c
PM _{2.5} (per 10 μg/m ³)	1.17 (1.10, 1.24) ^d	1.19 (1.02, 1.39)

^a Adjusted for sex of the infant, gestation age of infant, gestation age squared, maternal age, maternal race, maternal education level, and parity.

^b OR per interquartile range=1.03 (95% CI: 1.02, 1.04), IQR=1.96 μg/m³.

^c OR per interquartile range=1.05 (95% CI: 1.03, 1.08), IQR=1.96 μg/m³.

^d For all of LA County, including rural areas, OR=1.17 (1.10–1.24).

within the Supplementary Material (Table S1, Supplementary Material).

3.2.2. Multilevel spatial model

The multilevel spatial model provides PM_{2.5} coefficients on TLBW at a global county-wide level (Table 2) and at the census tract neighborhood level. The overall mean PM_{2.5} exposure–response estimate for our multilevel spatial model was similar in magnitude to the fixed effect logistic regression result (OR_{spatial}=1.19 versus OR_{fixed}=1.17). The two maps presented in Fig. 1 and Fig. 3 present the probability that a given census tract air pollution effect (with outcome on log-odds scale) is above zero (Fig. 1) and the probability that a given census tract effect is above the estimated overall mean effect (Fig. 3), while Fig. 2 presents the mean PM_{2.5} random effect per census tract.

For the probability effect map in Fig. 1, the census tracts in dark brown have a > 95% probability of an effect that is above zero ($P(\beta_j > 0)$). Thus, these areas represent census tracts where the PM_{2.5} exposure–response with TLBW is most likely to be positive. The dark brown neighborhoods in Fig. 3 have a > 95% probability for an effect above the county-wide (or “global”) mean effect. Hence, these areas represent census tracts that are most likely to exhibit a PM_{2.5} exposure–response that is greater in magnitude compared to the estimated mean exposure–response relationship, which we are considering to be 'hotspots' within the context of our study. The hotspots appear to be concentrated in census tracts within central and south-central LA County (Fig. 3).

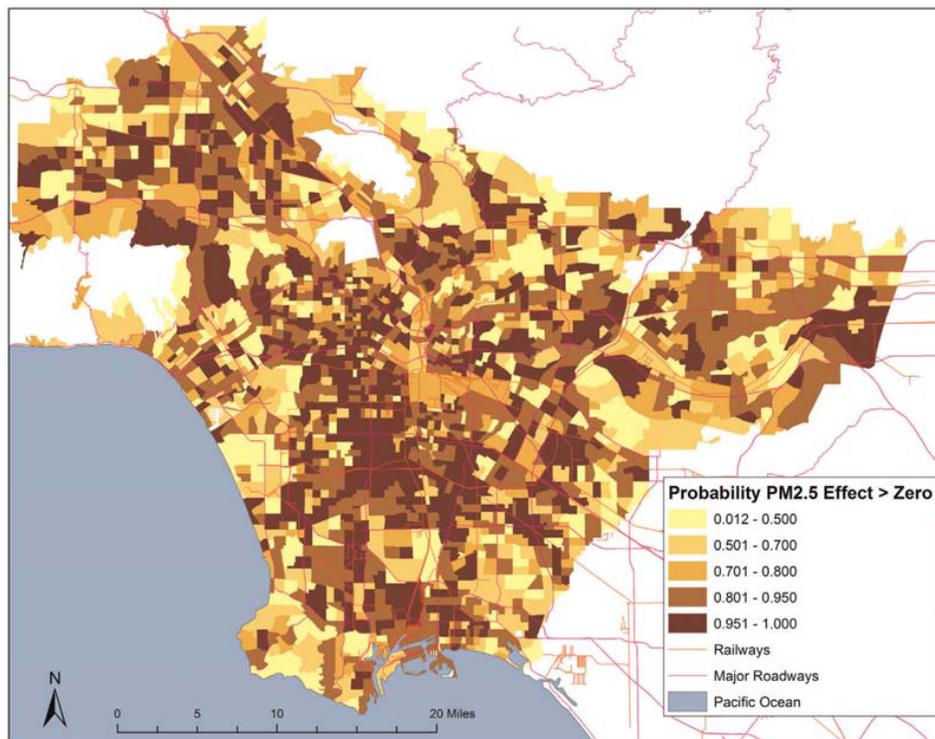


Fig. 1. Probability map for census tract PM_{2.5} effects for TLBW ($P(\beta_j > 0)$) after adjusting for maternal age, race-ethnicity, education, parity, and infant gestation + gestation squared, and infant sex.

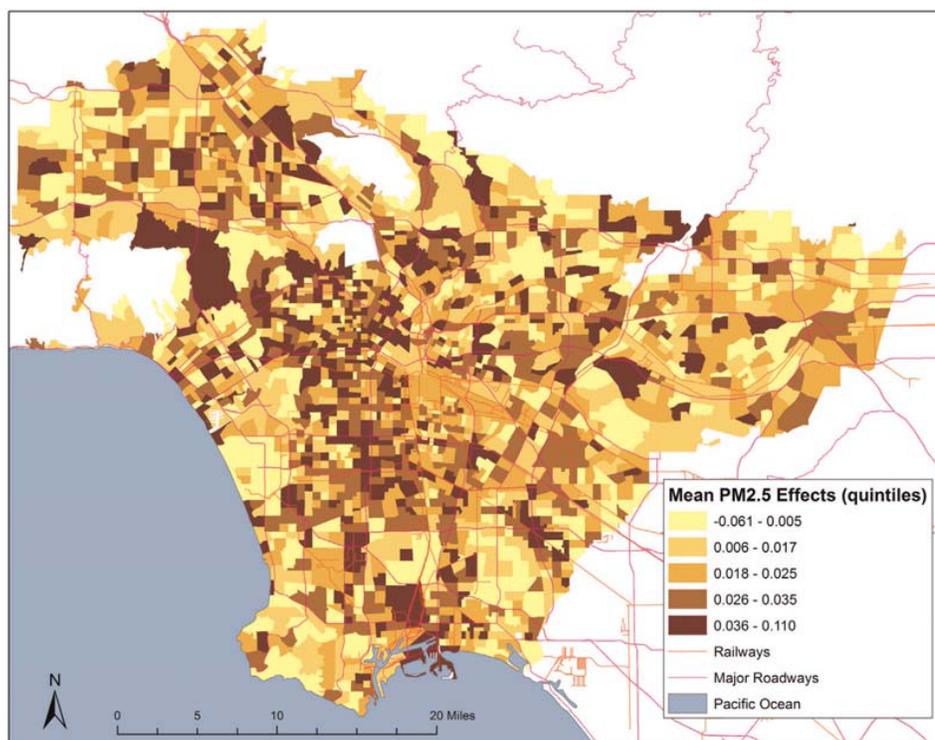


Fig. 2. Census tract PM_{2.5} effects for TLBW (mean) after adjusting for maternal age, race-ethnicity, education, parity, and infant gestation + gestation squared, and infant sex.

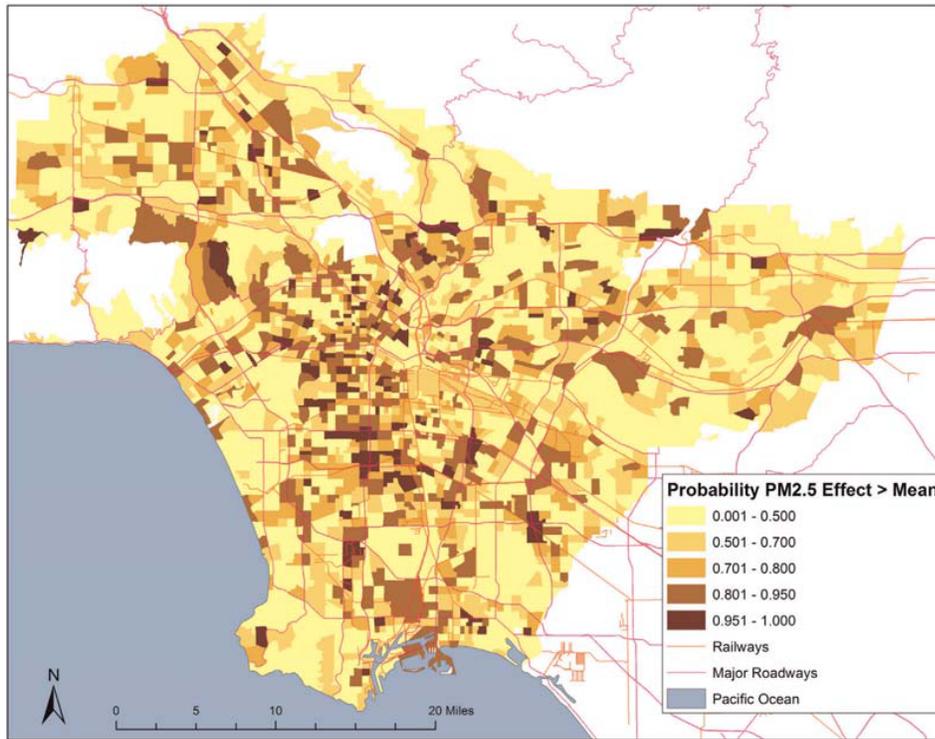


Fig. 3. Probability map for census tract PM_{2.5} effects for TLBW ($P(\beta_j > \bar{\beta}_j)$) after adjusting for maternal age, race-ethnicity, education, parity, and infant gestation + gestation squared, and infant sex.

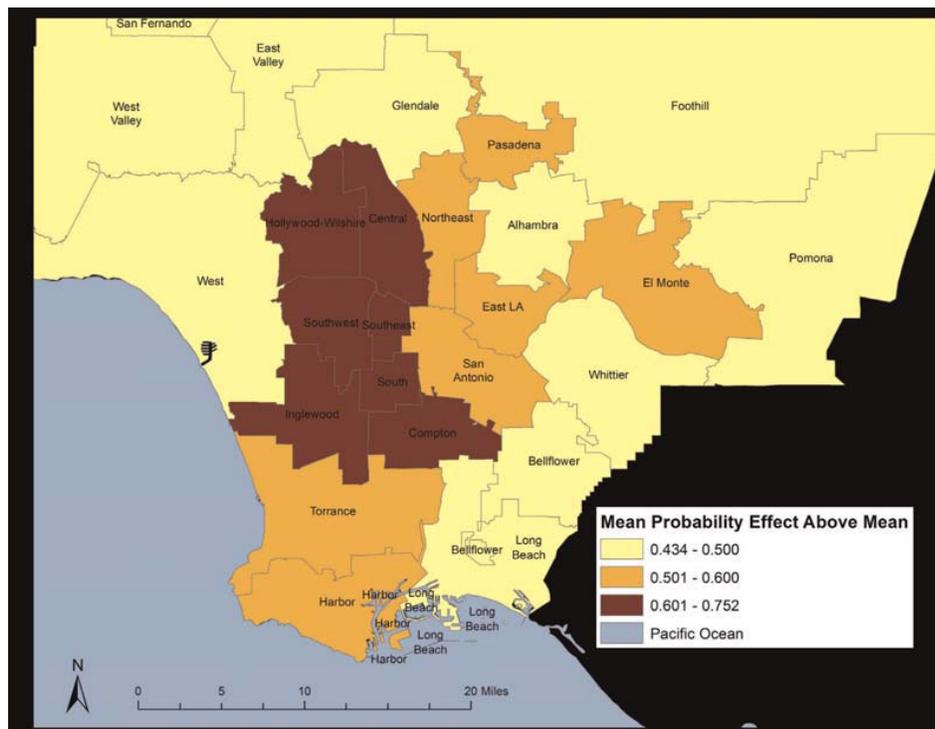


Fig. 4. Mean probabilities for census tract random effect above mean by LA County health districts.

