DISSERTATION

METHODS FOR EFFECT MODIFICATION WITH MULTIVARIATE ENVIRONMENTAL EXPOSURES

Submitted by

Danielle Dematteis

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2025

Doctoral Committee:

Advisor: Ander Wilson
Co-Advisor: Kayleigh Keller

Dan Cooley
Tianying Wang
Sheryl Magzamen

ABSTRACT

METHODS FOR EFFECT MODIFICATION WITH MULTIVARIATE ENVIRONMENTAL EXPOSURES

Humans are exposed to a multitude of environmental insults every day. Exposures such as air pollution, heat and extreme weather, heavy metals, and environmental chemicals are known to be linked to adverse health outcomes. There is interest in understanding how multivariate exposures, including repeated measures of the same exposure over time for the same observation and measures of multiple exposures at a single time point, impact health. Several statistical approaches have been proposed for the analysis of multivariate exposure data. Two commonly used methods are distributed lag models (DLMs) for repeated measures of exposure and Bayesian kernel machine regression (BKMR) for multiple exposures observed at a single time point. These methods and their variants are widely used in environmental health studies. However, they lack flexibility to estimate effect modification in most settings.

In this dissertation, I develop methods to include effect modification in both DLMs and BKMR. The first method is the distributed lag interaction model (DLIM) that extends the standard distributed lag framework to allow for modification of the exposure-time-response function by a single continuous variable. I use a cross-basis, or bi-dimensional function space, inspired by the distributed lag non-linear framework to simultaneously describe the temporal and modification structure. Next, I developed a distributed lag interaction model with index modification (DLIM-IM) that allows for modification of the exposure-time-response function by multiple modifiers via a data-derived modification index. I use a Bayesian hierarchical framework to simultaneously estimate the exposure-time-response function and a weighted modification index, and I allow for selection on the candidate modifiers. Finally, I propose and evaluate extensions of the BKMR framework to include effect modification by a categorical modifier. I propose a new version of

BKMR with a separable covariance function that allows for increased flexibility to estimate effect modification as well a comparing alternative ways to apply BKMR for assessing modification

For each of these methods, I validated these methods through simulation and applied them to multiple data sets to demonstrate their application. I have made open-source software for the methods publicly available on CRAN and GitHub.

ACKNOWLEDGEMENTS

I would like to thank the people who helped and guided me along the way to be able to write this dissertation. First, I want to thank my former educators who supported me through mathematics education: Maureen Peruzzi, Wendy Clayton, Robert Rotante, Michelle Glucksnis, Jennifer Connor, Joseph Arfin, Michael Bixler, Karen Quiggley, Michele Colon, Judit Kardos, Jeremy Russell, Jana Gevertz, Andrew Clifford, Cindy Curtis, Sharon Navard, and Tom Hagedorn. A special thanks to John Wnek for piquing my interest in research and to Michael Ochs for encouraging me to attend graduate school. Second, I would like to thank CSU faculty who were most meaningful in my graduate education: Jennifer Hoeting, Bailey Fosdick, Mary Meyer, Kayleigh Keller, Ander Wilson, Tianying Wang, Dan Cooley, and Sheryl Magzamen. Third, I would like to thank friends who supported me leading up to and throughout graduate school: Kamile Lukosiute, Alina Kuvelkar, and Maddie Rainey. A special thanks to Grant Bilker for reviewing this dissertation and to Kate Laviolette for creating a supportive community at Live Beyond Limit on which I leaned. Fourth, I would like to thank the people who inspired me to pursue a career in non-animal research: Ed Winters, Jennifer Litzau, Cara Thorsen, Solomon Carpenter, Cordelia Stone, Aniket Tomar, Cindy Gordon, and Liz Eisenhower. Finally, I would like to thank family, including Janet, Steve, Joesph, Anthony Demateis, Barbara Studley, and Tim Maley, who have been with me through this entire journey. A special thanks to my partner Chris Hartley and my little girls, Alaska who is an expert in consolation and Cuddles who is an inspirational beacon of perseverance.

Research reported in this publication was supported by the National Institute Of Environmental Health Sciences of the National Institutes of Health under Award Numbers R01ES029943 and R01ES035735. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

These data were supplied by the Center for Health and Environmental Data Vital Statistics Program of the Colorado Department of Public Health and Environment, which specifically disclaims responsibility for any analyses, interpretations, or conclusions it has not provided.

This work utilized the Alpine high performance computing resource at the University of Colorado Boulder. Alpine is jointly funded by the University of Colorado Boulder, the University of Colorado Anschutz, and Colorado State University and with support from NSF grants OAC-2201538 and OAC-2322260.

DEDICATION

I dedicate this dissertation to animals exploited in the name of science. My motivation for writing this dissertation came from the dream of creating an animal-free future for research.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS iv
DEDICATION vi
LIST OF TABLES ix
LIST OF FIGURES xii

Chapter 1 Introduction 1
1.1 Distributed Lag Model 2
1.2 Bayesian Kernel Machine Regression 3
1.3 Overview 4

Chapter 2 Penalized Distributed Lag Interaction Model 5
2.1 Introduction 5
2.2 Colorado Birth Cohort Data 9
2.3 Modeling Framework 11
2.3.1 Model Approach 11
2.3.2 Parameterization of exposure-time-response functions 12
2.3.3 Parameter Estimation 13
2.3.4 Inference 14
2.3.5 Alternative Approaches for Penalization 14
2.3.6 Alternative Approach without Penalization 15
2.3.7 Alternative Approach with Linear Interaction 15
2.3.8 Test for Interaction 16
2.3.9 Repeated Measures Among Individuals 16
2.4 Simulations 17
2.4.1 Simulation Design and Data Generation 17
2.4.2 Performance Evaluation 19
2.4.3 Simulation Results 20
2.4.4 Hyperparameter Selection and additional simulations 23
2.5 Analysis of Colorado Birth Cohort Data 24
2.5.1 Data and Model Fitting 24
2.5.2 Results 25
2.6 Discussion 29

Chapter 3 Distributed Lag Interaction Model with Index Modification 31
3.1 Introduction 31
3.2 Methods 33
3.2.1 Model 33
3.2.2 Parameterizing the Exposure-Time-Response Functions 35
3.2.3 Prior Specification 36
3.2.4 Parameter Estimation 37

3.3 Simulation Study . . . 38
3.3.1 Simulation Design and Data Generation . . . 39
3.3.2 Performance Evaluation . . . 40
3.3.3 Simulation Results . . . 41
3.4 Data Analyses . . . 43
3.4.1 Air Pollution, EnviroScreen and Birth Weight in Colorado . . . 44
3.4.2 Air Pollution, Maternal Stress, and Lipids in PROGRESS . . . 47
3.5 Discussion . . . 50

Chapter 4 Subgroup Analyses and Effect Modification with Bayesian Kernel Machine Regression . . . 55
4.1 Introduction . . . 55
4.2 Methods . . . 57
4.2.1 Standard BKMR . . . 57
4.2.2 Stratified BKMR . . . 58
4.2.3 Modifier-in-Kernel BKMR . . . 59
4.2.4 Group-separable BKMR . . . 60
4.2.5 Interpretation of the Exposure-Response Functions . . . 62
4.2.6 Bangladesh Reproductive Cohort Data . . . 63
4.2.7 Analysis Methods . . . 63
4.2.8 Simulation Methods . . . 64
4.3 Results . . . 67
4.3.1 Analysis Results . . . 67
4.3.2 Simulation Results . . . 68
4.4 Discussion . . . 73
4.5 Conclusions . . . 76

Chapter 5 Discussion . . . 77
5.1 Overview . . . 77
5.2 Future Work . . . 79

Appendix A Penalized Distributed Lag Interaction Model . . . 96
A.1 Additional Simulations . . . 96
A.2 Additional Figures and Tables . . . 105

Appendix B Distributed Lag Interaction Model with Index Modification . . . 111
B.1 Additional Methods . . . 111
B.2 Additional Simulation Table and Figures . . . 114
B.3 Additional Analysis Table and Figures . . . 121

Appendix C Subgroup Analyses and Effect Modification with Bayesian Kernel Machine Regression . . . 134
C.1 Additional Methods . . . 134
C.1.1 Prior specification for models . . . 134
C.1.2 Simulation Methods . . . 134
C.2 Additional Figures . . . 136

LIST OF TABLES

2.1 Cumulative and point-wise performance metrics for penalized models averaged over all 200 simulated data sets for scenarios 1 and 2 and three signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear. The table also includes the proportion of rejection ($\alpha = 0.05$) across simulated data sets under each null model (DLM or DLIM-linear) using the model comparison procedure described in Section 2.3.8. 20

2.2 Cumulative and point-wise performance metrics for penalized models averaged over all 200 simulated data sets for scenarios 3 and 4 and three signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear. The table also includes the proportion of rejection ($\alpha = 0.05$) across simulated data sets under each null model (DLM or DLIM-linear) using the model comparison procedure described in Section 2.3.8. 21

3.1 Performance metrics for the weighted modifier index, cumulative effects, and point-wise effect for DLIM-IMs without selection (DLIM-IM), the DLIM-IM with selection (DLIM-IM sel.), and the fixed-index model (fixed-index) averaged over all 200 simulated data sets with non-linear modification for all three weight scenarios and three signal-to-noise (SNR) levels. The low SNR is 0.1, medium SNR is 0.5, and the high SNR is 1. The number of basis functions for the exposure-time basis was 5 for both DLIM-IMs and 20 for the fixed-index model. 53

3.2 Modifier index weight estimates, standard deviations, and posterior inclusion probabilities (PIP) for each of the 14 included indicators of the EnviroScreen HSF score obtained in the Colorado birth cohort analysis using our proposed model with 3 degrees of freedom for both the modifier and exposure-time bases. Results are summarized across 5 chains. Indicators marked with (*) have PIPs over 0.5, suggesting that these indicators contribute to modification of the exposure-time-response relationship. . . . 54

4.1 Assumptions for models with modification considered in this manuscript: modifier-in-kernel (mod-in-kernel), first-approach group-separable (GS 1), second-approach group-separable models (GS 2), and stratified. The ✓ symbol indicates that the model does have, and the X symbol indicates that the model does not have, the corresponding assumption. 59

4.2 Performance metrics for BKMR without the modifier (exposure-only), BKMR with the modifier in the kernel (mod-in-kernel), group-separable approach 1 (group-sep. 1) with shared overall smoothing parameter τ, group-separable approach 2 (group-sep. 2) with group-specific overall smoothing $\boldsymbol{\tau}$ parameters, and stratified models fit for 200 simulated data sets across three simulation scenarios. Root mean squared error (RMSE), coverage (Cvg.), and average CI width (Width) were computed for each parameter estimate and averaged over the 200 data sets. 71

A.1 Cumulative and point-wise performance metrics for penalized and non-penalized (-ns) models averaged over all 200 simulated data sets for scenario 1 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear, and the number of basis functions for the exposure-time basis was 5 for both the DLM-ns and DLIM-linear-ns. 97
A.2 Cumulative and point-wise performance metrics for penalized and non-penalized (-ns) models averaged over all 200 simulated data sets for scenario 2 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear, and the number of basis functions for the exposure-time basis was 5 for both the DLM-ns and DLIM-linear-ns. 99
A.3 Cumulative and point-wise performance metrics for penalized and non-penalized (-ns) models averaged over all 200 simulated data sets for scenario 3 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear, and the number of basis functions for the exposure-time basis was 5 for both the DLM-ns and DLIM-linear-ns. 100
A.4 Cumulative and point-wise performance metrics for penalized and non-penalized (-ns) models averaged over all 200 simulated data sets for scenario 4 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear, and the number of basis functions for the exposure-time basis was 5 for both the DLM-ns and DLIM-linear-ns. 101
A.5 Cumulative and point-wise performance metrics for DLIM(20,20) models averaged over all 200 simulated data sets for scenarios 1 and 2 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. Alternative penalization methods are PS(2,1), PS(1,2), and CR and are compared to the penalization method in the main simulation results, PS(2,2). 102
A.6 Cumulative and point-wise performance metrics for DLIM(20,20) models averaged over all 200 simulated data sets for scenarios 3 and 4 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. Alternative penalization methods are PS(2,1), PS(1,2), and CR and are compared to the penalization method in the main simulation results, PS(2,2). 103
A.7 Cumulative and point-wise performance metrics resulting from repeating the simulation study described in the main text, with modifiers drawn from N(0.5,0.25) distribution instead of a Uniform(0,1) distribution. 104
A.8 Cumulative and point-wise performance metrics for DLIM(20,20) and standard DLM models averaged over 100 simulated data sets for all scenarios. Models were fit using Poisson regression on simulated count data. The DLIM(20,20) was penalized using PS(2,2) penalization and the standard DLM was penalized using P-splines. 105
A.9 Variable descriptions for the outcome, birth weight gestational age z-score (BWGAZ), and covariates used in the data analysis of this paper. 110

B.1 Extending Table 3.1 in the main text to include other weight scenarios. The different 10 scenario is the following weights: 0.3, 0.2, 0.1, 0.1, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05. The equal 10 scenario is 10 weights that are all 0.1. The sparse 10 scenario is the following weights: 0.5, 0.4, 0.1, 0, 0, 0, 0, 0, 0, 0. 117
B.2 Weight estimates and posterior inclusion probabilities (PIPs) for a DLIM-IM with modifier selection using ν_{mod} degrees of freedom in the modifier dimension and ν_{time} degrees of freedom in the exposure-time dimension, denoted (ν_{mod}, ν_{time}). The flag (*) indicates that the PIPs disagree in terms of selection, and the flag (**) indicates that the estimated weights are not similar. 133

LIST OF FIGURES

2.1 Summaries of Colorado EnviroScreen health and social factors (HSF) scores for census tracts selected in this study. Histogram of HSF scores for census tracts included in this study, with quartiles at scores 32, 46, and 59 (left). Map of the state of Colorado shaded by census tract. Tracts not included in our study are shaded in gray (right). . . . 9
2.2 Each curve is an exposure-time-response function for an individual, each with different modifying values. In this example, there are $n = 4$ individuals, with modifier values 0.1, 0.5, 0.6, and 0.9, respectively. Each scenario is described in Section 2.4.1. 19
2.3 Estimates of cumulative effects of ambient $PM_{2.5}$ exposure during gestation on BWGAZ from a standard DLM, DLIM-linear, and DLIM(20,20) across the middle 98% of Colorado EnviroScreen health and social factors (HSF) scores. We include confidence bands of 95% around all estimates. 26
2.4 Pointwise effect estimates of ambient $PM_{2.5}$ on BWGAZ from a standard DLM with 20 exposure-time degrees of freedom and a DLIM-linear with Colorado EnviroScreen health and social factors (HSF) score as a modifier. The top panel shows estimated exposure-time-response functions for each quartile of the HSF scores: scores 32, 46, and 59. The bottom panel shows estimated point-wise effects for a week around the middle of each trimester, excluding the first and last 1% of HSF scores. We include confidence bands of 95% around all estimates. 27

3.1 Modifier index weight estimates from DLIM-IMs with selection and DLIM-IMs without selection for all 200 simulated data sets across all weight scenarios and signal-to-noise (SNR) settings. The number of basis functions for the exposure-time basis is 5 for both the DLIM-IMs. The black line represents the true weight value. For scenario 3 (sparse weight scenario), weights 1-3 are non-zero, and weight numbers 4-50 is averaged over all 47 zero weights. 42
3.2 Estimated cumulative effect and exposure-time-response curves from the fixed-index penalized DLIM with linear modification and proposed DLIM-IM with selection applied to the Colorado birth cohort. The fixed-index DLIM uses 20 basis functions in the modifier and exposure-time dimensions, and the DLIM-IM with selection uses 5 basis functions in each dimension. 47
3.3 Estimated exposure-time-response curves for the 10^{th} ($\hat{m}^* = 0.13$), 25^{th} ($\hat{m}^* = 0.25$), 30^{th} ($\hat{m}^* = 0.27$), 40^{th} ($\hat{m}^* = 0.31$), 75^{th} ($\hat{m}^* = 0.45$), and 90^{th} ($\hat{m}^* = 0.56$) percentiles of the estimated lifetime stress index from a DLIM-IM without selection 2 months before through 22 months after last menstrual period (LMP) at weighted stress modifier values in the PROGRESS cohort. The DLIM-IM used 3 degrees of freedom in the modifier and exposure-time dimensions. 49

4.1 Exposure-response surfaces used in simulation. The axes on the horizontal plane are the first two exposures, and the vertical axis is the mean response. The exposure-response surface is constant across the third exposure. 66

4.2 Estimated sex-specific single-exposure exposure-response functions for each metal exposure (rows) and each model (columns) in $n = 349$ children in the Bangladesh cohort. The line represents the posterior mean response for an individual in the specified group as exposure to one metal varies (x-axis) and the other metals are fixed at their medians. The ribbons represent 0.95 CIs for the curve, and GS indicated the group-separable models. 69
4.3 Estimated sex-specific and between-group differences in single-exposure effects of lead on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. Panel (a) shows the estimated estimated average change in BSID-III for an IQR change in lead, while arsenic and manganese are fixed at their medians, for both girls (green triangles) and boys (purple circles) separately. Panel (b) shows the estimated difference in this effect between girls and boys. 70

A.1 Pointwise effect estimates of ambient $PM_{2.5}$ on BWGAZ from a standard DLM with 10 exposure-time degrees of freedom and a DLIM(20,20) with Colorado EnviroScreen health and social factors (HSF) score as a modifier. The top panel shows estimated exposure-time-response functions for each quartile of the HSF scores: scores 32, 46, and 59. The bottom panel shows estimated point-wise effects for a week around the middle of each trimester, excluding the first and last 1% of HSF scores. We include confidence bands of 95% around all estimates. 106
A.2 Estimated exposure-time-response functions for each week of first trimester using both a DLIM-linear with 20 nominal degrees of freedom and standard DLM. 107
A.3 Estimated exposure-time-response functions for each week of second trimester using both a DLIM-linear with 20 nominal degrees of freedom and standard DLM. 108
A.4 Estimated exposure-time-response functions for each week of third trimester using both a DLIM-linear with 20 nominal degrees of freedom and standard DLM. 109

B.1 Runtime averaged across 5 data sets for different number of modifiers. Data sets were simulated using the same generating mechanisms described in the simulation section of the main text. 115
B.2 Posterior inclusion probabilities for each weight for DLIM-IM with modifier selection for all 200 simulated data sets with for all weight scenarios and signal-to-noise (SNR) settings. The red lines are fixed at 1 for non-zero weights and 0 for zero weights. The black line is fixed at the inclusion threshold 0.5. 116
B.3 Estimated exposure-time-response functions (gray lines) from the three models considered in the main text (DLIM-IM, DLIM-IM with selection, fixed index model) at 3 modifier index values (0.3, 0.5, 0.7) for 10 individual simulated data sets. The true exposure-time-response function is presented in black. The data sets were simulated using a signal-to-noise ratio of 1 in the equal weights scenario (scenario 1). 118
B.4 Estimated exposure-time-response functions (gray lines) from the three models considered in the main text (DLIM-IM, DLIM-IM with selection, fixed index model) at 3 modifier index values (0.3, 0.5, 0.7) for 10 individual simulated data sets. The true exposure-time-response function is presented in black. The data sets were simulated using a signal-to-noise ratio of 1 in the different weights scenario (scenario 2). 119

B.5 Estimated exposure-time-response functions (gray lines) from the three models considered in the main text (DLIM-IM, DLIM-IM with selection, fixed index model) at 3 modifier index values (0.3, 0.5, 0.7) for 10 individual simulated data sets. The true exposure-time-response function is presented in black. The data sets were simulated using a signal-to-noise ratio of 1 in the sparse weights scenario (scenario 3). 120
B.6 Heat map of the absolute correlations for the 14 candidate modifiers used in the Colorado EnviroScreen analysis. 122
B.7 Heat map of the absolute correlations for the 4 stress modifiers used in the PROGRESS analysis. 123
B.8 Estimated cumulative effect from DLIM-IMs with selection using the Colorado birth cohort data set with 3 and 4 degrees of freedom in both the modifier and exposure-time dimensions. 124
B.9 Estimated exposure-time-response curves for DLIM-IMs with selection using the Colorado birth cohort data set with 3 and 4 degrees of freedom in both the modifier and exposure-time dimensions. 125
B.10 Estimated exposure-time-response functions across a grid of modifier index values (HSF) from the DLIM-IM with selection on the 14 modifiers (Multi-modifier model) presented in the main text and a DLIM-IM with one modifier (single-modifier model). The asterisk denotes modifiers that were found to contribute to modification in the multi-modifier model in the main text analysis. 126
B.11 Estimated exposure-time-response functions across a grid of modifier index values (HSF) from the DLIM-IM with selection on the 14 modifiers (Multi-modifier model) presented in the main text and a DLIM-IM with one modifier (single-modifier model). The asterisk denotes modifiers that were found to contribute to modification in the multi-modifier model in the main text analysis. 127
B.12 Estimated exposure-time-response functions across a grid of modifier index values (HSF) from the DLIM-IM with selection on the 14 modifiers (Multi-modifier model) presented in the main text and a DLIM-IM with one modifier (single-modifier model). The asterisk denotes modifiers that were found to contribute to modification in the multi-modifier model in the main text analysis. 128
B.13 Cumulative effect estimates for LDL 48 months after parturition from a DLIM-IM without selection using the PROGRESS cohort data set fit with 3 degrees of freedom in both the modifier and exposure-time dimensions. 129
B.14 Estimated exposure-time-response curves for LDL 48 months after parturition from a DLIM-IM without selection using the PROGRESS cohort data set fit with 3 degrees of freedom in both the modifier and exposure-time dimensions, evaluated at the 10^{th} ($\hat{m}^* = 0.15$), 25^{th} ($\hat{m}^* = 0.27$), 30^{th} ($\hat{m}^* = 0.29$), 40^{th} ($\hat{m}^* = 0.32$), 75^{th} ($\hat{m}^* = 0.47$), and 90^{th} ($\hat{m}^* = 0.58$) percentiles of the estimated modifier index. 130
B.15 Cumulative effect estimates for HDL 48 months after parturition from a DLIM-IM without selection using the PROGRESS cohort data set fit with 3 degrees of freedom in both the modifier and exposure-time dimensions. 131

B.16 Estimated exposure-time-response functions from a DLIM-IM using the 4 stress modifiers (multi-modifier model in solid black) presented in the main text and a DLIM-IM with one of the stress modifiers (single-modifier model in dashed blue) from the PROGRESS data set. The estimated exposure-time-response curve is presented at the 25th, 40th, and 75th percentiles of the estimated modifier index from the multi-modifier model presented in the main text. 132

C.1 Estimated sex-specific single-exposure effects of lead, manganese, and arsenic on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. This is the estimated average change in BSID-III for an IQR change in exposure, while other exposures are fixed at their medians, for both girls (green triangles) and boys (purple circles) separately. 137

C.2 Estimated between-group differences (girls effect minus boys effect) in single-exposure effects of lead, manganese, and arsenic on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. This is the estimated difference between girls and boys average change in BSID-III for an IQR change in exposure, while other exposures are fixed at their medians. 138

C.3 Estimated sex-specific total effects of the mixture on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. This is the estimated average change in neurodevelopment score for a change in all exposures between their first quartile and median (facet 1) and between their third quartile and median (facet 2) for girls (green triangles) and boys (purple circles). 139

C.4 Estimated differences (girls effect minus boys effect) between girls and boys in total effects of the mixture on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. This is the estimated difference girls and boys average change in neurodevelopment score for a change in all exposures between their first quartile and median (facet 1) and between their third quartile and median (facet 2). 140

Chapter 1

Introduction

A wide body of literature supports an association between environmental pollutants and human health (Bosetti et al., 2010; Jacobs et al., 2017; Pan et al., 2024; Šrám et al., 2005; Stieb et al., 2012; Vuong et al., 2020). In this dissertation, I consider the association of environmental pollutants with birth and child health outcomes. Previous studies have found an association between fetal exposure to environmental pollutants (e.g., particulate matter, heavy metals, nitric oxides, per- and polyfluoroalkyl substances, and pesticides) and birth and child health outcomes, including asthma (Bose et al., 2017; Hsu et al., 2015; Lee et al., 2018), premature birth (Bosetti et al., 2010; Chang et al., 2015; Warren et al., 2012), low birth weight (Bosetti et al., 2010; Coker et al., 2016; Dai et al., 2023; Warren et al., 2013), neurodevelopmental issues (Liu et al., 2022; Vuong et al., 2020), risk of developing chronic diseases (Farzan et al., 2025; Mu et al., 2024), and other adverse outcomes (Chiu et al., 2016). Such studies frequently analyze multivariate exposures that exist in many forms and come with modeling challenges. In this dissertation, I consider two common forms of multivariate exposures. First, I consider repeated measures of the same exposure over time, which can be challenging to model due to correlation between repeated measures of exposure. Second, I consider measures of multiple exposures at a single time point, which is challenging to model due to correlation between exposures and the potential for complex associations. While there are numerous statistical methods for multivariate exposures, the vast majority are designed to estimate homogeneous associations across the population.

Research supports that the effects of environmental exposures on health can be heterogeneous (Arcaya et al., 2016; Schnake-Mahl et al., 2020). Epidemiologists are often interested in whether subgroups of the population have similar or different health effects of exposure. Formally, effect modification is the phenomenon in which there is a relationship between exposures and the response that varies across strata of another variable, which I refer to as the modifier. For multivariate exposure settings with effect modification, the cumulative effect of the exposures could

vary across levels of the modifier and/or variation across the modifier could be on the univariate level. The concept of effect modification is similar to interaction, though true effect modification implies that the modifier is not associated with the exposures, or if so, then all confounders are controlled for (VanderWeele, 2009). We will use the terms interchangeably, which is common in non-causal settings such as are discussed in this dissertation.

Without advancements in effect modification modeling methods, researchers are often forced to choose from approaches that limit inference. For example, a researcher may average repeated measures of the exposure across time and regress the response onto the averaged exposures interacted with the modifying variable (Martenies et al., 2022b). This allows for estimating how the effect varies by the modifier, but loses information on the importance of exposure timing through the averaged exposures. Another popular approach is to fit separate models for each level of the modifier (Niu et al., 2022). If the modifier is continuous, this first requires discretizing the modifier. Fitting stratified models limits inference on differences between modifier levels and is difficult or inefficient with modifiers that contain many levels or are continuous. In this dissertation, I present extensions for models in multivariate exposure settings that include effect modification to address these limitations.

1.1 Distributed Lag Model

Distributed lag models (DLMs) are a popular approach for modeling a response as a function of repeated measures of an exposure (Almon, 1965). Given a response y_i, exposure x_{it} at time point $t = 1, \ldots, T$, and a vector of covariates $\mathbf{z}_i$ for individual i, the standard DLM is

$$y_i = \sum_{t=1}^{T} x_{it}\beta_t + \mathbf{z}_i'\boldsymbol{\gamma} + \varepsilon_i, \tag{1.1}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent and normally distributed. Here, β_t is the linear association between the response and the exposure at exposure-time point t, and the vector $[\beta_1, \ldots, \beta_T]$ defines the exposure-time-response function. Due to correlation between exposure-time points, the model

is regularized by smoothing the exposure-response function. There are various approaches for regularization, including spline expansion (Schwartz, 2000), regression trees (Mork et al., 2024), and Gaussian processes (Warren et al., 2020).

In a DLM, inference is made through estimates of the exposure-time-response function [$\beta_1, \ldots,$ β_T]. The estimated exposure-time-response function can be used to identify both windows of susceptibility and cumulative effects of exposure on response. A window of susceptibility is a range of the exposure-time points t during which there is an increased association between exposure and response. The cumulative effect is the total effect of a one-unit difference in the exposure on the response across the exposure history, or the sum of β_t across the exposure time-points $t = 1, \ldots, T$. There are few options for including effect modification with the existing DLM framework; therefore, researchers are forced to use approaches that limit inference, such as stratification or aggregation across repeated measures.

1.2 Bayesian Kernel Machine Regression

Myriad machine learning and statistical models exist for modeling environmental mixtures (Joubert et al., 2022). Mixture exposures are multivariate exposures that include multiple environmental pollutants (e.g., fetal exposure to heavy metals *in utero*) measured for the same individual. Bayesian kernel machine regression (BKMR) is a popular choice of model for environmental mixture analyses due to its versatility (Bobb et al., 2015, 2018). BKMR is Gaussian process regression that models the response y_i as a flexible function of the multiple exposures and controls for covariates. To describe BKMR (Chapter 4 and Section 1.2), I switch notation by using $\mathbf{z}$ for the exposures and $\mathbf{x}$ for covariates; whereas, I use $\mathbf{x}$ for the exposure and $\mathbf{z}$ for covariates to describe DLMs (Chapters 2 and 3 and Section 1.1). This follows convention in the literature for both the DLM and BKMR. The BKMR model is

$$y_i = h(\mathbf{z}_i) + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \tag{1.2}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent and normally distributed. The first term $h(\mathbf{z}_i)$ is the exposure-response function, which represents the complex associations between the exposures and response. BKMR accommodates most goals of a mixture analysis, including identifying exposures associated with the response, overall effect estimation, allowing for nonlinearities and interactions between exposures (Joubert et al., 2022). Despite the plethora of statistical methods for the analysis of mixtures, there are limited methods for effect measure modification or modeling heterogeneity with mixture exposures. Instead, researchers often rely on stratified analyses, which can be inefficient and limit inference.

1.3 Overview

The standard modeling approaches detailed in Sections 1.1 and 1.2 assume that the effects of the exposure on the response are homogeneous and do not allow for identifying differences for sub-populations. In Chapters 2, 3, and 4, I extend these standard modeling approaches to include effect modification that allows researchers to identify differences across the population based on one or multiple modifiers. Chapter 2 develops methods for modification by a single continuous modifier, and Chapter 3 introduces modification by a weighted combination of multiple modifiers. In Chapter 4, I propose and evaluate methods to include categorical effect modification in BKMR.

Chapter 2

Penalized Distributed Lag Interaction Model

2.1 Introduction

A growing body of epidemiological research supports an association between maternal exposure to air pollution and birth and children's health outcomes (Šrám et al., 2005; Bosetti et al., 2010; Stieb et al., 2012; Lakshmanan et al., 2015; Jacobs et al., 2017). One goal of such studies is to estimate the exposure-time-response function, which quantifies the association between exposures at multiple time points and a subsequent outcome. Another goal is to identify windows of susceptibility—periods during gestation when increased exposure can alter fetal development. Previous studies have estimated the exposure-time-response function and identified critical windows of susceptibility to air pollution associated with increased risk for asthma (Hsu et al., 2015), premature birth (Warren et al., 2012; Chang et al., 2015), low birth weight (Warren et al., 2013), and other adverse outcomes (Chiu et al., 2016). Most studies focused on estimating the exposure-time-response relationship and windows of susceptibility assume homogeneity across the population. However, both the toxicity of chemical exposure and the timing of the windows of susceptibility may vary by individual- or neighborhood-level factors (Becklake and Kauffmann, 1999; Lakshmanan et al., 2015; Wilson et al., 2017a; Lee et al., 2020; Niu et al., 2022). Some studies have aimed to address this heterogeneity with models stratified by a single categorical factor such as child sex (Bose et al., 2017; Lee et al., 2018; Bose et al., 2018). However, there are still gaps in statistical methodology aiming to identify windows of susceptibility and estimate the exposure-time-response relationship with modification by continuous individual- and neighborhood-level variables.

Distributed lag models (DLMs) are a popular method to estimate the association between an exposure observed longitudinally, such as weekly or daily measures of exposure during pregnancy, and birth and children's health outcomes (Hsu et al., 2015; Chiu et al., 2016, 2017; Bose et al.,

2017). The objective of a DLM analysis is to simultaneously estimate the exposure-response relationship at each of the exposure time points, i.e., the exposure-time-response relationship. Compared to using exposure averaged over pre-specified windows, such as clinically defined trimesters, a DLM allows for data-driven identification of windows of susceptibility and has reduced bias for the time-specific exposure-response association (Wilson et al., 2017b). A standard DLM analysis assumes a linear exposure-response relationship between each exposure time and the outcome (Schwartz, 2000; Zanobetti, 2000). The linear effect of exposure at each time point is usually constrained to vary smoothly across exposure times to reduce variance of the estimator that results from high autocorrelation between repeated measures of exposure.