3.2.3. LA health district summaries

LA County is composed of 26 health districts created from aggregates of census tract boundaries for the purposes of health assessments. Therefore, to highlight the observed spatial patterns in Fig. 3, from the posterior distribution of the marginal random effects we calculated and mapped the average probabilities for LA County health districts with respect to tract-level probabilities above the overall mean PM_{2.5} coefficient. These numerical summaries are simply descriptive since they were acquired by calculating the mean tract-level probabilities. Health districts of LA urban core, including Central, Compton, Hollywood-Wilshire, Inglewood, South, Southeast, and Southwest health districts, are characterized by the highest probabilities that the air pollution effect coefficients are above the overall mean coefficient (Fig. 4). Thus the map suggests effect 'hotspots' are concentrated within these health districts, which are generally lower income and non-white in terms of race-ethnicity (see Supplemental Materials, Figs. S2 and S3).

4. Discussion

4.1. Key findings

We applied Bayesian multilevel spatial modeling to examine whether the exposure–response relationship between PM_{2.5} and TLBW varies spatially. Consistent with previous findings from LA County (Ghosh et al., 2012; Ritz et al., 2007; Wilhelm et al., 2011), we observed an overall relationship between increasing PM_{2.5} exposure and increasing risk of TLBW. More important, we observed substantive variations across census tracts within LA County in the exposure–response between PM_{2.5} and TLBW. Higher probabilities for positive PM_{2.5} effects were mostly concentrated in central LA and south central LA sub-regions. Relative to the mean regional PM_{2.5} effect on the log odds of TLBW, several census tracts located in central LA and south-central LA exhibited higher exposure–response relationships in terms of effect size and posterior probabilities for effects above the mean ($P(\beta_j > \bar{\beta}) > 0.95$). These observations suggested that PM_{2.5} related adverse effects on birth weight may be modified by place.

A number of plausible explanations may account for the spatial patterning in the exposure–response between PM_{2.5} exposures and TLBW observed in our study. Firstly, regionally varying and spatially correlated neighborhood contextual factors may enhance exposure gradients within an urban setting and other spatially structured individual factors may further create susceptibility to adverse birth outcome by interacting with PM_{2.5}. Regionally varying and overlapping aspects of neighborhoods with the potential to enhance exposure to air pollutants or susceptibility to air pollution related health effects may include (but are not limited to): built environment factors (i.e. age of homes, homes set back further from the curb along heavily trafficked roadways) (Ponce, 2005; Ramachandran et al., 2003); spatially correlated variation in the types of PM_{2.5} sources (e.g. large truck traffic) and thus PM_{2.5} component mixtures (Laurent et al., 2014; Wilhelm et al., 2011); the presence of older and higher pollution emitting vehicles, and neighborhood SES (Ponce, 2005). For example, Singer and Harley (2000) observed that older vehicles tended to emit higher air pollutant levels relative to newer vehicles within the LA area, and that vehicular emissions tended to be higher in low income areas compared to higher income areas (even for vehicles of the same age). Individual-level differences that display spatial clustering may also partially explain spatial patterns in birth outcomes risks; such as psychosocial (Ghosh et al., 2010), occupational (Horner and Mefford, 2007; Ritz et al., 2007), or nutritional factors (Jedrychowski et al., 2010; Lane et al., 2008), as well as individual

home environments (i.e. home insulation or access to air conditioning (Ghosh et al., 2013; Jerrett et al., 2005; Ponce, 2005)). For instance, Ritz et al. (2007) found that parous women in LA without an occupation outside the home during the last 6 weeks of the pregnancy who were highly exposed to traffic-related air pollution had higher odds for preterm birth than exposed parous women working outside the home, illustrating the potential impact of exposure misclassification when using a home address. In another study we conducted in LA (Ponce, 2005) individuals' access to health insurance and their race, as well as neighborhood level factors such as SES and the physical environment (i.e. proximity to air pollution-related traffic and winter season) acted in concert to increase susceptibility to adverse pregnancy outcomes across LA county census tracts. Taken together this suggests a rather complex set of individual- and neighborhood-level social, cultural and environmental contributors to adverse birth outcomes that vary over space and may act on different biologic pathways to impair growth of the fetus resulting in TLBW, as suggested by the spatially varying effects estimated in our study.

In addition to spatial clustering of neighborhood and individual determinants and effect measure modifiers for birth outcomes, multi-pollutant mixtures in urban areas may create gradients in effects between Sub-Regions (Levy et al., 2013; Novák et al., 2014). While multi-pollutant mixtures may be more toxic in terms of birth outcomes, our study did not explicitly account for pollutant mixtures. While inclusion of a spatial random effects term may have mitigated this limitation to some extent – since multiple pollutant profiles have been observed to be clustered spatially (Austin et al., 2012) – this is an important limitation of this study. Furthermore, it cannot be ruled out that neighborhood-level and individual-level susceptibility and pollutant mixtures co-occur and together contribute to the observed spatially varying effect estimates seen in our study. Within regions of CA, such geographic-based susceptibility may be particularly acute. For instance, countywide studies in three California counties (Alameda, LA and San Diego) found that, while concentrations of individual pollutants such as diesel PM, NO₂, and PM_{2.5} were statistically significantly higher within socio-economically disadvantaged compared to less disadvantaged communities, when cumulative exposures to diesel PM, NO₂, and PM_{2.5} were considered, the relationship between SES and exposure was stronger (Su et al., 2012). Overlap of environmental and SES risk factors that can enhance neighborhood-level susceptibility has been reported previously (Jerrett and Finkelstein, 2005; Morello-Frosch and Shenassa, 2006).

4.2. Spatial dependency, air pollution, and birth outcome studies

A multilevel spatial hierarchical modeling approach is established as a flexible means of addressing spatial structure in the exposure–response relationship between air pollution and health effects (Boehm Vock et al., 2014; Dominici et al., 2000; Lee et al., 2013) and may therefore highlight notable localized effects (Chakraborty, 2012; Dominici et al., 2000; Earnest et al., 2007). A major statistical advantage gained in using this approach to modeling a spatially-structured exposure–response relationship is to maximize statistical power by using data in all sub-regions to inform the analysis, rather than calculating separate regression models for each sub-region (Gelman and Hill, 2006). Multilevel modeling approaches which incorporate spatial smoothing allow information from nearby regions to potentially exert more weight and influence compared to distant regions (Banerjee et al., 2004; Zhuoqing, 2000).

A strength of our approach is the inclusion of individual-level pollutant effect estimates that are modeled with spatial structure at the census-tract level. Some air pollution and birth outcome studies have accounted for spatial dependency in the residuals, but still

assume a global effect due to exposure (Berrocal et al., 2011; Castelló et al., 2013; Thompson et al., 2014; Williams et al., 2007). A spatial correlated autoregressive (CAR) model has been applied by (Berrocal et al., 2011) to examine the effect of CT-level $PM_{2.5}$ on continuous birth weight in North Carolina. An important distinction between the present study and Berrocal et al (2011) is that we applied a spatially structured random air pollution effect term, whereas Berrocal et al (2011) implemented a random intercept and did not explore the possibility of geographic disparities in the PM exposure–response relationship. A study by Thompson et al. (2014) examined the exceedance probability of very LBW risks in relation to proximity to National Priorities List Superfund Sites in Texas by modeling the spatially structured error term using Poisson regression. This study, however, used aggregated outcomes for a given geographic area and did not include individual-level air pollution estimates of exposure. A study conducted in Spain that examined municipal-level risks of PTB and LBW with proximity to different types of industries modeled spatially varying effects using Poisson regression with a spatial error term and an unstructured error term (Castelló et al., 2013). A major difference in the Castelló et al. (2013) study is that these researchers, again, used aggregated outcome data and did not relate birth outcomes with individual-level estimates of air pollutant exposures. A study by Williams et al. (2007) applied a linear hierarchical random effects model with spatially unstructured random coefficients and found substantial variation across census tracts regarding the estimated effects of maternal residential proximity to hazardous air pollution sites for reducing average birth weight. Our results also found varying effects by census tract; however, Williams et al (2007) did not use air pollution estimates but rather the proxy measure of spatial proximity to hazardous air pollution emitting sites and did not apply spatial structure to the random coefficients. While it is clear from these studies that multilevel modeling is capable of revealing important spatial processes regarding air pollution-related reductions in birth weight; our work goes beyond previous findings by not only applying spatial structure to pollutant effects but illustrating spatially varying effects while adjusting for individual level confounders.

4.3. Study limitations

Our study is limited by the presence of unmeasured confounders. Most notably we lack information on maternal smoking or maternal exposure to indoor smoking. However, our previous research (Ritz et al., 2007) found that adjustment for maternal or household smoking did not alter the strength of air pollution effects on adverse birth outcomes in LA County. Our study also did not account for spatially varying housing characteristics (e.g. age of housing stock, substandard housing, or lack of air conditioning) that could potentially exacerbate gradients in intra-urban exposures; even between neighborhoods with similar ambient PM concentrations (Baxter et al., 2007; Burgos et al., 2013; Clougherty et al., 2011; Jerrett and Finkelstein, 2005; Lv and Zhu, 2013; Meng et al., 2005; Ramachandran et al., 2003; Reid et al., 2009). Additionally, PM-related birth outcome risks may be modified by individual-level or neighborhood-level susceptibility factors that are often spatially patterned, such as SES, racial–ethnic status, maternal body mass index, maternal nutrition status, and other adverse neighborhood conditions, e.g., poor access to healthy foods or green spaces (English et al., 2003; Hystad et al., 2014; Jedrychowski et al., 2010; Kannan et al., 2006; Lakshmanan et al., 2015; Lane et al., 2008; Laurent et al., 2014; Ponce, 2005; Schempf et al., 2009).