Many authors have proposed extensions to DLMs adding flexibility to address specific research questions or improve performance. Armstrong (2006) and Gasparrini et al. (2010) proposed the distributed lag non-linear model, which is of particular interest to this paper. Distributed lag non-linear models relax the linear assumption for DLMs by introducing a cross-basis, a bi-dimensional representation of both the exposure time and the exposure values. While this allows for non-linear exposure relationships at different time points, both the standard DLM and distributed lag non-linear model frameworks assume homogeneous associations across the population.

To estimate modification in the distributed lag model framework, Wilson et al. (2017a) proposed a Bayesian distributed lag interaction model (BDLIM). BDLIMs allow for interaction between repeated measures of exposure and a single categorical factor within the DLM framework. However, this approach is limited by a small set of categorical variables or by stratifying continuous variables to use as the effect modifier. Warren et al. (2020) developed a spatially-varying Gaussian process model for windows of susceptibility, which allows the exposure-time-response function to vary across areal units (e.g., census tract). However, this approach does not attribute variation in the exposure-time-response function to any specific area-level modifying variable. Mork et al. (2024) proposed a heterogeneous DLM that extends the DLM framework using Bayesian additive regression trees to estimate different exposure-time-response functions for various sub-groups across many covariates. The heterogeneous DLM allows for increased flexibility with multiple

continuous and categorical modifiers, but it comes at the cost of increased computational expense, is more challenging to interpret the resulting posterior sample, and does not impose a smoothness constraint on the modifying factor. Existing methodology is clearly lacking a flexible and computationally efficient method to assess modification attributable to a single continuous modifying variable.

Previous studies have aimed to identify effect modification based on a single continuous variable but were forced to choose alternative approaches due to the lack of methods for a single continuous modifier. One approach is discretizing the continuous modifier into categories. For example, Niu et al. (2022) used BDLIMs to investigate the association between air pollution and birth weight stratified by individual stress measurements and the California EnviroScreen composite score, a neighborhood-level measure of burden from pollution and population vulnerability. Several other studies have used BDLIMs with dichotomized versions of continuous predictors, including antioxidant intake (Lee et al., 2020) and maternal stress (Bose et al., 2017; Lee et al., 2018). Such an approach is limited in identifying windows of susceptibility by pre-determined strata of the study population. Another approach is using time-averaged measures of exposure (e.g., pregnancy average exposure) modified by a continuous factor. Martenies et al. (2022b) use a neighborhood-level score modeled after the California EnviroScreen score as a continuous modifier but fit a linear regression model using annual average exposure. The aggregation in this approach prevents identifying windows of susceptibility during pregnancy. Neither of these approaches can identify windows of susceptibility and estimate modification by a continuous variable.

We investigate the potential for the association between maternal exposure to ambient air pollution and birth weight to be modified by area-level factors that may influence vulnerability. Evidence shows that a myriad of neighborhood-level factors (e.g., gentrification, poverty rate, green space, residential segregation, and healthcare accessibility) can modify the association between environmental exposures and health outcomes (Arcaya et al., 2016; Schnake-Mahl et al., 2020; Martenies et al., 2022a). Estimating the modification of many neighborhood factors is challenging because neighborhood-level factors can be highly correlated with each other, making it difficult to

differentiate the modification attributable to separate correlated factors. Therefore, several studies have used an index aggregating multiple neighborhood factors as a candidate modifying factor to study the effect of air pollution on health. Most of these studies considered averaged exposure (Martenies et al., 2022b), while others have used DLMs for repeated measures of exposure stratified by a dichotomized version of a composite index (Niu et al., 2022). We consider the Colorado EnviroScreen health and social factors score, which is a measure of neighborhood-level health and social vulnerabilities in Colorado, as a modifier of the association between ambient air pollution and birth weight (Colorado Department of Public Health and Environment (CDPHE), 2022; Colorado EnviroScreen Tool Team, 2022).

In this paper, we propose a distributed lag interaction model that allows each individual to have a personalized exposure-time-response function based on a continuous modifying variable. Unlike the standard DLM, our model can estimate unique effects for each possible value of the modifying factor across all exposure time points. The proposed approach builds on the cross-basis approach for the distributed lag non-linear model framework of Gasparrini et al. (2010) but adapts it to capture the modified exposure-time-response relationship. Specifically, the approach uses basis expansions in two dimensions. Expansion in the modifier dimension allows the exposure-time-response function to vary for different levels of the modifier. Expansion in the exposure-time dimension constrains the exposure-time-response function to vary smoothly across exposure time points, providing the necessary regularization in the presence of multicollinearity between repeated measures of exposure. We estimated the association between maternal exposure to ambient fine particulate matter ($PM_{2.5}$) and birth weight using a large administrative data set derived from the Colorado Birth Registry. In our analysis, we accounted for modification with the Colorado EnviroScreen health and social factors score to assess how maternal neighborhoods influence vulnerability to ambient $PM_{2.5}$. We identified windows of susceptibility and were also able to identify neighborhoods with heightened sensitivity. Our software is available in the `R` package `dlim`. The package is publicly available on GitHub.

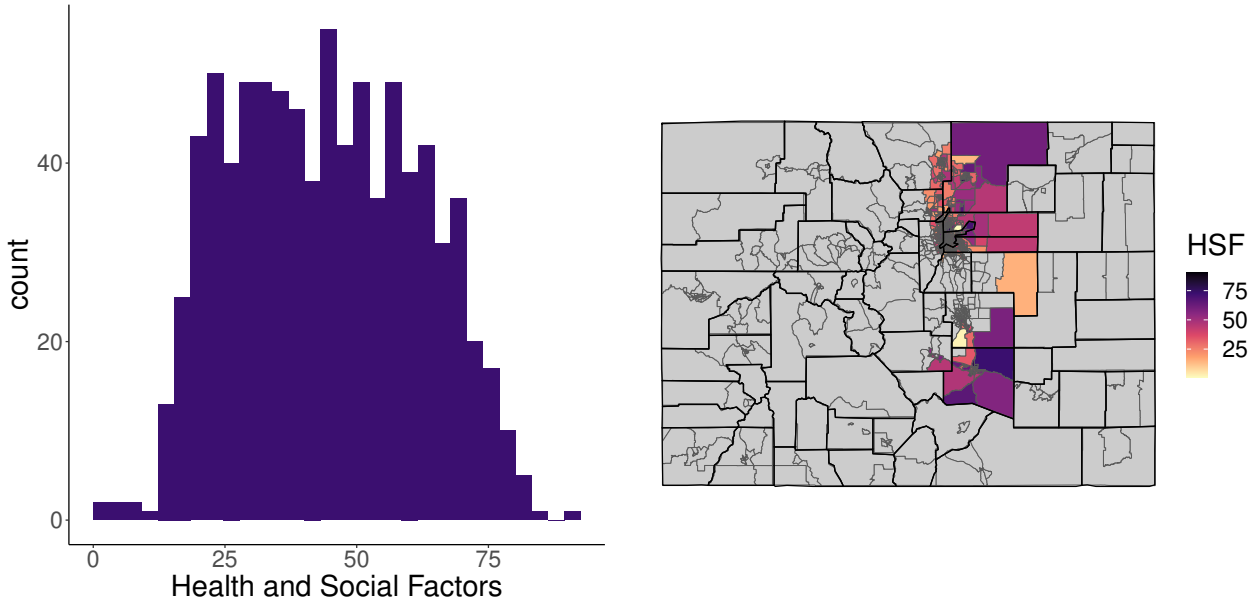


Figure 2.1: Summaries of Colorado EnviroScreen health and social factors (HSF) scores for census tracts selected in this study. Histogram of HSF scores for census tracts included in this study, with quartiles at scores 32, 46, and 59 (left). Map of the state of Colorado shaded by census tract. Tracts not included in our study are shaded in gray (right).

2.2 Colorado Birth Cohort Data

Our Colorado birth cohort data set contains all live births in Colorado, USA with dates of birth between 2007 and 2018. We limited our analysis to full-term (estimated gestational age of 37 weeks or longer), singleton births with estimated dates of conception between 2007 and 2017. Exposure measurement error may be greater in the mountainous regions of the state, and confounding by elevation is associated with lower birth weight (Yip, 1987). To reduce the impact of both exposure measurement error and confounding by elevation, we further limited to births in counties constituting Colorado's Front Range region (extending east of the Rocky Mountains from Colorado Springs to the Wyoming border) and in census tracts at elevation, based on the centroid, of 6000 feet or lower (Figure 2.1).

The data set includes birth weight, along with measurements before pregnancy of the pregnant person's age, body mass index (BMI), height, weight, race and ethnicity, education, income, and marital status. The data set also includes maternal self-reported prenatal care habits and smoking habits during pregnancy. See Appendix Table A.9 for more information on these variables. The

outcome of interest is birth weight for gestational age z-score (BWGAZ) constructed using the Fenton growth charts (Fenton, 2003).

We obtained environmental justice scores from Colorado EnviroScreeen on the census-tract level (Colorado EnviroScreen Tool Team, 2022; Colorado Department of Public Health and Environment (CDPHE), 2022). Colorado EnviroScreen synthesizes environmental exposures and effects, climate vulnerability, sensitive populations, and demographics across multiple years to create a proxy for long-term environmental injustice in an area. Scores are percentiles, ranging from 0 to 100, with a higher score indicating that the area is more likely to experience environmental health injustices. We used the health and social factors score, which quantifies a population's long-term susceptibility to environmental changes and vulnerability to social or economic adversities in an area. The health and social factors score comprises two component scores: a sensitive population score and demographics score. The sensitive population component score is an aggregation of variables on the census-tract level: asthma hospitalization rate, prevalence of cancer and diabetes, rate of heart disease in adults, life expectancy, rate of low birth weight, prevalence of poor mental health conditions, and the percent of elderly and young children for an area. The demographics component score combines the percent of the total population in a census tract that are burdened by housing costs, have disabilities, have less than high school education, have limited English proficiency, are people of color, and have low income. We linked the EnviroScreen health and social factors score to the Colorado birth cohort based on census tract of maternal residence at birth.

We also obtained daily ambient $PM_{2.5}$ concentrations for each census tract from the downscaled models published by the US Environmental Protection Agency, which combine measurements from regulatory monitoring networks and output from atmospheric chemistry models (www.epa.gov/hesc/rsig-related-downloadable-data-files). Using the census tract of maternal residence at birth, we linked exposure data to each birth and computed average exposure for each week of gestation. We also linked ambient daily maximum temperature at the census-tract level from Abatzoglou (2013). Our final analysis data set included 393,205 births in 786 census tracts across 13 counties.

2.3 Modeling Framework

2.3.1 Model Approach

For individual $i = 1 \dots n$, y_i is the response, x_{it} is the exposure at time t for $t = 1, \dots, T$, and $\mathbf{z}_i$ is a covariate vector including the intercept. Assuming y_i follows an exponential family distribution, a standard DLM with no modification is

$$g[E(y_i|\mathbf{x}_i, \mathbf{z}_i)] = \sum_{t=1}^{T} x_{it}\beta_t + \mathbf{z}_i'\boldsymbol{\gamma}, \tag{2.1}$$

where g is the link function, β_t is the linear effect of exposure at time t on the response, and $\boldsymbol{\gamma}$ is a vector of regression coefficients. The vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_T)'$ is the exposure-time-response function. The exposure-time-response function of a standard DLM is the same for all individuals.

We are interested in estimating how the exposure-time-response function varies as a function of a scalar, continuous modifying variable m_i. Throughout, we assume that m_i is also included in the covariate vector $\mathbf{z}_i$. Our proposed distributed lag interaction model is

$$g[E(y_i|\mathbf{x}_i, \mathbf{z}_i, m_i)] = \sum_{t=1}^{T} x_{it}\beta_t(m_i) + \mathbf{z}_i'\boldsymbol{\gamma}. \tag{2.2}$$

Unlike the standard DLM formulation, the linear effect of exposure on the response at exposure time t for individual i, denoted $\beta_t(m_i)$ in model (3.1), depends on the value of the modifier m_i for that individual. In our analysis, m_i is the health and social factors score for the census tract of maternal residence for birth i.

We target two parameters from model (3.1) for inference. First, we are interested in the exposure-time-response function for the i^{th} individual with modifier value m_i, which is defined as $\boldsymbol{\beta}(m_i) = [\beta_1(m_i), \dots, \beta_T(m_i)]'$. The t^{th} element of $\boldsymbol{\beta}(m_i)$, $\beta_t(m_i)$, is the linear effect of exposure at time point t for individual i with modifier value m_i. Second, we are interested in the cumulative effect of exposure for each individual. In model (3.1), the cumulative effect for indi-

vidual i is the expected difference in the response associated with one unit higher exposure at all exposure time points, defined as $\delta(m_i) = \sum_{t=1}^{T} \beta_t(m_i)$.

2.3.2 Parameterization of exposure-time-response functions

We allow $\beta_t(m_i)$ in (3.1) to vary smoothly in both the exposure-time and modifier dimensions through a penalized spline approach. Following the penalized cross-basis formulation proposed by Gasparrini et al. (2017), we create a cross-basis for the modifier-time space.

To allow for the linear effect of exposure at each time point to vary smoothly as a function of the modifier, we use B-spline basis expansions of the modifier values. The basis expansion for modifier value m_i is denoted $\mathbf{b}(m_i) = [b_1(m_i), \ldots, b_{\nu_{mod}}(m_i)]'$, where ν_{mod} is the user-defined number of basis functions for the modifier basis expansion. We can write $\beta_t(m_i)$ as a linear combination of modifier basis values,

$$\beta_t(m_i) = \sum_{k=1}^{\nu_{mod}} b_k(m_i)\eta_{tk}. \tag{2.3}$$

Then, we use B-splines to create a basis expansion of the exposure time points to add a smoothness constraint in the time dimension and to regularize the model, reducing the effect of multicollinearity between repeated measures of exposure. The exposure-time basis expansion for time point t is denoted $\mathbf{c}(t) = [c_1(t), \ldots, c_{\nu_{time}}(t)]'$, where ν_{time} is the user defined number of basis functions for the exposure-time basis expansion. We can write η_{tk} from (2.3) as a linear combination of exposure-time basis values,

$$\eta_{tk} = \sum_{j=1}^{\nu_{time}} c_j(t)\theta_{jk}, \tag{2.4}$$

where θ_{jk} for $k = 1, \ldots, \nu_{mod}$ and $j = 1, \ldots, \nu_{time}$ are regression parameters.

Using the parameterizations (2.3) and (2.4), the first term on the right-hand side of (3.1) can be written in terms of $\{\theta_{jk}\}_{j=1,k=1}^{\nu_{mod},\nu_{time}}$ as

$$\begin{aligned}\sum_{t=1}^{T} x_{it}\beta_t(m_i) &= \sum_{k=1}^{\nu_{mod}}\sum_{t=1}^{T} x_{it}b_k(m_i)\sum_{j=1}^{\nu_{time}} c_j(t)\theta_{jk} \\ &= \sum_{j=1}^{\nu_{time}}\sum_{k=1}^{\nu_{mod}}\left[\sum_{t=1}^{T} x_{it}b_k(m_i)c_j(t)\right]\theta_{jk} \\ &= \sum_{j=1}^{\nu_{time}}\sum_{k=1}^{\nu_{mod}} w_{jk}(m_i,\mathbf{x}_i)\theta_{jk},\end{aligned} \tag{2.5}$$

where $w_{jk}(m_i,\mathbf{x}_i)$ is an element of the cross-basis for individual i representing the j^{th} element in the expansion of the exposure-time basis and the k^{th} element in the expansion of the modifier basis. Incorporating the representation in (2.5) into (3.1), we have the following reparameterized model

$$g[E(y_i|\mathbf{x}_i,\mathbf{z}_i,m_i)] = \sum_{j=1}^{\nu_{time}}\sum_{k=1}^{\nu_{mod}} w_{jk}(m_i,\mathbf{x}_i)\theta_{jk} + \mathbf{z}_i'\boldsymbol{\gamma}. \tag{2.6}$$

Since $\beta_t(m_i)$ for each $t = 1,\ldots,T$ is a linear combination of $\{\theta_{jk}\}_{j=1,k=1}^{\nu_{mod},\nu_{time}}$, we obtain all estimates of the exposure-time-response function and cumulative effects using linear transformations of the estimates of $\{\theta_{jk}\}_{j=1,k=1}^{\nu_{mod},\nu_{time}}$ (see Section 2.3.4 for more details).

2.3.3 Parameter Estimation

Various basis expansions and estimation procedures can be applied to estimate the model presented in Section 2.3.2. In this paper, we focus on a penalized spline approach to allow for data-adaptive smoothness in the exposure-time-response surface. We discuss alternative penalty methods in Section 2.3.5. Let $l(\boldsymbol{\theta},\boldsymbol{\gamma})$ be the likelihood for our unpenalized model, where $\boldsymbol{\theta} = [\theta_{11},\ldots,\theta_{\nu_{time}1},\theta_{12},\ldots,\theta_{\nu_{time}2},\ldots,\theta_{1\nu_{mod}},\ldots,\theta_{\nu_{time}\nu_{mod}}]'$. Let $\mathbf{D}_\nu$ be a $(v-2)\times v$ second order difference matrix and $\mathbf{I}_\nu$ be a $\nu\times\nu$ diagonal matrix. Define a $\nu\times\nu$ squared difference matrix $\mathbf{S}_\nu^* = \mathbf{D}_\nu'\mathbf{D}_\nu$. To obtain penalty matrices $\mathbf{S}_{\nu_{time}}$ and $\mathbf{S}_{\nu_{mod}}$, we divide $\mathbf{S}_{\nu_{time}}^*$ and $\mathbf{S}_{\nu_{mod}}^*$ by their first eigenvalues, respectively, to ensure each applies equally weighted penalties. Let $\otimes$ denote the

Kronecker product. The penalized likelihood is

$$l_p(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = l(\boldsymbol{\theta}, \boldsymbol{\gamma}) - \frac{1}{2}\boldsymbol{\theta}'\{\lambda_{mod}(\mathbf{S}_{\nu_{mod}} \otimes \mathbf{I}_{\nu_{time}}) + \lambda_{time}(\mathbf{I}_{\nu_{mod}} \otimes \mathbf{S}_{\nu_{time}})\}\boldsymbol{\theta}, \tag{2.7}$$

where $\boldsymbol{\lambda} = [\lambda_{mod}, \lambda_{time}]$ are penalty parameters that control smoothness. We fit the model with REML (Laird and Ware, 1982) using the `gam` function from the `mgcv` package in `R` (Wood, 2017).

2.3.4 Inference

We conduct inference on both the exposure-time-response function and the cumulative effect for each individual (see Section 2.3.1). Our estimator of the exposure-time-response function for individual i is $\hat{\boldsymbol{\beta}}(m_i) = [\mathbf{b}'(m_i) \otimes \mathbf{C}]\hat{\boldsymbol{\theta}}$, where $\mathbf{C} = [\mathbf{c}(1) \dots \mathbf{c}(T)]'$. The t^{th} element of $\hat{\boldsymbol{\beta}}(m_i)$ can also be written as $\hat{\beta}_t(m_i) = \sum_{j=1}^{\nu_{time}} \sum_{k=1}^{\nu_{mod}} b_k(m_i)c_j(t)\hat{\theta}_{jk}$. The estimator $\hat{\beta}_t(m_i)$ has variance $V[\hat{\beta}_t(m_i)] = \mathbf{1}'_{\nu_{time}\nu_{mod}}\mathbf{A}_t(m_i)\mathbf{1}_{\nu_{time}\nu_{mod}}$, where $\mathbf{A}_t(m_i)$ is the $\nu_{time}\nu_{mod} \times \nu_{time}\nu_{mod}$ matrix

$$\begin{aligned}\mathbf{A}_t(m_i) =& [\mathbf{c}'(t) \otimes \mathbf{1}_{\nu_{mod} \times \nu_{time}\nu_{mod}}] \odot [\mathbf{1}_{\nu_{time} \times \nu_{time}\nu_{mod}} \otimes \mathbf{b}(m_i)] \\ & \odot [\mathbf{c}'(t) \otimes \mathbf{1}_{\nu_{mod} \times \nu_{time}\nu_{mod}})]' \odot [\mathbf{1}_{\nu_{time} \times \nu_{time}\nu_{mod}} \otimes \mathbf{b}(m_i)]' \odot V[\hat{\boldsymbol{\theta}}].\end{aligned} \tag{2.8}$$

The operator $\odot$ represents element-wise multiplication and $V[\hat{\boldsymbol{\theta}}]$ is the covariance matrix for the estimated coefficients. We estimate $V[\hat{\boldsymbol{\theta}}]$ using empirical Bayes estimators within the `mgcv` package in `R` (Marra and Wood, 2012).

Our estimator of the cumulative effect for individual i is $\hat{\delta}(m_i) = \sum_{t=1}^{T} \hat{\beta}_t(m_i)$, or equivalently, $\hat{\delta}(m_i) = \mathbf{w}'_*(m_i)\hat{\boldsymbol{\theta}}$, where $\mathbf{w}_*(m_i)$ is the vectorization by column of $[\mathbf{1}'_T \otimes \mathbf{b}(m_i)]\mathbf{C}$. The variance of $\hat{\delta}(m_i)$ is $V[\hat{\delta}(m_i)] = \mathbf{w}'_*(m_i)V[\hat{\boldsymbol{\theta}}]\mathbf{w}_*(m_i)$.

2.3.5 Alternative Approaches for Penalization

There are potential alternative penalization structures, several of which are explored in Gasparrini et al. (2017). As described in Section 2.3.3, we can impose a penalty in both the modifier and exposure-time dimensions by using difference matrices. The order of the difference matrix is

related to the amount of smoothness imposed. In our main approach, we consider second order difference matrix penalization in both dimensions. We could instead use a first order difference matrix in either dimension to impose less smoothness for that particular dimension. We refer to this method as PS(d_{time}, d_{mod}) penalization, where d_{time} and d_{mod} are the difference matrix orders for the exposure-time and modifier dimensions, respectively.

Another option is to impose penalties on the second derivative in each dimension of the exposure-time-response basis functions (Green and Silverman, 1994). Here, we consider using natural cubic splines with the added constraint of allowing only one spline to be non-zero at each knot for computational convenience (Wood, 2017). We refer to this method as CR penalization.

2.3.6 Alternative Approach without Penalization

We can also perform model estimation without penalization using natural splines. However, the use of natural splines requires *a priori* specification of the degrees of freedom or a data-driven selection process, such as cross-validation, that can be computationally burdensome and difficult in practice. We recommend using AIC to choose degrees of freedom for non-penalized models.

2.3.7 Alternative Approach with Linear Interaction

The proposed model allows for flexible modification of the exposure-time-response function. However, there may be situations where a simpler form of modification is preferable. We can construct the modifier basis using polynomial functions, particularly a polynomial function of degree 1 (i.e., linear), instead of penalized B-splines. The modifier basis expansion for modifier value m_i is then $[1, m_i]$. Using this modifier basis expansion for the cross-basis described in Section 2.3.2 is a distributed lag model with an additional linear interaction term, equivalent to the model $g[E(y_i|x_{it}, m_i, \mathbf{z}_i')] = \sum_{t=1}^{T} x_{it}\beta_t + \sum_{t=1}^{T} x_{it}m_i\beta_{1t} + \mathbf{z}_i'\boldsymbol{\gamma}$. We refer to this reduced model as a distributed lag model with linear interaction. Increasing the degree of the polynomial basis for the modifier increases the additional number of interaction terms. Estimation and inference in Sections 2.3.3 and 2.3.4 proceed the same. We impose two penalty matrices in the lag dimension. The first penalty matrix is $diag(1, 0) \otimes \mathbf{S}_{\nu_{time}}$, which penalizes regression coefficient estimates con-

tributing to estimates of the β_t. The second penalty matrix is $diag(0,1) \otimes \mathbf{S}_{\nu_{time}}$, which penalizes regression coefficient estimates contributing to estimates of the β_{1t}.

2.3.8 Test for Interaction

To test for effect modification, we developed a bootstrap likelihood ratio test using the approach in Roca-Pardiñas et al. (2005). The null model is a standard DLM, as in (2.1), a model without effect modification. The full model is a model with effect modification, either our proposed model in (3.1) or alternative model presented in Section 2.3.7. First, we fit the null model and obtain fitted values $\tilde{y}_i$. For $b = 1, \ldots, B$, we first generate a sample y_i^{*b} for $i = 1, \ldots, n$, with y_i^{*b} drawn from the assumed exponential family distribution built around $\tilde{y}_i$. For a Gaussian model, we sample the residuals $(y_i - \tilde{y}_i)$ with replacement and add them to $\tilde{y}_i$. For non-Gaussian GLM, we draw from the distribution parameterized by $\tilde{y}_i$ (e.g., $y_i^{*b} \sim \text{Pois}(\tilde{y}_i)$). We then fit both the full and reduced model to each bootstrap sample and calculate a likelihood ratio test statistic denoted as $\hat{T}^{*b}$. This results in an empirical null distribution for our test. The critical value for specified α level is the $1 - \alpha$ quantile of the bootstrap sample of test statistics $\hat{T}^{*1}, \ldots, \hat{T}^{*B}$. We reject the null hypothesis of no modification when the test statistic fit on the original data is greater than the empirical critical value. We can also reject the null hypothesis based on an empirical bootstrap p-value, which we define as the proportion of bootstrap test statistics that are greater than the test statistic calculated from the original data. We can also apply a similar test to compare the DLIM with nonlinear modification presented in Section 2.3.1 to the distributed lag model with linear interaction presented in Section 2.3.7.

2.3.9 Repeated Measures Among Individuals

We also propose a mixed model representation of the DLIM. For example, one may consider exposure in the weeks prior to each visit in a longitudinal cohort (Peng et al., 2017). Consider the longitudinal study with individuals $i = 1, \ldots, N$ and visits $l = 1, \ldots, n_i$ for individual i. We account for repeated measures among individuals by adding an individual-specific intercept α_i to (3.1). The mixed model DLIM is

$$g[E(y_{il}|\mathbf{x}_{il}, \mathbf{z}_{il}, m_{il})] = \sum_{t=1}^{T} x_{il,t}\beta_t(m_{il}) + \mathbf{z}'_{il}\boldsymbol{\gamma} + \alpha_i. \tag{2.9}$$

We assume the individual-specific intercepts are independent of each other and the residuals, and are distributed normal with mean zero and constant variance. The modifier m_{il} can be constant over time or time dependent. We fit the model with REML and the `mgcv` package as described in Section 2.3.3.

2.4 Simulations

We validated our proposed DLIM with a simulation under four scenarios and three different signal-to-noise ratios. We evaluated the method's ability to estimate the exposure-time-response functions and cumulative effects. We compared the operating characteristics of the proposed model to a DLM with linear interaction (DLIM-linear) as described in Section 2.3.7 and a standard DLM model with no effect modification. We additionally evaluated the empirical performance of the bootstrap test for modification. We provide results for a DLIM(20,20) with PS(2,2) penalization, i.e., a DLIM with 20 basis functions and penalty with a second order difference matrix in the exposure-time basis and 20 basis functions and penalty with a second order difference matrix in the modifier basis, and a DLIM-linear and standard DLM with nominal 10 degrees of freedom for the exposure-time basis and penalty with a second order difference matrix. We discuss selection of degrees of freedom and sensitivity in Section 2.4.4, and we show results for models with different number of basis functions and additional models and data generating mechanisms in Appendix A. Code for implementing these simulations is provided as a companion file to this manuscript.

2.4.1 Simulation Design and Data Generation

We randomly generated 1000 birth dates and constructed 37 weeks of exposure data for each birth from observed ambient $PM_{2.5}$ data. For each birth, we simulated modifier value m_i from a continuous uniform distribution on $(0, 1)$ because the modifier of interest in this paper and other motivating papers is a bounded index (Martenies et al., 2022b; Niu et al., 2022).

We considered four scenarios with increasing complexity of the exposure-time-response function and modification. Figure 2.2 illustrates the four scenarios. We based the data generation scenarios on the function $f(t, c) := 2.5\phi[(t - c)/5]$, where $\phi(\cdot)$ is the normal probability density function. We generated data for the scenarios as follows.

1. **No Modification**: The peak effect and its timing are the same for each individual. To simulate this scenario, we let $\beta_t(m_i) = f(t, 20)$.

2. **Linear Modification**: The timing of peak effect is the same for each individual, but the peak effect is linearly scaled by each individual's modifying value. To simulate this scenario, we let $\beta_t(m_i) = m_i * f(t, 20)$.

3. **Non-linear Shift Modification**: The peak effect is the same for each individual but the timing of peak effect is shifted according to a non-linear function of each individual's modifying value. To simulate this scenario, we allowed the centers of the exposure-time-response function normal curves, ranging from 10 to 30, to be a non-linear function of the modifier value, $\beta_t(m_i) = f(t, 37[1 + \exp\{-20(m_i - 0.5)\}]^{-1})$.

4. **Complex Modification**: The peak effect and its timing are functions of each individual's modifying value. To simulate this scenario, we combined scenarios 2 and 3 to obtain $\beta_t(m_i) = m_i * f(t, 37[1 + \exp\{-20(m_i - 0.5)\}]^{-1})$.

For all scenarios, we generated 3 covariates and their corresponding regression coefficients from a standard normal distribution. To simulate response values, we assumed a Gaussian distribution with variance σ^2 and used an identity link function g in (3.1). We fixed the regression coefficient for the main effect of the modifier at 1. We investigated model fit for increasing signal-to-noise ratios, including 0.5 (low), 1 (medium), and 10 (high). Here, the signal is the standard deviation of the first term across $i = 1, \ldots, n$ in (3.1), and noise is σ. We simulated 200 data sets for each scenario and signal-to-noise combination.

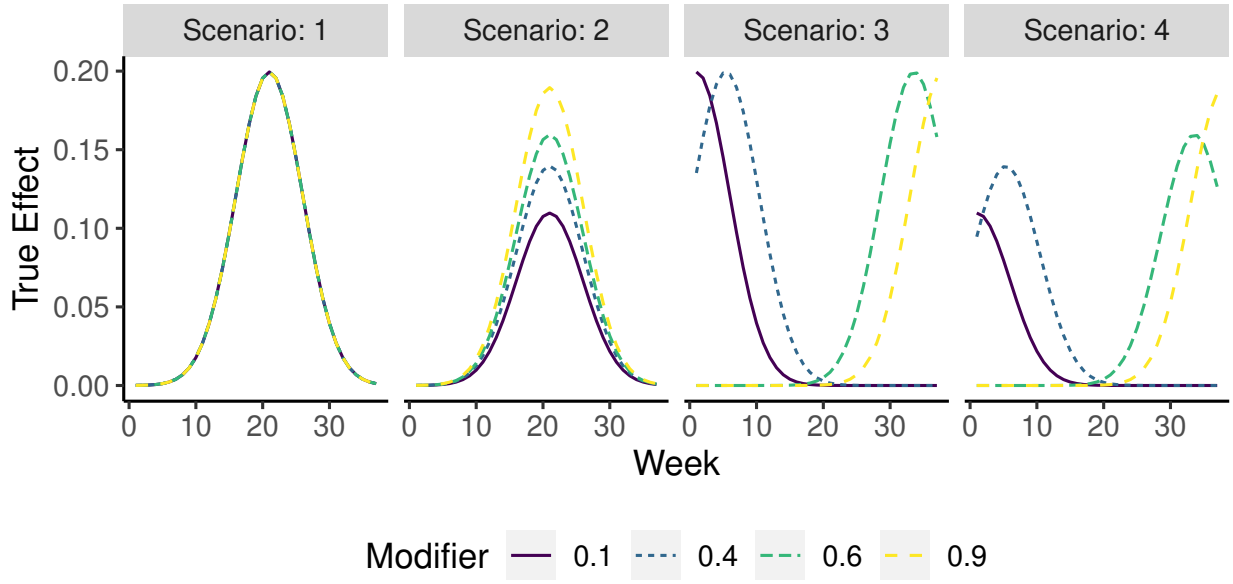


Figure 2.2: Each curve is an exposure-time-response function for an individual, each with different modifying values. In this example, there are $n = 4$ individuals, with modifier values 0.1, 0.5, 0.6, and 0.9, respectively. Each scenario is described in Section 2.4.1.

2.4.2 Performance Evaluation

For each simulated data set, we calculated the root mean squared error (RMSE) for the cumulative effect as $[n^{-1}\sum_{i=1}^{n}\{\delta(m_i) - \hat{\delta}(m_i)\}^2]^{1/2}$ and cumulative coverage as $n^{-1}\sum_{i=1}^{n}\mathbb{I}[|\delta(m_i) - \hat{\delta}(m_i)| \leq \Phi^{-1}(1-\alpha/2)\hat{V}^{1/2}\{\hat{\delta}(m_i)\}]$. Here, $\mathbb{I}(A) = 1$ if event A occurs and $\mathbb{I}(A) = 0$ if event A does not occur. We also calculated RMSE for point-wise effects as $[n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\hat{\beta}_t(m_i) - \beta_t(m_i)\}^2]^{1/2}$ and point-wise coverage as $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{I}[|\beta_t(m_i) - \hat{\beta}_t(m_i)| \leq \Phi^{-1}(1-\alpha/2)\hat{V}^{1/2}\{\hat{\beta}_t(m_i)\}]$. We averaged RMSE and coverage across the 200 simulated data sets.

To evaluate our testing procedure we conducted three tests on each data set. These tests compared: 1) DLIM(20,20) to DLM, 2) DLIM-linear to DLM, and 3) DLIM(20,20) to DLIM-linear. We report the proportion of times that the null was rejected for each test.

2.4.3 Simulation Results

Parameter Estimation

Table 2.1 includes results for simulation scenarios 1 and 2, and Table 2.2 includes results for simulation scenarios 3 and 4. Overall, our simulation demonstrates that the DLIM(20,20) can capture complex effect modification patterns. Further, using the DLIM over the standard DLM framework results in substantially less bias and more accurate inference when there is effect modification and only a small loss in efficiency when there is no modification.

Table 2.1: Cumulative and point-wise performance metrics for penalized models averaged over all 200 simulated data sets for scenarios 1 and 2 and three signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear. The table also includes the proportion of rejection ($\alpha = 0.05$) across simulated data sets under each null model (DLM or DLIM-linear) using the model comparison procedure described in Section 2.3.8.

		Cumulative		Point-wise		Model Comparison	
SNR	Model	RMSE	Coverage	RMSE	Coverage	DLM	DLIM-linear
Scenario 1: No Modification							
low	DLIM(20,20)	0.263	0.95	0.026	0.92	0.04	0.05
	DLIM-linear	0.257	0.95	0.024	0.92	0.06	-
	DLM	0.169	0.96	0.020	0.91	-	-
medium	DLIM(20,20)	0.131	0.95	0.015	0.96	0.04	0.05
	DLIM-linear	0.129	0.95	0.013	0.94	0.05	-
	DLM	0.084	0.96	0.011	0.94	-	-
high	DLIM(20,20)	0.013	0.95	0.003	0.98	0.04	0.04
	DLIM-linear	0.013	0.95	0.002	0.90	0.04	-
	DLM	0.008	0.96	0.002	0.90	-	-
Scenario 2: Linear Modification							
low	DLIM(20,20)	0.331	0.95	0.030	0.90	0.20	0.04
	DLIM-linear	0.329	0.95	0.029	0.89	0.30	-
	DLM	0.438	0.70	0.028	0.83	-	-
medium	DLIM(20,20)	0.166	0.95	0.017	0.93	0.67	0.06
	DLIM-linear	0.166	0.95	0.017	0.91	0.80	-
	DLM	0.383	0.40	0.019	0.78	-	-
high	DLIM(20,20)	0.017	0.95	0.003	0.98	1.00	0.05
	DLIM-linear	0.016	0.95	0.002	0.94	1.00	-
	DLM	0.361	0.06	0.014	0.55	-	-

Table 2.2: Cumulative and point-wise performance metrics for penalized models averaged over all 200 simulated data sets for scenarios 3 and 4 and three signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear. The table also includes the proportion of rejection ($\alpha = 0.05$) across simulated data sets under each null model (DLM or DLIM-linear) using the model comparison procedure described in Section 2.3.8.

		Cumulative		Point-wise		Model Comparison	
SNR	Model	RMSE	Coverage	RMSE	Coverage	DLM	DLIM-linear
Scenario 3: Non-linear Shift Modification							
low	DLIM(20,20)	0.404	0.94	0.049	0.82	1.00	1.00
	DLIM-linear	0.620	0.79	0.053	0.73	0.52	-
	DLM	0.555	0.72	0.065	0.49	-	-
medium	DLIM(20,20)	0.206	0.95	0.036	0.81	1.00	1.00
	DLIM-linear	0.514	0.69	0.048	0.62	0.98	-
	DLM	0.494	0.55	0.063	0.33	-	-
high	DLIM(20,20)	0.025	0.92	0.010	0.91	1.00	1.00
	DLIM-linear	0.470	0.58	0.045	0.55	1.00	-
	DLM	0.469	0.38	0.062	0.23	-	-
Scenario 4: Complex Modification							
low	DLIM(20,20)	0.320	0.94	0.037	0.82	1.00	1.00
	DLIM-linear	0.481	0.79	0.040	0.74	0.52	-
	DLM	0.442	0.71	0.049	0.52	-	-
medium	DLIM(20,20)	0.163	0.95	0.028	0.81	1.00	1.00
	DLIM-linear	0.395	0.69	0.036	0.63	0.98	-
	DLM	0.394	0.54	0.047	0.35	-	-
high	DLIM(20,20)	0.020	0.92	0.008	0.91	1.00	1.00
	DLIM-linear	0.359	0.59	0.034	0.55	1.00	-
	DLM	0.374	0.42	0.046	0.24	-	-

For simulation scenario 1, in which each individual's exposure-time-response function does not depend on its modifier value, all models performed similarly to each other. The standard DLM had cumulative RMSE lower than the DLIMs for all signal-to-noise settings, and slightly lower point-wise RMSE for the low signal-to-noise setting. We expected this because the simulated structure does not include modification to the true exposure-time-response functions. Our results demonstrate that when there is no modification, the DLM is slightly more efficient than both DLIMs. However, the gain in efficiency is relatively small.

For scenario 2, in which each individual's exposure-time-response function is linearly scaled by their modifying value, the DLIMs generally performed better than the standard DLM, which is mis-specified in this scenario. In the high signal-to-noise setting of scenario 2, the DLIM(20,20) and DLIM-linear had much lower cumulative RMSE (0.017 and 0.016, respectively) than the standard DLM (0.361). The DLIM(20,20) and DLIM-linear had much lower point-wise RMSE (0.003 and 0.002, respectively) than the standard DLM's point-wise RMSE (0.014). The DLIM(20,20) and DLIM-linear also had cumulative coverage (0.95) at the nominal level; whereas, the standard DLM's cumulative coverage (0.06) was well below the nominal level. The DLIM(20,20) and DLIM-linear also had point-wise coverage (0.98 and 0.94, respectively) near the nominal level; whereas, the standard DLM's point-wise coverage (0.55) was well below the nominal level. Our results demonstrate that when the signal-to-noise ratio is high, a model accounting for linear interaction in the estimated exposure-time-response function is more favorable than the standard DLM. In the medium and low signal-to-noise settings for scenario 2, the DLIMs out-performed the standard DLM in terms of cumulative effect RMSE and coverage. For point-wise metrics, all three models performed similarly in terms of RMSE, but the DLIM(20,20) and DLIM-linear had coverage near the nominal level (0.90 and 0.89, respectively) while the standard DLM had coverage lower than the nominal level (0.83). Our results show that if true modification to the exposure-time-response function is linear, then it is more appropriate to fit a DLIM. Fitting a standard DLM can result in low coverage.

For the third and fourth simulation scenarios, in which each individual's exposure-time-response function also depends on its modifier value, a DLIM(20,20) was more accurate than the other models. For all signal-to-noise settings, the DLIM(20,20) had lower cumulative and point-wise RMSE than the other models. The DLIM(20,20) also had cumulative and point-wise coverage much closer to the nominal level than the other models. For example, cumulative coverage in the high signal-to-noise setting for scenario 3 was 0.92 for the DLIM(20,20); whereas, the cumulative coverage was 0.58 for the DLIM-linear and 0.38 for the standard DLM in the same setting. In the lower signal-to-noise ratio settings, point-wise coverage for the DLIM(20,20) model is closer to the nominal

level than the other models, but is below the nominal level. For example in the low signal-to-noise setting for scenario 4, the point-wise coverage was 0.82; whereas, the point-wise coverage for the DLIM-linear was 0.74 and 0.52 for the standard DLM. We attribute the low point-wise coverage for these scenarios to attenuated estimates of the peak effect and over-smoothing of modification patterns in the low signal-to-noise settings. This persisted across various choices of basis constructions, number of basis functions, and penalty structures (see Appendix Tables A.1-A.6). As the signal-to-noise increases, our model can better estimate the peak effects and over-smoothing is reduced, resulting in smaller RMSE and coverage closer to the nominal level. Our results demonstrate that when modification to the exposure-time-response function is present, a DLIM(20,20) can accurately capture the complex modification patterns.

Testing Modification

Using the proposed bootstrap test, we were able to accurately identify scenarios with modification and without modification in the simulation study. Tables 2.1 and 2.2 show the proportion of times that the null model was rejected at the $\alpha = 0.05$ level using the test. For scenario 1 with no modification present, the tests comparing the DLIM(20,20) to the standard DLM and comparing the DLIM-linear to the standard DLM both had type I error rates between 0.04 and 0.06 across all signal-to-noise ratios. For scenario 2 with linear modification, the test for comparing the DLIM-linear to the standard DLM had power of 0.30 in the low signal setting, 0.80 in the medium signal setting, and 1.00 in the high signal setting. Additionally, the test for comparing the DLIM(20,20) to the DLIM-linear maintained a type I error rate near the nominal level, ranging from 0.04 to 0.06 across the signal-to-noise levels. For scenarios 3 and 4, the test accurately identified the DLIM(20,20) as the preferred model. The power for testing the DLIM(20,20) compared to both the DLIM-linear and standard DLM was 1.00 across all signal-to-noise settings.

2.4.4 Hyperparameter Selection and additional simulations

The DLIMs with PS(2,2) penalization generally performed better than DLIMs with alternative penalization methods discussed in Section 2.3.5. In particular, the DLIM(20,20) with PS(2,2)

penalization performed better than the DLIM(20,20) with CR, PS(1,2), or PS(2,1) penalization. See Appendix Tables A.5 and A.6 for a comparison to the PS(2,2) penalization method used in the simulation study.

The penalized DLIMs generally performed better with a larger number of basis functions. Overall, our simulation study shows that using a large number of basis functions, e.g., 20, for both the exposure-time and modifier dimensions is best. Both the penalized DLM and penalized DLIM-linear did not benefit from an increased number of basis functions; therefore, we only include results for these models with 10 nominal degrees of freedom in the exposure-time dimension. All non-penalized models performed poorly and were sensitive to the choice of degrees of freedom. See Appendix Tables A.1-A.4 for a comparison between DLIMs fit with few and many basis functions and for non-penalized model results.

We considered an additional simulation with a normally distributed modifier. We present results in Appendix Table A.7, which were consistent with results presented in the main text.

We also demonstrate performance for a GLM version of the model. We present results for a simulation using count data in Appendix Table A.8, which were similar to results in Section 2.4.3 and suggest the use of our proposed model over the standard DLM.

2.5 Analysis of Colorado Birth Cohort Data

2.5.1 Data and Model Fitting

We applied our penalized model to the Colorado birth cohort data set. Our analysis included BWGAZ as the response, weekly exposure to ambient $PM_{2.5}$ for the first 37 weeks of gestation as the exposure of interest, and all covariates described in Section 2.2. In addition, we include categorical variables for maternal county of residence, and year and month of conception (Leung et al., 2023). We used the Colorado EnviroScreen health and social factors score as the modifier. Health and social factors scores quantify neighborhood-level biological susceptibility and social vulnerabilities. Variables for race, ethnicity, and education are included into our model on both the

individual level as covariates and on the neighborhood level as part of the health and social factors score as described in Section 2.2.

We fit a penalized DLM with nominal 20 degrees of freedom, a penalized DLIM-linear with nominal 20 degrees of freedom, and a penalized DLIM(20,20) and performed the model comparison test described in Section 2.3.8. We performed all analyses using our `dlim` package in `R` version 4.3.0.

2.5.2 Results

Model comparisons with the bootstrap test supported using the DLIM-linear. When comparing the DLIM-linear to the DLM, the empirical bootstrap p-value from the test was 0.019, indicating strong support for modification. The empirical bootstrap p-value was 0.077 when comparing the DLIM(20,20) to the DLM and was 0.402 when comparing the DLIM(20,20) to the DLIM-linear. This indicates that the data does not support the need for the more flexible DLIM(20,20) over the DLIM-linear. Furthermore, the results from the DLIM(20,20) show evidence of modification consistent with the DLIM-linear. Figure 2.3 shows the estimated cumulative effect by individual health and social factors score using each method. Cumulative effect estimates from the DLIM(20,20) are similar to those from the DLIM-linear, further supporting that the added flexibility of the DLIM(20,20) is not needed. Therefore, we chose the DLIM-linear for the analysis. See Appendix Figure A.1 for plots of point-wise effect estimates from the DLIM(20,20).

Figure 2.3 shows the cumulative effect for each individual health and social factors score using each method. With a standard DLM, we estimated a negative cumulative effect of ambient $PM_{2.5}$ on BWGAZ for all values of the health and social factors score. With a DLIM-linear, we estimated a negative effect only for low health and social factors scores. Specifically, we identified an association between birth weight and increased maternal exposure to ambient $PM_{2.5}$ for health and social factors scores lower than 50. The cumulative effect for those health and social factors scores ranged from a decrease of 0.014 to 0.034 in BWGAZ for a 1 μg/m^3 difference in ambient $PM_{2.5}$.

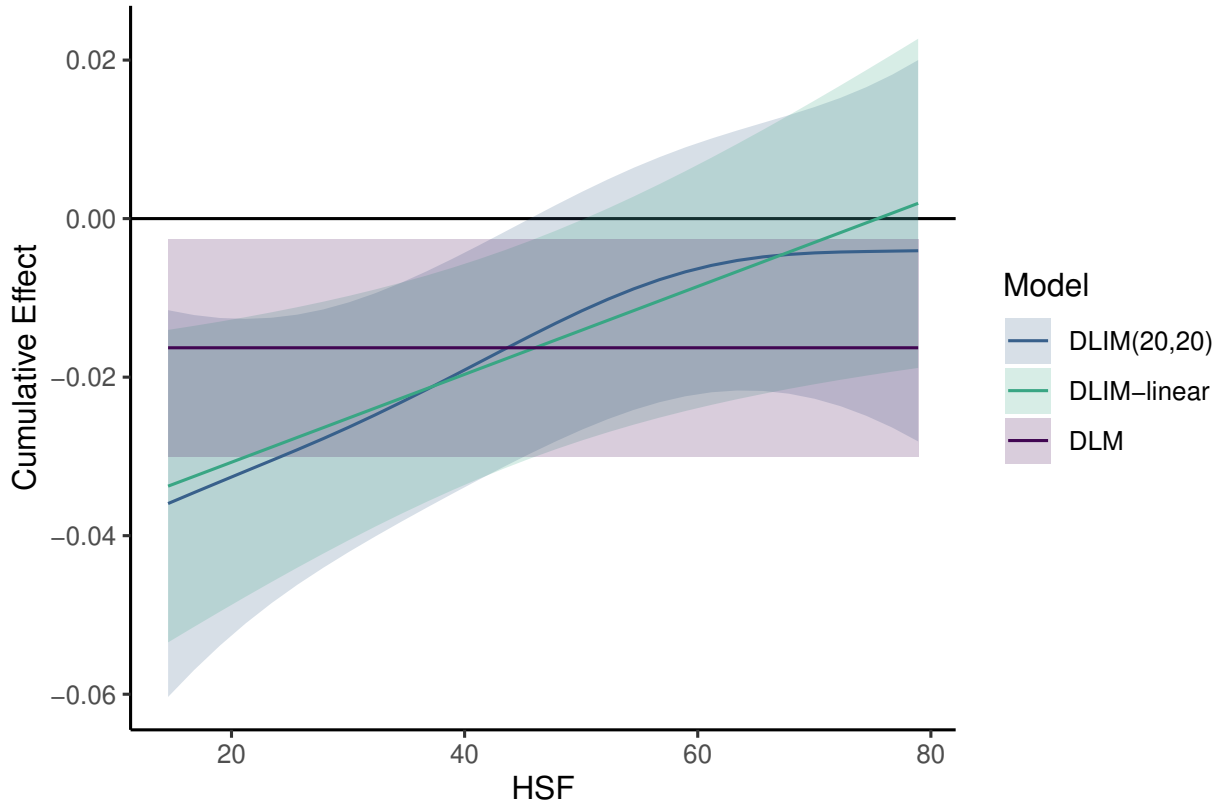


Figure 2.3: Estimates of cumulative effects of ambient $PM_{2.5}$ exposure during gestation on BWGAZ from a standard DLM, DLIM-linear, and DLIM(20,20) across the middle 98% of Colorado EnviroScreen health and social factors (HSF) scores. We include confidence bands of 95% around all estimates.

Our results demonstrate that the effect, cumulative across gestation, of ambient $PM_{2.5}$ on BWGAZ may be modified by neighborhood-level vulnerability of the population in an area.

The top panel in Figure 2.4 shows the estimated exposure-time-response functions for the 25th, 50th and 75th percentiles of the health and social factors scores: 32, 46, and 59, respectively. The estimated exposure-time-response functions are the same across health and social factors scores for the standard DLM but vary for the DLIM-linear. For a health and social factors score of 32, we identified a window of susceptibility during the the first and third trimesters. Specifically, we identified a small window from weeks 7 to 11 and another window from weeks 27 to 37. The peak effect was a decrease of 0.0007 in BWGAZ during the first window and a decrease of 0.0012 in BWGAZ during the second window. The windows we identified are similar in timing to those in some previous work; however, results on the specific timing of the windows are highly variable

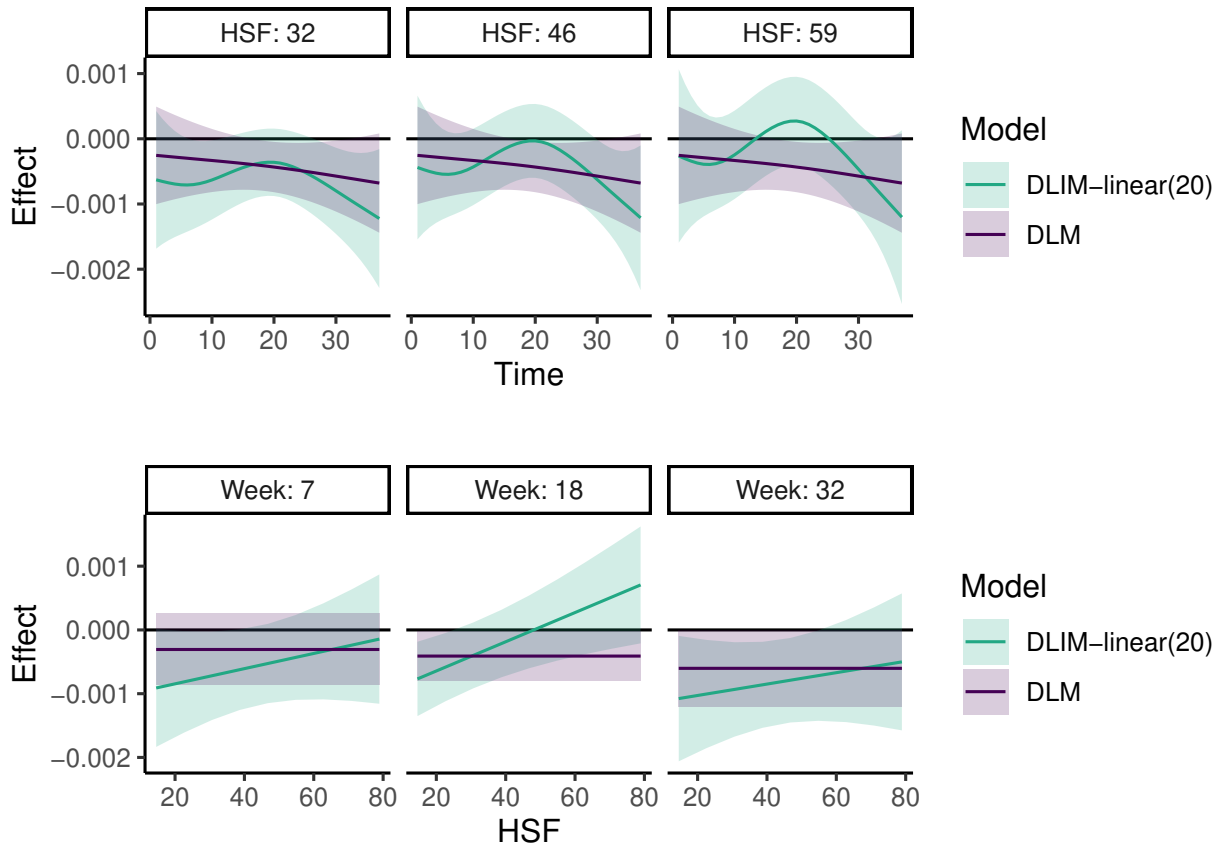


Figure 2.4: Pointwise effect estimates of ambient $PM_{2.5}$ on BWGAZ from a standard DLM with 20 exposure-time degrees of freedom and a DLIM-linear with Colorado EnviroScreen health and social factors (HSF) score as a modifier. The top panel shows estimated exposure-time-response functions for each quartile of the HSF scores: scores 32, 46, and 59. The bottom panel shows estimated point-wise effects for a week around the middle of each trimester, excluding the first and last 1% of HSF scores. We include confidence bands of 95% around all estimates.

across studies (Bosetti et al., 2010; Stieb et al., 2012; Mork et al., 2024; Niu et al., 2022; Johnson et al., 2022; Gong and Zhan, 2022).

The bottom panel of Figure 2.4 shows the estimated linear effect of exposure as a function of health and social factors scores at approximately the middle of each trimester: weeks 7, 18, and 32. See Appendix Figures A.2-A.4 for plots at all time points. In Figure 2.4, the effect for each week is constant across health and social factors scores for the standard DLM but varies for the DLIM-linear. We estimated a negative effect for some health and social factors scores with the DLIM-linear. During the first trimester, we estimated a negative effect for people who live in areas with health and social factors between 18 and 37. We estimated a negative effect during the middle of the second trimester for people who live in areas with health and social factors scores less than 25, and we estimated a negative effect during the middle of the third trimester for people who live in areas with health and social factors scores less than 55. Our findings are also reflected in the top panel. For example, we estimated a negative effect for people in an area with a health and social factors score of 46 during week 32, which is shown in both the second plot of the top panel and the third plot in the bottom panel. We did not estimate an effect of ambient $PM_{2.5}$ on BWGAZ with the standard DLM for any values of the health and social factors score at any time during gestation. Our results provide evidence that the effect of ambient $PM_{2.5}$ on BWGAZ at specific weeks during gestation may be associated with the susceptibility and vulnerability of the population in an individual's area of residence.

Overall, we found a negative association between ambient $PM_{2.5}$ exposure and BWGAZ consistent with a large body of literature (Bosetti et al., 2010; Jacobs et al., 2017; Lakshmanan et al., 2015; Stieb et al., 2012; Šrám et al., 2005). Further, we estimated this effect among individuals living in census tracts with low health and social factors scores and identified a small window of susceptibility for individuals with low health and social factors scores. The modification pattern is contrary to the expected results, as a low health and social factors score is intended to indicate decreased vulnerability to environmental exposures. Our analysis is the first observational study to use the Colorado EnviroScreen data. Hence, there is not a body of work for comparison. We

propose a few explanations. First, similar inverse relationships have been found in some studies. For example, Millett (2016) reviewed literature supporting immigrant children having a health advantage over native-born children of the same socio-economic status, and Chu et al. (2022) found that foreign-born black and Latina women's children can have higher birth weight than native-born women's children in the United States. Second, area-level modification may be present, but our modifier may not be weighted correctly or may be misspecified. Colorado EnviroScreen uses geometric means to weight indicators into component scores, which our area-level modifier may not correctly capture. Finally, our conditioning on full-term births and gestational age (Neophytou et al., 2021), along with live-birth bias (Raz et al., 2018) could explain this inverse relationship.

2.6 Discussion

In this paper, we proposed a distributed lag interaction model (DLIM). The proposed model extends the DLM framework to allow for individualized exposure-time-response functions based on a continuous modifying variable. The DLIM allows for estimation of personalized exposure-time-response functions by smoothing over both the exposure-time dimension and modifier dimension and for formal testing for interaction. Our proposed model is versatile, can be fit as GLMs and in mixed models, and is computationally efficient. Fitting and interpreting our DLIM is user-friendly through the development of our `dlim` package in `R`.

Our proposed DLIM framework fills an existing gap in the literature. Many researchers aim to estimate the effect of air pollution on birth outcomes modified by a single continuous covariate, e.g., maternal stress and neighborhood vulnerability. Without the DLIM methodology, researchers must partition the continuous modifier into categorical factors and fit separate DLMs or BDLIMs for each subgroup (Bose et al., 2017, 2018; Lee et al., 2018; Niu et al., 2022). While this approach provides an approximation of the effect over each subgroup of a continuous modifier, researchers cannot obtain a unique estimate for any value of the modifier without fitting a DLIM. Such approaches also include the potential for bias due to misspecification resulting from categorizing a continuous variable.

Through simulation, we considered the possible methods of modification explored by Wilson et al. (2017a) and verified that a DLIM can accurately estimate the exposure-time-response function and cumulative effect whether or not there is modification. A standard DLM is more efficient when no modification is present, but only by a small amount. When modification is present, a standard DLM is not able to accurately capture the modification structure, which leads to bias in both the exposure-time-response function and cumulative effect estimates. While the gain in efficiency is small when fitting a standard DLM if there is no modification, the risk of fitting a standard DLM when there is modification is large.

We applied a DLIM to analyze the effect of ambient $PM_{2.5}$ on BWGAZ in Colorado modified by Colorado EnviroScreen's health and social factors score, a quantification of neighborhood-level vulnerability and susceptibility. We estimated a negative cumulative effect for lower health and social factors scores. This indicates that the susceptibility and vulnerability of pregnant people's areas of residence may modify the total effect of ambient $PM_{2.5}$ on BWGAZ. Our results provide evidence that pregnant people who live in areas that are less susceptible and vulnerable to social, economic, environmental, and climate adversities are at risk of having pregnancies with lower BWGAZ when exposed to higher levels of ambient air pollution.

Our DLIM framework is novel as it allows for smooth modification of an exposure-time-response function by a continuous factor. Existing methodology only allows for modification to the exposure-time-response function based on a categorical modifier. While our paper focuses on maternal exposure to air pollution and birth outcomes, the proposed methodology is broadly applicable in environmental health and other fields. For example, DLMs are often used in time series studies of air pollution on hospitalization or mortality rates (Schwartz, 2000), as well as in other fields (Schliep et al., 2021; Yang et al., 2022). An added level of novelty in this paper is using an index from Colorado EnviroScreen as a neighborhood-level modifying variable. Using this index as a modifier allows for modification based on 15 factors through a parsimonious index while fitting a single model that is computationally efficient and readily interpretable.

Chapter 3

Distributed Lag Interaction Model with Index Modification

3.1 Introduction

A wide body of literature supports an association between exposure to air pollution during pregnancy and birth and children's health outcomes (Šrám et al., 2005; Bosetti et al., 2010; Warren et al., 2012; Stieb et al., 2012; Chang et al., 2015; Hsu et al., 2015; Lakshmanan et al., 2015; Chiu et al., 2016; Jacobs et al., 2017). Studies often use a distributed lag model (DLM) to estimate the association between air pollution exposure and health outcomes (Chiu et al., 2016; Hsu et al., 2015). The DLM framework regresses the response onto repeated measures of exposure to form an exposure-time-response function of linear associations between the time-varying exposure and response (Almon, 1965). The exposure-time-response function can be used to identify windows of susceptibility, periods when there is an association. However, associations between environmental exposures and health outcomes can vary across populations (Arcaya et al., 2016; Martenies et al., 2022b; Schnake-Mahl et al., 2020). Therefore, effect modification in a DLM framework is of epidemiological interest.

Numerous extensions to DLMs exist to address methodological gaps by incorporating different forms of flexibility, including effect heterogeneity for pre-specified factors. Wilson et al. (2017a) developed a Bayesian distributed lag interaction model that includes interaction between the repeated measures of exposure and a single categorical variable. Warren et al. (2012) developed a spatially-varying Gaussian process model for windows of susceptibility that allows the exposure-time-response function to vary by areal units (e.g., census tracts). Dematteis et al. (2024b) developed a distributed lag interaction model (DLIM) to allow for heterogeneous effect modification across a population using a single continuous variable. A DLIM extends the exposure-time-

response function to vary continuously across the population based on a single continuous modifier using a bi-dimensional function space or cross-basis (Dematteis et al., 2024b; Gasparrini et al., 2010). Mork et al. (2024) developed a heterogeneous DLM that uses additive regression trees to estimate DLMs for subgroups of the population based on multiple continuous and categorical variables using a non-parametric regression tree method. In this paper, we consider extending the DLIM to multiple modifiers as a parsimonious and interpretable model.

Many studies have considered single or multiple index models to study the association between environmental mixture exposures and health outcomes. This is in part because linear index models are interpretable and can reduce the effects of multicolinearity that arise from correlated predictors. The environmental health literature has widely adopted weighted quantile sum regression and quantile g-computation to evaluate the impact of environmental mixtures on health outcomes by creating a mixture index (Czarnota et al., 2015; Daniel et al., 2020; Keil et al., 2020; Wheeler et al., 2021; Wu et al., 2021) while more recent innovations have developed multiple index models (McGee et al., 2023b). Epidemiological researchers have also adopted indices for effect modification. For example, EnviroScreen metrics aggregate multiple continuous variables measured at the census-tract level into a single continuous index aiming to quantify environmental stressors in areas of residence. Niu et al. (2022) and Martenies et al. (2022b) used California EnviroScreen scores as modifiers of exposure-time-response functions. In other contexts, multiple measures of physical or behavioral health may be combined to form a risk index, e.g., a mental health index (Kopta and Lowry, 2002) or a quality of life index (Kai et al., 1991), to succinctly summarize multiple related factors. However, these analyses that consider modification through an index all rely on fixed, pre-specified index weights and do not allow for data-driven weight estimation or modifier selection within the index.

In settings with repeated measures of exposure, creating a single index from many similar modifiers would allow for modification of the exposure-time-response function based on the cumulative impact of all modifiers. However, using an existing index with fixed weights can be problematic for two reasons. First, the fixed weights may not represent the true underlying structure of difference in

susceptibility, leading to an incorrectly estimated exposure-time-response function. Second, this approach does not allow for identifying which of the candidate modifiers are driving the differences in the exposure-time-response function. Clearly, there is a need to simultaneously estimate modifier index weights and the exposure-time-response function in a distributed lag framework.

We propose a distributed lag interaction model with index modification (DLIM-IM). The proposed approach extends the single modifier model of Demateis et al. (2024b) to multiple modifiers through a data-derived modifier index. The methodology in this paper uses a Bayesian hierarchical framework to simultaneously estimate the exposure-time-response function and the index weightings for multiple continuous modifiers to create a single continuous index that modifies the exposure-time-response function. We applied our proposed method in two data analyses. The first uses a Colorado administrative birth cohort data set along with Colorado EnviroScreen to estimate the association between exposure to ambient $PM_{2.5}$ (particulate matter less than 2.5 microns in diameter) during gestation and birth weight for gestational age z-score (BWGAZ) modified by measures of neighborhood-level vulnerability. The second uses the Mexican Programming Research in Obesity, Growth, Environment, and Social Stressors (PROGRESS) cohort data set to estimate the association between exposure to ambient $PM_{2.5}$ surrounding pregnancy and birthing-parent cardio-metabolic endpoints modified by time after parturition. Our R package `dlimIM` is publicly available on GitHub.

3.2 Methods

3.2.1 Model

For individual $i = 1, \ldots, n$, let y_i be the response, $\mathbf{x}_i = [x_{i1}, \ldots, x_{iT}]'$ be a vector of exposures measured at time points $t = 1, \ldots, T$, and $\mathbf{m}_i = [m_{i1}, \ldots, m_{iL}]'$ be a vector of L candidate modifying variables. We assume the modifying variables to be scaled to the unit interval and to operate in the same direction. Also, let $\mathbf{z}_i$ be a vector of p covariates, including 1 for the intercept and the candidate modifier values so their main effects are included in the model.

We make three key assumptions about the exposure-response relationship. First, we assume that the effect of exposure on the response at each time point is linear. Second, we assume that the exposure-response relationship varies smoothly across exposure-time. These first two assumptions are widely assumed across the DLM literature (Almon, 1965). Third, we assume that the relationship between repeated measures of the exposure and the response is modified by a linear combination of candidate modifiers.