While the $PM_{2.5}$ LUR estimates in our study best represents the spatial contrasts of chronic exposures at maternal residences throughout LA county, our estimates lacked the temporal resolution to consider exposures during specific pregnancy time periods. This limitation may obscure important biologic differences with regard

to birth outcome risks associated with different trimester exposure windows. Studies that have relied upon nearest site monitors for $PM_{2.5}$ estimation (Ghosh et al., 2012; Wilhelm et al., 2011) are better equipped to capture the temporal contrasts in maternal exposures, however, these studies lacked the spatial resolution to assess spatially varying effects of $PM_{2.5}$. For instance, while $PM_{2.5}$ may be fairly homogenous over a large region, it is likely that local-scale sources of $PM_{2.5}$ pollution carry greater importance when examining spatially varying TLBW effects (Laurent et al., 2014, 2013). Therefore, it was determined that the value in obtaining high spatial resolution was an acceptable temporal tradeoff, given the nature of our research question. Furthermore, we are confident in the ability of our exposure model to assess TLBW risks since our overall fixed effect $PM_{2.5}$ exposure–response estimate was consistent in terms of effect size when compared with previous research findings (Dadvand et al., 2013; Ghosh et al., 2012; Hyder et al., 2014; Laurent et al., 2014; Stieb et al., 2012; Wilhelm et al., 2011). For example, in the present study, we found an OR of 1.03 per IQR increase in maternal $PM_{2.5}$ exposure (Table 2). Ghosh et al. (2012) estimated maternal $PM_{2.5}$ concentrations, using an inverse distance weighting procedure based on governmental air monitoring stations for the years 2000–2006 in LA County, and found an OR of 1.04 per interquartile range (IQR) increase for entire pregnancy $PM_{2.5}$ exposure. Recently, Laurent et al. (2014) estimated an OR of 1.025 per IQR increase in maternal $PM_{2.5}$ exposure for LA County births between 2001 and 2008. Notably, Laurent et al. (2014) found that gasoline $PM_{2.5}$ exposure imparted the highest risk of TLBW compared to all other sources of $PM_{2.5}$ within LA. In a separate $PM_{2.5}$ and birth outcomes study, (Dadvand et al., 2013) pooled multiple $PM_{2.5}$ and TLBW analyses from seven different country study sites, despite large heterogeneity between the country-specific $PM_{2.5}$ effect estimates, they estimated a 10% (95%CI: 3%, 18%) adjusted increased odds of TLBW for a 10-unit increase in $PM_{2.5}$ exposure, which is comparable to our finding of a 17% increase per 10-unit increase in $PM_{2.5}$ exposure (Table 2).

4.4. Public health implications

Findings from our research is highly relevant to environmental health disparities and regulatory policy. First of all, our study implies that uniform regulatory standards geared towards reducing public health impacts from air pollution may not be sufficiently protective of susceptible sub-populations, and that such policies may need to be spatially tailored to protect these sub-populations. Secondly, our approach could identify 'hotspots' to help guide spatially targeted public health interventions intended to protect susceptible sub-populations from outdoor air pollution health effects (e.g., for example, by installing HEPA filters and air conditioning to reduce indoor exposures). Lastly, while our study found large within-county differences in effect estimates and thus the potential for $PM_{2.5}$ effect 'hotspots', additional data on potential modifying factors by neighborhood (i.e. $PM_{2.5}$ composition or neighborhood food environment) are needed to more fully explain the causes for this apparent spatial variation in the exposure–response relationship between $PM_{2.5}$ and TLBW.

5. Conclusion

We found that maternal exposure to $PM_{2.5}$ was associated with higher odds of TLBW in LA County. Moreover, our results indicate that the spatial patterning of the exposure–response relationship for $PM_{2.5}$ and TLBW needed to be considered. While previous research conducted in LA County has found variation of pollutant effects on adverse birth outcomes based on neighborhood factors

such as SES, our results take these previous findings a step further by identifying neighborhood TLBW 'hotspots' most likely to be affected negatively by air pollution. Also, compared to global effect estimates, our findings suggest the potential value of modeling spatial random air pollution effect coefficients in identifying disproportionately impacted communities as well the relative probability of localized exposure–response estimates. Finally, additional research is needed in hotspot areas to explore which spatially-based factors may help to better understand these differences between neighborhoods.

Acknowledgments

Research described in this article was conducted under contract to the Health Effects Institute (HEI), an organization jointly funded by the United States Environmental Protection Agency (EPA) (Assistance Award No. R-82811201) and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI, or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.envres.2015.06.044>.

References

- Austin, E., Coull, B., Thomas, D., Koutrakis, P., 2012. A framework for identifying distinct multipollutant profiles in air pollution data. *Environ. Int.* 45, 112–121. <http://dx.doi.org/10.1016/j.envint.2012.04.003>.
- Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC, Boca Raton, Florida.
- Baxter, L.K., Clougherty, J.E., Laden, F., Levy, J.L., 2007. Predictors of concentrations of nitrogen dioxide, fine particulate matter, and particle constituents inside of lower socioeconomic status urban homes. *J. Expo. Sci. Environ. Epidemiol.* 17, 433–444. <http://dx.doi.org/10.1038/sj.jes.7500532>.
- Beckerman, B.S., Jerrett, M., Martin, R.V., van Donkelaar, A., Ross, Z., Burnett, R.T., 2013. Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California. *Atmos. Environ.* 77, 172–177. <http://dx.doi.org/10.1016/j.atmosenv.2013.04.024>.
- Bennett, A., Yukich, J., Miller, J.M., Vounatsou, P., Hamainza, B., Ingwe, M.M., Moonga, H.B., Kamuliwo, M., Keating, J., Smith, T.A., Steketee, R.W., Eisele, T.P., 2014. A methodological framework for the improved use of routine health system data to evaluate national malaria control programs: evidence from Zambia. *Popul. Health Metr.* 12, 1–11. <http://dx.doi.org/10.1186/s12963-014-0030-0>.
- Berrocal, V.J., Gelfand, A.E., Holland, D.M., Burke, J., Miranda, M.L., 2011. On the use of a $PM_{2.5}$ exposure simulator to explain birthweight. *Environmetrics* 22, 553–571. <http://dx.doi.org/10.1002/env.1086>.
- Besag, J., York, J.C., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Stat. Math.* 43, 1–59.
- Bivand, R., Hauke, J., Kossowski, T., 2013. Computing the Jacobian in Gaussian spatial autoregressive models: an illustrated comparison of available methods: computing the Jacobian in spatial autoregressive models. *Geogr. Anal.* 45, 150–179. <http://dx.doi.org/10.1111/gean.12008>.
- Bivand, R., Piras, G., 2015. Comparing implementations of estimation methods for spatial econometrics. *J. Stat. Softw.* 63, 1–36.
- Boehm Vock, L.F., Reich, B.J., Fuentes, M., Dominici, F., 2014. Spatial variable selection methods for investigating acute health effects of fine particulate matter components. *Biometrics* 71, 167–177. <http://dx.doi.org/10.1111/biom.12254>.
- Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., Karr, C., 2008. A Cohort study of traffic-related air pollution impacts on birth outcomes. *Environ. Health Perspect.* 116, 680–686. <http://dx.doi.org/10.1289/ehp.10952>.
- Burgos, S., Ruiz, P., Koifman, R., 2013. Changes to indoor air quality as a result of relocating families from slums to public housing. *Atmos. Environ.* 70, 179–185. <http://dx.doi.org/10.1016/j.atmosenv.2012.12.044>.
- Castelló, A., Río, I., García-Pérez, J., Fernández-Navarro, P., Waller, L.A., Clennon, J.A., Bolívar, F., López-Abente, G., 2013. Adverse birth outcomes in the vicinity of industrial installations in Spain 2004–2008. *Environ. Sci. Pollut. Res.* 20, 4933–4946. <http://dx.doi.org/10.1007/s11356-012-1444-5>.
- Chakraborty, J., 2012. Cancer risk from exposure to hazardous air pollutants: spatial and social inequities in Tampa Bay, Florida. *Int. J. Environ. Health Res.* 22, 165–183. <http://dx.doi.org/10.1080/09603123.2011.628643>.
- Choi, J., Fuentes, M., Reich, B.J., 2009. Spatial–temporal association between fine particulate matter and daily mortality. *Comput. Stat. Data Anal.* 53, 2989–3000. <http://dx.doi.org/10.1016/j.csda.2008.05.018>.
- Clougherty, J.E., Houseman, E.A., Levy, J.L., 2011. Source apportionment of indoor residential fine particulate matter using land use regression and constrained factor analysis: Indoor-source apportionment using LUR and factor analysis. *Indoor Air* 21, 53–66. <http://dx.doi.org/10.1111/j.1600-0668.2010.00682.x>.
- Dadvand, P., Parker, J., Bell, M.L., Bonzini, M., Brauer, M., Darrow, L.A., Gehring, U., Glinianaia, S.V., Gouveia, N., Ha, E., Leem, J.H., van den Hooven, E.H., Jalaludin, B., Jesdale, B.M., Lepeule, J., Morello-Frosch, R., Morgan, G.G., Pesatori, A.C., Pierik, F.H., Pless-Mulloli, T., Rich, D.Q., Sathyanarayana, S., Seo, J., Slama, R., Strickland, M., Tamburic, L., Wartenberg, D., Nieuwenhuijsen, M.J., Woodruff, T. J., 2013. Maternal exposure to particulate air pollution and term birth weight: a multi-country evaluation of effect and heterogeneity. *Environ. Health Perspect.* 121, 267–373. <http://dx.doi.org/10.1289/ehp.1205575>.
- Dominici, F., Samet, J.M., Zeiger, S., 2000. Combine evidence on air pollution and daily mortality from the 20 largest us cities: a hierarchical modelling strategy. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* 163, 263–302.
- Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R., Beard, J., 2007. Evaluating the effect of neighbourhood weight matrices on smoothing properties of conditional autoregressive (CAR) models. *Int. J. Health Geogr.* 6, 54. <http://dx.doi.org/10.1186/1476-072X-6-54>.
- English, P.B., Kharrazi, M., Davies, S., Scalf, R., Waller, L., Neutra, R., 2003. Changes in the spatial pattern of low birth weight in a southern California county: the role of individual and neighborhood level factors. *Soc. Sci. Med.* 56, 2073–2088. [http://dx.doi.org/10.1016/S0277-9536\(02\)00202-2](http://dx.doi.org/10.1016/S0277-9536(02)00202-2).
- Everitt, B., Hothorn, T., 2010. A Handbook of Statistical Analyses Using R, 2nd ed. CRC Press, Boca Raton.
- Fleischer, N.L., Merialedi, M., van Donkelaar, A., Vadillo-Ortega, F., Martin, R.V., Betran, A.P., Souza, J.P., O'Neill, M.S., 2014. Outdoor air pollution, preterm birth, and low birth weight: analysis of the world health organization global survey on maternal and perinatal health. *Environ. Health Perspect.* 122, 425–430. <http://dx.doi.org/10.1289/ehp.1306837>.
- Fuentes, M., Song, H.-R., Ghosh, S.K., Holland, D.M., Davis, J.M., 2006. Spatial association between speciated fine particles and mortality. *Biometrics* 62, 855–863. <http://dx.doi.org/10.1111/j.1541-0420.2006.00526.x>.
- Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models, 1st edition. Cambridge University Press, Cambridge; New York.
- Ghosh, J.K.C., Wilhelm, M.H., Dunkel-Schetter, C., Lombardi, C.A., Ritz, B.R., 2010. Paternal support and preterm birth, and the moderation of effects of chronic stress: a study in Los Angeles County mothers. *Arch. Women's Ment. Health* 13, 327–338. <http://dx.doi.org/10.1007/s00737-009-0135-9>.
- Ghosh, J.K.C., Wilhelm, M., Ritz, B., 2013. Effects of residential indoor air quality and household ventilation on preterm birth and term low birth weight in Los Angeles County, California. *Am. J. Public Health* 103, 686–694. <http://dx.doi.org/10.2105/AJPH.2012.300987>.
- Ghosh, J.K.C., Wilhelm, M., Su, J., Goldberg, D., Cockburn, M., Jerrett, M., Ritz, B., 2012. Assessing the influence of traffic-related air pollution on risk of term low birth weight on the basis of land-use-based regression models and measures of air toxics. *Am. J. Epidemiol.* 175, 1262–1274. <http://dx.doi.org/10.1093/aje/kwr469>.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1998. Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, Boca Raton, FL.
- Goldberg, D.W., Wilson, J.P., Knoblock, C.A., Ritz, B., Cockburn, M.G., 2008. An effective and efficient approach for manually improving geocoded data. *Int. J. Health Geogr.* 7, 60. <http://dx.doi.org/10.1186/1476-072X-7-60>.
- Grilli, L., Metelli, S., Rampichini, C., 2014. Bayesian estimation with integrated nested Laplace approximation for binary logit mixed models. *J. Stat. Comput. Simul.*, 1–9. <http://dx.doi.org/10.1080/00949655.2014.935377>.
- Hajat, A., Diez-Roux, A.V., Adar, S.D., Auchincloss, A.H., Lovasi, G.S., O'Neilles, M.S., Sheppard, L., Kaufman, J.D., 2013. Air pollution and individual and neighborhood socioeconomic status: evidence from the multi-ethnic study of atherosclerosis (MESA). *Environ. Health Perspect.* 121, 1325–1333. <http://dx.doi.org/10.1289/ehp.1206337>.
- Hao, Y., Strosnider, H., Balluz, L., Qualters, J.R., 2015. Geographic variation in the association between ambient fine particulate matter ($PM_{2.5}$) and term low birth weight in the United States. *Environ. Health Perspect.* [Advanced Publication], 1–28. <http://dx.doi.org/10.1289/ehp.1408798>.
- Harris, G., Thompson, W.D., Fitzgerald, E., Wartenberg, D., 2014. The association of $PM_{2.5}$ with full term low birth weight at different spatial scales. *Environ. Res.* 134, 427–434. <http://dx.doi.org/10.1016/j.envres.2014.05.034>.
- Horner, M.W., Mefford, J.N., 2007. Investigating urban spatial mismatch using job – housing indicators to model home – work separation. *Environ. Plan. A* 39, 1420–1440. <http://dx.doi.org/10.1068/a37443>.
- Hyder, A., Lee, H.J., Ebisu, K., Koutrakis, P., Belanger, K., Bell, M.L., 2014. $PM_{2.5}$ exposure and birth outcomes: use of satellite- and monitor-based data. *Epidemiology* 25, 58–67. <http://dx.doi.org/10.1097/EDE.000000000000027>.
- Hystad, P., Davies, H.W., Frank, L., Van Loon, J., Gehring, U., Tamburic, L., Brauer, M., 2014. Residential greenness and birth outcomes: evaluating the influence of spatially correlated built-environment factors. *Environ. Health Perspect.* 122, 1095–1102. <http://dx.doi.org/10.1289/ehp.1308049>.
- Jedrychowski, W., Perera, F., Mrozek-Budzyn, D., Flak, E., Mroz, E., Sochacka-Tatara, E., Jacek, R., Kaim, I., Skolicki, Z., Spengler, J.D., 2010. Higher fish consumption in