Let $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_L]'$ be the vector of index weights for the L candidate modifiers such that $\rho_l \geq 0, l = 1, \ldots, L$, and $\sum_{l=1}^{L} \rho_l = 1$. While other constraints have been used for single-index models, e.g., constraining one element of the index to be positive (Yu and Ruppert, 2002) or L_2 constraint (McGee et al., 2023b), we chose to impose directional homogeneity because this is common for single-index models used in environmental epidemiology (Carrico et al., 2015). The weighted modifier index for individual i is $m_i^* = \mathbf{m}_i'\boldsymbol{\rho}$. Our proposed model is

$$g[E(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{m}_i)] = \sum_{t=1}^{T} x_{it}\beta_t(m_i^*) + \mathbf{z}_i'\boldsymbol{\gamma}, \tag{3.1}$$

where g is the link function, $\beta_t(m_i^*)$ is the linear effect of exposure at time t on the response for an individual with modifier index value m_i^*, and $\boldsymbol{\gamma}$ is a vector of regression coefficients for the covariates and intercept. When m^* is replaced by a scalar modifying factor, the model in (3.1) is the distributed lag interaction model presented by Dematteis et al. (2024b). Here, we extend the model to include a data-driven index of multiple candidate modifiers.

From (3.1), we focus inference on three parameters. First, the exposure-time-response function for weighted modifier m^* is $[\beta_1(m^*), \ldots, \beta_T(m^*)]'$. The exposure-time-response function quantifies how the effect of exposure on response varies both as a function of exposure-time and the modifier index, which allows us to identify windows of susceptibility and ranges of the modifier index associated with susceptibility. We further discuss the parameterization of the exposure-time-response function in Section 3.2.2. Second, the cumulative effect for an individual with modifier value m^* is $\text{CE}(m^*) = \sum_{t=1}^{T} \beta_t(m^*)$. For a model with an identity link function, the cumulative

effect is the expected aggregated change in response per unit of exposure at all time points across the exposure history. Third, the modifier index weights $\boldsymbol{\rho}$ provide insight into which modifiers contribute to altering the exposure-time-response function and their level of contribution. We propose methodology for a model with a Gaussian family and identity link function and a model with a binomial family and logit link function.

3.2.2 Parameterizing the Exposure-Time-Response Functions

We parameterize the exposure-time-response function as a smooth function of both exposure time and the weighted modifier index. We parameterize the exposure-time-response vector $[\beta_1(m_i^*), \ldots, \beta_T(m_i^*)]$ using a cross-basis, which is a bi-dimensional function space that describes both the modified exposure-time-response associations and repeated measures across time. To create the cross-basis, we use splines to smooth in the modifier and exposure-time dimensions similar to the approach in Demateis et al. (2024b). We use natural splines that include an intercept to expand the weighted modifier index with ν_{mod} degrees of freedom. Scaling $[m_{1l}, \ldots, m_{nl}]$ for each $l = 1, \ldots, L$ to lie within the unit interval allows us to set boundary knots at 0 and 1 to ensure all possible modifier index values are within boundary knots. We denote the modifier index basis evaluated at value m_i^* as $\mathbf{b}(m_i^*) = [b_1(m_i^*), \ldots, b_{\nu_{mod}}(m_i^*)]'$ for $i = 1, \ldots, n$. We also use natural splines that include an intercept and boundary knots at the first and last exposure-time points to expand in the exposure-time dimension with ν_{time} degrees of freedom. Including an intercept in the modifier basis allows us to include a main effect of the exposure history. Without the intercept, a reference point for the modifier index is necessary for interpreting changes in the exposure-time-response function per modifier index value (Gasparrini et al., 2010). The basis evaluated at exposure-time t for $t = 1, \ldots, T$ is $\mathbf{c}(t) = [c_1(t), \ldots, c_{\nu_{time}}(t)]'$. Combining these two basis expansions, the cross-basis for individual i is $w_{jk}(m_i^*, \mathbf{x}_i) = \sum_{t=1}^{T} x_{it} b_k(m_i^*) c_j(t)$ for $k = 1, \ldots, \nu_{mod}$ and $j = 1, \ldots, \nu_{time}$.

Using this parameterization, the sum $\sum_{t=1}^{T} x_{it}\beta_t(m_i^*)$ in (3.1) becomes $\sum_{k=1}^{\nu_{mod}} \sum_{j=1}^{\nu_{time}} \theta_{jk} w_{jk}(m_i^*, \mathbf{x}_i)$. We denote the $n \times \nu_{time}\nu_{mod}$ design matrix for the cross-basis as $\mathbf{W}(\mathbf{X}, \mathbf{M}, \boldsymbol{\rho})$ and the

$\nu_{time}\nu_{mod}$ cross-basis regression coefficients as $\boldsymbol{\theta}$. Let $\mathbf{U}(\boldsymbol{\rho}) = [\mathbf{W}(\mathbf{X}, \mathbf{M}, \boldsymbol{\rho}), \mathbf{Z}]$ and $\boldsymbol{\Psi} = [\boldsymbol{\theta}', \boldsymbol{\gamma}']'$. Under this parameterization, (3.1) for all individuals is

$$g[E\{\mathbf{y}|\mathbf{W}(\mathbf{X}, \mathbf{M}, \boldsymbol{\rho}), \mathbf{Z}, \mathbf{M}\}] = \mathbf{U}(\boldsymbol{\rho})\boldsymbol{\Psi}. \tag{3.2}$$

3.2.3 Prior Specification

We consider prior specifications for modeling the modifier index weights $\boldsymbol{\rho}$ with and without selection. For the model without modifier selection, we specify a Dirichlet prior on the weights, $\boldsymbol{\rho} \sim \text{Dir}(\mathbf{q})$, where $\mathbf{q} = [q_1, \ldots, q_L]'$ is a vector of hyper-parameters controlling the manner and strength of influence that the prior has on the posterior (McGee et al., 2023a). We parameterize the Dirichlet distribution using its relation to independent, normalized Gamma random variables. We let $a_l \sim \text{Gamma}(q_l, 1), l = 1, \ldots, L$ and define $\rho_l = a_l / \sum_{l=1}^{L} a_l$. This ensures $\rho_l \geq 0$ for $l = 1, \ldots, L$, $\sum_{l=1}^{L} \rho_l = 1$, and $\boldsymbol{\rho} \sim \text{Dir}(\mathbf{q})$.

For a high-dimensional modifier space, we may be interested in a more parsimonious model via variable selection on the candidate modifiers. We again set $\rho_l = a_l / \sum_{l=1}^{L} a_l$, but now impose a spike-and-slab prior on the weights to perform selection on the modifiers included in the modification index (Koslovsky, 2023; McGee et al., 2023a). The spike-and-slab prior

$$a_l|\eta_l \sim \eta_l \times \text{Gamma}(q_l, 1) + (1 - \eta_l)\delta_0 \tag{3.3}$$

is a mixture distribution of a Dirac-delta function δ_0 (spike) and a Gamma distribution (slab). The indicator for inclusion of weight l in the prior is $\eta_l|\nu_l \sim$Bernoulli(ν_l), and ν_l is the prior inclusion probability. We fix $\nu_l = 0.5$. Adjusting ν_l controls sparsity, which may be helpful in low signal settings.

The Dirichlet prior for the weights can be made more or less informative through changes in q_l. A more informative Dirichlet prior could be centered on values based on prior information. For example, McGee et al. (2023a) center a Dirichlet prior for weights for components of an environmental mixture on toxicologically derived values, which could be helpful in low signal

settings. A less informative Dirichlet prior puts equal weights on all modifiers and should be used if no prior information is available.

We use a flat prior for the intercept, and we use independent normal priors with mean zero and variance τ^2 for the cross-basis regression coefficients $\boldsymbol{\theta}$. For the covariate regression coefficients excluding the intercept, we use independent normal priors with mean zero and variance ξ^2. We assume an inverse-Gamma prior on the variance σ^2.

3.2.4 Parameter Estimation

We use Markov chain Monte Carlo (MCMC) to sample from the posterior distribution. The model has closed form full conditional distributions for regression coefficients $\boldsymbol{\Psi}$ and error variance but not for the modifier index weights. Therefore, we use a Metropolis-Hastings within Gibbs sampler to sample from the posterior distribution. We sample regression coefficients $\boldsymbol{\Psi}$ and error variance σ^2 from a Gibbs sampler, and we sample the modifier index weights using a Metropolis-Hastings algorithm. The algorithm for our MCMC sampler is outlined in Appendix B Algorithm 1. For a logistic model, we use Pólya-Gamma augmentation in a Gibbs sampler as in Polson et al. (2013) for the regression coefficients and associated latent variables. See Appendix Section B.1 for more details.

For a model without modifier selection, we use the proposal distribution $a_l^*|a_l^{(s-1)} \sim \text{FoldedN}\left(a_l^{(s-1)}, \zeta_l^2\right)$ to propose an update a_l^* for the l^{th} weight based on the current weight $a_l^{(s-1)}$ at iteration $s-1$, where ζ_l is an adaptively tuned hyperparameter. We accept the proposal a_l^* with probability $\min(r, 1)$, where

$$r = \frac{(a_l^*)^{q_l - 1}\exp(-a_l^*)}{(a_l^{(s-1)})^{q_l-1}\exp(-a_l^{(s-1)})} \frac{p(\mathbf{y}|\mathbf{a}^*, \boldsymbol{\Psi}^{(s)}, \cdot)}{p(\mathbf{y}|\mathbf{a}^{(s-1)}, \boldsymbol{\Psi}^{(s)}, \cdot)}. \tag{3.4}$$

Here, $\mathbf{a}^{(s-1)}$ is the current vector of un-normalized weights, $\mathbf{a}^*$ is the same as $\mathbf{a}^{(s-1)}$ with the l^{th} weight replaced by a_l^*, and $\boldsymbol{\Psi}^{(s)}$ is the vector of regression coefficients at iteration s. In (3.4), p is the normal density for a Gaussian model or the binomial density for a binomial model, where $\cdot$ represents other model parameters that are family-dependent (e.g., error variance for a Gaussian model).

When performing modifier selection, we first propose a_l^* as an update for un-normalized weight l from

$$a_l^* | a_l^{(s-1)} \sim \left\{ \begin{array}{ll} \text{Gamma}(q_l, 1), & \text{if } a_l^{(s-1)} = 0 \\ \delta_0, & \text{if } a_l^{(s-1)} \neq 0 \end{array} \right\}$$

and accept that proposed update with probability $\min(r_{\text{select}}, 1)$, where

$$r_{\text{select}} = \left\{ \begin{array}{ll} \frac{1-\nu_l}{\nu_l} \frac{p(\mathbf{y}|\mathbf{a}^*, \boldsymbol{\Psi}^{(s)}, \cdot)}{p(\mathbf{y}|\mathbf{a}^{(s-1)}, \boldsymbol{\Psi}^{(s)}, \cdot)}, & \text{if } a_l^* = 0, a_l^{(s-1)} \neq 0 \\ \frac{\nu_l}{1-\nu_l} \frac{p(\mathbf{y}|\mathbf{a}^*, \boldsymbol{\Psi}^{(s)}, \cdot)}{p(\mathbf{y}|\mathbf{a}^{(s-1)}, \boldsymbol{\Psi}^{(s)}, \cdot)}, & \text{if } a_l^* \neq 0, a_l^{(s-1)} = 0 \end{array} \right\}. \tag{3.5}$$

If the proposal is $a^* = 0$ and is rejected, we then propose a new non-zero value from FoldedN $\left(a_l^{(s-1)}, \zeta_l^2\right)$ and accept or reject that value using the acceptance ratio in (3.4).

To obtain marginal posterior samples for the exposure-time-response associations and cumulative effects, we use linear transformations of the marginal posterior samples for $\boldsymbol{\theta}$ at each iteration s. To obtain posterior samples of the exposure-time-response associations for modifier index m^*, we transform the posterior samples of $\boldsymbol{\theta}$ at each iteration s using the transformation $[\mathbf{b}(m^*) \otimes \mathbf{C}]\boldsymbol{\theta}^{(s)}$ to obtain $[\beta_1(m^*)^{(s)}, \ldots, \beta_T(m^*)^{(s)}]'$. Here, $\mathbf{C}$ is the $T \times \nu_{time}$ exposure-time basis. To obtain posterior samples of the cumulative effect for modifier index m^*, we transform the posterior samples of $\boldsymbol{\theta}$ at each iteration s using the transformation $\mathbf{w}_*(m^*)\boldsymbol{\theta}^{(s)}$ to obtain $\text{CE}(m^*)^{(s)}$. Here, $\mathbf{w}_*(m^*)$ is the vectorization, by column, of $[\mathbf{1}' \otimes \mathbf{b}(m^*)]\mathbf{C}$. We use the posterior means and intervals of these transformed parameters to perform inference.

3.3 Simulation Study

We evaluated the proposed DLIM-IM's ability to estimate cumulative effects and exposure-time-response functions, as well as identify correct modifier index weights. We provide results for our proposed DLIM-IM with and without modifier selection using 5 degrees of freedom in both the modifier and exposure-time dimensions. For prior distributions, we use $q_l = 1$ for all l for the weights prior, $\tau^2 = 100$ for the cross-basis regression coefficients' priors, $\xi^2 = 110$ for the covariate regression coefficients' priors, and we use 1 and 0.001 for the shape and scale parameters

of the error variance prior. We compare to results from a penalized DLIM with 20 basis functions in each dimension as described in Demateis et al. (2024b), fixing the weights equally in each scenario, which we refer to as the fixed-index model. Code for implementing this simulation is provided as a companion file to this manuscript and is available on GitHub.

3.3.1 Simulation Design and Data Generation

We simulated data using $n = 1000$ vectors of real $PM_{2.5}$ concentration measurements over 37 weeks. We generated modifier values from multivariate normal distributions with zero means, variances of 1, and covariances ranging from 0.4 to 0.7. To do this, we generated a covariance matrix with 1s on the diagonal and sampled from a uniform distribution on the off-diagonal elements, forcing the matrix to be both symmetric and positive definite. For each simulated data set, we scaled each of the L modifiers for individuals $i = 1, \ldots, n$ to range from 0 to 1.

We considered three modifier index weighting scenarios. Scenario 1 uses equal weights $\rho_1 = \rho_2 = \rho_3 = 1/3$ as a baseline for comparison to a fixed-index. Scenario 2 has 3 different weights (0.5, 0.4, 0.1). Scenario 3 has 50 weights, 3 of which are 1/3 and the rest are zero. See Appendix Table B.1 for more weight scenarios.

We generated 3 covariates from multivariate normal distributions with the mean of the first week of exposure as the mean and the following covariance structure: $\text{cov}(\mathbf{z}_1, \mathbf{z}_2) = 0.5$, $\text{cov}(\mathbf{z}_1, \mathbf{z}_3) = 0.6$, $\text{cov}(\mathbf{z}_2, \mathbf{z}_3) = 0.7$, $\text{Var}(\mathbf{z}_1) = \text{Var}(\mathbf{z}_2) = \text{Var}(\mathbf{z}_3) = 1$. We generated regression coefficients for the modifiers from a uniform distribution on $(-1, 1)$ and regression coefficients for the covariates from independent standard normal distributions. The exposure-time-response function for the data generating mechanism was $\beta_t(m_i^*) = m_i^* \times f(t, 37[1 + \exp\{-20(m_i^* - 0.5)\}]^{-1})$, where $f(t, c) := 2.5\phi[(t - c)/5]$ and $\phi(\cdot)$ is the normal probability density function. We simulated response values using a normal distribution with variance σ^2. We considered three signal-to-noise settings: low (0.1), medium (0.5), and high (1). The signal is the standard deviation of the simulated exposure-time-response function, and noise is σ. We simulated 200 data sets for each com-

bination of weight scenarios and signal-to-noise settings. See Appendix Figure B.1 for the scaling of computation time as a function of sample size.

3.3.2 Performance Evaluation

To evaluate modifier index weight estimation, we provide box plots for weight estimates across 200 simulated data sets for our proposed models. In Appendix Figure B.2, we provide posterior inclusion probabilities (PIPs) for the modifier index weights from our proposed modifier selection models. Using the estimated modifier index weights, we constructed the weighted modifier index for each simulated data set and computed the root mean squared error (RMSE) as $[n^{-1}\sum_{i=1}^{n}\{m_i^* - \hat{m}_i^*\}^2]^{1/2}$ and average absolute bias as $n^{-1}\sum_{i=1}^{n}|m_i^* - \hat{m}_i^*|$, where $\hat{m}_i^* = \mathbf{m}_i'\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\rho}}$ is the posterior mean of $\boldsymbol{\rho}$. We averaged both RMSE and average absolute bias over the 200 simulated data sets.

We also evaluated average RMSE, coverage, and credible or confidence interval (CI) width for effect estimates. For cumulative effect estimates for each data set, we calculated RMSE as $[n^{-1}\sum_{i=1}^{n}\{\text{CE}(m_i) - \hat{\text{CE}}(m_i)\}^2]^{1/2}$, and we calcualted coverage as $n^{-1}\sum_{i=1}^{n}\mathbb{I}[Q_{\alpha/2}\{\text{CE}(m_i)|\mathbf{y}\} \leq \text{CE}(m_i) \leq Q_{1-\alpha/2}\{\text{CE}(m_i)|\mathbf{y}\}]$ for credible intervals and as $n^{-1}\sum_{i=1}^{n}\mathbb{I}[|\text{CE}(m_i) - \hat{\text{CE}}(m_i)| \leq \Phi^{-1}(1-\alpha/2)\hat{\text{var}}\{\hat{\text{CE}}(m_i)\}]$ for confidence intervals. The function $\mathbb{I}(A)$ is an indicator of event A's occurrence, and $Q_a(.|\mathbf{y})$ is the a^{th}-quantile of the sampled posterior distribution for a given parameter. For point-wise effect estimates for each data set, we calculated RMSE as $[n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\beta_t(m_i) - \hat{\beta}_t(m_i)\}^2]^{1/2}$, and we calculated coverage as $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{I}[Q_{\alpha/2}\{\beta_t(m_i)|\mathbf{y}\} \leq \beta_t(m_i) \leq Q_{1-\alpha/2}\{\beta_t(m_i)|\mathbf{y}\}]$ for credible intervals and as $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{I}[|\beta_t(m_i) - \hat{\beta}_t(m_i)| \leq \Phi^{-1}(1-\alpha/2)\hat{\text{var}}\{\hat{\beta}_t(m_i)\}]$ for confidence intervals. We also compared average credible interval width and average confidence interval width for the proposed and fixed-index models, respectively. For cumulative effect estimates, we calculated average credible interval width as $n^{-1}\sum_{i=1}^{n}[Q_{1-\alpha/2}\{\text{CE}(m_i)|\mathbf{y}\} - Q_{\alpha/2}\{\text{CE}(m_i)|\mathbf{y}\}]$ and average confidence interval width as $n^{-1}\sum_{i=1}^{n}2\Phi^{-1}(1-\alpha/2)[\hat{\text{var}}\{\hat{\text{CE}}(m_i)\}]^{1/2}$ for each data set. For point-wise effect estimates, we calculated average credible interval width as $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}[Q_{1-\alpha/2}\{\beta_t(m_i)|\mathbf{y}\} - Q_{\alpha/2}\{\beta_t(m_i)|\mathbf{y}\}]$

and average confidence interval width as $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} 2\Phi^{-1}(1-\alpha/2)[\hat{\text{var}}\{\hat{\beta}_t(m_i)\}]^{1/2}$. We averaged RMSE, coverage, and average CI width across the 200 simulated data sets.

3.3.3 Simulation Results

Figure 3.1 shows summaries of the modifier index weight estimates. Table 3.1 includes a summary of the cumulative and point-wise effect estimates for the scenarios in our simulation study. In the Appendix Figures B.3-B.5, we provide figures of the estimated exposure-time-response curve for individual data sets. Our proposed DLIM-IM produces more accurate modifier index weight estimates than assuming equal weights, and the proposed DLIM-IM with modifier selection is able to identify factors that modify the distributed lag function. The DLIM-IMs performed better than the fixed-index model in terms of cumulative effect performance metrics and performed better in terms of some point-wise performance metrics as well.

The proposed DLIM-IM produces more accurate weight estimates than assuming equal weights as shown in Figure 3.1, which translates to more accurate estimates of the weighted modifier index. In scenarios 2 and 3 where the fixed model weights are mis-specified, the RMSE and absolute bias for the estimated weighted modifier index are better for the DLIM-IMs compared to the fixed-index model. In the high signal-to-noise setting for scenario 2, the RMSE was 0.004 for the DLIM-IMs and 0.027 for the fixed index model, and the absolute bias was 0.003 for the DLIM-IMs and 0.022 for the fixed-index model. In scenario 3, the DLIM-IM with selection produced more accurate weighted modifier index estimates than the DLIM-IM without selection. In the high signal-to-noise setting, the RMSE was 0.0129 for the selection model and 0.0255 for the model without selection. The absolute bias was 0.0103 for the selection model and 0.0204 for the model without selection. The model with selection produces PIPs above 0.5 for non-zero weights and below 0.5 for zero weights. Our results demonstrate that the proposed DLIM-IM can estimate modifier index weights and select modifiers in various settings.

The DLIM-IMs performed better than the fixed-index model in terms of cumulative effect estimate metrics. In all three scenarios, cumulative coverage was similar among the three models, but

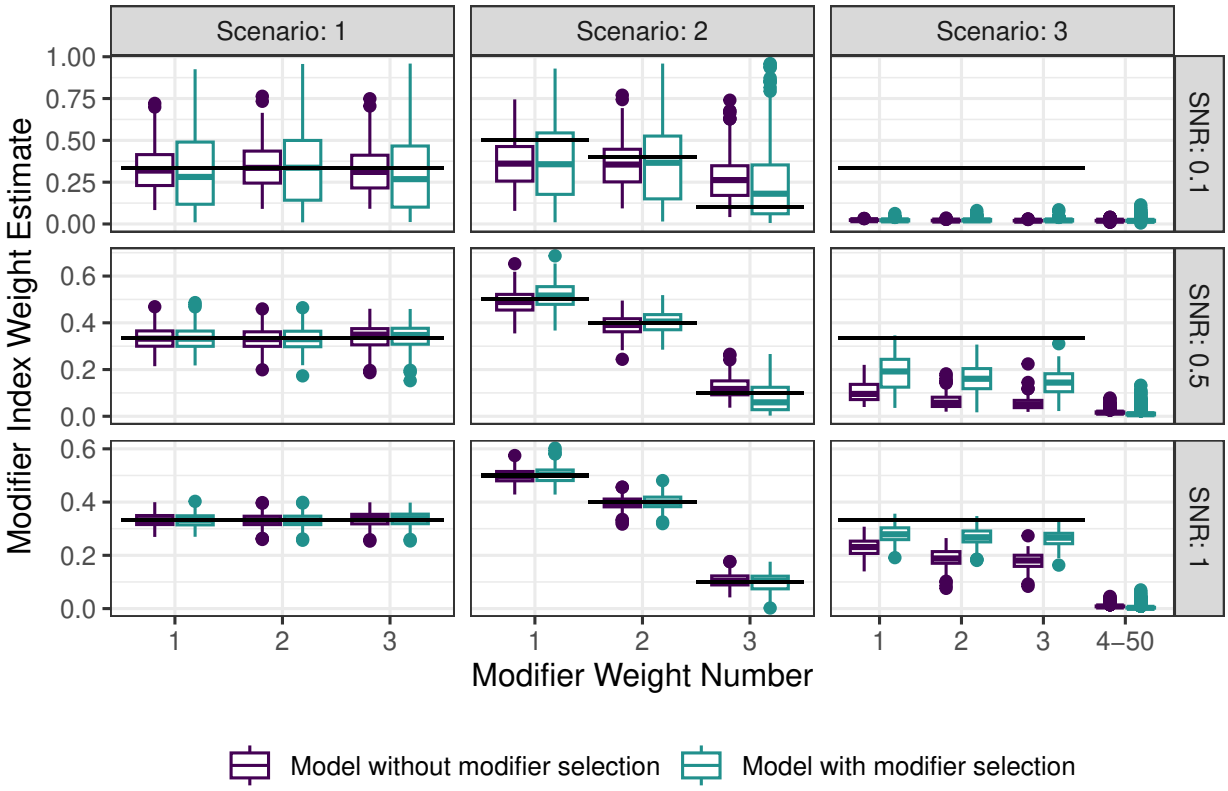


Figure 3.1: Modifier index weight estimates from DLIM-IMs with selection and DLIM-IMs without selection for all 200 simulated data sets across all weight scenarios and signal-to-noise (SNR) settings. The number of basis functions for the exposure-time basis is 5 for both the DLIM-IMs. The black line represents the true weight value. For scenario 3 (sparse weight scenario), weights 1-3 are non-zero, and weight numbers 4-50 is averaged over all 47 zero weights.

the DLIM-IMs had smaller CI widths on average. For example in the high signal-to-noise setting for scenario 1, the cumulative coverage was 0.95 for the three models; however, the average CI width was 0.74 for the fixed-index model and 0.79 for both DLIM-IMs with and without selection. In all three scenarios, cumulative RMSE was similar or lower for the DLIM-IMs than the fixed-index model. In scenario 3, the DLIM-IM with selection had better performance metrics than the DLIM-IM without selection. For example in the medium signal-to-noise setting for scenario 3, the average CI width for the DLIM-IM with and without selection was 1.52 and 1.63, respectively, and the RMSE was 0.32 and 0.36, respectively. Our results show that the DLIM-IM can estimate cumulative effects more efficiently and with coverage closer to the nominal level than the fixed-index

model, and the DLIM-IM with selection estimates cumulative effects better than the DLIM-IM without selection when weights are sparse.

The DLIM-IMs performed better than the fixed-index model in terms of point-wise effect estimate metrics. In all three scenarios and signal-to-noise settings, except the high signal-to-noise setting for scenario 1, the DLIM-IM had point-wise coverage similar or closer to the nominal level than the fixed-index model. For example in the medium signal-to-noise setting for scenario 2, the point-wise coverage was 0.89 for the DLIM-IMs with and without selection and 0.70 for the fixed-index model. However, point-wise RMSE was either similar or lower for the fixed-index model. We attribute this to the penalization of splines in the fixed-index model. In scenario 3, point-wise coverage, RMSE, and average CI width were similar for the DLIM-IMs. Overall, the DLIM-IMs had point-wise effect coverage closer to the nominal level than the fixed-index model and RMSE and average CI width similar to or better than the fixed-index model, while performance metrics for the two DLIM-IMs were similar.

3.4 Data Analyses

We used our proposed model to analyze two data sets. In the first application, we demonstrate the DLIM-IM's ability to perform selection on multiple continuous modifiers instead of including all modifiers with fixed weights. We use our proposed model with selection to extend the analysis of Colorado administrative birth data performed in Demateis et al. (2024b), which used a pre-defined index of 15 modifiers. Here, we extend the analysis to identify which factors contribute to modification of the association between exposure to ambient $PM_{2.5}$ during pregnancy and BWGAZ. In the second application, we demonstrate the DLIM-IM's ability to combine multiple measures of birthing-parent stress during gestation and lifetime instead of performing separate analyses on each modifier or arbitrarily choosing weights. We use the DLIM-IM without selection to create a weighted lifetime stress index using 4 birthing-parent stress variables. We use this lifetime stress index as a modifier of the association between exposure to ambient $PM_{2.5}$ surrounding pregnancy and cholesterol levels from the PROGRESS cohort data set. See Appendix Figures B.6

and B.7 for information on modifier correlation for each analysis. For prior distributions, we use $\tau^2 = 100$ for the cross-basis regression coefficients priors, $\xi^2 = 110$ for the covariate regression coefficients priors, and we use 1 and 0.001 for the shape and scale parameters of the error variance prior.

3.4.1 Air Pollution, EnviroScreen and Birth Weight in Colorado

We used a Colorado administrative birth cohort data set containing all live, full-term, singleton births with dates of conception from 2007-2018 in census tracts in the Colorado Front Range with elevation lower than 6,000 feet (Demateis et al., 2024b). We matched the data with weekly average ambient $PM_{2.5}$ concentration during the gestational period obtained from the down-scaled models published by the US Environmental Protection Agency (https://www.epa.gov/hesc/rsig-related-downloadable-data-files) at birthing-parent residences. We regressed BWGAZ, constructed using Fenton growth charts (Fenton, 2003), onto repeated measures of ambient $PM_{2.5}$ measured for the first 37 weeks of gestation. The final data set we used included 393,204 births in 786 census tracts across 13 counties.

For candidate modifiers, we used the continuous indicators included in the Colorado EnviroScreen health and social factors (HSF) score, excluding the indicator for low birth weight (Colorado Department of Public Health and Environment (CDPHE), 2022). The HSF score from Colorado EnviroScreen is constructed using the following tract-level percentile variables: percent of people of color, disability rate, percent of households linguistically isolated, percent of households that are housing-cost burdened, percent of people with less than a high school education, percent of people with low income, diabetes prevalence, asthma hospitalization rate, rate of poor mental health, percent of people over 64 years of age, percent of people under 5 years of age, cancer prevalence, rate of heart disease in adults, and estimated life expectancy. Indicators are directionally adjusted following standard practice within Colorado Enviroscreen. For example, the percentile values for average life expectancy were reversed so that a higher percentile corresponds with shorter life expectancy. We divide all candidate modifiers by 100 to the unit interval. We

also include all candidate modifiers as covariates in the model by including the linear main effect of each component of EnviroScreen. In addition, we control for each pregnant person's age, body mass index (BMI), height, and weight at conception. We also control for each pregnant person's race and ethnicity, education, income, and marital status before pregnancy and self-reported prenatal care habits and smoking habits during pregnancy.

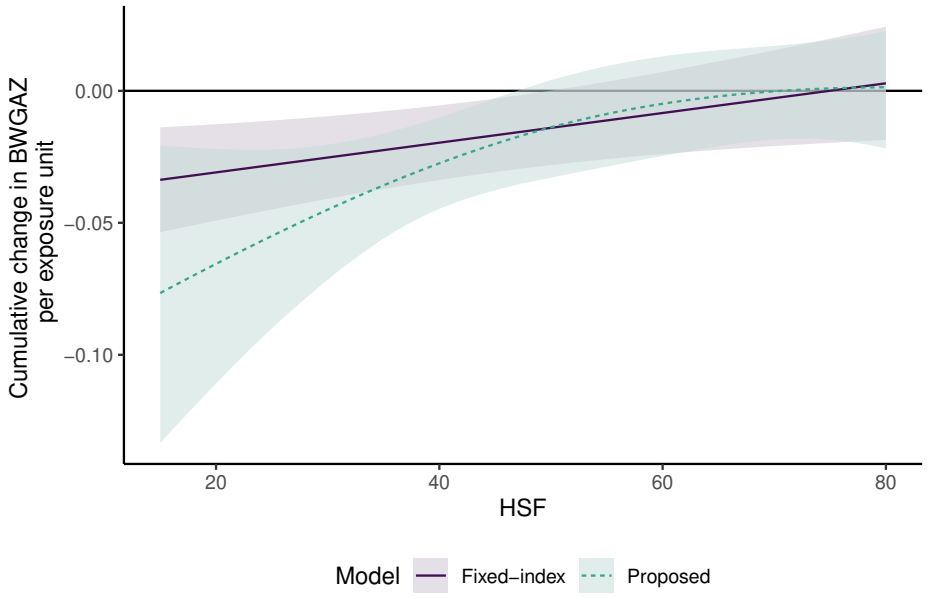
The analysis in Demateis et al. (2024b) fit a DLIM using the HSF score constructed by geometrically averaging the candidate modifiers *a priori* as a single continuous modifier and presented estimated exposure-time-response functions for the HSF score quartiles: 32, 46, and 59. We fit our proposed model and estimated the weights of the candidate modifiers constructing the HSF score to 1) compare estimated exposure-time-response functions using fixed and estimated weights for the candidate modifiers and 2) identify which of the 14 candidate modifiers contributed to modification of the exposure-time-response function.

We fit our proposed model with modifier selection using 3 degrees of freedom in both the exposure-time and modifier dimensions (Appendix Figures B.8 and B.9 and Table B.2), and we used $q_1 = \cdots = q_{14} = 1$ as the hyperparameter for the Dirichlet prior. We ran our MCMC sampler with 5 chains with a 2.4GHz AMD Milan processor with 3.8GB of RAM for 50,000 iterations and discarded the first 30,000 for burn-in based on visual inspection of the parameters' trace plots. The runtime ranged from 3-5 days for each chain. We compare to the penalized DLIM with 20 B-spline basis functions in the exposure-time dimension and a linear modifier basis fit in Demateis et al. (2024b). We refer to this as the fixed-index model. We compare exposure-time-response functions estimated at Colorado EnviroScreen HSF scores to corresponding estimated modifier index value. For example, we compare the exposure-time-response curve estimated for Colorado EnviroScreen HSF score 32 to the estimated curve at estimated modifier index score of 0.32. We refer to both the Colorado EnviroScreen HSF score 32 and estimated modifier index value 0.32 as HSF 32 for simplicity.