- pregnancy may confer protection against the harmful effect of prenatal exposure to fine particulate matter. *Ann. Nutr. Metab.* 56, 119–126. <http://dx.doi.org/10.1159/000275918>.
- Jerrett, M., Burnett, R.T., Beckerman, B.S., Turner, M.C., Krewski, D., Thurston, G., Martin, R.V., van Donkelaar, A., Hughes, E., Shi, Y., Gapstur, S.M., Thun, M.J., Pope, C.A., 2013. Spatial analysis of air pollution and mortality in California. *Am. J. Respir. Crit. Care Med.* 188, 593–599. <http://dx.doi.org/10.1164/rccm.201303-0609OC>.
- Jerrett, M., Burnett, R.T., Ma, R., Pope, C.A., Krewski, D., Newbold, K.B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E.E., Thun, M.J., 2005. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 16, 727–736. <http://dx.doi.org/10.1097/01.ede.0000181630.15826.7d>.
- Jerrett, M., Finkelstein, M., 2005. Geographies of risk in studies linking chronic air pollution exposure to health outcomes. *J. Toxicol. Environ. Health A* 68, 1207–1242. <http://dx.doi.org/10.1080/15287390590936085>.
- Kannan, S., Misra, D.P., Dvovich, J.T., Krishnakumar, A., 2006. Exposures to airborne particulate matter and adverse perinatal outcomes: a biologically plausible mechanistic framework for exploring potential effect modification by nutrition. *Environ. Health Perspect.* 114, 1636–1642.
- Krewski, D., Jerrett, M., Burnett, R.T., Ma, R., Hughes, E., Shi, Y., Turner, M.C., Pope, C.A., Thurston, G., Calle, E.E., Thun, M.J., Beckerman, B., DeLuca, P., Finkelstein, N., Ito, K., Moore, D.K., Newbold, K.B., Ramsay, T., Ross, Z., Shin, H., Tempalski, B., 2009. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. *Res. Rep. Health Eff. Inst.* 140, 5–114, discussion 115–136.
- Lakshmanan, A., Chiu, Y.-H.M., Coull, B.A., Just, A.C., Maxwell, S.L., Schwartz, J., Gryparis, A., Kloog, I., Wright, R.J., Wright, R.O., 2015. Associations between prenatal traffic-related air pollution exposure and birth weight: modification by sex and maternal pre-pregnancy body mass index. *Environ. Res.* 137, 268–277. <http://dx.doi.org/10.1016/j.envres.2014.10.035>.
- Lane, S.D., Keefe, R.H., Rubinstein, R., Levandowski, B.A., Webster, N., Cibula, D.A., Boahene, A.K., Dele-Michael, O., Carter, D., Jones, T., Wojtowycz, M., Brill, J., 2008. Structural violence, urban retail food markets, and low birth weight. *Health Place* 14, 415–423. <http://dx.doi.org/10.1016/j.healthplace.2007.08.008>.
- Laurent, O., Hu, J., Li, L., Cockburn, M., Escobedo, L., Kleeman, M.J., Wu, J., 2014. Sources and contents of air pollution affecting term low birth weight in Los Angeles County, California, 2001–2008. *Environ. Res.* 134, 488–495. <http://dx.doi.org/10.1016/j.envres.2014.05.003>.
- Laurent, O., Wu, J., Li, L., Chung, J., Bartell, S., 2013. Investigating the association between birth weight and complementary air pollution metrics: a cohort study. *Environ. Health* 12, 18. <http://dx.doi.org/10.1186/1476-069X-12-18>.
- Lee, D., Mitchell, R., 2014. Controlling for localised spatio-temporal autocorrelation in long-term air pollution and health studies. *Stat. Methods Med. Res.* <http://dx.doi.org/10.1177/0962280214527384>.
- Lee, D., Rushworth, A., Sahu, S., 2013. A Bayesian localised conditional auto-regressive model for estimating the health effects of air pollution. University of Glasgow, University of Southampton.
- Levy, I., Mihele, C., Lu, G., Narayan, J., Brook, J.R., 2013. Evaluating Multipollutant Exposure and Urban Air Quality: Pollutant Interrelationships, Neighborhood Variability, and Nitrogen Dioxide as a Proxy Pollutant. *Environ. Health Perspect.* <http://dx.doi.org/10.1289/ehp.1306518>.
- Lv, J., Zhu, L., 2013. Effect of central ventilation and air conditioner system on the concentration and health risk from airborne polycyclic aromatic hydrocarbons. *J. Environ. Sci. China* 25, 531–536.
- Martino, S., Rue, H., 2009. Implementing Approximate Bayesian Inference using Integrated Nested Laplace Approximation: a manual for the inla program.
- McConnell, R., Islam, T., Shankardass, K., Jerrett, M., Lurmann, F., Gilliland, F., Gauderman, J., Avol, E., Künzli, N., Yao, L., Peters, J., Berhane, K., 2010. Childhood incident asthma and traffic-related air pollution at home and school. *Environ. Health Perspect.* 118, 1021–1026. <http://dx.doi.org/10.1289/ehp.0901232>.
- Meng, Q.Y., Turpin, B.J., Korn, L., Weisel, C.P., Morandi, M., Colome, S., Zhang, J.J., Stock, T., Spector, D., Winer, A., Zhang, L., Lee, J.H., Giovanetti, R., Cui, W., Kwon, J., Alimokhtari, S., Shendell, D., Jones, J., Farrar, C., Maberti, S., 2005. Influence of ambient (outdoor) sources on residential indoor and personal PM_{2.5} concentrations: analyses of RIOPA data. *J. Expo. Anal. Environ. Epidemiol.* 15, 17–28. <http://dx.doi.org/10.1038/sj.jea.7500378>.
- Messer, L.C., Kaufman, J.S., Dole, N., Herring, A., Laria, B.A., 2006. Violent crime exposure classification and adverse birth outcomes: a geographically-defined cohort study. *Int. J. Health Geogr.* 5, 22. <http://dx.doi.org/10.1186/1476-072X-5-22>.
- Molitor, J., Jerrett, M., Chang, C.C., Molitor, N.T., Gauderman, J., Berhane, K., McConnell, R., Lurmann, F., Wu, J., Winer, A., Thomas, D., 2007. Assessing uncertainty in spatial exposure models for air pollution health effects assessment. *Env. Health Perspect.* 115, 1147–1153.
- Molitor, J., Su, J.G., Molitor, N.-T., Rubio, V.G., Richardson, S., Hastie, D., Morello-Frosch, R., Jerrett, M., 2011. Identifying vulnerable populations through an examination of the association between multipollutant profiles and poverty. *Environ. Sci. Technol.* 45, 7754–7760. <http://dx.doi.org/10.1021/es104017x>.
- Morello-Frosch, R., Shenassa, E.D., 2006. The environmental “Riskscape” and social inequality: implications for explaining maternal and child health disparities. *Environ. Health Perspect.* 114, 1150–1153. <http://dx.doi.org/10.1289/ehp.8930>.
- Novák, J., Hilscherová, K., Landlová, L., Čupr, P., Kohút, L., Giesy, J.P., Klánová, J., 2014. Composition and effects of inhalable size fractions of atmospheric aerosols in the polluted atmosphere. Part II: in vitro biological potencies. *Environ. Int.* 63, 64–70. <http://dx.doi.org/10.1016/j.envint.2013.10.013>.
- Padula, A.M., Mortimer, K., Hubbard, A., Lurmann, F., Jerrett, M., Tager, I.B., 2012. Exposure to traffic-related air pollution during pregnancy and term low birth weight: estimation of causal associations in a semiparametric model. *Am. J. Epidemiol.* 176, 815–824. <http://dx.doi.org/10.1093/aje/kws148>.
- Parker, J.D., Rich, D.Q., Glinianaia, S.V., Leem, J.H., Warthenberg, D., Bell, M.L., Bonzini, M., Brauer, M., Darrow, L., Gehring, U., Gouveia, N., Grillo, P., Ha, E., van den Hooven, E.H., Jalaludin, B., Jesdale, B.M., Lepeule, J., Morello-Frosch, R., Morgan, G.G., Slama, R., Pierik, F.H., Pesatori, A.C., Sathyanarayana, S., Seo, J., Strickland, M., Tamburic, L., Woodruff, T.J., 2011. The international collaboration on air pollution and pregnancy outcomes: initial results. *Environ. Health Perspect.* 119, 1023–1028. <http://dx.doi.org/10.1289/ehp.1002725>.
- Pedersen, M., Gehring, U., Beelen, R., Wang, M., Giorgis-Allemand, L., Andersen, A.M.N., Basagaña, X., Bernard, C., Cirach, M., Forastiere, F., de Hoogh, K., Gražulevičienė, R., Gruzjeva, O., Hoek, G., Jedynska, A., Klümper, C., Kooter, I.M., Krämer, U., Kukkonen, J., Porta, D., Postma, D.S., Raaschou-Nielsen, O., van Rossem, L., Sunyer, J., Sørensen, M., Tsai, M.-Y., Vrijlkotte, T.G., Wilhelm, M., Nieuwenhuijsen, M.J., Pershagen, G., Brunekreef, B., Kogevinas, M., Slama, R., 2015. Elemental constituents of particulate matter and newborn’s size in eight European cohorts. *Environ. Health Perspect.* <http://dx.doi.org/10.1289/ehp.1409546>.
- Ponce, N.A., 2005. Preterm birth: the interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods. *Am. J. Epidemiol.* 162, 140–148. <http://dx.doi.org/10.1093/aje/kwi173>.
- Proietti, E., Röösli, M., Frey, U., Latzin, P., 2013. Air pollution during pregnancy and neonatal outcome: a review. *J. Aerosol Med. Pulm. Drug Deliv.* 26, 9–23. <http://dx.doi.org/10.1089/jamp.2011.0932>.
- Ramachandran, G., Adgate, J.L., Pratt, G.C., Sexton, K., 2003. Characterizing Indoor and Outdoor 15 Minute Average PM 2.5 Concentrations in Urban Neighborhoods. *Aerosol Sci. Technol.* 37, 33–45. <http://dx.doi.org/10.1080/02786820300889>.
- Reid, C., O’Neill, M., Gronlund, C., Brines, S., Brown, D., Diez-Roux, A., Schwartz, J., 2009. Mapping community determinants of heat vulnerability. *Environ. Health Perspect.* <http://dx.doi.org/10.1289/ehp.0900683>.
- Ritz, B., Wilhelm, M., 2008. Ambient air pollution and adverse birth outcomes: methodologic issues in an emerging field. *Basic Clin. Pharmacol. Toxicol.* 102, 182–190. <http://dx.doi.org/10.1111/j.1742-7843.2007.00161.x>.
- Ritz, B., Wilhelm, M., Hoggatt, K.J., Ghosh, J.K.C., 2007. Ambient air pollution and preterm birth in the environment and pregnancy outcomes study at the university of California, Los Angeles. *Am. J. Epidemiol.* 166, 1045–1052. <http://dx.doi.org/10.1093/aje/kwm181>.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71, 319–392. <http://dx.doi.org/10.1111/j.1467-9868.2008.00700.x>.
- Rue, H., Martino, S., Lindgren, F., Simpson, D., Riebler, A., Krainski, E., 2015. INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation.
- Rue, H., Martino, S., Lindgren, F., Simpson, D., Riebler, A., Krainski, E., 2014. Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation [WWW Document]. URL (<http://inla.googlecode.com/hg-history/default/rinla/DESCRIPTION>).
- Samoli, E., Analitis, A., Touloumi, G., Schwartz, J., Anderson, H.R., Sunyer, J., Bisanti, L., Zmirou, D., Vonk, J.M., Pekkanen, J., Goodman, P., Paldy, A., Schindler, C., Katsouyanni, K., 2004. Estimating the exposure-response relationships between particulate matter and mortality within the APHEA multicity project. *Environ. Health Perspect.* 113, 88–95. <http://dx.doi.org/10.1289/ehp.7387>.
- Schempf, A., Strobino, D., O’Campo, P., 2009. Neighborhood effects on birthweight: an exploration of psychosocial and behavioral pathways in Baltimore, 1995–1996. *Soc. Sci. Med.* 68, 100–110. <http://dx.doi.org/10.1016/j.socscimed.2008.10.006>.
- Shah, P.S., Balkhair, T., 2011. Air pollution and birth outcomes: a systematic review. *Environ. Int.* 37, 498–516. <http://dx.doi.org/10.1016/j.envint.2010.10.009>.
- Shankardass, K., McConnell, R., Jerrett, M., Milam, J., Richardson, J., Berhane, K., 2009. Parental stress increases the effect of traffic-related air pollution on childhood asthma incidence. *Proc. Natl. Acad. Sci. USA* 106, 12406–12411. <http://dx.doi.org/10.1073/pnas.0812910106>.
- Singer, B.C., Harley, R.A., 2000. A fuel-based inventory of motor vehicle exhaust emissions in the Los Angeles area during summer 1997. *Atmos. Environ.* 34, 1783–1795. [http://dx.doi.org/10.1016/S1352-2310\(99\)00358-1](http://dx.doi.org/10.1016/S1352-2310(99)00358-1).
- Stieb, D.M., Chen, L., Eshoul, M., Judek, S., 2012. Ambient air pollution, birth weight and preterm birth: a systematic review and meta-analysis. *Environ. Res.* 117, 100–111. <http://dx.doi.org/10.1016/j.envres.2012.05.007>.
- Su, J.G., Jerrett, M., Morello-Frosch, R., Jesdale, B.M., Kyle, A.D., 2012. Inequalities in cumulative environmental burdens among three urbanized counties in California. *Environ. Int.* 40, 79–87. <http://dx.doi.org/10.1016/j.envint.2011.11.003>.
- Thompson, J.A., Bissett, W.T., Sweeney, A.M., 2014. Evaluating geostatistical modeling of exceedance probability as the first step in disease cluster investigations: very low birth weights near toxic Texas sites. *Environ. Health Glob. Access Sci. Source* 13, 47. <http://dx.doi.org/10.1186/1476-069X-13-47>.
- Van Donkelaar, A., Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., Villeneuve, P.J., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ. Health Perspect.* 118, 847–855. <http://dx.doi.org/10.1289/ehp.0901623>.
- Wilhelm, M., Ghosh, J.K., Su, J., Cockburn, M., Jerrett, M., Ritz, B., 2011. Traffic-related air toxics and term low birth weight in Los Angeles County, California. *Environ. Health Perspect.* 120, 132–138. <http://dx.doi.org/10.1289/ehp.1103408>.