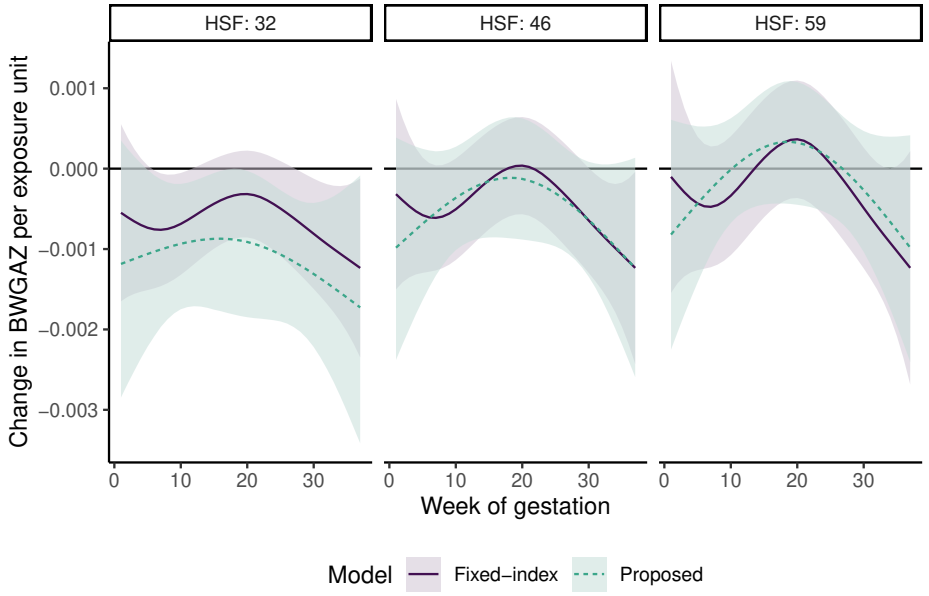
We estimated similar cumulative effects and different windows of susceptibility using both models. Figure 3.2 shows the cumulative effect estimates and estimated exposure-time-response

curves from both the fixed-index and the proposed models. We estimated a negative cumulative effect of exposure for individuals with HSF scores below 46 using the proposed model and below 50 using the fixed-index model. Using the proposed model with data-derived modifier index, we estimated a slightly larger negative effect, although the intervals are wider and overlapping. We also estimated similar windows of susceptibility using both models. For HSF 32 with the fixed-index model, we estimated a window during the first and third trimester; whereas with the proposed model, we estimated a window from weeks 5 to 18 and another window from weeks 20-37. For HSF 46, we estimated a window during the third trimester with the fixed-index model, and we did not estimate any windows with the proposed model due to larger intervals in the more complex index modification model. For HSF 59, we did not estimate any windows with either model.

Using the DLIM-IM, we were able to additionally identify which of the candidate modifiers contribute to modification. Table 3.2 shows that 5 of the 14 candidate modifiers were selected into the model with PIPs greater than 0.5. Percent of people of color ($\hat{\rho} = 0.220$, PIP $= 0.807$), percent of people with less than high school education ($\hat{\rho} = 0.194$, PIP $= 0.749$), and asthma hospitalization rate ($\hat{\rho} = 0.135$, PIP $= 0.743$) were the top three factors contributing to modification. This indicates that these variables contributed most, per unit change, to the modification of the effect of exposure to $PM_{2.5}$ during pregnancy on BWGAZ. Because the modifier index weights are different from one another, this means that each of the modifiers has a different contribution to modification of the exposure-time-response function. While modifiers that were selected into the model (PIP > 0.5) generally had estimated weights larger than the average of equal weights ($1/14$), and modifiers not selected generally had weight estimates that were smaller, we still use PIPs as a basis of selection because it is plausible that this is not always the case. See Appendix Figures B.10-B.12 for a comparison of the estimated exposure-time response curve from the model presented here and models with only one of the modifiers. Our results demonstrate how the proposed model can be used in an analysis with multiple modifiers to identify important factors that contribute to modification of the exposure-time-response function.



(a) Estimated cumulative effects across health and social factor modifier values.



(b) Estimated exposure-time-response curves across 37 weeks of gestation for the quartiles of the Colorado EnviroScreen health and social factors scores.

Figure 3.2: Estimated cumulative effect and exposure-time-response curves from the fixed-index penalized DLIM with linear modification and proposed DLIM-IM with selection applied to the Colorado birth cohort. The fixed-index DLIM uses 20 basis functions in the modifier and exposure-time dimensions, and the DLIM-IM with selection uses 5 basis functions in each dimension.

3.4.2 Air Pollution, Maternal Stress, and Lipids in PROGRESS

The PROGRESS cohort data set is a longitudinal birth cohort study with a total of 948 birthing-parent-infant pairs in Mexico City, Mexico that were recruited between 2007 and 2011 (Braun

et al., 2014; Burris et al., 2014). We obtained daily $PM_{2.5}$ concentrations based on a validated model as described in Gutiérrez-Avila et al. (2022) at each of the participant's residential addresses and aggregated the concentrations monthly. As responses, we used high-density lipoprotein (HDL) measured in plasma by enzymatic photometric assays and low-density lipoprotein (LDL) derived from the Friedewald equation (Friedewald et al., 1972). Using a DLIM-IM, we regressed birthing-parent HDL and LDL, measured 4 years after parturition onto monthly average $PM_{2.5}$ starting at 2 months before through 22 months after each birthing-parent's last menstrual period (LMP). We controlled for each birthing-parent's age, BMI, assessed socio-economic status (lower, medium, and higher), and marital status at the second trimester. We also controlled for their smoking habits (passive or active) and parity at baseline, alcohol intake and cardiometabolic medications 2 years after parturition, and the season (November to February, March and April, and May to October) during LMP. After removing observations with incomplete data, the number of observations is 272.

We used 4 different measures of birthing-parent stress to create a data-derived lifetime stress index as a modifier of the exposure-time-response function. The stress modifiers we used are the Edinburgh Postnatal Depression scale (EPDS) as a measure of depression during gestation (Levis et al., 2020), the Crisis in Family Systems (CRISYS) to quantify the impact of negative life events (Sherlock et al., 2023), the Perceived Stress Scale as a measure of self-reported stress during gestation (Cohen et al., 1983), and the State-Trait Anxiety Inventory (STAI) as a measure of anxiety during gestation (Spielberger et al., 1983). Higher values for each modifier correspond to higher stress levels. Each of the four modifiers was measured once for each individual during gestation. To create a common scale across stress variables we rescale each modifier to range from 0 to 1.

We fit a DLIM-IM without selection using 3 degrees of freedom in both the exposure-time and modifier dimensions, and we used $q_1 = \cdots = q_4 = 5$ as the hyperparameter for the Dirichlet prior. We ran our MCMC sampler with 5 chains for 30,000 iterations and discarded the first 10,000 for burn-in based on visual inspection of the trace plots, which had a runtime of about 5 minutes for each chain. We present results in the main text for only the HDL response as we did not estimate

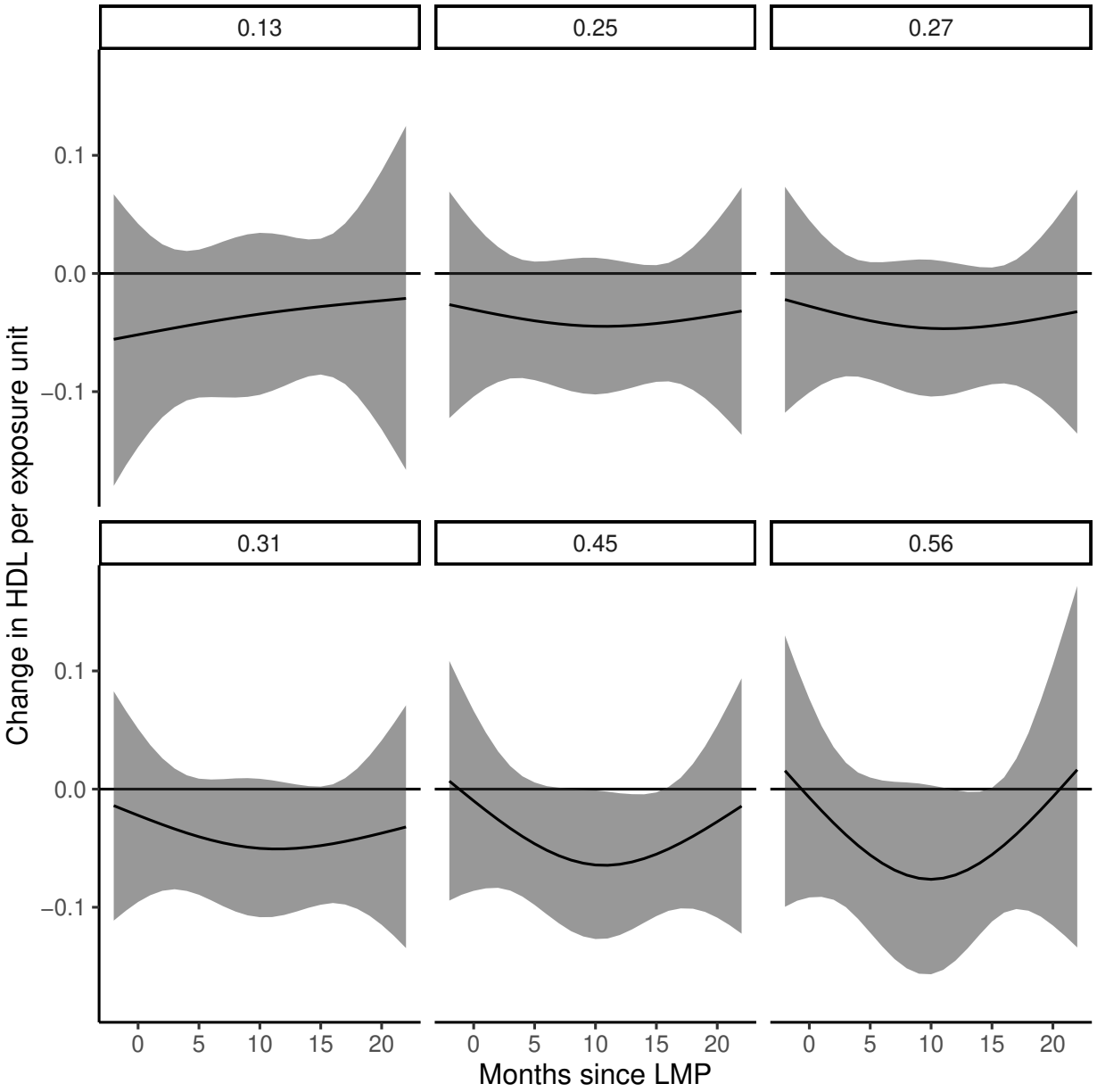


Figure 3.3: Estimated exposure-time-response curves for the 10^{th} ($\hat{m}^* = 0.13$), 25^{th} ($\hat{m}^* = 0.25$), 30^{th} ($\hat{m}^* = 0.27$), 40^{th} ($\hat{m}^* = 0.31$), 75^{th} ($\hat{m}^* = 0.45$), and 90^{th} ($\hat{m}^* = 0.56$) percentiles of the estimated lifetime stress index from a DLIM-IM without selection 2 months before through 22 months after last menstrual period (LMP) at weighted stress modifier values in the PROGRESS cohort. The DLIM-IM used 3 degrees of freedom in the modifier and exposure-time dimensions.

any effect of exposure to $PM_{2.5}$ on LDL but present results in Appendix Figures B.13 and B.14. The weighted lifetime stress index we estimate combines various indicators of birthing-parent stress, with higher values indicating more stress.

We estimated a small critical window of susceptibility and quantified the importance of each stress modifier in contributing to modification of the exposure-time-response function. Figure 3.3 shows the exposure-time-response function at the middle 90% of the estimated weighted lifetime stress modifier index values. We estimated small critical windows of susceptibility for individuals who experienced moderate to high levels of stress. Specifically, we estimated a window between months 12 to 16 at the 75^{th} percentile of the estimated stress index ($\hat{m}^* = 0.45$) and another window between months 14 to 15 at the 90^{th} percentile of the estimated stress index ($\hat{m}^* = 0.56$). We did not estimate a cumulative effect for any weighted lifetime stress modifier index value (Appendix Figure B.15). We estimated the weight for CRYSIS as 0.283 (SD = 0.096), EPDS as 0.249 (SD = 0.091), STAI as 0.236 (SD = 0.086), and perceived stress as 0.232 (SD = 0.088). Our results suggest that the effect of ambient $PM_{2.5}$ exposure on HDL 48 months after parturition may be modified by lifetime stress, with each stress modifier contributing similarly to the modification. See Appendix Figure B.16 for a comparison of the estimated exposure-time response curve from the model presented here and models with only one of the modifiers. Further, our analysis demonstrates how the proposed methodology can be used to combine multiple measures of a single construct into a cohesive analysis rather than being forced to arbitrarily combine measures *a priori*, select a single measure, or run repeated analysis with different candidate modifiers.

3.5 Discussion

In this paper, we propose a Bayesian distributed lag interaction model for multiple continuous modifiers, an extension of the framework for a single continuous modifier in Demateis et al. (2024b). Our proposed framework simultaneously estimates the exposure-time-response functions varying by a modifier index and the weights of that modifier index. By including a spike-and-slab prior on the weights, we are able to estimate the weights and perform selection on the modifiers, allowing for both identification of the importance of each modifier and selection among candidate modifiers. Our model is versatile and can be applied for both a Gaussian and binomial model.

Code for our simulations and analyses can be found as a companion file to this manuscript, and our package is available on GitHub.

Our proposed model fills a methodological gap in the distributed lag framework by allowing for modification of the exposure-time-response function by multiple continuous modifiers. This framework is an extension of the DLIM presented in Demateis et al. (2024b), which allows for modification by only a single continuous variable. Our proposed framework also builds on index-based models, e.g., weighted quantile sum regression and quantile g-computation, commonly used in environmental health studies (Czarnota et al., 2015; Keil et al., 2020). By creating a weighted modifier index, our proposed model has advantages over models that also allow for modification by multiple modifiers, such as the heterogeneous DLM, because our posterior samples are easily interpretable. Continuous interaction between multiple modifiers and the exposure is a valuable addition to the distributed lag framework.

Through simulation, we considered various modifier index weight scenarios and demonstrated that our proposed DLIM-IM is able to produce more accurate modifier index weight estimates than assuming equal weights. Furthermore, our model is able to estimate the exposure-time-response functions more accurately when compared to the model in Demateis et al. (2024b) with mis-specified fixed modifier index weights. This demonstrates that there is a benefit to using a model that estimates modifier index weights instead of arbitrarily fixing weights *a priori*. Our DLIM-IM with selection produced modifier index weight estimates that were more accurate than those of the DLIM-IM without selection when modifier index weights were truly sparse.

We applied our proposed model to two data sets to highlight various applications. The first application extended the analysis in Demateis et al. (2024b) by estimating weights of the various Colorado EnviroScreen HSF components to create a data-derived modifier index quantifying neighborhood-level vulnerability and susceptibility. While we estimated similar windows of susceptibility using our DLIM-IM, we were able to further identify which of the HSF components contributed to modification of the exposure-time-response function using PIPs obtained through our modifier selection approach. In the second application, we created a weighted lifetime stress

modifier index from four variables all measuring different aspects of stress, including depression during gestation, negative life events, perceived stress during gestation, and anxiety during gestation. Instead of performing multiple analyses with each stress modifier or arbitrarily weighting them together, we were able to estimate a weighted lifetime stress index and identify windows of susceptibility based on the weighted lifetime stress index.

There are several limitations to our proposed model. First, the weights are constrained to be constant across time. While statistically simple, a model with time-varying weights would be highly parametric and require massive datasets to estimate in practice. The proposed model does allow for modifying variables to change over time. For example, the model could account for a birthing parent who moved during pregnancy. Second, identifiability issues are possible when including highly correlated modifiers, and weight estimation accuracy decreases as the number of potential modifiers increases, as shown in the simulation study. If modifier weight estimation is of interest with a large or highly correlated set of modifiers, we recommend using a pre-selection approach based on existing literature. Third, real data sets often have low signal, and the simulation study demonstrated that modifier index weight estimation may be less accurate. Incorporating prior information on weight can improve estimates.

Our proposed methodology is novel because it allows for modification of the exposure-time-response function by a data-derived index constructed from multiple modifiers. This work extends the distributed lag framework, builds upon index models in environmental health, and fills a critical gap in epidemiology literature.

Table 3.1: Performance metrics for the weighted modifier index, cumulative effects, and point-wise effect for DLIM-IMs without selection (DLIM-IM), the DLIM-IM with selection (DLIM-IM sel.), and the fixed-index model (fixed-index) averaged over all 200 simulated data sets with non-linear modification for all three weight scenarios and three signal-to-noise (SNR) levels. The low SNR is 0.1, medium SNR is 0.5, and the high SNR is 1. The number of basis functions for the exposure-time basis was 5 for both DLIM-IMs and 20 for the fixed-index model.

		Index	Cumulative Effect			Point-wise Effect		
SNR	Model	RMSE	RMSE	Coverage	Width	RMSE	Coverage	Width
Scenario 1: Equal Weights								
low	DLIM-IM	0.0221	1.10	0.95	4.96	0.23	0.96	0.88
	DLIM-IM sel.	0.0381	1.09	0.95	4.96	0.23	0.96	0.88
	Fixed-index	0	1.77	0.94	7.29	0.12	0.92	0.42
med.	DLIM-IM	0.0081	0.28	0.96	1.29	0.06	0.89	0.19
	DLIM-IM sel.	0.0082	0.28	0.96	1.29	0.06	0.89	0.19
	Fixed-index	0	0.36	0.95	1.51	0.04	0.72	0.10
high	DLIM-IM	0.0041	0.18	0.95	0.74	0.04	0.77	0.10
	DLIM-IM sel.	0.0041	0.18	0.95	0.74	0.04	0.77	0.10
	Fixed-index	0	0.19	0.95	0.79	0.03	0.87	0.09
Scenario 2: Different Weights								
low	DLIM-IM	0.0295	1.11	0.95	4.98	0.23	0.96	0.89
	DLIM-IM sel.	0.0398	1.10	0.95	4.98	0.23	0.96	0.89
	Fixed-index	0.0273	1.78	0.94	7.32	0.12	0.92	0.42
med.	DLIM-IM	0.0078	0.28	0.96	1.27	0.06	0.89	0.19
	DLIM-IM sel.	0.0094	0.27	0.96	1.26	0.06	0.89	0.18
	Fixed-index	0.0273	0.37	0.95	1.53	0.04	0.70	0.10
high	DLIM-IM	0.0042	0.18	0.95	0.73	0.04	0.77	0.10
	DLIM-IM sel.	0.0049	0.18	0.95	0.73	0.04	0.77	0.10
	Fixed-index	0.0273	0.20	0.95	0.83	0.03	0.77	0.08
Scenario 3: Sparse Weights								
low	DLIM-IM	0.0603	1.18	0.98	6.19	0.26	0.96	0.97
	DLIM-IM sel.	0.0600	1.12	0.98	6.04	0.26	0.96	0.99
	Fixed-index	0.0606	1.78	0.96	7.76	0.13	0.91	0.45
med.	DLIM-IM	0.0485	0.36	0.96	1.63	0.07	0.89	0.22
	DLIM-IM sel.	0.0333	0.32	0.97	1.52	0.07	0.90	0.21
	Fixed-index	0.0606	0.39	0.94	1.66	0.05	0.67	0.10
high	DLIM-IM	0.0255	0.20	0.94	0.84	0.04	0.80	0.11
	DLIM-IM sel.	0.0129	0.18	0.94	0.79	0.04	0.79	0.10
	Fixed-index	0.0606	0.23	0.93	0.95	0.04	0.59	0.07

Table 3.2: Modifier index weight estimates, standard deviations, and posterior inclusion probabilities (PIP) for each of the 14 included indicators of the EnviroScreen HSF score obtained in the Colorado birth cohort analysis using our proposed model with 3 degrees of freedom for both the modifier and exposure-time bases. Results are summarized across 5 chains. Indicators marked with (*) have PIPs over 0.5, suggesting that these indicators contribute to modification of the exposure-time-response relationship.

Indicator	Posterior mean	Posterior SD	PIP
Percent of people of color (*)	0.220	0.179	0.807
Percent of people with less than high school ed. (*)	0.194	0.192	0.749
Asthma hospitalization rate (*)	0.135	0.119	0.743
Percent of people over age 64 (*)	0.097	0.131	0.591
Percent of households linguistically isolated (*)	0.069	0.100	0.531
Percent of households housing-cost burdened	0.060	0.095	0.495
Disability rate	0.055	0.097	0.460
Diabetes prevalence	0.046	0.071	0.472
Rate of poor mental health	0.029	0.059	0.374
Cancer prevalence	0.034	0.059	0.410
Percent of people under age 5	0.020	0.047	0.308
Rate of heart disease in adults	0.015	0.047	0.236
Life expectancy	0.013	0.034	0.254
Percent of people with low income	0.013	0.034	0.261

Chapter 4

Subgroup Analyses and Effect Modification with Bayesian Kernel Machine Regression

4.1 Introduction

There is widespread interest in understanding how exposure to environmental mixtures impacts human health (Dominici et al., 2010; Greenbaum and Shaikh, 2010; Vedal and Kaufman, 2011). Challenges inherent to the epidemiological analysis of exposure to environmental mixtures include high correlation between components in a mixture, high-dimensionality of the mixture, and complex exposure-response relationships, including interactions between pollutants and non-linear associations with the response (Gibson et al., 2019b). To address these and other challenges, numerous methods have been proposed for the analysis of environmental mixtures data (Gibson et al., 2019b; Joubert et al., 2022; Zhu et al., 2024; Taylor et al., 2016). Several reviews have characterized the methods that can address common research goals faced in mixture analyses, including identifying toxic pollutants, estimating the effects of single pollutants and/or the whole mixture on a health response, and identifying interactions between pollutants (Gibson et al., 2019a; Taylor et al., 2016). However, a notable gap in the literature for applying mixture methods is guidance on assessing effect modification.

Research supports that environmental effects on health outcomes are not homogeneous across the population (Bauer et al., 2020; Bell et al., 2013; Chakravartty et al., 2014; Choi et al., 2017; Goodman et al., 2023; Son et al., 2019). When estimating the association between a single environmental pollutant and a health outcome, it is common to include an interaction term in a parametric regression model to allow for different effects among subgroups. However, modeling effect modification is more complicated with sophisticated statistical methods for mixture analyses. This is especially true for modification by a categorical factor, such as child sex or smoking status (Dai

et al., 2023; Farzan et al., 2025; Letellier et al., 2022; Li et al., 2025; Mu et al., 2024; Wang et al., 2022, 2024), as most statistical approaches for mixture analyses are designed for multiple continuous predictors and may not be appropriate for binary or categorical factors. Therefore, it is common for researchers to fit models stratified by a categorical variable that may alter the association between the exposures and response. However, the sratification approach has limitations. First, the sample size for each model is reduced in a stratified analysis. This can reduce the precision of estimates and decrease statistical power (Wang et al., 2024). Second, it is challenging to quantify differences in exposure effects between groups and to assess uncertainty in these differences. Thus, there is a need for mixture analysis methods that specifically allow for modification by categorical factors, including estimating subgroup-specific analyses and between-group differences with quantification of uncertainty.

Bayesian kernel machine regression (BKMR) is a widely used method for mixture analyses in epidemiological studies (Bobb et al., 2015, 2018). BKMR is a semi-parametric regression method that can estimate a flexible, multi-dimensional exposure-response surface that represents the joint association between the exposures and the outcome. To allow for complex associations while also reducing overfitting, BKMR uses a kernel function in a hierarchical Bayesian framework to estimate the exposure-response surface. While BKMR can address many common research questions for environmental mixtures, there is currently no guidance in the literature on how to allow for modification by a categorical factor. Several studies have applied BKMR separately to different subgroups in stratified analyses (Dai et al., 2023; Farzan et al., 2025; Mu et al., 2024; Wang et al., 2022, 2024). Alternatively, it is theoretically feasible to include a categorical factor in the kernel function to assess modification (Bobb et al., 2018). However, this approach has not, to our knowledge, been rigorously evaluated.

We consider BKMR analyses of environmental mixtures with modification by a categorical variable. We consider four approaches, including two that rely on the standard BKMR model and software and two new variants introduced in this paper. We evaluate the operating characteristics of each method through a simulation study to give practical advice on how to conduct a mixture

analysis using BKMR with effect modification by a categorical factor. We also illustrate the approaches through an analysis of a metal mixture and infant neurodevelopment from a Bangladesh cohort.

4.2 Methods

For individual $i = 1, \ldots, n$, let y_i be a continuous response, $\mathbf{z}_i = [z_{i1}, \ldots, z_{iM}]'$ be a vector of M different continuous exposures in a mixture, $\mathbf{x}_i$ be a vector of covariates, and w_i be a categorical variable with P categories. We denote the levels of the modifier as $w_i \in \{1, \ldots, P\}$. For completeness in notation, we define $\mathbf{w} = [w_1, \ldots, w_n]'$. First, we detail the standard BKMR model and existing methods of effect modification. Then, we present two novel approaches.

4.2.1 Standard BKMR

The standard BKMR model (Bobb et al., 2015) with no effect modification is

$$y_i = h(\mathbf{z}_i) + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \tag{4.1}$$

where $h(\mathbf{z}_i)$ is a nonparametric exposure-response function for individual i that may include non-linearities and interactions, $\boldsymbol{\beta}$ is a vector of regression coefficients for the covariates, and $\varepsilon_1, \ldots, \varepsilon_n$ are residual terms that are assumed to be independent and normally distributed with mean zero and variance σ^2.

BKMR parameterizes the exposure-response function $h(\cdot)$ with a Gaussian process. This allows for a high-dimensional set of exposures and induces smoothness and allows for non-linear associations and interactions. For a vector of outcomes $\mathbf{y} = [y_1, \ldots, y_n]'$, matrix of exposures $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]'$, and matrix of covariates $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]'$, BKMR can be expressed as the

hierarchical Bayesian model

$$\mathbf{y}|\mathbf{h}, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{h} + \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \tag{4.2}$$

$$\mathbf{h}|\mathbf{Z}, \boldsymbol{\rho}, \tau \sim N[\mathbf{0}, \tau\mathbf{K}(\mathbf{Z}, \boldsymbol{\rho})], \tag{4.3}$$

where $\mathbf{h} = [h_1, \ldots, h_n]'$ is the exposure-response vector, $\mathbf{K}(\mathbf{Z}, \boldsymbol{\rho})$ is a covariance matrix defined by a kernel function of the exposures, $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_M]'$ is a vector of parameters that control smoothness of the exposure-response surface separately in each exposure dimension, and τ is a variance component that controls the overall smoothness of the exposure-response surface. BKMR is most commonly applied with a Gaussian kernel function. The (ij)-element of the Gaussian kernel matrix $\mathbf{K}(\mathbf{Z}, \boldsymbol{\rho})$ is

$$K(\mathbf{z}_i, \mathbf{z}_j, \boldsymbol{\rho}) = \exp\left[-\sum_{m=1}^{M} \frac{(z_{im} - z_{jm})^2}{\rho_m}\right]. \tag{4.4}$$

A standard BMKR model can be fit with Markov chain Monte Carlo (MCMC) methods using the `bkmr` R package (Bobb, 2022). After fitting with MCMC, we can estimate $h(z_1, \ldots, z_M)$ for any exposure profile $z_1, \ldots, z_M$.

4.2.2 Stratified BKMR

The stratified approach to effect modification in BKMR fits a separate BKMR model to each level of the modifier w, yielding different exposure-response surfaces for each group. This approach can be fit with the existing `bkmr` R package (Bobb, 2022). Table 4.1 summarizes the key model assumptions of this method and the others discussed herein. This stratified approach to effect modification assumes no sharing of information among groups, which limits inference. Most importantly, the exposure-response surfaces are estimated independently for each group, so the exposure-response surface is smoothed differently for each group, and there is no penalization to impose similar shape or smoothness in the surfaces among the groups. In addition, the effects of covariates ($\boldsymbol{\beta}$) and error variance (σ^2) are different for each group. When fitting a stratified

analysis, the modifier is not included in the covariates because there is no variation in that variable within groups.

Table 4.1: Assumptions for models with modification considered in this manuscript: modifier-in-kernel (mod-in-kernel), first-approach group-separable (GS 1), second-approach group-separable models (GS 2), and stratified. The ✓ symbol indicates that the model does have, and the X symbol indicates that the model does not have, the corresponding assumption.

Assumptions	Mod-in-kernel	GS 1	GS 2	Stratified
Pooled analysis	✓	✓	✓	X
Parameters σ^2 and $\boldsymbol{\beta}$ shared among groups	✓	✓	✓	X
Exposure-specific smoothing parameters ($\boldsymbol{\rho}$) shared among groups	✓	✓	✓	X
Overall exposure-response surface smoothing parameter (τ) shared among groups	✓	✓	X	X
Penalty on exposure-response shape among groups	✓	X	X	X

4.2.3 Modifier-in-Kernel BKMR

The second approach we consider is to include the modifier in the kernel along with the exposures. This approach can also be estimated with the existing BKMR package (Bobb, 2022). When there are more than two levels to the categorical modifier, and it is not assumed to be ordinal, the modifier should first be converted to indicator variables for group membership. Let $\mathbf{d}_i = [d_{i1}, \ldots, d_{i(P-1)}]'$ contain indicator variables for modifier w_i. We append $\mathbf{d}_i$ to $\mathbf{z}_i$ and fit a standard BKMR, which has the effect of changing the (ij)-element of the kernel matrix to

$$K^{\text{mk}}(\mathbf{z}_i, \mathbf{z}_j, \mathbf{d}_i, \mathbf{d}_j, \boldsymbol{\rho}^{\text{mk}}) = \exp\left[-\left\{\sum_{m=1}^{M} \frac{(z_{im} - z_{jm})^2}{\rho_m^{\text{mk}}} + \sum_{p=M+1}^{M+P-1} \frac{(d_{ip} - d_{jp})^2}{\rho_p^{\text{mk}}}\right\}\right]. \tag{4.5}$$

The superscript "mk" is placed on parameters in this model to denote the modifier-in-kernel model. The indicator variables $\mathbf{d}_i$ for each individual are also included in the model as covariates.

Of the methods discussed here, this approach makes the strongest assumptions about sharing information among levels of the modifier (Table 4.1). This approach shares the same smoothing parameters ($\rho_1^{\text{mk}}, \ldots, \rho_{M+P-1}^{\text{mk}}$ and τ) for all subgroups, resulting in approximately the same level of smoothness for the exposure-response surface in each group. A defining characteristic of this model is that the exposure-response surface has an additional dimension for each indicator variable. This imposes a penalty that induces similarity in the exposure-response function between each group and the reference group. Additionally, this method shares the same regression coefficients and error variance between groups.

4.2.4 Group-separable BKMR

The novel approaches for effect modification with BKMR that we present here use a separable kernel function. This allows for a level of information sharing among groups that spans the gap between the stratified and modifier-in-kernel approaches. We present two variants on the approach. Both can be fit with our `bkmrGS` package, available on GitHub, that uses a similar MCMC algorithm as the standard BKMR model. For the first group-separable model, the prior on the exposure-response function is $\mathbf{h}_{\mathbf{w}}^{\text{gs}_1}$,

$$\mathbf{h}_{\mathbf{w}}^{\text{gs}_1} | \mathbf{Z}, \boldsymbol{\rho}^{\text{gs}_1}, \tau^{\text{gs}_1}, \mathbf{w} \sim N[\mathbf{0}, \tau^{\text{gs}_1} \mathbf{K}^{\text{gs}_1}(\mathbf{Z}, \mathbf{w}, \boldsymbol{\rho}^{\text{gs}_1})], \tag{4.6}$$

where the (ij)-element of the kernel matrix $\mathbf{K}^{\text{gs}_1}(\mathbf{Z}, \mathbf{w}, \boldsymbol{\rho}^{\text{gs}_1})$ is

$$K^{\text{gs}_1}(\mathbf{z}_i, \mathbf{z}_j, w_i, w_j, \boldsymbol{\rho}^{\text{gs}_1}) = \begin{cases} \exp\left[-\sum_{m=1}^{M} \frac{(z_{im}-z_{jm})^2}{\rho_m^{\text{gs}1}}\right], & \text{if } w_i = w_j \\ 0, & \text{if } w_i \neq w_j \end{cases} \tag{4.7}$$

The superscript "gs_1" is placed on parameters in this model to denote the first group-separable approach. The kernel matrix defined this way has a block-diagonal structure, meaning that the matrix is constructed with a group of non-zero elements for each modifier level and with blocks of zeros for the correlations between observations with different levels of the modifier.

The key difference between the group-separable approach and the modifier-in-kernel approach is that the modifier-in-kernel approach penalizes the exposure-response functions to be to have a similar shape for all groups; whereas, the group-separable approach does not (Table 4.1). As such, the group-separable approach has greater flexibility to estimate differently shaped exposure-response functions. Like the modifier-in-kernel approach, the group-separable approach shares parameters among groups. The covariate effects ($\boldsymbol{\beta}$), error variance (σ^2), and smoothing parameters ($\boldsymbol{\rho}$ and τ) are the same for all subgroups. This implies that the smoothness of the exposure-response surface will be similar for all groups, but the shape has freedom to vary among groups.

The second group-separable approach also blocks the kernel by modifier group but additionally allows for different levels of smoothness in the exposure-response surface for each level of the modifier. This is achieved by letting the smoothing parameter τ vary among the subgroups. The prior on the exposure-response surface is

$$\mathbf{h}_{\mathbf{w}}^{\text{gs}_2}|\mathbf{Z}, \boldsymbol{\rho}^{\text{gs}_2}, \boldsymbol{\tau}^{\text{gs}_2}, \mathbf{w} \sim N[\mathbf{0}, \mathbf{K}^{\text{gs}_2}(\mathbf{Z}, \mathbf{w}, \boldsymbol{\rho}^{\text{gs}_2}, \boldsymbol{\tau}^{\text{gs}_2})], \tag{4.8}$$

where $\boldsymbol{\tau}^{\text{gs}_2}$ is a vector of length P containing a unique smoothing parameter for each modifier group. The kernel function for the (ij)-element of kernel matrix $\mathbf{K}^{\text{gs}_2}(\mathbf{Z}, \mathbf{w}, \boldsymbol{\rho}^{\text{gs}_2}, \boldsymbol{\tau}^{\text{gs}_2})$ is

$$K^{\text{gs}_2}(\mathbf{z}_i, \mathbf{z}_j, w_i, w_j, \boldsymbol{\tau}^{\text{gs}_2}, \boldsymbol{\rho}^{\text{gs}_2}) = \begin{cases} \tau_{w_i}^{\text{gs}_2} \times \exp\left[-\sum_{m=1}^{M} \frac{(z_{im}-z_{jm})^2}{\rho_m^{\text{gs}_2}}\right], & w_i = w_j \\ 0, & w_i \neq w_j \end{cases} \tag{4.9}$$

where $\tau_{w_i}^{\text{gs}_2}$ is the smoothing parameter for modifier group w_i for individual i. The superscript "gs_2" is placed on parameters in this model to denote the modifier-in-kernel approach. This kernel matrix has the same block-diagonal structure as (4.7), but instead of scaling each group's block by the same variance component as in the first group-separable approach, this second group-separable approach, scales each modifier's block by a group-specific variance component $\tau_{w_i}^{\text{gs}_2}$. The indicator variables $\mathbf{d}_i$ for each individual are included in the model as covariates for both group-separable models.

4.2.5 Interpretation of the Exposure-Response Functions

There are multiple types of exposure-response effects that are of interest in a BKMR model. We focus on two commonly used summaries of the exposure-response surface for BKMR as previously described (Bobb et al., 2018): the total effect of the mixture and single-exposure effects. We define effects for models with and without effect modification. For a model with effect modification, we define these effects as either group-specific or between-group differences in the relevant effect.

For a standard BKMR model without modification, the total effect of the mixture on the response is the difference in mean outcome when all exposures are fixed at a specific quantile (q_1) compared to when all exposures are fixed at a different quantile (q_2). Let z_m^q denote the q^{th} quantile of $[z_{1m}, \ldots, z_{nm}]'$. Formally, the total mixture effect is $\Delta_{tot}(q_1, q_2) = h(z_1^{q_2}, \ldots, z_M^{q_2}) - h(z_1^{q_1}, \ldots, z_M^{q_1})$. For a model with modification, the total mixture effect specific to the p^{th} modifier group ($w = p$) is $\Delta_{tot}(q_1, q_2|p) = h(z_1^{q_2}, \ldots, z_M^{q_2}|w = p) - h(z_1^{q_1}, \ldots, z_M^{q_1}|w = p)$. Additionally, the between-group difference of the total mixture effect between modifier groups p and p' is the change in total mixture effect between modifier groups p and p', which is $\Delta_{tot}(q_1, q_2|p, p') = \Delta_{tot}(q_1, q_2|p') - \Delta_{tot}(q_1, q_2|p)$.

For a standard BKMR model without modification, the single-exposure effect of a given exposure on the response is the difference in mean outcome between when one exposure is fixed at the q_1 quantile versus when it is fixed at the q_2 quantile and all other all other exposures are fixed at q. For example, the single-exposure effect of $\mathbf{z}_1$ is $\Delta_1(q_1, q_2, q) = h(z_1^{q_2}, z_2^q, \ldots, z_M^q) - h(z_1^{q_1}, z_2^q, \ldots, z_M^q)$. For a model with modification, the single-exposure effect of a given exposure on the response in a modifier group $w = p$ is the change in the mean outcome between quantiles q_1 and q_2 for that exposure, while all other exposures are fixed at q and only the exposure-response surface for modifier group $w = p$ is considered. For example, the single-exposure effect of the first exposure in modifier group p is $\Delta_1(q_1, q_2, q|p) = h(z_1^{q_2}, z_2^q, \ldots, z_M^q|w = p) - h(z_1^{q_1}, z_2^q, \ldots, z_M^q|w = p)$. The between-group difference for the single-exposure effect between modifier levels p and p' is $\Delta_1(q_1, q_2, q|p, p') = \Delta_1(q_1, q_2, q|p') - \Delta_1(q_1, q_2, q|p)$.

To estimate between-group differences in effects using the stratified model, we matched posterior samples from each modifier group and performed inference directly on those matched posterior samples.

4.2.6 Bangladesh Reproductive Cohort Data

We obtained data from The Bangladesh Reproductive Cohort Study (Project Jeebon) for 351 children born between 2008 and 2013 in the Pabna Upazilas (administrative regions) of Bangladesh. The data have been previously described and made publicly available in a partially masked dataset (Liu et al., 2022). We consider a mixture of three metals during gestation: lead, manganese, and arsenic. Cumulative gestational metal exposures were measured using umbilical cord venous blood samples at delivery. The dependent variable in our analysis is infant neurodevelopment quantified using the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III) (Del Rosario et al., 2021) measured at 20-40 weeks of infant age. We consider child sex as a single binary modifier in our analyses. The data also contains 20 covariates as described, and adjusted for, in Liu et al. (Liu et al., 2022). The names of these covariates are masked in the publicly available data but were previously reported to include: parent's demographic status (age, education, smoking history), infant's biometric measurements at birth (sex, birth weight, length, head circumference, birth order and gestational age), and quality of early childhood home environment (HOME score (Totsika and Sylva, 2004), maternal depression scale, and maternal IQ).

4.2.7 Analysis Methods

We applied BKMR models described in this paper in a data analysis to demonstrate model performance in practice. From the Bangladesh cohort data set (Liu et al., 2022), we used the BSID-III neurodevelopment score as the response, the three metals as exposures, infant sex as a modifier, and controlled for all other covariates in the data set. We discretized obvious categorical variables and removed two observations that were the only observations in a level of a categorical variable. This resulted in a final analysis sample of 349.

We fit five BKMR models: standard BKMR without modification (referred to as the exposure-only), modifier-in-kernel, group-separable approach 1, group-separable approach 2, and stratified. We used BKMR default priors presented in the Appendix C “Additional Methods" section for all models and tuned hyper-parameters for convergence. We ran the models for 50,000 iterations and discarded 20,000 iterations for warm-up, and thinned by 10. We chose warm-up periods based on visual assessments of trace plots for parameters, and we thinned to reduce computational time. We compare estimates of the exposure-response surface along with group-specific effects and between-group differences in effect estimates among the five models.

4.2.8 Simulation Methods

We evaluated the operating characteristics of the existing stratified and modifier-in-kernel methods and novel group-separable methods. We compared performance of these models along with a standard BKMR model without modification (exposure-only).

Simulation design and data generation

To ensure a realistic correlation structure in the simulated data, we used the real exposures and modifier from the Bangladesh cohort in our simulation study. We included all $M = 3$ metal exposures. Of the three exposures, we simulated an association between only two of the exposures and the response. Figure 4.1 shows the three exposure-response surfaces that we used to construct modification scenarios. We assume that the three surfaces h_1, h_2, and h_3 are constant across the third exposure. To ensure some correlation between the modifier and other variables, we used child sex as the modifier in the two-group scenario and one of the HOME score components as the modifier in the three-group scenario. The scenarios are as follows.

1. **Two-group scenario 1**: The outcomes in the boy group ($n = 180$) are simulated from exposure-response surface h_1, and the outcomes in the girl group ($n = 171$) are simulated from exposure-response surface h_2, which has less of an effect.

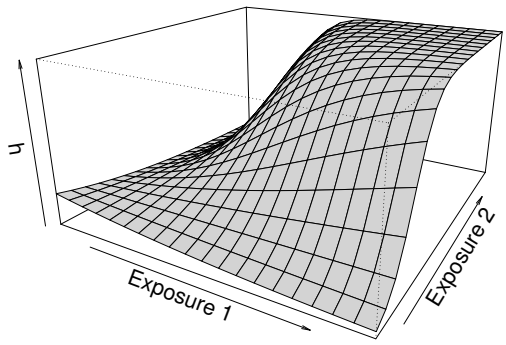
2. **Two-group scenario 2**: The outcomes in the boy group are simulated from exposure-response surface h_1. The observations in the girl group are simulated from exposure-response surface h_3, which has no effect.

3. **Three-group scenario**: The outcomes in group one ($n = 45$) are simulated from exposure-response surface h_1, the observations in group two ($n = 233$) are simulated from exposure-response surface h_2, and the observations in group three are simulated from h_3 ($n = 73$).

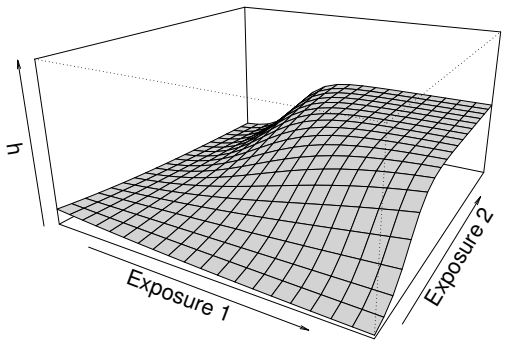
We used 10 covariates from the Bangladesh cohort data set, treated as continuous predictors, and generated their regression coefficients from a standard normal distribution. The main effect for each modifier group, with respect to the reference group, is set to 1. To generate response values, we use a Gaussian likelihood with variance σ^2 and use a signal-to-noise ratio of 1, where signal is the standard deviation of the mean response $\mathbf{h} + \mathbf{X}\boldsymbol{\beta}$ and noise is σ in (4.2). We simulated 200 data sets for each scenario. We used BKMR default priors listed in the Appendix C "Additional Methods" section. We tuned hyper-parameters, i.e., step size and starting values, for model convergence. We ran all models, except for the second group-separable model, for 10,000 MCMC iterations, removing 5,000 as warm-up. We ran the second group-separable model for 20,000 MCMC iterations and removed 5,000 as warm-up, thinning by 5 to retain the same number of posterior samples as the other models. The sample size for the simulation was $n = 351$.

Performance evaluation

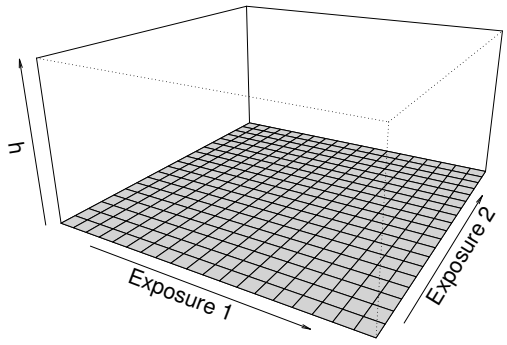
We considered the following parameters and evaluation metrics to assess model performance. The parameters we estimated were the exposure-response surface, group-specific total mixture effects, group-specific single-exposure effects, between-group differences in total mixture effects, and between-group differences in single-exposure effects. To assess performance, we computed root mean squared error (RMSE) of the posterior mean along with coverage and average width for 0.95 probability credible intervals (CIs) for each parameter. For each simulated data set, we average these metrics across exposure ranges defined in the Appendix C "Simulation Methods"



(a) h_1



(b) h_2



(c) h_3

Figure 4.1: Exposure-response surfaces used in simulation. The axes on the horizontal plane are the first two exposures, and the vertical axis is the mean response. The exposure-response surface is constant across the third exposure.

section. Additionally, we average RMSE across the modifier groups for group-specific effects and the exposure-response surface. We average metrics for each effect across the 200 data sets.

4.3 Results

4.3.1 Analysis Results

The analysis includes 179 boys and 170 girls. All models are adjusted for potential confounders as described in the Methods section. Figure 4.2 shows the estimated sex-specific exposure-response functions from each of the five models for each exposure. The models are arranged from least flexible on the left (exposure-only BKMR) to most flexible on the right (stratified BKMR). Figure 4.3 shows estimated sex-specific single-exposure effects and estimated differences in single-exposure effects between sexes for lead.

All model estimates of the exposure-response function for lead in Figure 4.2 showed evidence of a negative relationship between lead and BSID-III. For the exposure-only model, this relationship was for both sexes combined because the exposure-only model only includes child sex as a covariate. Estimates of the exposure-response function from the modifier-in-kernel model were similar to estimates from the exposure-only model for both sexes, with almost no difference between the two estimated sex-specific exposure-response functions. In contrast, the estimates from the group-separable and stratified models, which are more flexible, demonstrated evidence of a negative relationship between lead and BSID-III for boys but no relationship for girls. These more flexible models had larger CIs, with the stratified model having the largest CIs.

For manganese, the estimated sex-specific exposure-response functions in Figure 4.2 from all models showed a positive relationship with BSID-III for boys. Overall, the CIs are larger for manganese, providing little evidence of an exposure-response relationships statistically distinguishable from no effect. The estimated exposure-response functions from the modifier-in-kernel model were nearly identical for boys and girls. In contrast, the group-separable and stratified models showed evidence of effect modification. The estimated exposure-response functions for boys showed a positive relationship with BSID-III, while the estimated functions for girls did not show evidence

of a relationship. For arsenic, the estimated functions from all models show no evidence of effect modification and flatter estimated exposure-response functions.

Figure 4.3 shows the sex-specific estimated single-exposure effect, or average difference in BSID-III per IQR change in lead (4.3a), and the estimated difference of that effect between boys and girls (4.3b) from each of the models. Figure 4.3a shows evidence of a negative association between lead and BSID-III from the group-separable and stratified models among boys, but no association among girls. Figure 4.3b shows moderate evidence of a difference in the lead association between boys and girls. Again, there are larger CIs for the stratified model than the group-separable models. In contrast, we estimated smaller attenuated effects from the modifier-in-kernel and exposure-only model and no evidence of differences in these effects between boys and girls. For both manganese and arsenic, we found no evidence of sex-specific single-exposure effects on BSID-III (Appendix Figure C.1), and we found no evidence of a difference in single-exposure effect between sexes (Appendix Figure C.2). Furthermore, we found no evidence of an effect of the total mixture on BSID-III (Appendix Figures C.3 and C.4).

4.3.2 Simulation Results

Table 4.2 shows performance metrics for the five models we consider in the simulation study described above.

Exposure-response surface

For the estimated exposure-response surface across all three scenarios, the group-separable BKMR models had similar RMSE than the other models. RMSE for the stratified model was similar to RMSE for the group-separable models in the two-group scenarios but larger in the three-group scenario. Across all scenarios, the coverage was similar for all models. However, the average CI widths varied substantially. The exposure-only and modifier-in-kernel models had small average CI widths, and the stratified model had the largest average CI widths.

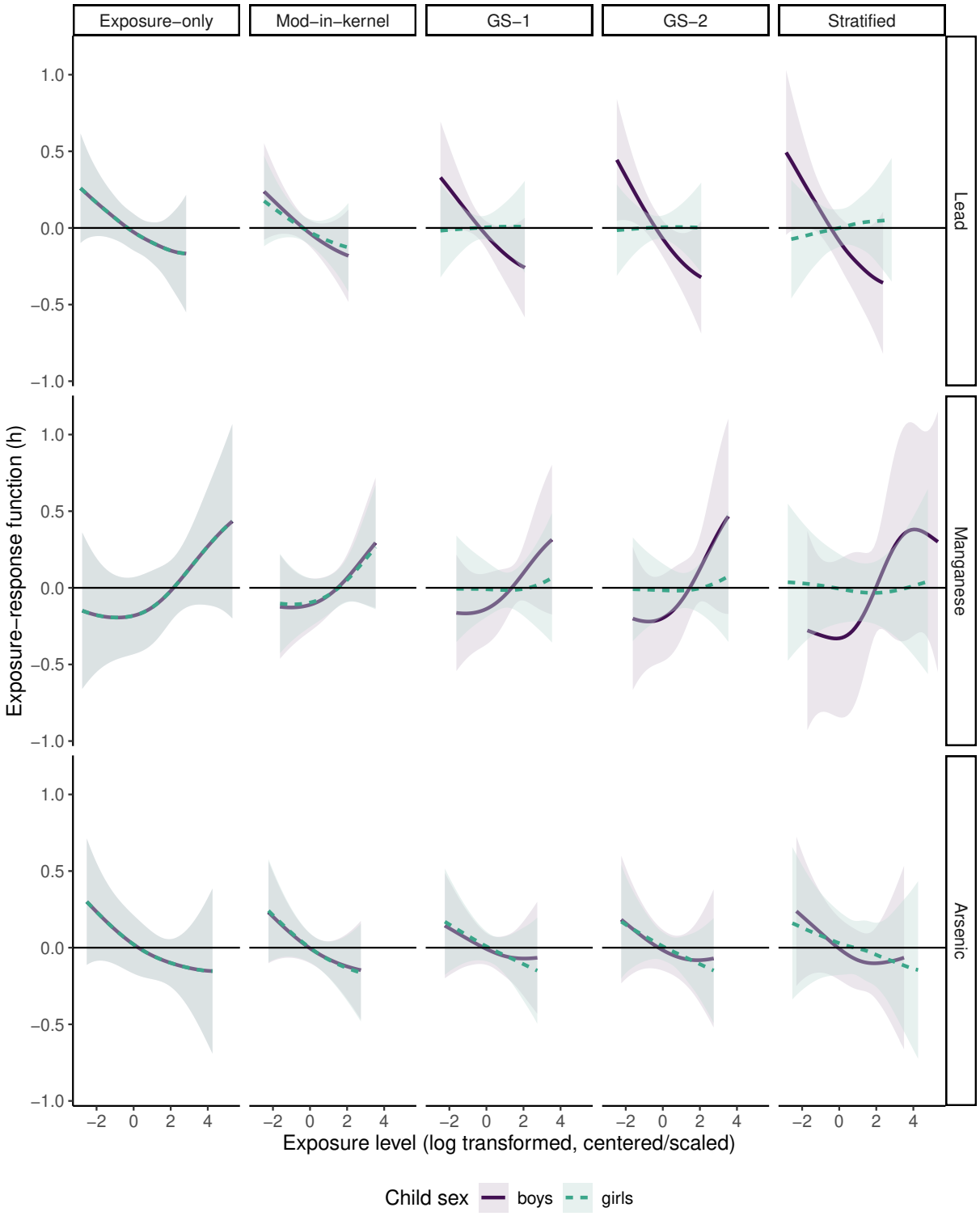
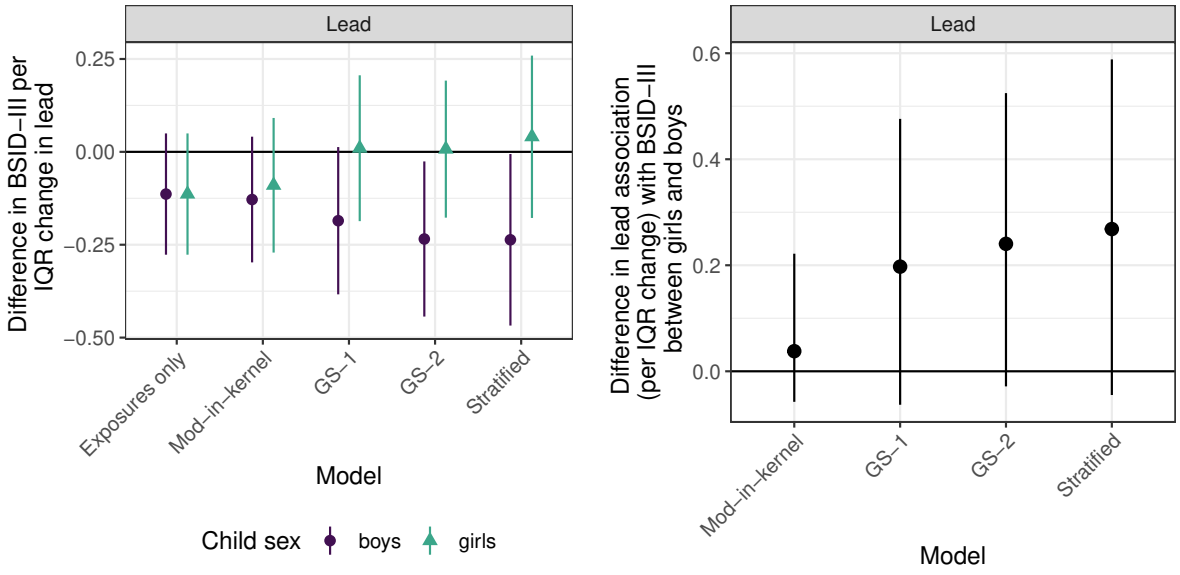


Figure 4.2: Estimated sex-specific single-exposure exposure-response functions for each metal exposure (rows) and each model (columns) in $n = 349$ children in the Bangladesh cohort. The line represents the posterior mean response for an individual in the specified group as exposure to one metal varies (x-axis) and the other metals are fixed at their medians. The ribbons represent 0.95 CIs for the curve, and GS indicated the group-separable models.



(a) Group-specific lead association estimates

(b) Between-group differences in lead association estimates

Figure 4.3: Estimated sex-specific and between-group differences in single-exposure effects of lead on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. Panel (a) shows the estimated estimated average change in BSID-III for an IQR change in lead, while arsenic and manganese are fixed at their medians, for both girls (green triangles) and boys (purple circles) separately. Panel (b) shows the estimated difference in this effect between girls and boys.

Table 4.2: Performance metrics for BKMR without the modifier (exposure-only), BKMR with the modifier in the kernel (mod-in-kernel), group-separable approach 1 (group-sep. 1) with shared overall smoothing parameter τ, group-separable approach 2 (group-sep. 2) with group-specific overall smoothing $\boldsymbol{\tau}$ parameters, and stratified models fit for 200 simulated data sets across three simulation scenarios. Root mean squared error (RMSE), coverage (Cvg.), and average CI width (Width) were computed for each parameter estimate and averaged over the 200 data sets.

	Two-group Scenario 1			Two-group Scenario 2			Three-group Scenario		
Model	RMSE	Cvg.	Width	RMSE	Cvg.	Width	RMSE	Cvg.	Width
Exposure-response surface									
Exposure-only	0.19	0.99	1.06	0.26	0.96	1.10	0.22	0.96	1.03
Mod-in-kernel	0.18	1.00	1.27	0.22	0.99	1.35	0.21	1.00	1.48
Group-sep. 1	0.19	1.00	1.36	0.18	1.00	1.41	0.21	1.00	1.78
Group-sep. 2	0.19	1.00	1.38	0.18	1.00	1.43	0.23	1.00	1.88
Stratified	0.20	1.00	1.43	0.19	1.00	1.47	0.27	1.00	2.10
Group-specific total mixture effect									
Exposure-only	0.23	0.79	0.53	0.33	0.58	0.54	0.25	0.75	0.51
Mod-in-kernel	0.20	0.87	0.58	0.26	0.82	0.64	0.25	0.83	0.59
Group-sep. 1	0.20	0.91	0.65	0.19	0.93	0.66	0.22	0.95	0.75
Group-sep. 2	0.19	0.91	0.66	0.18	0.95	0.67	0.22	0.97	0.86
Stratified	0.20	0.92	0.72	0.18	0.96	0.74	0.25	0.98	1.08
Group-specific single-exposure effect									
Exposure-only	0.35	0.91	1.23	0.47	0.80	1.26	0.41	0.84	1.18
Mod-in-kernel	0.34	0.92	1.31	0.40	0.90	1.43	0.41	0.87	1.33
Group-sep. 1	0.36	0.94	1.45	0.35	0.95	1.48	0.40	0.95	1.67
Group-sep. 2	0.36	0.95	1.49	0.34	0.97	1.52	0.43	0.97	1.93
Stratified	0.38	0.95	1.61	0.35	0.97	1.60	0.50	0.98	2.43
Between-group differences in total mixture effect									
Mod-in-kernel	0.22	0.71	0.48	0.39	0.56	0.63	0.35	0.45	0.49
Group-sep. 1	0.18	0.98	0.88	0.24	0.94	0.91	0.27	0.96	1.05
Group-sep. 2	0.19	0.98	0.91	0.22	0.96	0.93	0.28	0.97	1.22
Stratified	0.20	0.98	1.02	0.23	0.97	1.06	0.32	0.99	1.60
Between-group differences in single-exposure effect									
Mod-in-kernel	0.31	0.83	0.99	0.57	0.73	1.27	0.48	0.65	1.01
Group-sep. 1	0.38	0.98	1.91	0.46	0.95	1.98	0.47	0.96	2.22
Group-sep. 2	0.40	0.98	2.00	0.46	0.97	2.06	0.53	0.98	2.61
Stratified	0.44	0.97	2.28	0.49	0.96	2.31	0.64	0.99	3.44

Group-specific effects

For both the group-specific total mixture effect estimates and single-exposure effect estimates, the group-separable models had RMSE as low or lower than the other models in most cases. The group-separable and stratified models had similar RMSE in the two-group scenarios, but the group-separable models had lower RMSE in the three-group scenario. For group-specific single-exposure effect estimates in the three-group scenario, RMSE for the group-separable models was 0.40 for the first approach and 0.43 for the second, but RMSE for the stratified model was 0.50. The group-separable and stratified models had similar coverage, but the group-separable models had smaller average CI widths. For group-specific single-exposure effect estimates in the three-group scenario, average CI width for the group-separable models was 1.67 for the first approach and 1.93 for the second, but average CI width for the stratified model was 2.43. While the two group-separable models had similar RMSE and coverage, the first-approach group-separable model, with smoothing parameter τ shared among groups, had smaller average CI widths.

The modifier-in-kernel and exposure-only models generally had higher RMSE than the group-separable models and had coverage well below the nominal coverage level. For example, in the second two-group scenario, the total mixture effect RMSE was higher for the exposure-only (0.33) and modifier-in-kernel (0.26) models than for the group-separable and stratified models (0.18 to 0.19). Coverage for the exposure-only (0.58) and modifier-in-kernel models (0.82) was well below the nominal level of 0.95. In the three-group scenario for single-exposure effects, the exposure-only and modifier-in-kernel models had similar RMSE compared to the other models. However, this low RMSE was accompanied by low coverage for the exposure-only (0.84) and modifier-in-kernel (0.87) models.

Between-group differences in effects

For the between-group differences in both the total mixture effect and the single-exposure effect estimates, the group-separable models had RMSE as low or lower than stratified model. While the group-separable and stratified models had similar coverage, the group-separable models had smaller CI widths on average compared to the stratified model. The two group-separable models

had similar RMSE and coverage, but the first group-separable model with shared smoothing parameter τ had narrower CIs on average. The modifier-in-kernel model generally had higher RMSE than the group-separable models and had coverage below the nominal level (0.45 to 0.83). The exposure-only model cannot estimate between-group differences.

4.4 Discussion

There is substantial interest in estimating the health effects of environmental mixtures, and BKMR has emerged as a popular tool for such analyses. However, there is little guidance on how to conduct subgroup analyses and estimate between-group differences in effects with BKMR. In this paper, we presented and evaluated methods for assessing effect modification by a categorical variable with BKMR. The first method we considered fits multiple BKMR models stratified by the modifier's levels, and the second includes the modifier directly into the kernel function used to model the exposure-response function. These first two methods can be implemented with the existing BKMR framework and software. The other two methods we consider are new variants of the BKMR framework that involve constructing a group-separable kernel with two approaches to control smoothness. This structure allows for increased flexibility in the group-specific exposure-response functions while allowing for coherent inference on both subgroup-specific total mixture effects and between-group differences.

Table 4.1 highlights the differences in assumptions for each approach. The least flexible model for modification is the modifier-in-kernel model, which includes the modifier in the kernel of a standard BKMR model, along with the exposures. This approach shares all model parameters among groups, including smoothing parameters for the exposure-response functions, the covariates' effects, and residual variance. Including the modifier in the kernel this way forces a similar shape in the exposure-response function for all groups. At the other extreme is the stratified BKMR approach, which fits multiple models stratified by the modifier levels. This approach does not share any model parameters across groups. While this provides the most flexibility in allowing for effect modification, not sharing information across groups can increase the variance of the estimators and

limit inferences that can be made. This is especially true when some levels of the modifier have small sample sizes.

The two new models fit BKMR with a separable covariance function. Both variants pool observations and share exposure-specific smoothing, covariate effects, and residual variance parameters across groups. However, the variants differ in how they smooth the exposure-response surface. The first variant shares the overall exposure-response surface smoothing parameter, similar to the modifier-in-kernel model, but does not force similarity in exposure-response surface shapes among groups like the modifier-in-kernel approach does. The second variant allows for more flexibility by using group-specific smoothing parameters for the exposure-response surface, similar to the stratified model. The tradeoff between assumptions of existing models allows the new models to address limitations of existing approaches.

The results from the analysis of the Bangladesh cohort illustrate the differences in these methods. Using the modifier-in-kernel method, we estimated exposure-response functions for each metal that were nearly identical for both boys and girls. In contrast, we estimated different exposure-response functions for girls and boys using the two group-separable and stratified models. The results in this analysis demonstrate the progression of these methods where the more flexible methods have increased capacity to estimate group-specific effects and between-group differences but have larger variances.

Results from the simulation study demonstrate similar differences among the methods. In most scenarios, the modifier-in-kernel model had RMSE similar to, or larger than, the group-separable models, and often had coverage below the nominal level and narrower intervals. This indicates that while the model had lower variance, the RMSE was larger due to increased bias that resulted from the restrictive nature of the imposed penalty, which forces the exposure-response surface to be similar between groups. In fact, the modifier-in-kernel model often performed similarly to the exposure-only model, which did not include modification. Similarly, the stratified model had larger RMSE than the group-separable models, but while coverage was close to the nominal level, CIs were wider. This combination of lower RMSE but increased CI widths, implies the method has

lower bias but larger variance than the modifier-in-kernel method. The group-separable models had comparatively low RMSE and smaller interval widths compared to both the modifier-in-kernel and stratified approaches. The reduction in variance compared to the stratified model is most apparent for between-group differences effects estimates and in the scenario with three subgroups due to smaller sample size in each stratified model. This implies that, relative to the other methods, both the bias and variance are low using the group-separable approaches.

The group-separable models introduced here are “Goldilocks" models, combining the lower variance of a pooled analysis that uses parameter sharing with the flexibility in the exposure-response functions of a stratified model. Because the group-separable models do not impose penalization on the shape of the exposure-response surface among the groups, the model has increased flexibility to better estimate group-specific effects and between-group differences. Due to parameter sharing among the groups, the group-separable model had lower variance than the stratified model, particularly when some subgroups had small sample sizes. The second group-separable approach (with group-specific smoothing parameter τ) adds an additional flexibility from the first approach to allow different levels of overall exposure-response smoothing in each group at the cost of only slightly increased variance.