- Williams, B.L., Pennock-Román, M., Suen, H.K., Magsumbol, M.S., Ozdenerol, E., 2007. Assessing the impact of the local environment on birth outcomes: a case for HLM. *J. Expo. Sci. Environ. Epidemiol.* 17, 445–457. <http://dx.doi.org/10.1038/sj.jes.7500537>.
- Wu, J., Wilhelm, M., Chung, J., Ritz, B., 2011. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. *Environ. Res.* 111, 685–692. <http://dx.doi.org/10.1016/j.envres.2011.03.008>.
- Zhuoqiong, He, Sun, Dongchu, 2000. Hierarchical Bayes estimation of hunting success rates with spatial correlations. *Biometrics* 56, 360–367. <http://dx.doi.org/10.1111/j.0006-341X.2000.00360.x> 360–7.

ABOUT THE AUTHORS

John Molitor is a biostatistician and mathematical epidemiologist with specializations in the modeling of associations between air pollution exposures and health outcomes and in the field of genetic epidemiology. He has particular expertise in the development and application of advanced statistical methods to examine the joint effects of multiple pollutants on health. He has a Ph.D. in mathematical statistics from the University of Missouri, completed post-doctoral work at the University of Southern California, worked as a Lecturer at Imperial College, London, and is currently an associate professor of biostatistics in the College of Public Health at Oregon State University.

Eric Coker has several years of experience in the field of environmental epidemiology. He is currently a Ph.D. candidate at Oregon State University; his academic research examines air pollution effects on birth outcomes. Mr. Coker concurrently works as a practicing epidemiologist at the New Mexico Department of Health, Santa Fe, where his work involves data analysis for the National Environmental Public Health Tracking Network (Centers for Disease Control and Prevention). He also has an M.S. degree from the University of Washington, where his research focused on developing a reference spectral library for the measurement of air pollutants (NO₂, SO₂, and VOCs) using ultraviolet differential optical absorption spectroscopy. For his M.S. degree from the University of California–San Francisco, Coker examined the relationship between placental malaria infection and low-birth-weight infants in Uganda.

Michael Jerrett was the first to graduate from the collaborative M.A. program in political science and environmental studies at the University of Toronto in 1987. He subsequently completed a Ph.D. in geography at the University of Toronto and then worked for two years as a postdoctoral fellow in environmental health with John Eyles at McMaster University. Building on his specialties, Michael currently assesses air pollution–health associations in the United States and Canada, with special reference to geographic exposure models and social–spatial effect modifiers. He also pursues research in environmental accounting focusing on the determinants of and evaluation of environmental costs and benefits. He has designed and analyzed local, provincial, state, and national-level health and environment databases in North America, Europe, and Asia. He is currently Professor and Chair, Department of Environmental Health Sciences, UCLA Fielding School of Public Health.

Beate Ritz joined the faculty of the School of Public Health at the University of California–Los Angeles (UCLA) in 1995 and is currently professor and vice chair of the Epidemiology department. She also holds co-appointments in the Environmental Health department at the UCLA School of Public Health and in the Neurology department at the UCLA School of Medicine. Ritz received her M.D. and a Ph.D. in medical sociology from the University of Hamburg, Germany, in 1983 and 1987; she was a research fellow and resident at the Psychiatric University-Hospital in Hamburg from 1987–1989, and received doctoral training and a Ph.D. degree in epidemiology in 1995 from UCLA.

Her research focuses on the health effects of occupational and environmental toxins such as pesticides, ionizing radiation, and air pollution on chronic diseases including neurodegenerative disorders (Parkinson's disease), cancers, and adverse birth outcomes and asthma.

Arthur Li is a biostatistician who works at the City of Hope National Cancer Center research hospital in Los Angeles, California. He is a cofounder of Stats and More, a company devoted to statistical analysis. He has in-depth experience in the analysis of biomedical data and is an expert in data management. He is author of the book *Handbook of SAS DATA Step programming* (2013).