Some of the model assumptions can be relaxed. The group-separable model assumes a common covariate effect for all groups. This assumption can be relaxed by including an interaction between covariates and the modifier. Additionally, the group-separable models assume equal variance among the groups, which could be relaxed by including group-specific variance parameters. While BKMR is flexible, limitations also exist. First, it is not possible to test all possible effect estimates in a simulation study. Second, we do not include selection on the exposures. While this limitation could be addressed, it is not obvious whether variable selection should be done separately for each subgroup or together. Third, we do not explore all choices of kernels. Adding more flexibility of smoothing the exposure-response function is possible with other kernel choices, such as allowing for group-specific smoothing parameters in each exposure dimension.

4.5 Conclusions

This work extends the BKMR framework and fills a critical methodological gap in epidemiological research. We demonstrated that we can include effect modification in BKMR. In light of our simulation and data analysis results, we recommend using our novel group-separable models to include categorical effect modification in BKMR. For studies with a large sample sizes within each subgroup, the stratified model is also a sufficient choice. In such cases, the stratified model can be expected to perform similarly to the group-separable models but at a lower computational cost. For smaller sample sizes or samples with some small subgroups, the parameter sharing of the group-separable model will result in better performance. We do not recommend the modifier-in-kernel approach. The stratified model can be fit with the existing `bkmr` R package. The group separable models can be fit with the package `bkmrGS`, which is built upon the `bkmr` package. The `bkmrGS` R package is available on GitHub.

Chapter 5

Discussion

5.1 Overview

Identifying effect modification is a key goal in many environmental epidemiological studies. The work in this dissertation builds upon existing methods in environmental health research to incorporate effect modification into models for multivariate environmental exposures. I considered two common forms of multivariate environmental exposures: repeated measures of a single exposure across an exposure history and mixture exposures. A popular method to use for repeated measures of exposures is a distributed lag model (DLM), which regresses the response onto repeated measures of the exposures with regularization to account for correlation between exposure-time points. A popular method to use for mixture exposures is Bayesian kernel machine regression (BKMR), which models the response as a flexible function of the multiple exposures and represents the complex exposure-response relationship via a kernel function in a Bayesian hierarchical framework. In this dissertation, I presented new approaches for including effect modification in these two multivariate environmental exposure models. The extensions I developed allow epidemiologists to estimate and quantify uncertainty for interactions between the exposure data and the modifying variable or variables.

In Chapter 2, I considered exposure settings with repeated measures of a single exposure and presented a novel extension of the distributed lag framework to include effect modification by a single continuous variable. I developed a distributed lag interaction model (DLIM) that allows the exposure-time-response function to vary by a single continuous modifier. I showed in simulation that a DLIM out-performs a standard DLM when modification of the exposure-time-response function is truly present. The estimated exposure-time-response functions can be used to estimate different windows of susceptibility based on the modifier. I applied this method to a Colorado birth cohort and estimated different exposure-time-response curves based on the Colorado Enviro-

Screen health and social factors score, a census-tract-level modifier quantifying health and social vulnerabilities. Compared to the standard DLM, this method allowed for insight into the variation of the association between birth weight and exposure to $PM_{2.5}$ across the Front Range region of Colorado. The novel application of this method using an EnviroScreen metric can aid in the flexibility of similar future studies. I published this work (Demateis et al., 2024b) in *Environmetrics* and published the `R` package `dlim` on CRAN (Demateis et al., 2024a) to make performing analyses user-friendly.

I performed an additional analysis with the method described in Chapter 2 with collaborators at the Icahn School of Medicine at Mount Sinai. For this analysis, I extended the DLIM method to a mixed model and applied the model to repeated measures of an outcome. This allowed the DLIM to be fit for trajectory outcomes. Specifically, I fit a DLIM to estimate the effect of exposure to $PM_{2.5}$ 2 months before through 22 months after the birthing parent's last menstrual period on maternal lipid measures, modified by time after child birth. This analysis demonstrates a different application of effect modification because it allows for variation of the exposure-time-response curve longitudinally instead of across the population. This work is published in *Environmental Research* (India Aldana et al., 2025).

In Chapter 3, I continued the extension of the DLM framework from Chapter 2 to include modification of multiple modifiers via a modification index. I developed a distributed lag interaction model with index modification (DLIM-IM) that allows the exposure-time-response function to vary by multiple modifiers through a modification index. I showed in simulation that a DLIM-IM with multiple modifiers out-performs a DLIM weighting those modifiers into a single index *a priori* using equal weights. Developing this method was motivated by the common practice in environmental health research of combining multiple modifiers into an index and was also motivated by the limitation of fitting a DLIM with a fixed-index modifier, e.g., the EnviroScreen health and social factors score. Using this new approach, epidemiologists can simultaneously estimate a modifier index using multiple potential modifiers and also estimate the exposure-time-response function modified by that estimated index. I extended the analysis from Chapter 2 and applied a

DLIM-IM to construct an estimated health and social vulnerabilities index, while simultaneously estimating how the association between birth weight and exposure to $PM_{2.5}$ varies based on that estimated index. This novel approach is built upon the DLM (Chiu et al., 2016; Hsu et al., 2015) and single-index model literature (Czarnota et al., 2015; Daniel et al., 2020; Keil et al., 2020; Wheeler et al., 2021; Wu et al., 2021) in environmental health research to allow for data-derived index modification.

In Chapter 4, I considered exposure settings with multiple exposures and evaluated methods of including effect modification in BKMR. I considered four approaches: direct inclusion of the modifier in the kernel along with the exposures (modifier-in-kernel), two novel group-separable approaches that uses the modifier to construct a kernel blocked by modifier group, and a stratified model. In simulation, I compared these approaches and showed that the modifier-in-kernel approach has poor performance in practice and the group-separable models are preferable over the stratified approach. This allows researchers to perform analyses similar to fitting stratified models with the added benefits of fitting a single model, which allows for sharing of information between groups and more efficient between-group comparisons. I applied the stratified, modifier-in-kernel, and group-separable models to a Bangladesh cohort to estimate effects of exposure to metals during gestation on infant neurodevelopment based on infant sex. The proposed group-separable model produced estimates similar to the stratified approach but with smaller credible intervals that still maintained the nominal coverage level. I estimated differences in associations with neurodevelopment between boys and girls more efficiently with the proposed group-separable models than the stratified models. The proposed models for assessing categorical effect modification with BKMR allows epidemiologists to identify effect modification more efficiently.

5.2 Future Work

There are several potential extensions of the methods I proposed in this dissertation. The standard DLM assumes a linear association between the exposure and response, and I retain that assumption for the DLIMs developed. However, the standard DLM framework was extended to

allow for non-linear associations (Gasparrini et al., 2010). Allowing for non-linear associations in the DLIM framework would allow for more flexibility, though it comes with a set of challenges. The extension of non-linear associations in a DLM required the development of a cross-basis, which I implement in the DLIM framework to allow for modification. Because I already represent the exposure-response surface using a bi-dimensional function space, I would have to introduce a further dimension to allow for non-linear associations. This would be highly parametric, require a large data set, decrease computational speed, and introduce a challenge to interpretation of the results.

The DLIM-IM has limitations that, if improved upon, would greatly increase the usability of this model. The primary short-coming is the computational cost of fitting the model. Unlike the DLIM, the DLIM-IM uses MCMC for estimation. Using approximate Bayesian computation methods such as Laplacian approximation could greatly increase the computational speed without hindering inference. A second limitation of the DLIM-IM is the lack of flexibility to impose penalization on the exposure-time-response surface. I impose a ridge penalty on the regression coefficients through my choice of prior, but I could have used different priors. Smoothness could also be controlled directly by modeling the exposure-time-response surface using a Gaussian process model instead of the cross-basis structure.

Additionally, a test for interaction in the DLIM-IM framework would allow researchers to support estimated effect modification of the exposure-time-response function with a formal test. Finally, I chose to use the same modification index at each exposure-time point by not allowing the modifier index weights to vary by time. Allowing the modifier index weights to vary by time would increase flexibility of the model. However, a model with time-varying weights would be highly parametric and require a massive data set to estimate parameters.

Extensions of categorical effect modification in BKMR include variation in kernel construction. I considered two kernel constructions: one with the modifier in the kernel and the other using the modifier to block the kernel. The blocked kernel construction could have many variations. One that I did not consider is allowing the smoothing parameter ρ_m for exposure m to be specific for group

p, i.e., ρ_{mp}. This would allow for the estimated exposure-response function for each category to have different levels of smoothing. Additionally, a formal test for modification of the exposure-response function would be helpful for researchers to use in publications. Finally, my work can be extended to allow for selection among the exposures but would need to be tested to determine how to perform selection would be each group.

Bibliography

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1):121–131.

Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*, 33(1):178.

Arcaya, M. C., Tucker-Seeley, R. D., Kim, R., Schnake-Mahl, A., So, M., and Subramanian, S. (2016). Research on neighborhood effects on health in the United States: A systematic review of study characteristics. *Social Science & Medicine*, 168:16–29.

Armstrong, B. (2006). Models for the Relationship Between Ambient Temperature and Daily Mortality. *Epidemiology*, 17(6):624–631.

Bauer, J. A., Fruh, V., Howe, C. G., White, R. F., and Claus Henn, B. (2020). Associations of Metals and Neurodevelopment: a Review of Recent Evidence on Susceptibility Factors. *Current Epidemiology Reports*, 7(4):237–262.

Becklake, M. R. and Kauffmann, F. (1999). Gender differences in airway behaviour over the human life span. *Thorax*, 54(12):1119–1138.

Bell, M. L., Zanobetti, A., and Dominici, F. (2013). Evidence on Vulnerability and Susceptibility to Health Risks Associated With Short-Term Exposure to Particulate Matter: A Systematic Review and Meta-Analysis. *American Journal of Epidemiology*, 178(6):865–876.

Bobb, J. F. (2022). bkmr: Bayesian Kernel Machine Regression. R package version 0.2.2.

Bobb, J. F., Claus Henn, B., Valeri, L., and Coull, B. A. (2018). Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. *Environmental Health*, 17(1):67.

Bobb, J. F., Valeri, L., Claus Henn, B., Christiani, D. C., Wright, R. O., Mazumdar, M., Godleski, J. J., and Coull, B. A. (2015). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*, 16(3):493–508.

Bose, S., Chiu, Y.-H. M., Hsu, H.-H. L., Di, Q., Rosa, M. J., Lee, A., Kloog, I., Wilson, A., Schwartz, J., Wright, R. O., Cohen, S., Coull, B. A., and Wright, R. J. (2017). Prenatal Nitrate Exposure and Childhood Asthma. Influence of Maternal Prenatal Stress and Fetal Sex. *American Journal of Respiratory and Critical Care Medicine*, 196(11):1396–1403.

Bose, S., Rosa, M. J., Mathilda Chiu, Y.-H., Leon Hsu, H.-H., Di, Q., Lee, A., Kloog, I., Wilson, A., Schwartz, J., Wright, R. O., Morgan, W. J., Coull, B. A., and Wright, R. J. (2018). Prenatal nitrate air pollution exposure and reduced child lung function: Timing and fetal sex effects. *Environmental Research*, 167:591–597.

Bosetti, C., Nieuwenhuijsen, M. J., Gallus, S., Cipriani, S., La Vecchia, C., and Parazzini, F. (2010). Ambient particulate matter and preterm birth or birth weight: a review of the literature. *Archives of Toxicology*, 84(6):447–460.

Braun, J. M., Wright, R. J., Just, A. C., Power, M. C., Tamayo Y Ortiz, M., Schnaas, L., Hu, H., Wright, R. O., and Tellez-Rojo, M. M. (2014). Relationships between lead biomarkers and diurnal salivary cortisol indices in pregnant women from Mexico City: a cross-sectional study. *Environmental Health*, 13(1):50.

Burris, H. H., Baccarelli, A. A., Motta, V., Byun, H.-M., Just, A. C., Mercado-Garcia, A., Schwartz, J., Svensson, K., Téllez-Rojo, M. M., and Wright, R. O. (2014). Association between length of gestation and cervical DNA methylation of *PTGER2* and LINE 1-HS. *Epigenetics*, 9(8):1083–1091.

Carrico, C., Gennings, C., Wheeler, D. C., and Factor-Litvak, P. (2015). Characterization of Weighted Quantile Sum Regression for Highly Correlated Data in a Risk Analysis Setting. *Journal of Agricultural, Biological, and Environmental Statistics*, 20(1):100–120.

Chakravartty, D., Wiseman, C. L. S., and Cole, D. C. (2014). Differential environmental exposure among non-Indigenous Canadians as a function of sex/gender and race/ethnicity variables: A scoping review. *Canadian Journal of Public Health*, 105(6):e438–e444.

Chang, H. H., Warren, J. L., Darrow, L. A., Reich, B. J., and Waller, L. A. (2015). Assessment of critical exposure and outcome windows in time-to-event analysis with application to air pollution and preterm birth study. *Biostatistics*, 16(3):509–521.

Chiu, Y.-H. M., Hsu, H.-H. L., Coull, B. A., Bellinger, D. C., Kloog, I., Schwartz, J., Wright, R. O., and Wright, R. J. (2016). Prenatal particulate air pollution and neurodevelopment in urban children: Examining sensitive windows and sex-specific associations. *Environment International*, 87:56–65.

Chiu, Y.-H. M., Hsu, H.-H. L., Wilson, A., Coull, B. A., Pendo, M. P., Baccarelli, A., Kloog, I., Schwartz, J., Wright, R. O., Taveras, E. M., and Wright, R. J. (2017). Prenatal particulate air pollution exposure and body composition in urban preschool children: Examining sensitive windows and sex-specific associations. *Environmental Research*, 158:798–805.

Choi, J., Knudsen, L. E., Mizrak, S., and Joas, A. (2017). Identification of exposure to environmental chemicals in children and older adults using human biomonitoring data sorted by age: Results from a literature review. *International Journal of Hygiene and Environmental Health*, 220(2):282–298.

Chu, M. T., Ettinger de Cuba, S., Fabian, M. P., Lane, K. J., James-Todd, T., Williams, D. R., Coull, B. A., Carnes, F., Massaro, M., Levy, J. I., Laden, F., Sandel, M., Adamkiewicz, G., and Zanobetti, A. (2022). The immigrant birthweight paradox in an urban cohort: Role of immigrant enclaves and ambient air pollution. *Journal of Exposure Science & Environmental Epidemiology*, 32(4):571–582.

Cohen, S., Kamarck, T., and Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, 24(4):385.

Coker, E., Liverani, S., Ghosh, J. K., Jerrett, M., Beckerman, B., Li, A., Ritz, B., and Molitor, J. (2016). Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles County. *Environment International*, 91:1–13.

Colorado Department of Public Health and Environment (CDPHE) (2022). Colorado EnviroScreen tool. Technical report. Publication Title: Colorado EnviroScreen Tool.

Colorado EnviroScreen Tool Team (2022). Colorado EnviroScreen tool technical document. Technical report, Colorado EnviroScreen tool team., Denver, CO, USA.

Czarnota, J., Gennings, C., and Wheeler, D. C. (2015). Assessment of Weighted Quantile Sum Regression for Modeling Chemical Mixtures and Cancer Risk. *Cancer Informatics*, 14s2:CIN.S17295.

Dai, Y., Zhang, J., Wang, Z., Ding, J., Xu, S., Zhang, B., Guo, J., Qi, X., Chang, X., Wu, C., and Zhou, Z. (2023). Per- and polyfluoroalkyl substances in umbilical cord serum and body mass index trajectories from birth to age 10 years: Findings from a longitudinal birth cohort (SMBCS). *Environment International*, 180:108238.

Daniel, S., Balalian, A. A., Whyatt, R. M., Liu, X., Rauh, V., Herbstman, J., and Factor-Litvak, P. (2020). Perinatal phthalates exposure decreases fine-motor functions in 11-year-old girls: Results from weighted Quantile sum regression. *Environment International*, 136:105424.

Del Rosario, C., Slevin, M., Molloy, E. J., Quigley, J., and Nixon, E. (2021). How to use the Bayley Scales of Infant and Toddler Development. *Archives of disease in childhood - Education & practice edition*, 106(2):108–112.

Demateis, D., Keller, K., and Wilson, A. (2024a). dlim: Distributed Lag Interaction Model. R package version 0.1.0.

Demateis, D., Keller, K. P., Rojas-Rueda, D., Kioumourtzoglou, M., and Wilson, A. (2024b). Penalized distributed lag interaction model: Air pollution, birth weight, and neighborhood vulnerability. *Environmetrics*, 35(4):e2843.

Dominici, F., Peng, R. D., Barr, C. D., and Bell, M. L. (2010). Protecting Human Health From Air Pollution: Shifting From a Single-pollutant to a Multipollutant Approach. *Epidemiology*, 21(2):187–194.

Farzan, S. F., Niu, Z., Guo, F., Shahriar, M., Kibriya, M. G., Jasmine, F., Sarwar, G., Jackson, B. P., Ahsan, H., and Argos, M. (2025). Exposure to metal mixtures and telomere length in Bangladeshi children. *American Journal of Epidemiology*, 194(1):35–43.

Fenton, T. R. (2003). A new growth chart for preterm babies: Babson and Benda's chart updated with recent data and a new format. *BMC Pediatrics*, 3(1):13.

Friedewald, W. T., Levy, R. I., and Fredrickson, D. S. (1972). Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, Without Use of the Preparative Ultracentrifuge. *Clinical Chemistry*, 18(6):499–502.

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224–2234.

Gasparrini, A., Scheipl, F., Armstrong, B., and Kenward, M. G. (2017). A penalized framework for distributed lag non-linear models: Penalized DLNMs. *Biometrics*, 73(3):938–948.

Gibson, E. A., Goldsmith, J., and Kioumourtzoglou, M.-A. (2019a). Complex Mixtures, Complex Analyses: an Emphasis on Interpretable Results. *Current Environmental Health Reports*, 6(2):53–61.

Gibson, E. A., Nunez, Y., Abuawad, A., Zota, A. R., Renzetti, S., Devick, K. L., Gennings, C., Goldsmith, J., Coull, B. A., and Kioumourtzoglou, M.-A. (2019b). An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length. *Environmental Health*, 18(1):76.

Gong, X. and Zhan, F. B. (2022). A method for identifying critical time windows of maternal air pollution exposures associated with low birth weight in offspring using massive geographic data. *Environmental Science and Pollution Research*, 29(22):33345–33360.

Goodman, C. V., Green, R., DaCosta, A., Flora, D., Lanphear, B., and Till, C. (2023). Sex difference of pre- and post-natal exposure to six developmental neurotoxicants on intellectual abilities: a systematic review and meta-analysis of human studies. *Environmental Health*, 22(1):80.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a rougnhness penalty approch*. Number 58 in Monographs on statistics and applied probability. Chapman & Hall, London Glasgow New York [etc.].

Greenbaum, D. and Shaikh, R. (2010). First Steps Toward Multipollutant Science for Air Quality Decisions. *Epidemiology*, 21(2):195–197.

Gutiérrez-Avila, I., Arfer, K. B., Carrión, D., Rush, J., Kloog, I., Naeger, A. R., Grutter, M., Páramo-Figueroa, V. H., Riojas-Rodríguez, H., and Just, A. C. (2022). Prediction of daily mean and one-hour maximum PM2.5 concentrations and applications in Central Mexico using satellite-based machine-learning models. *Journal of Exposure Science & Environmental Epidemiology*, 32(6):917–925.

Hsu, H.-H. L., Mathilda Chiu, Y.-H., Coull, B. A., Kloog, I., Schwartz, J., Lee, A., Wright, R. O., and Wright, R. J. (2015). Prenatal Particulate Air Pollution and Asthma Onset in Urban Children. Identifying Sensitive Windows and Sex Differences. *American Journal of Respiratory and Critical Care Medicine*, 192(9):1052–1059.

India Aldana, S., Demateis, D., Valvi, D., Just, A. C., Gutiérrez-Avila, I., Estrada-Gutierrez, G., Téllez Rojo, M. M., Wright, R. O., Baccarelli, A. A., Wu, H., Keller, K. P., Wilson, A., and Colicino, E. (2025). Windows of susceptibility to air pollution during and surrounding pregnancy in relation to longitudinal maternal measures of adiposity and lipid profiles. *Environmental Research*, 274:121198.

Jacobs, M., Zhang, G., Chen, S., Mullins, B., Bell, M., Jin, L., Guo, Y., Huxley, R., and Pereira, G. (2017). The association between ambient air pollution and selected adverse pregnancy outcomes in China: A systematic review. *Science of The Total Environment*, 579:1179–1192.

Johnson, M., Shin, H. H., Roberts, E., Sun, L., Fisher, M., Hystad, P., Van Donkelaar, A., Martin, R. V., Fraser, W. D., Lavigne, E., Clark, N., Beaulac, V., and Arbuckle, T. E. (2022). Critical Time Windows for Air Pollution Exposure and Birth Weight in a Multicity Canadian Pregnancy Cohort. *Epidemiology*, 33(1):7–16.

Joubert, B. R., Kioumourtzoglou, M.-A., Chamberlain, T., Chen, H. Y., Gennings, C., Turyk, M. E., Miranda, M. L., Webster, T. F., Ensor, K. B., Dunson, D. B., and Coull, B. A. (2022). Powering Research through Innovative Methods for Mixtures in Epidemiology (PRIME) Program: Novel and Expanded Statistical Methods. *International Journal of Environmental Research and Public Health*, 19(3):1378.

Kai, I., Ohi, G., Kobayashi, Y., Ishizaki, T., Hisata, M., and Kiuchi, M. (1991). Quality of Life: A Possible Health Index for the Elderly. *Asia Pacific Journal of Public Health*, 5(3):221–227.

Keil, A. P., Buckley, J. P., O'Brien, K. M., Ferguson, K. K., Zhao, S., and White, A. J. (2020). A Quantile-Based g-Computation Approach to Addressing the Effects of Exposure Mixtures. *Environmental Health Perspectives*, 128(4):047004.

Kopta, S. and Lowry, J. (2002). Psychometric Evaluation of the Behavioral Health Questionnaire-20: A Brief Instrument for Assessing Global Mental Health and the Three Phases of Psychotherapy Outcome. *Psychotherapy Research*, 12(4):413–426.

Koslovsky, M. D. (2023). A Bayesian Zero-Inflated Dirichlet-Multinomial Regression Model for Multivariate Compositional Count Data. *Biometrics*, 79(4):3239–3251.

Laird, N. M. and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4):963.

Lakshmanan, A., Chiu, Y.-H. M., Coull, B. A., Just, A. C., Maxwell, S. L., Schwartz, J., Gryparis, A., Kloog, I., Wright, R. J., and 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. *Environmental Research*, 137:268–277.

Lee, A., Leon Hsu, H.-H., Mathilda Chiu, Y.-H., Bose, S., Rosa, M. J., Kloog, I., Wilson, A., Schwartz, J., Cohen, S., Coull, B. A., Wright, R. O., and Wright, R. J. (2018). Prenatal fine particulate exposure and early childhood asthma: Effect of maternal stress and fetal sex. *Journal of Allergy and Clinical Immunology*, 141(5):1880–1886.

Lee, A. G., Cowell, W., Kannan, S., Ganguri, H. B., Nentin, F., Wilson, A., Coull, B. A., Wright, R. O., Baccarelli, A., Bollati, V., and Wright, R. J. (2020). Prenatal particulate air pollution and newborn telomere length: Effect modification by maternal antioxidant intakes and infant sex. *Environmental Research*, 187:109707.

Letellier, N., Zamora, S., Yang, J.-A., Sears, D. D., Jankowska, M. M., and Benmarhnia, T. (2022). How do environmental characteristics jointly contribute to cardiometabolic health? A quantile g-computation mixture analysis. *Preventive Medicine Reports*, 30:102005.

Leung, M., Rowland, S. T., Coull, B. A., Modest, A. M., Hacker, M. R., Schwartz, J., Kioumourtzoglou, M.-A., Weisskopf, M. G., and Wilson, A. (2023). Bias Amplification and Variance Inflation in Distributed Lag Models Using Low-Spatial-Resolution Data. *American Journal of Epidemiology*, 192(4):644–657.

Levis, B., Negeri, Z., Sun, Y., Benedetti, A., and Thombs, B. D. (2020). Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. *BMJ*, page m4022.

Li, Z.-H., Hu, C.-Y., Dai, S.-W., Ma, H.-Y., Zhang, S.-Y., Sun, C., Li, J.-H., Huang, K., Chen, M.-L., Gao, G.-P., and Zhang, X.-J. (2025). Sex-specific associations between maternal exposure to metal mixtures and fetal growth trajectories: A prospective birth cohort study. *Science of The Total Environment*, 959:178291.

Liu, J. Z., Deng, W., Lee, J., Lin, P.-i. D., Valeri, L., Christiani, D. C., Bellinger, D. C., Wright, R. O., Mazumdar, M. M., and Coull, B. A. (2022). A Cross-Validated Ensemble Approach

to Robust Hypothesis Testing of Continuous Nonlinear Interactions: Application to Nutrition-Environment Studies. *Journal of the American Statistical Association*, 117(538):561–573.

Marra, G. and Wood, S. N. (2012). Coverage Properties of Confidence Intervals for Generalized Additive Model Components: Coverage properties of GAM intervals. *Scandinavian Journal of Statistics*, 39(1):53–74.

Martenies, S. E., Hoskovec, L., Wilson, A., Moore, B. F., Starling, A. P., Allshouse, W. B., Adgate, J. L., Dabelea, D., and Magzamen, S. (2022a). Using non-parametric Bayes shrinkage to assess relationships between multiple environmental and social stressors and neonatal size and body composition in the Healthy Start cohort. *Environmental Health*, 21(1):111.

Martenies, S. E., Zhang, M., Corrigan, A. E., Kvit, A., Shields, T., Wheaton, W., Bastain, T. M., Breton, C. V., Dabelea, D., Habre, R., Magzamen, S., Padula, A. M., Him, D. A., Camargo, C. A., Cowell, W., Croen, L. A., Deoni, S., Everson, T. M., Hartert, T. V., Hipwell, A. E., McEvoy, C. T., Morello-Frosch, R., O'Connor, T. G., Petriello, M., Sathyanarayana, S., Stanford, J. B., Woodruff, T. J., Wright, R. J., and Kress, A. M. (2022b). Associations between combined exposure to environmental hazards and social stressors at the neighborhood level and individual perinatal outcomes in the ECHO-wide cohort. *Health & Place*, 76:102858.

McGee, G., Wilson, A., Coull, B. A., and Webster, T. F. (2023a). Incorporating biological knowledge in analyses of environmental mixtures and health. *Statistics in Medicine*, 42(17):3016–3031.

McGee, G., Wilson, A., Webster, T. F., and Coull, B. A. (2023b). Bayesian multiple index models for environmental mixtures. *Biometrics*, 79(1):462–474.

Millett, L. S. (2016). The Healthy Immigrant Paradox and Child Maltreatment: A Systematic Review. *Journal of Immigrant and Minority Health*, 18(5):1199–1215.

Mork, D., Kioumourtzoglou, M.-A., Weisskopf, M., Coull, B. A., and Wilson, A. (2024). Heterogeneous Distributed Lag Models to Estimate Personalized Effects of Maternal Exposures to Air Pollution. *Journal of the American Statistical Association*, 119(545):14–26.

Mu, C., Lin, M., Shao, Y., Liao, Q., Liang, J., Yu, C., Wu, X., Chen, M., Tang, Y., Zhou, L., Qiu, X., Pan, D., and Huang, D. (2024). Associations between maternal serum neonicotinoid pesticide exposure during pregnancy and newborn telomere length: Effect modification by sampling season. *Ecotoxicology and Environmental Safety*, 273:116164.

Neophytou, A. M., Kioumourtzoglou, M.-A., Goin, D. E., Darwin, K. C., and Casey, J. A. (2021). Educational note: addressing special cases of bias that frequently occur in perinatal epidemiology. *International Journal of Epidemiology*, 50(1):337–345.

Niu, Z., Habre, R., Chavez, T. A., Yang, T., Grubbs, B. H., Eckel, S. P., Berhane, K., Toledo-Corral, C. M., Johnston, J., Dunton, G. F., Lerner, D., Al-Marayati, L., Lurmann, F., Pavlovic, N., Farzan, S. F., Bastain, T. M., and Breton, C. V. (2022). Association Between Ambient Air Pollution and Birth Weight by Maternal Individual- and Neighborhood-Level Stressors. *JAMA Network Open*, 5(10):e2238174.

Pan, S., Li, Z., Rubbo, B., Quon-Chow, V., Chen, J. C., Baumert, B. O., Garcia, E., Aung, M. T., Conti, D. V., and Chatzi, L. (2024). Applications of mixture methods in epidemiological studies investigating the health impact of persistent organic pollutants exposures: a scoping review. *Journal of Exposure Science & Environmental Epidemiology*.

Peng, C., Sanchez-Guerra, M., Wilson, A., Mehta, A. J., Zhong, J., Zanobetti, A., Brennan, K., Dereix, A. E., Coull, B. A., Vokonas, P., Schwartz, J., and Baccarelli, A. A. (2017). Short-term effects of air temperature and mitochondrial DNA lesions within an older population. *Environment International*, 103:23–29.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables. *Journal of the American Statistical Association*, 108(504):1339–1349.

Raz, R., Kioumourtzoglou, M.-A., and Weisskopf, M. G. (2018). Live-Birth Bias and Observed Associations Between Air Pollution and Autism. *American Journal of Epidemiology*, 187(11):2292–2296.

Roca-Pardiñas, J., Cadarso-Suárez, C., and González-Manteiga, W. (2005). Testing for interactions in generalized additive models: Application to SO2 pollution data. *Statistics and Computing*, 15(4):289–299.

Schliep, E. M., Schafer, T. L. J., and Hawkey, M. (2021). Distributed lag models to identify the cumulative effects of training and recovery in athletes using multivariate ordinal wellness data. *Journal of Quantitative Analysis in Sports*, 17(3):241–254.

Schnake-Mahl, A. S., Jahn, J. L., Subramanian, S., Waters, M. C., and Arcaya, M. (2020). Gentrification, Neighborhood Change, and Population Health: a Systematic Review. *Journal of Urban Health*, 97(1):1–25.

Schwartz, J. (2000). The Distributed Lag between Air Pollution and Daily Deaths:. *Epidemiology*, 11(3):320–326.

Sherlock, P., Shalowitz, M. U., Berry, C., Cella, D., Blackwell, C. K., Cowell, W., Rodriguez, K. M. R., Wright, R. J., and program collaborators for Environmental Influences on Child Health Outcomes (2023). A short form of the Crisis in Family Systems (CRISYS) in a racially diverse sample of pregnant women. *Current Psychology*, 42(10):8393–8401.

Son, J.-Y., Liu, J. C., and Bell, M. L. (2019). Temperature-related mortality: a systematic review and investigation of effect modifiers. *Environmental Research Letters*, 14(7):073004.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., and Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, Palo Alto, CA.

Stieb, D. M., Chen, L., Eshoul, M., and Judek, S. (2012). Ambient air pollution, birth weight and preterm birth: A systematic review and meta-analysis. *Environmental Research*, 117:100–111.

Taylor, K. W., Joubert, B. R., Braun, J. M., Dilworth, C., Gennings, C., Hauser, R., Heindel, J. J., Rider, C. V., Webster, T. F., and Carlin, D. J. (2016). Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology: Lessons from an Innovative Workshop. *Environmental Health Perspectives*, 124(12).

Totsika, V. and Sylva, K. (2004). The Home Observation for Measurement of the Environment Revisited. *Child and Adolescent Mental Health*, 9(1):25–35.

VanderWeele, T. J. (2009). On the Distinction Between Interaction and Effect Modification. *Epidemiology*, 20(6):863–871. Publisher: Ovid Technologies (Wolters Kluwer Health).

Vedal, S. and Kaufman, J. D. (2011). What Does Multi-Pollutant Air Pollution Research Mean? *American Journal of Respiratory and Critical Care Medicine*, 183(1):4–6.

Vuong, A. M., Yolton, K., Braun, J. M., Lanphear, B. P., and Chen, A. (2020). Chemical mixtures and neurobehavior: a review of epidemiologic findings and future directions. *Reviews on Environmental Health*, 35(3):245–256.

Wang, J., Wang, W., Zhang, W., Wang, J., Huang, Y., Hu, Z., Chen, Y., Guo, X., Deng, F., and Zhang, L. (2022). Co-exposure to multiple air pollutants and sleep disordered breathing in patients with or without obstructive sleep apnea: A cross-sectional study. *Environmental Research*, 212:113155.