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Coker E, Liverani S, Ghosh JK, Jerrett M, Beckerman B, Li A, Ritz B, Molitor J. 2016. Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles County. *Environ Int* 91:1–13; doi:10.1016/j.envint.2016.02.011. Available: www.sciencedirect.com/science/article/pii/S0160412016300460.

Coker E, Ghosh J, Jerrett M, Gomez-Rubio V, Beckerman B, Cockburn M, et al. 2015. Modeling spatial effects of PM_{2.5} on term low birth weight in Los Angeles County. *Environ Res* 142:354–364.

Molitor J, Su JG, Molitor N-T, Rubio VG, Richardson S, Hastie D, et al. 2011. Identifying vulnerable populations through an examination of the association between multipollutant profiles and poverty. *Environ Sci Technol* 45:7754–7760; doi:10.1021/es104017x.

ABBREVIATIONS AND OTHER TERMS

CI	credible interval
CO	carbon monoxide
CT	census tract
CBG	census block group
IQR	interquartile range
LA	Los Angeles
LUR	land use regression
MCMC	Markov chain Monte Carlo
NO	nitric oxide
NO ₂	nitrogen dioxide
NO _x	nitrogen oxides
NRC	National Research Council
PM	particulate matter
PM _{2.5}	particulate matter ≤ 2.5 μm in aerodynamic diameter
QA	quality assurance
RFA	request for applications
R-INLA	integrated nested Laplace approximations (in R)
SES	socioeconomic status
TLBW	term low birth weight
UCLA	University of California–Los Angeles
U.S. EPA	U.S. Environmental Protection Agency

Research Report 183, Part 3, *Modeling of Multipollutant Profiles and Spatially Varying Health Effects with Applications to Indicators of Adverse Birth Outcomes*, J. Molitor et al.

INTRODUCTION

Air pollution is a complex mixture of gaseous, liquid, and solid components, which varies greatly in composition and concentration across time and space owing to differences and proximity to sources, weather, and topography. Although it is clear that people are exposed to complex mixtures of pollutants emitted by diverse sources, the U.S. Clean Air Act and air quality guidelines and standards worldwide are geared toward control of individual, or small sets of, pollutants. Consequently, most epidemiologic studies of air pollution and health have focused on estimating the adverse effects associated with ambient exposure to a single pollutant; in some cases, results are adjusted for exposure to other pollutants as possible confounders, mostly in two- and three-pollutant models. Single-pollutant research is relatively easier to conduct, and the results are easier to interpret in comparison to multipollutant approaches, which pose many challenges (Dominici et al. 2010). Of particular interest is the combined effect of various constituents of an air pollution mixture, and whether the combined effect differs from the effects of those individual pollutants within the mixture: combined pollutants may elicit health effects that are synergistic, additive, or less than additive.

Employing multipollutant models using conventional statistical approaches frequently produces results that are difficult to interpret because air pollutant levels are often highly correlated. Therefore, more sophisticated statistical methods are needed to investigate the health effects of air pollution mixtures.

Dr. John Molitor's 2-year study, "Modeling of Multi-Pollutant Profiles with applications to RIOPA data and to indicators of adverse birth outcomes using data from the UCLA Environment and Pregnancy Outcome Study (EPOS)" began in September 2010. Total expenditures were \$232,000. The draft Investigators' Report from Molitor and colleagues was received for review in December 2014. A revised report, received in May 2015, was accepted for publication in June 2015. During the review process, the HEI Health Review Committee and the investigators had the opportunity to exchange comments and to clarify issues in both the Investigators' Report and the Review Committee's Critique.

This document has not been reviewed by public or private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views of these parties, and no endorsements by them should be inferred.

To address these important questions, HEI issued a Request for Applications (RFA*) 09-1, "Methods to Investigate the Effects of Multiple Air Pollution Constituents," which solicited research proposals that would address these methodologic challenges through the development of innovative statistical methods. The RFA primarily sought proposals in which existing statistical approaches (including those from fields outside of epidemiology) could be modified, extended, or combined, and their usefulness illustrated by application to real-life data, rather than the development of purely theoretical approaches and use of simulations only. (See the Preface for more detail on the scientific background for the RFA development.)

Three studies were funded under RFA 09-1 that represent a variety of statistical approaches and applications. The studies by Dr. Brent Coull and Dr. Eun Sug Park and their colleagues are described in Research Report 183, Parts 1 and 2. For the current study, Dr. John Molitor and colleagues proposed to develop and apply statistical methods to examine associations between spatial patterns of correlated air pollutants and outcomes of health and poverty.

The development of new methods typically follows a series of steps before the methods can enter general use (see sidebar — Process of Statistical Methods Development and Evaluation). The work conducted by Molitor and colleagues addressed the first three steps in this methods development process, along with having an application in complex real-world settings. Their report focuses on these real-world applications.

Originally, Molitor and colleagues proposed using two datasets to demonstrate the methods, the UCLA Environment and Pregnancy Outcome Study (Ritz et al. 2007), and RIOPA (Turpin et al. 2007; Weisel et al. 2005).

Over the course of the study, they decided to focus their applications solely on measures of poverty and adverse birth outcomes in Los Angeles County using census and birth certificate data.

This critique provides the HEI Health Review Committee's evaluation of the study. It is intended to aid the sponsors of HEI and the public by highlighting both the strengths and limitations of the study and by placing the Investigators' Report into scientific and regulatory perspective.

* A list of abbreviations and other terms appears at the end of the Investigators' Report.

Process for Statistical Methods Development and Evaluation

- Formulate problem
- Develop conceptual framework for statistical models and specify parameters of interest including, where appropriate, development of statistical theory
- Write software to estimate parameters based on the conceptual framework
- Conduct preliminary tests in simulated data sets with known attributes
- Conduct preliminary tests in a well-studied data set, if available
- Test in a simplified real-world setting
- Test in a complex real-world setting
- Other investigators apply methods in settings that differ from those with which the method was developed and tested

APPROACH

The aims of the study by Dr. Molitor and colleagues were:

1. Develop Bayesian clustering methods to characterize profiles of pollutants or other patterns of covariates (e.g., socioeconomic status [SES]) relevant to the study of health effects.
2. Associate multipollutant profiles found in Aim 1 to health outcomes using cluster assignments as random effects in a regression model, while accounting for relevant confounders and taking into account uncertainty in the clustering process.
3. Develop ways to assess a *best* clustering and then assess the uncertainty related to this best clustering by Bayesian model averaging.
4. Analyze real datasets assessing associations between multipollutant profiles and measures of poverty and analyze associations between pollutant profiles with adverse birth outcomes, all in Los Angeles (LA) County.

The investigators built on their previous work (Molitor et al. 2010) to develop Bayesian clustering methods to identify spatial clusters of air pollution exposures — and of other covariates such as SES — and to estimate the association of health outcomes with those clusters. They use the term *profile* to define a set of pollutants (or more

generally exposures). Their approach has three components: a prior for cluster allocation, a profile assignment submodel to assign exposure profiles to clusters, and a health effects submodel to link clusters of exposure profiles to the health outcome. The Bayesian models described by Molitor and colleagues are mostly fit using Markov chain Monte Carlo techniques. Their Bayesian framework allows a supervised (joint) estimation (meaning that they allowed the relationship between health outcomes and exposures to inform the formation of the clusters).

An important feature of these clustering methods is that they are flexible. For example, the number of clusters is allowed to vary (up to a prespecified maximum number), while most other clustering approaches require a pre-defined and fixed number of clusters in order to proceed with the estimation. In addition, these methods quantify the uncertainty related to the clustering allocation and propagate it in the health analyses. To group exposure profiles into clusters, Molitor and colleagues used Dirichlet-process mixture modeling techniques and combined the resulting clusters with multilevel regression models to estimate health outcomes. Subsequently, they developed postprocessing Bayesian model-averaging techniques to find clusters that best represent the data and to assess uncertainty in the cluster allocation. They set the maximum number of clusters at 20, and the methods allow clusters to be empty.

Parts of the methods are implemented in the PReMiuM R statistical software package by Liverani and colleagues (2015); others use WinBUGS, or R-INLA (the R package for integrated nested Laplace approximations).

Molitor and colleagues conducted analyses using three applications to demonstrate these methods using different data sources in LA County (see Critique Table 1 for an overview). The scientific questions considered in each application differ per application because of differences in exposure and health outcomes, but also because of the level of analysis and the exposure contrasts. In application 1, the study team investigated associations between exposure clusters based on four pollutants and the number of people living below the poverty level at the census tract (CT) level. In application 2, the study team investigated the associations between individual-level PM_{2.5} (particulate matter $\leq 2.5 \mu\text{m}$ in aerodynamic diameter) and term low birth weight using a multilevel spatial logistic regression model to assess whether the associations vary within CTs. Note that this application does not represent multipollutant modeling, and is therefore included as an appendix at the suggestion of the HEI Review Committee. In application 3, the study team investigated associations between exposure clusters based on three pollutants and

Critique Table 1. Overview of Molitor’s Applications in LA County

Application #	Exposure Information	Outcome Information	Unit of Analysis
1	<p>NO₂: land use regression model, using 181 sites from a two-week period in summer 2006 and winter 2007 in Los Angeles (Su et al. 2009).</p> <p>PM_{2.5}: universal kriging methods using 23 regulatory sites in LA for 2000 (Jerrett et al. 2005).</p> <p>On- and off-road diesel: using a U.S. EPA based computer simulation model ASPEN (U.S. EPA 2005).</p> <p>NO₂ and PM_{2.5} predictions targeted at census tract centroids.</p>	<p>Number of people living below poverty level</p> <p>Census data 2000</p>	Census Tract (CT) ^a (N = 2,038)
2 (In Appendix A)	<p>PM_{2.5}: land use regression model using 112 regulatory sites for 1998–2002 in California (Jerrett et al. 2013).</p> <p>PM_{2.5} predictions targeted at individual residential address level.</p>	<p>Term low birth weight (< 2500 g)</p> <p>Birth certificate data for the years 1995–2006</p>	Individual (N = 1,356,304)
3	<p>NO₂ and NO: same land use regression model as in application 1.</p> <p>PM_{2.5}: same land use regression model as in application 2.</p> <p>Pollutant exposure predictions targeted at individual residential address level, averaged within census block group.</p> <p>Five contextual variables: percentage Hispanic, percentage nonHispanic black, percentage nonHispanic white, median household income, percentage of homes older than 1950 (Census data 2000).</p>	<p>Term low birth weight (< 2500 g)</p> <p>Birth certificate data for the years 2000–2006 (N = 804,726)</p>	Census Block Group (CBG) ^b (N = 6,280)

^a CT = Census Tract. This is a small, relatively stable geographic area for which the U.S. Census Bureau publishes sample data; it usually has a population between 2500 and 8000 people.

^b CBG = Census Block Group. This is the smallest geographical area for which the U.S. Census Bureau publishes sample data; it usually has a population between 600 and 3000 people.

term low birth weight at the census block group (CBG) level, taking into account contextual clusters as well (e.g., SES-related variables: race–ethnicity, median household income and percentages of homes older than 1950).

THE HEALTH REVIEW COMMITTEE’S EVALUATION

In its independent review of the study, the HEI Review Committee concluded that the investigators extended their cluster methods to include continuous exposures and

successfully implemented these methods to analyze multi-pollutant mixtures. The Committee welcomed that large parts of the methods have been implemented in R packages, which can be freely used. The availability of user-friendly software is a key component of any wider adoption of a new approach beyond the statistical community. The Committee noted that the approach taken by the authors in applications 1 and 3 addressed an important question in multipollutant research, that is, what are the combined effects of various constituents of an air pollution mixture. Their approach was aimed at identifying spatial clusters of air pollution exposures — and other covariates

such as SES in application 3 — and estimating health outcomes associated with those clusters. Their analyses demonstrated that their approach can be applied to real-world data sets and that they produced results that were largely concordant with a priori expectations. Results indicate that the effects of pollutants, as well as SES variables, vary spatially and that they vary in a complex interconnected manner. The Committee thought the difficult subject matter was made much more accessible through the investigators' approach to presenting their results. For example, the Committee liked the spatially-varying maps of effects, which they believe are a useful and effective tool to communicate results.

The Committee appreciated the flexibility of the clustering approach; for example, the fact that the number of clusters does not need to be predefined (other than the specification of a maximum allowable number of clusters) and that the clustering algorithm was incorporated into a Bayesian framework. This approach gives the investigators a way to quantitatively estimate the uncertainty of the clustering approach. This has advantages over many other clustering methods because it removes some of the ad hoc aspects associated with this type of modeling (Hastie et al. 2009).

The approach developed by the Molitor team considers different scientific questions than most other air pollution studies capturing within-city spatial variability in exposures. For instance, in application 3, it explicitly included spatially-varying contextual factors (e.g., SES variables) as inputs to the clusters, in a similar way as air pollutants were treated. The effects of individual and neighborhood SES on health are increasingly well documented, as it is understood that there are relationships between air pollution and SES, making it an important factor in air pollution epidemiologic studies. In many settings, low-SES communities are disproportionately exposed to air pollution, and those communities may be more susceptible to air pollution owing to other underlying disparities (Clark et al. 2014; O'Neill et al. 2003). Note that some studies report opposite associations: higher SES has been associated with higher air pollution levels, for example in New York and Rome, highlighting the importance of investigating the SES–air pollution associations in a specific setting (Hajat et al. 2013; Cesaroni et al. 2010). Typically, air pollution studies treat SES as a confounder and adjust for it in a fairly simple way, by using an individual or neighborhood SES variable and treating this as a fixed covariate effect in the health model. Somewhat more complex methods are sometimes used to combine different SES variables, for example, using principal component or factor analysis (Cesaroni et al. 2010; Shmool et al. 2014). The use of spatially-varying contextual variables as inputs to the clusters in the current

study is unique; it can potentially provide new insight into understanding vulnerable and susceptible populations. It is notable that in application 3 the investigators incorporated contextual variables into their risk analyses in two different ways: by estimating separate contextual clusters in addition to the air pollutant clusters, and by clustering jointly on both the contextual and pollutant measures. These two different approaches led to different maps of effects, suggesting that there is still more to be learned about vulnerable and susceptible populations.

The methods developed by Molitor and colleagues are complex and computationally demanding. The investigators have put their models in a unified Bayesian framework as one way to allow a supervised (joint) estimation (meaning that they allowed the relationship between health outcomes and exposures to inform the formation of the clusters). The other two studies funded under RFA 09-1 (Coull et al. 2015; Park et al. 2015) also have used Bayesian methods in supervised modeling approaches. However, in general, there are several important practical features of supervised modeling approaches that are worth considering. Typically, they are computationally demanding. Computational demands, in fact, led the Molitor team to decide to not pursue a full Bayesian supervised modeling approach in application 3 where the clustering was done before the health outcomes were incorporated (while a full Bayesian supervised modeling approach was conducted in application 1). Computational feasibility is a key component of any wider adoption of a new approach and can be especially difficult to achieve when modeling multiple pollutants. Another consideration of supervised models is that the clusters identified are dependent on the health outcome, and changing the health outcome will generally change the definition of the clusters to some extent. This feature may be less intuitive than unsupervised models because those assign one exposure per participant that can then be used to model a number of different health outcomes. Finally, the potential for feedback between the exposure and health models must be acknowledged.

While feedback is an inherent feature of all supervised models, its presence could influence the health effect estimates when the amount of data available for the health outcome and the exposure data is unbalanced, in concert with misspecification of either or of both the exposure and health models. Unsupervised approaches, in which exposures (or in this case, the clusters) are estimated first and then incorporated in subsequent health models, avoid feedback. In many air pollution epidemiology settings, the amount of exposure data available is far less than the amount of health data, and this imbalance can exacerbate potential bias from feedback caused by model misspecification. Indeed, the

number of exposure measurements was small compared with the size of the study population in the applications. The Committee noted that in the current analyses this imbalance of exposure and outcome data was not a feature of the supervised modeling itself because the investigators aligned the exposure predictions with the health data before clustering. However, they did not explicitly acknowledge the impact of exposure measurement error due to using exposure predictions in their health analyses. In recent literature dealing with correction for exposure measurement error in epidemiologic studies, many investigators have intentionally chosen an unsupervised (e.g., two-stage) approach when their exposure modeling focus is on spatial prediction of pollutants from spatially misaligned monitors (e.g., Bergen et al. 2013; Gryparis et al. 2009; Szpiro et al. 2013). The exposure measurement error correction methodology for spatially-varying pollutants in multipollutant research is in its infancy (Bergen et al. 2016), thus it is not surprising that Molitor and colleagues did not address this in their work. Similarly, the importance of, and implications for supervised modeling in studies that focus on using clustered exposures for inferences about health effects is an important future research topic.

Often in applications of complex methods, data simplifications are required to make the problem tractable, and there is always the potential that important information is lost with such steps. One feature the Committee noted was the aggregation of air pollution exposure from the individual level to the CT (application 1) or CBG level (application 3). The scientific question considered in application 1 lends itself to a CT-level analysis because the outcome is defined at that level, so the unknown impact of exposure aggregation in application 1 is whether the approach to summarizing residence-level pollutant predictions adopted by the investigators affects the findings. In application 3, the investigators stated that their interest was in between-neighborhood effects, but the data are all available at the individual level. The Committee wondered about the magnitude of potential bias from aggregation of the exposure to the CBG level, especially for air pollutants that have documented substantial variation at small spatial scales (e.g., NO, NO₂). Another simplification adopted by the investigators in application 3 was to precluster the pollutants and contextual variables prior to fitting the multilevel health analysis. In addition, the number of joint contextual–multipollutant clusters reported appears to be the maximum number the investigators allowed in the clustering algorithm, so several key properties of the clustering methods — namely that the number of and specific details of the clusters are variable and that the cluster selection is supervised by the outcome — were not present

in this example. The Committee noted that the effects of the various data simplifications were not studied.

The Committee thought that this project was unusual as a statistical methods development project because the report focuses exclusively on the applications of the methods; it gives limited attention to the methods themselves and does not follow the typical methods development process steps (see sidebar introduction). Most of the methods were previously developed by the investigators and can be found in a previous publication (Molitor et al. 2010). In addition, the published version of application 1 includes a few more modeling details (Molitor et al. 2011). The primary advances in the current study include the types of covariates that can be included in the profile regression, use of R-INLA for some of the computations, and the approaches to presenting the results. While some simulation studies were done as part of Molitor and colleagues (2010), none were presented for the applications presented in the current report. The Committee thought it would have been worthwhile to understand how the methods perform under known conditions and to compare the methods to traditional statistical methods for which the research community has already developed a deep understanding of their properties and performance. Currently, it remains unclear how well the methods used by Molitor and colleagues perform. Without such a comparison it is also difficult to quantify the degree to which their work has led to the identification of important new insights.

Partly because of the unusual focus, the Committee came to the overall conclusion that the methods developed by Molitor and colleagues show promise, but that the full extent to which they will be useful remains to be seen. The Committee did not expect that this work alone would resolve what practical improvements in understanding would be achievable with the application of this methodology in different settings. Yet, there remain open questions about the relative merits of the methodology as well as the specific data requirements and the amount of complexity the methods can or cannot handle. The Committee also noted that the available data had a limited set of pollutants the investigators considered (four pollutants at most), which captures only part of the air pollutant mixture. Moreover, some of the included pollutants have generally similar spatial patterns (e.g., NO₂ and NO). Future work is necessary to further evaluate the methods in this report. In particular, it would be useful to apply them in a broader range of settings and pollutants, including in locations other than Los Angeles, which is characterized by general high levels of air pollution as well as high spatial contrasts within the city.

SUMMARY AND CONCLUSIONS

Dr. Molitor and colleagues developed methods to address an important question in multipollutant research, that is, what are the combined effects of various constituents of an air pollution mixture. Their approach was aimed at identifying spatial clusters of air pollution exposures — and other covariates such as SES — and estimating health outcomes associated with those clusters. The investigators' analyses demonstrated that the methods can be applied to real-world data sets and that they produced results that were largely concordant with a priori expectations. In its independent review of the study, the HEI Review Committee concluded that the investigators extended their cluster methods to include continuous exposures and successfully implemented these methods to analyze multipollutant mixtures. They appreciated the flexibility of the clustering approach; for example, the number of clusters does not need to be predefined and uncertainty related to cluster allocation is accounted for. The explicit inclusion of spatially-varying contextual factors (e.g., SES variables) as inputs to the clusters, in a treatment similar to that of air pollutants, was considered unique and can potentially provide new insight into understanding vulnerable and susceptible populations. The Committee concluded that the multipollutant methods developed show promise, but that the full extent to which they will be useful remains to be seen. Future work is necessary to fully evaluate these methods, including simulation studies, comparison to traditional statistical methods, application in other settings, and inclusion of more pollutants. Such analyses could help to determine the degree to which these new methods will lead to a better understanding of how pollutant mixtures contribute to health effects, and ultimately, to better decisions about how to control them.

ACKNOWLEDGMENTS

The Health Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigators' Report. The Committee is also grateful to Kate Adams for her oversight of the study, to Katy Walker for preparing the initial review, to Hanna Boogaard for her assistance in preparing its Critique, to Carol Moyer for science editing of this Report and its Critique, and to Hope Green, Fred Howe, and Ruth Shaw for their roles in preparing this Research Report for publication.

REFERENCES

- Bergen S, Sheppard L, Sampson PD, Kim SY, Richards M, Vedal S, et al. 2013. A national prediction model for PM_{2.5} component exposures and measurement error-corrected health effect inference. *Environ Health Perspect* 121:1017–1025.
- Bergen S, Sheppard L, Kaufman JD, Szpiro AA. 2016. Multipollutant measurement error in air pollution epidemiology studies arising from predicting exposures with penalized regression splines. *J R Stat Soc C*. doi: 10.1111/rssc.12144.
- Cesaroni G, Badaloni C, Romano V, Donato E, Perucci CA, Forastiere F. 2010. Socioeconomic position and health status of people who live near busy roads: the Rome longitudinal study (RoLS). *Environ Health* 9:41.
- Clark LP, Millet DB, Marshall JD. 2014. National patterns in environmental injustice and inequality: Outdoor NO₂ air pollution in the United States. *PLoS One* 9:e94431.
- Coull BA, Bobb JF, Wellenius GA, Kioumourtzoglou M-A, Mittleman MA, Koutrakis P, et al. 2015. Part 1. Statistical Learning Methods for the Effects of Multiple Air Pollution Constituents. In: *Development of Statistical Methods for Multipollutant Research*. Research Report 183. Boston, MA:Health Effects Institute.
- Dominici F, Peng RD, Barr CD, Bell ML. 2010. Protecting human health from air pollution: Shifting from a single-pollutant to a multipollutant approach. *Epidemiology* 21:187–194.
- Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. 2009. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics* 10:258–274.
- Hajat A, Diez-Roux AV, Adar SD, Auchincloss AH, Lovasi GS, O'Neill MS, et al. 2013. Air pollution and individual and neighborhood socioeconomic status: evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). *Environ Health Perspect* 121:1325–1333.
- Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning* (2nd edition). New York, NY:Springer-Verlag.
- Jerrett M, Burnett RT, Ma R, Pope CA, 3rd, Krewski D, Newbold KB, et al. 2005. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 16:727–736.
- Jerrett M, Burnett RT, Beckerman BS, Turner MC, Krewski D, Thurston G, et al. 2013. Spatial analysis of air pollution

- and mortality in California. *Am J Respir Crit Care Med* 188:593–599.
- Liverani S, Hastie DI, Azizi L, Papathomas M, Richardson S. 2015. PReMiuM: An R Package for Profile Regression Mixture Models Using Dirichlet Processes. *J Stat Soft*, 64(7):1–30.
- Molitor J, Papathomas M, Jerrett M, Richardson S. 2010. Bayesian profile regression with an application to the national survey of children’s health. *Biostatistics* 11:484–498.
- Molitor J, Su JG, Molitor NT, Rubio VG, Richardson S, Hastie D, et al. 2011. Identifying vulnerable populations through an examination of the association between multi-pollutant profiles and poverty. *Environ Sci Technol* 45:7754–7760.
- O’Neill MS, Jerrett M, Kawachi I, Levy JI, Cohen AJ, Gouveia N, et al. 2003. Health, wealth, and air pollution: advancing theory and methods. *Environ Health Perspect* 111:1861–1870.
- Park ES, Symanski E, Han D, Spiegelman C. 2015. Part 2. Development of Enhanced Statistical Methods for Assessing Health Effects Associated with an Unknown Number of Major Sources of Multiple Air Pollutants. In: *Development of Statistical Methods for Multipollutant Research*. Research Report 183. Boston, MA:Health Effects Institute.
- Ritz B, Wilhelm M, Hoggatt KJ, Ghosh JK. 2007. Ambient air pollution and preterm birth in the environment and pregnancy outcomes study at the University of California, Los Angeles. *Am J Epidemiol* 166(9):1045–1052.
- Shmool JL, Kubzansky LD, Newman OD, Spengler J, Shepard P, Clougherty JE. 2014. Social stressors and air pollution across New York City communities: a spatial approach for assessing correlations among multiple exposures. *Environ Health* 13:91.
- Su JG, Jerrett M, Beckerman B, Wilhelm M, Ghosh JK, Ritz B. 2009. Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. *Environ Res* 109:657–670.
- Szpiro AA, Paciorek CJ. 2013. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics* 24:501–517.
- Turpin BJ, Weisel CP, Morandi M, Colome S, Stock T, Eisenreich S, et al. 2007. Relationships of Indoor, Outdoor, and Personal Air (RIOPA): Part II. Analyses of concentrations of particulate matter species. Research Report 130. Boston, MA:Health Effects Institute.
- U.S. Environmental Protection Agency. 2005. National Air Toxics Assessment. Washington DC:U.S. Environmental Protection Agency.
- Weisel CP, Zhang J, Turpin B, Morandi MT, Colome S, Stock TH, et al. 2005. Part I. Collection methods and descriptive analyses. In: *Relationships of Indoor, Outdoor, and Personal Air (RIOPA)*. Research Report 130. Boston, MA:Health Effects Institute.

RELATED HEI PUBLICATIONS: MULTIPOLLUTANTS, BIRTH OUTCOMES, AND METHODS

Number	Title	Principal Investigator	Date*
Research Reports			
189	Ambient Air Pollution and Adverse Pregnancy Outcomes in Wuhan, China	Z. Qian	In Press
188	Adverse Reproductive Health Outcomes and Exposure to Gaseous and Particulate-Matter Air Pollution in Pregnant Women	J. Wu	In Press
183	Development of Statistical Methods for Multipollutant Research		
	<i>Part 1.</i> Statistical Learning Methods for the Effects of Multiple Air Pollution Constituents	B.A. Coull	2015
	<i>Part 2.</i> Development of Enhanced Statistical Methods for Assessing Health Effects Associated with an Unknown Number of Major Sources of Multiple Air Pollutants	E.S. Park	2015
181	Modeling and Analysis of Personal Exposure to Mixtures of Volatile Organic Compounds: Further Analysis of the RIOPA Data	S. Batterman	2014
178	National Particle Component Toxicity (NPACT) Initiative Report on Cardiovascular Effects	S. Vedal	2013
177	National Particle Component Toxicity (NPACT) Initiative: Integrated Epidemiologic and Toxicologic Studies of the Health Effects of Particulate Matter Components	M. Lippmann	2013
175	New Statistical Approaches to Semiparametric Regression with Application to Air Pollution Research	J.M. Robins	2013
161	Assessment of the Health Impacts of Particulate Matter Characteristics	M.L. Bell	2012
152	Evaluating Heterogeneity in Indoor and Outdoor Air Pollution Using Land-Use Regression and Constrained Factor Analysis	J.I. Levy	2010
140	Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality	D. Krewski	2009
139	Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: The NLCS-AIR Study	B. Brunekreef	2009
130	Relationships of Indoor, Outdoor, and Personal Air (RIOPA)		
	<i>Part I.</i> Collection Methods and Descriptive Analyses	C.P. Weisel	2005
	<i>Part II.</i> Analyses of Concentrations of Particulate Matter Species	B.J. Turpin	2007
HEI Special Reports			
17	Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects		2010
16	Mobile-Source Air Toxics: A Critical Review of the Literature on Exposure and Health Effects		2007

* Copies of these reports can be obtained from HEI; pdf's are available for free downloading at <http://pubs.healtheffects.org>

HEI BOARD, COMMITTEES, and STAFF

Board of Directors

Richard F. Celeste, Chair *President Emeritus, Colorado College*

Sherwood Boehlert *Of Counsel, Accord Group; Former Chair, U.S. House of Representatives Science Committee*

Enriqueta Bond *President Emerita, Burroughs Wellcome Fund*

Purnell W. Choppin *President Emeritus, Howard Hughes Medical Institute*

Michael T. Clegg *Professor of Biological Sciences, University of California–Irvine*

Jared L. Cohon *President Emeritus and Professor, Civil and Environmental Engineering and Engineering and Public Policy, Carnegie Mellon University*

Stephen Corman *President, Corman Enterprises*

Linda Rosenstock *Dean Emerita and Professor of Health Policy and Management, Environmental Health Sciences and Medicine, University of California–Los Angeles*

Henry Schacht *Managing Director, Warburg Pincus; Former Chairman and Chief Executive Officer, Lucent Technologies*

Warren M. Washington *Senior Scientist, National Center for Atmospheric Research; Former Chair, National Science Board*

Health Research Committee

David L. Eaton, Chair *Dean and Vice Provost of the Graduate School, University of Washington–Seattle*

Jeffrey R. Brook *Senior Research Scientist, Air Quality Research Division, Environment Canada, and Assistant Professor, University of Toronto, Canada*

David Christiani *Elkan Blout Professor of Environmental Genetics, Harvard T.H. Chan School of Public Health*

Francesca Dominici *Professor of Biostatistics and Senior Associate Dean for Research, Harvard T.H. Chan School of Public Health*

David E. Foster *Phil and Jean Myers Professor Emeritus, Department of Mechanical Engineering, Engine Research Center, University of Wisconsin–Madison*

Uwe Heinrich *Professor, Hannover Medical School; Executive Director, Fraunhofer Institute for Toxicology and Experimental Medicine, Hanover, Germany*

Amy H. Herring *Carol Remmer Angle Distinguished Professor of Children's Environmental Health, and Associate Chair, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina–Chapel Hill*

Barbara Hoffmann *Professor of Environmental Epidemiology and Head of Environmental Epidemiology of Aging, IUF-Leibniz Research Institute for Environmental Medicine, and Professor, Medical Faculty, Heinrich Heine University of Düsseldorf, Germany*

Allen L. Robinson *Raymond J. Lane Distinguished Professor and Head, Department of Mechanical Engineering, and Professor, Department of Engineering and Public Policy, Carnegie Mellon University*

Ivan Rusyn *Professor, Department of Veterinary Integrative Biosciences, Texas A&M University*

HEI BOARD, COMMITTEES, and STAFF

Health Review Committee

James A. Merchant, Chair *Professor and Founding Dean Emeritus, College of Public Health, University of Iowa*

Michael Brauer *Professor, School of Environmental Health, University of British Columbia, Canada*

Bert Brunekreef *Professor of Environmental Epidemiology, Institute of Risk Assessment Sciences, University of Utrecht, the Netherlands*

Mark W. Frampton *Professor of Medicine and Environmental Medicine, University of Rochester Medical Center*

Jana B. Milford *Professor, Department of Mechanical Engineering and Environmental Engineering Program, University of Colorado–Boulder*

Roger D. Peng *Associate Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

Lianne Sheppard *Professor of Biostatistics, School of Public Health, University of Washington–Seattle*

Officers and Staff

Daniel S. Greenbaum *President*

Robert M. O’Keefe *Vice President*

Rashid Shaikh *Director of Science*

Jacqueline C. Rutledge *Director of Finance and Administration*

April Rieger *Corporate Secretary*

Zachary Abbott *Research Assistant*

Sarah Benckart *Science Administration Assistant*

Hanna Boogaard *Staff Scientist*

Aaron J. Cohen *Principal Scientist*

Maria G. Costantini *Principal Scientist*

Philip J. DeMarco *Compliance Manager*

Hope Green *Publications Associate*

Anny Luu *Executive Assistant*

Hilary Selby Polk *Managing Editor*

Evan Rosenberg *Staff Accountant*

Robert A. Shavers *Operations Manager*

Annemoon M.M. van Erp *Managing Scientist*

Donna J. Vorhees *Senior Scientist*

Katherine Walker *Senior Scientist*



HEALTH
EFFECTS
INSTITUTE

75 Federal Street, Suite 1400
Boston, MA 02110, USA
+1-617-488-2300
www.healtheffects.org

RESEARCH
REPORT

Number 183
Part 3
April 2016