Wang, Y., Meng, Z., Wei, S., Li, X., Su, Z., Jiang, Y., Wu, H., Pan, H., Wang, J., Zhou, Q., Qiao, Y., and Fan, Y. (2024). Urinary volatile organic compound metabolites and COPD among US adults: mixture, interaction and mediation analysis. *Environmental Health*, 23(1):45.

Warren, J., Fuentes, M., Herring, A., and Langlois, P. (2012). Spatial-Temporal Modeling of the Association between Air Pollution Exposure and Preterm Birth: Identifying Critical Windows of Exposure. *Biometrics*, 68(4):1157–1167.

Warren, J. L., Fuentes, M., Herring, A. H., and Langlois, P. H. (2013). Air Pollution Metric Analysis While Determining Susceptible Periods of Pregnancy for Low Birth Weight. *ISRN Obstetrics and Gynecology*, 2013:1–9.

Warren, J. L., Luben, T. J., and Chang, H. H. (2020). A spatially varying distributed lag model with application to an air pollution and term low birth weight study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(3):681–696.

Wheeler, D. C., Rustom, S., Carli, M., Whitehead, T. P., Ward, M. H., and Metayer, C. (2021). Assessment of Grouped Weighted Quantile Sum Regression for Modeling Chemical Mixtures and Cancer Risk. *International Journal of Environmental Research and Public Health*, 18(2):504.

Wilson, A., Chiu, Y.-H. M., Hsu, H.-H. L., Wright, R. O., Wright, R. J., and Coull, B. A. (2017a). Bayesian distributed lag interaction models to identify perinatal windows of vulnerability in children's health. *Biostatistics*, 18(3):537–552.

Wilson, A., Chiu, Y.-H. M., Hsu, H.-H. L., Wright, R. O., Wright, R. J., and Coull, B. A. (2017b). Potential for Bias When Estimating Critical Windows for Air Pollution in Children's Health. *American Journal of Epidemiology*, 186(11):1281–1289.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman & Hall/CRC texts in statistical science. CRC Press/Taylor & Francis Group, Boca Raton, second edition edition.

Wu, C.-J., Ho, A.-C., Chen, S.-Y., Pan, C.-H., Chuang, H.-C., Chen, W.-L., Wang, C.-C., and Lai, C.-H. (2021). Occupational Exposure to Metals and Serum Adiponectin Levels Among Shipyard Workers: A Cohort Study with Weighted Quantile Sum Regression Analyses.

Yang, D., Chen, L., Yang, Y., Shi, J., Huang, Z., Li, M., Yang, Y., and Ji, X. (2022). Effect of PM2.5 exposure on Vitamin D status among pregnant women: A distributed lag analysis. *Ecotoxicology and Environmental Safety*, 239:113642.

Yip, R. (1987). Altitude and birth weight. *The Journal of Pediatrics*, 111(6):869–876.

Yu, Y. and Ruppert, D. (2002). Penalized Spline Estimation for Partially Linear Single-Index Models. *Journal of the American Statistical Association*, 97(460):1042–1054.

Zanobetti, A. (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics*, 1(3):279–292.

Zhu, G., Wen, Y., Cao, K., He, S., and Wang, T. (2024). A review of common statistical methods for dealing with multiple pollutant mixtures and multiple exposures. *Frontiers in Public Health*, 12:1377685.

Šrám, R. J., Binková, B., Dejmek, J., and Bobak, M. (2005). Ambient Air Pollution and Pregnancy Outcomes: A Review of the Literature. *Environmental Health Perspectives*, 113(4):375–382.

Appendix A

Penalized Distributed Lag Interaction Model

A.1 Additional Simulations

Table A.1: Cumulative and point-wise performance metrics for penalized and non-penalized (-ns) models averaged over all 200 simulated data sets for scenario 1 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear, and the number of basis functions for the exposure-time basis was 5 for both the DLM-ns and DLIM-linear-ns.

		Cumulative			Point-wise		
SNR	Model	RMSE	Coverage	CI Width	RMSE	Coverage	CI Width
low	DLIM(20,20)	0.26	0.95	1.08	0.03	0.92	0.10
	DLIM(10,10)	0.26	0.95	1.08	0.03	0.89	0.09
	DLIM-linear	0.26	0.95	1.07	0.02	0.92	0.09
	DLM	0.17	0.96	0.78	0.02	0.91	0.07
	DLIM-ns(5,5)	0.27	0.95	1.10	0.04	0.92	0.14
	DLIM-linear-ns	0.26	0.95	1.08	0.03	0.89	0.09
	DLM-ns	0.17	0.95	0.78	0.02	0.85	0.06
medium	DLIM(20,20)	0.13	0.95	0.54	0.01	0.96	0.06
	DLIM(10,10)	0.13	0.95	0.54	0.01	0.93	0.05
	DLIM-linear	0.13	0.95	0.54	0.01	0.94	0.05
	DLM	0.08	0.96	0.39	0.01	0.94	0.04
	DLIM-ns(5,5)	0.14	0.95	0.55	0.02	0.85	0.07
	DLIM-linear-ns	0.13	0.95	0.54	0.02	0.75	0.04
	DLM-ns	0.08	0.95	0.39	0.02	0.65	0.03
high	DLIM(20,20)	0.01	0.95	0.06	0.00	0.98	0.01
	DLIM(10,10)	0.01	0.95	0.06	0.00	0.92	0.01
	DLIM-linear	0.01	0.95	0.05	0.00	0.90	0.01
	DLM	0.01	0.96	0.04	0.00	0.90	0.01
	DLIM-ns(5,5)	0.02	0.82	0.07	0.01	0.17	0.01
	DLIM-linear-ns	0.02	0.82	0.07	0.01	0.13	0.01
	DLM-ns	0.02	0.71	0.05	0.01	0.11	0.00

We used the exposure data from the simulation study in the main text, simulated modifier values from a Uniform(0,1) distribution and based data generation on each of the four simulation scenarios described in the main text. For each simulation scenario, we scaled $\beta_t(m_i)$ by 0.1 to obtain reasonable count responses. To simulate response values, we assumed a Poisson distribution and used the log link function. We fixed the regression coefficient for the main effect of the modifier at 0.1. We simulated 100 data sets. We fit the penalized DLIM and DLM with REML and used 20 nominal degrees of freedom in both the exposure-time and modifier dimensions.

Table A.8 includes simulation results for a Poisson model. Overall, the DLIM is better able to capture modification patterns than the DLM. As expected, the DLIM and DLM perform similarly in simulation scenario 1. In the other simulation scenarios, the DLIM performs better than the DLM. While the confidence intervals are wider for the DLIM, the coverage is largely closer to the nominal level than the DLM and the RMSE is smaller. For example, the cumulative coverage for the DLIM in scenario 3 is much closer to the nominal level than the DLM (0.93 and 0.16, respectively). While the larger interval width contributes to this (0.072 and 0.044, respectively), the RMSE is much smaller for the DLIM than the DLM (0.018 and 0.057, respectively). Our results demonstrate that when there is modification in a count data setting, the DLIM is more efficient than the standard DLM.

Table A.2: Cumulative and point-wise performance metrics for penalized and non-penalized (-ns) models averaged over all 200 simulated data sets for scenario 2 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear, and the number of basis functions for the exposure-time basis was 5 for both the DLM-ns and DLIM-linear-ns.

		Cumulative			Point-wise		
SNR	Model	RMSE	Coverage	CI Width	RMSE	Coverage	CI Width
low	DLIM(20,20)	0.33	0.95	1.36	0.03	0.90	0.10
	DLIM(10,10)	0.33	0.95	1.36	0.03	0.89	0.10
	DLIM-linear	0.33	0.95	1.35	0.03	0.89	0.10
	DLM	0.44	0.70	0.98	0.03	0.83	0.08
	DLIM-ns(5,5)	0.34	0.95	1.39	0.05	0.94	0.17
	DLIM-linear-ns	0.33	0.95	1.36	0.03	0.92	0.11
	DLM-ns	0.44	0.70	0.99	0.03	0.84	0.08
medium	DLIM(20,20)	0.17	0.95	0.68	0.02	0.93	0.06
	DLIM(10,10)	0.17	0.95	0.68	0.02	0.91	0.06
	DLIM-linear	0.17	0.95	0.68	0.02	0.91	0.06
	DLM	0.38	0.40	0.50	0.02	0.78	0.05
	DLIM-ns(5,5)	0.17	0.95	0.69	0.03	0.91	0.09
	DLIM-linear-ns	0.16	0.95	0.68	0.02	0.87	0.06
	DLM-ns	0.38	0.40	0.50	0.02	0.70	0.04
high	DLIM(20,20)	0.02	0.95	0.07	0.00	0.98	0.01
	DLIM(10,10)	0.02	0.95	0.07	0.00	0.94	0.01
	DLIM-linear	0.02	0.95	0.07	0.00	0.94	0.01
	DLM	0.36	0.06	0.07	0.01	0.55	0.01
	DLIM-ns(5,5)	0.02	0.89	0.08	0.01	0.28	0.01
	DLIM-linear-ns	0.02	0.89	0.07	0.01	0.18	0.01
	DLM-ns	0.36	0.06	0.08	0.02	0.16	0.01

Table A.3: Cumulative and point-wise performance metrics for penalized and non-penalized (-ns) models averaged over all 200 simulated data sets for scenario 3 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear, and the number of basis functions for the exposure-time basis was 5 for both the DLM-ns and DLIM-linear-ns.

		Cumulative			Point-wise		
SNR	Model	RMSE	Coverage	CI Width	RMSE	Coverage	CI Width
low	DLIM(20,20)	0.40	0.94	1.63	0.05	0.82	0.13
	DLIM(10,10)	0.41	0.94	1.62	0.05	0.80	0.13
	DLIM-linear	0.62	0.79	1.71	0.05	0.73	0.11
	DLM	0.56	0.72	1.25	0.06	0.49	0.08
	DLIM-ns(5,5)	0.42	0.93	1.64	0.07	0.88	0.21
	DLIM-linear-ns	0.62	0.79	1.74	0.06	0.82	0.15
	DLM-ns	0.55	0.73	1.26	0.07	0.62	0.10
medium	DLIM(20,20)	0.21	0.95	0.84	0.04	0.81	0.09
	DLIM(10,10)	0.21	0.93	0.83	0.04	0.78	0.08
	DLIM-linear	0.51	0.69	1.05	0.05	0.62	0.07
	DLM	0.49	0.55	0.77	0.06	0.33	0.05
	DLIM-ns(5,5)	0.24	0.90	0.85	0.04	0.79	0.11
	DLIM-linear-ns	0.52	0.70	1.06	0.05	0.73	0.09
	DLM-ns	0.49	0.57	0.77	0.06	0.42	0.06
high	DLIM(20,20)	0.02	0.92	0.10	0.01	0.91	0.03
	DLIM(10,10)	0.07	0.77	0.18	0.02	0.70	0.03
	DLIM-linear	0.47	0.58	0.70	0.04	0.55	0.05
	DLM	0.47	0.38	0.52	0.06	0.23	0.04
	DLIM-ns(5,5)	0.13	0.61	0.26	0.03	0.51	0.03
	DLIM-linear-ns	0.47	0.59	0.70	0.04	0.62	0.06
	DLM-ns	0.47	0.40	0.52	0.06	0.28	0.04

Table A.4: Cumulative and point-wise performance metrics for penalized and non-penalized (-ns) models averaged over all 200 simulated data sets for scenario 4 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. The number of basis functions for the exposure-time basis was 10 for both the DLM and DLIM-linear, and the number of basis functions for the exposure-time basis was 5 for both the DLM-ns and DLIM-linear-ns.

		Cumulative			Point-wise		
SNR	Model	RMSE	Coverage	CI Width	RMSE	Coverage	CI Width
low	DLIM(20,20)	0.32	0.94	1.29	0.04	0.82	0.10
	DLIM(10,10)	0.32	0.94	1.28	0.04	0.80	0.10
	DLIM-linear	0.48	0.79	1.35	0.04	0.74	0.08
	DLM	0.44	0.71	0.98	0.05	0.52	0.06
	DLIM-ns(5,5)	0.34	0.93	1.30	0.05	0.89	0.16
	DLIM-linear-ns	0.48	0.80	1.37	0.04	0.82	0.12
	DLM-ns	0.44	0.72	0.99	0.05	0.64	0.08
medium	DLIM(20,20)	0.16	0.95	0.66	0.03	0.81	0.07
	DLIM(10,10)	0.17	0.93	0.66	0.03	0.78	0.06
	DLIM-linear	0.40	0.69	0.81	0.04	0.63	0.05
	DLM	0.39	0.54	0.60	0.05	0.35	0.04
	DLIM-ns(5,5)	0.19	0.90	0.67	0.03	0.80	0.08
	DLIM-linear-ns	0.40	0.70	0.83	0.04	0.74	0.07
	DLM-ns	0.39	0.54	0.60	0.05	0.45	0.05
high	DLIM(20,20)	0.02	0.92	0.08	0.01	0.91	0.02
	DLIM(10,10)	0.05	0.77	0.14	0.02	0.71	0.02
	DLIM-linear	0.36	0.59	0.53	0.03	0.55	0.04
	DLM	0.37	0.42	0.39	0.05	0.24	0.03
	DLIM-ns(5,5)	0.10	0.62	0.20	0.02	0.51	0.02
	DLIM-linear-ns	0.36	0.59	0.54	0.03	0.63	0.04
	DLM-ns	0.37	0.43	0.40	0.05	0.30	0.03

Table A.5: Cumulative and point-wise performance metrics for DLIM(20,20) models averaged over all 200 simulated data sets for scenarios 1 and 2 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. Alternative penalization methods are PS(2,1), PS(1,2), and CR and are compared to the penalization method in the main simulation results, PS(2,2).

			Cumulative		Point-wise	
Scenario	SNR	Method	RMSE	Coverage	RMSE	Coverage
1	low	PS(2,2)	0.263	0.95	0.026	0.92
1	low	PS(2,1)	0.172	0.95	0.020	0.94
1	low	PS(1,2)	0.262	0.95	0.024	0.98
1	low	CR	0.263	0.95	0.026	0.92
1	medium	PS(2,2)	0.131	0.95	0.015	0.96
1	medium	PS(2,1)	0.086	0.96	0.011	0.97
1	medium	PS(1,2)	0.131	0.95	0.015	0.99
1	medium	CR	0.131	0.95	0.015	0.96
1	high	PS(2,2)	0.013	0.95	0.003	0.98
1	high	PS(2,1)	0.009	0.95	0.002	0.98
1	high	PS(1,2)	0.013	0.95	0.003	0.97
1	high	CR	0.013	0.95	0.003	0.98
2	low	PS(2,2)	0.331	0.95	0.030	0.90
2	low	PS(2,1)	0.425	0.73	0.027	0.87
2	low	PS(1,2)	0.329	0.95	0.026	0.97
2	low	CR	0.330	0.95	0.030	0.90
2	medium	PS(2,2)	0.166	0.95	0.017	0.93
2	medium	PS(2,1)	0.315	0.56	0.018	0.87
2	medium	PS(1,2)	0.165	0.95	0.016	0.99
2	medium	CR	0.165	0.95	0.017	0.93
2	high	PS(2,2)	0.017	0.95	0.003	0.98
2	high	PS(2,1)	0.019	0.92	0.003	0.99
2	high	PS(1,2)	0.017	0.95	0.004	0.98
2	high	CR	0.017	0.95	0.003	0.98

Table A.6: Cumulative and point-wise performance metrics for DLIM(20,20) models averaged over all 200 simulated data sets for scenarios 3 and 4 and all signal-to-noise (SNR) levels. The low SNR is 0.5, medium SNR is 1, and the high SNR is 10. Alternative penalization methods are PS(2,1), PS(1,2), and CR and are compared to the penalization method in the main simulation results, PS(2,2).

			Cumulative		Point-wise	
Scenario	SNR	Method	RMSE	Coverage	RMSE	Coverage
3	low	PS(2,2)	0.404	0.94	0.049	0.82
3	low	PS(2,1)	0.361	0.96	0.052	0.77
3	low	PS(1,2)	0.402	0.94	0.050	0.85
3	low	CR	0.403	0.94	0.049	0.81
3	medium	PS(2,2)	0.206	0.95	0.036	0.81
3	medium	PS(2,1)	0.202	0.95	0.036	0.84
3	medium	PS(1,2)	0.205	0.95	0.037	0.88
3	medium	CR	0.205	0.95	0.037	0.81
3	high	PS(2,2)	0.025	0.92	0.010	0.91
3	high	PS(2,1)	0.025	0.92	0.010	0.91
3	high	PS(1,2)	0.025	0.92	0.011	0.94
3	high	CR	0.025	0.92	0.010	0.90
4	low	PS(2,2)	0.320	0.94	0.037	0.82
4	low	PS(2,1)	0.288	0.95	0.040	0.77
4	low	PS(1,2)	0.318	0.94	0.038	0.84
4	low	CR	0.319	0.94	0.038	0.81
4	medium	PS(2,2)	0.163	0.95	0.028	0.81
4	medium	PS(2,1)	0.160	0.95	0.028	0.83
4	medium	PS(1,2)	0.162	0.95	0.029	0.88
4	medium	CR	0.162	0.95	0.028	0.81
4	high	PS(2,2)	0.020	0.92	0.008	0.91
4	high	PS(2,1)	0.020	0.92	0.008	0.91
4	high	PS(1,2)	0.020	0.92	0.009	0.94
4	high	CR	0.020	0.92	0.008	0.90

Table A.7: Cumulative and point-wise performance metrics resulting from repeating the simulation study described in the main text, with modifiers drawn from N(0.5,0.25) distribution instead of a Uniform(0,1) distribution.

			Cumulative		Point-wise	
Scenario	SNR	Model	RMSE	Coverage	RMSE	Coverage
1	low	DLIM(20,20)	0.263	0.95	0.026	0.92
1	low	DLIM-linear	0.256	0.95	0.024	0.92
1	low	DLM	0.169	0.96	0.020	0.91
1	medium	DLIM(20,20)	0.131	0.95	0.015	0.96
1	medium	DLIM-linear	0.128	0.95	0.013	0.94
1	medium	DLM	0.084	0.96	0.011	0.94
1	high	DLIM(20,20)	0.013	0.95	0.003	0.98
1	high	DLIM-linear	0.013	0.95	0.002	0.90
1	high	DLM	0.008	0.96	0.002	0.90
2	low	DLIM(20,20)	0.330	0.95	0.030	0.90
2	low	DLIM-linear	0.328	0.95	0.029	0.89
2	low	DLM	0.437	0.70	0.028	0.83
2	medium	DLIM(20,20)	0.165	0.95	0.017	0.93
2	medium	DLIM-linear	0.165	0.95	0.017	0.91
2	medium	DLM	0.381	0.40	0.019	0.78
2	high	DLIM(20,20)	0.017	0.95	0.003	0.98
2	high	DLIM-linear	0.016	0.95	0.002	0.94
2	high	DLM	0.359	0.06	0.014	0.55
3	low	DLIM(20,20)	0.403	0.94	0.049	0.81
3	low	DLIM-linear	0.628	0.77	0.054	0.72
3	low	DLM	0.566	0.70	0.065	0.49
3	medium	DLIM(20,20)	0.205	0.95	0.037	0.81
3	medium	DLIM-linear	0.524	0.68	0.048	0.62
3	medium	DLM	0.506	0.53	0.063	0.33
3	high	DLIM(20,20)	0.025	0.92	0.010	0.90
3	high	DLIM-linear	0.481	0.57	0.046	0.55
3	high	DLM	0.481	0.37	0.062	0.23
4	low	DLIM(20,20)	0.319	0.94	0.038	0.81
4	low	DLIM-linear	0.487	0.78	0.041	0.74
4	low	DLM	0.447	0.70	0.049	0.52
4	medium	DLIM(20,20)	0.162	0.95	0.028	0.81
4	medium	DLIM-linear	0.404	0.68	0.037	0.63
4	medium	DLM	0.400	0.53	0.047	0.35
4	high	DLIM(20,20)	0.020	0.92	0.008	0.90
4	high	DLIM-linear	0.368	0.57	0.035	0.55
4	high	DLM	0.381	0.42	0.047	0.24

Table A.8: Cumulative and point-wise performance metrics for DLIM(20,20) and standard DLM models averaged over 100 simulated data sets for all scenarios. Models were fit using Poisson regression on simulated count data. The DLIM(20,20) was penalized using PS(2,2) penalization and the standard DLM was penalized using P-splines.

		Cumulative			Point-wise		
Scenario	Model	RMSE	Coverage	Width	RMSE	Coverage	Width
1	DLIM	0.010	0.94	0.040	0.001	0.97	0.005
1	DLM	0.005	0.95	0.029	0.001	0.97	0.004
2	DLIM	0.014	0.93	0.056	0.001	0.96	0.006
2	DLM	0.040	0.31	0.038	0.002	0.78	0.005
3	DLIM	0.018	0.93	0.072	0.003	0.90	0.010
3	DLM	0.057	0.16	0.044	0.007	0.16	0.004
4	DLIM	0.021	0.93	0.085	0.003	0.82	0.008
4	DLM	0.044	0.38	0.055	0.005	0.24	0.004

A.2 Additional Figures and Tables

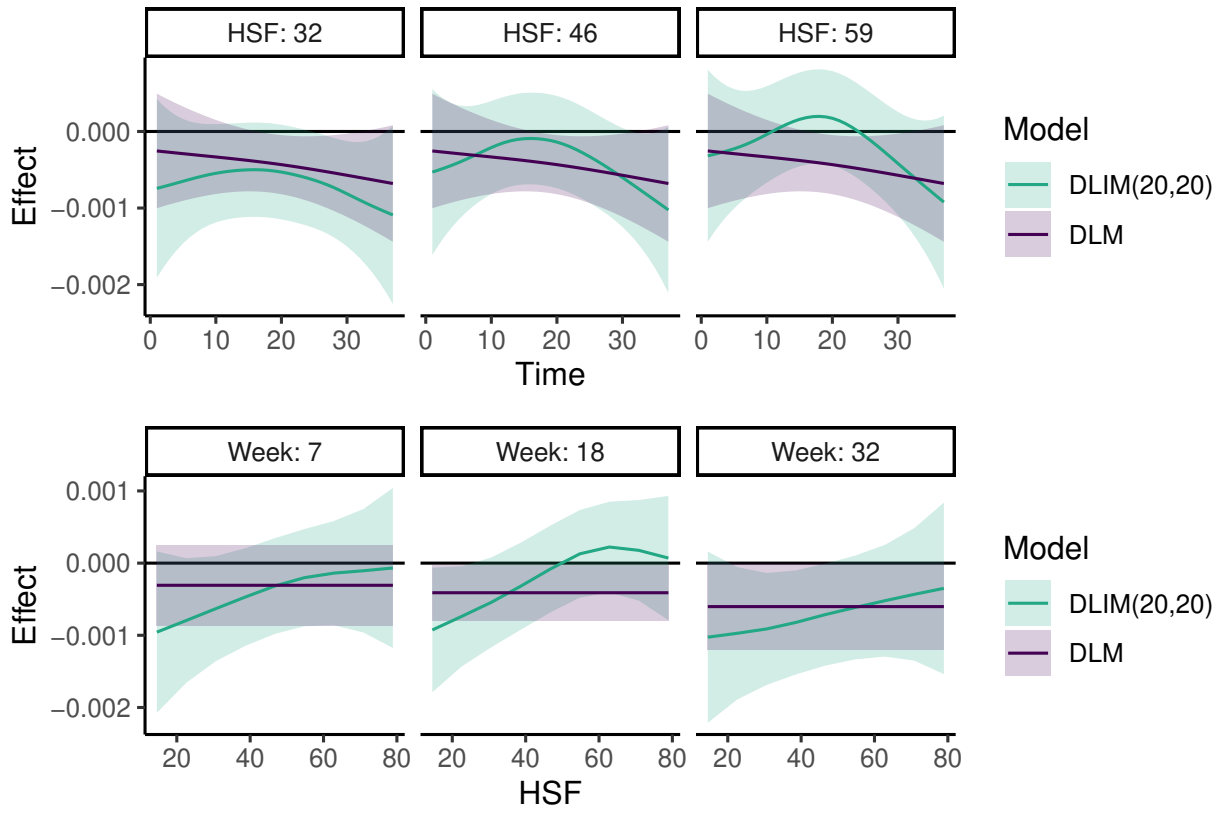


Figure A.1: Pointwise effect estimates of ambient $PM_{2.5}$ on BWGAZ from a standard DLM with 10 exposure-time degrees of freedom and a DLIM(20,20) with Colorado EnviroScreen health and social factors (HSF) score as a modifier. The top panel shows estimated exposure-time-response functions for each quartile of the HSF scores: scores 32, 46, and 59. The bottom panel shows estimated point-wise effects for a week around the middle of each trimester, excluding the first and last 1% of HSF scores. We include confidence bands of 95% around all estimates.

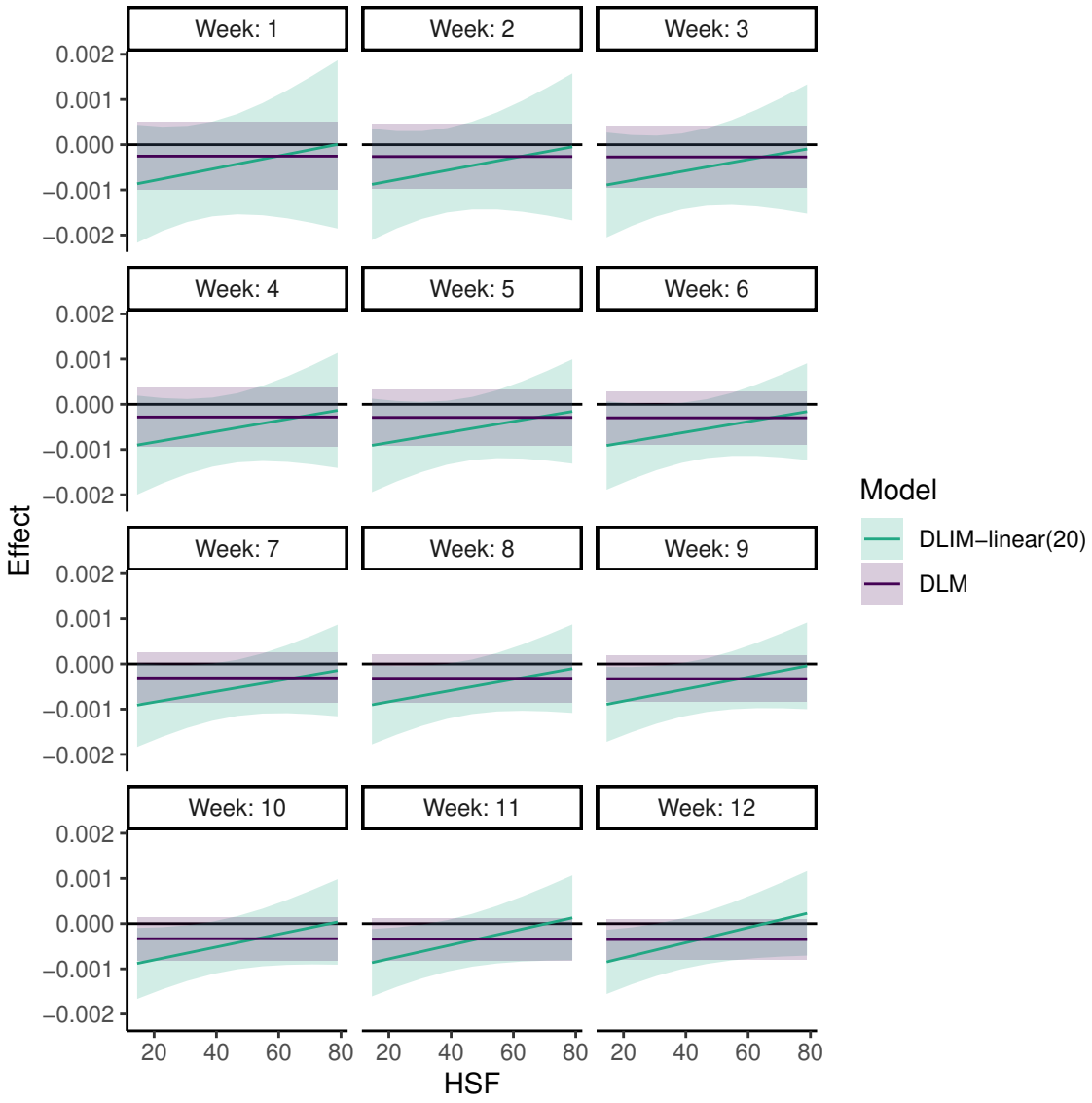


Figure A.2: Estimated exposure-time-response functions for each week of first trimester using both a DLIM-linear with 20 nominal degrees of freedom and standard DLM.

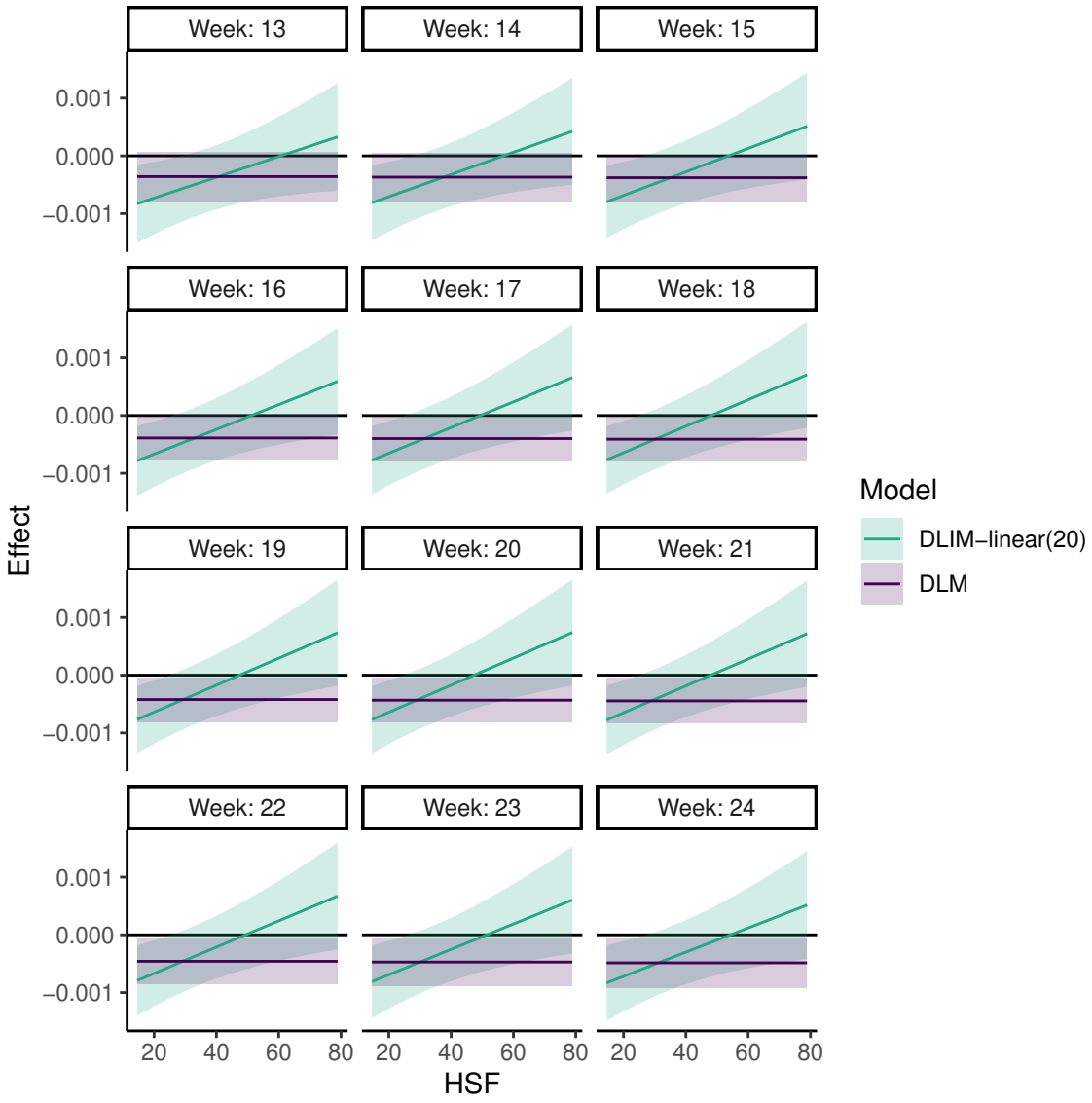


Figure A.3: Estimated exposure-time-response functions for each week of second trimester using both a DLIM-linear with 20 nominal degrees of freedom and standard DLM.

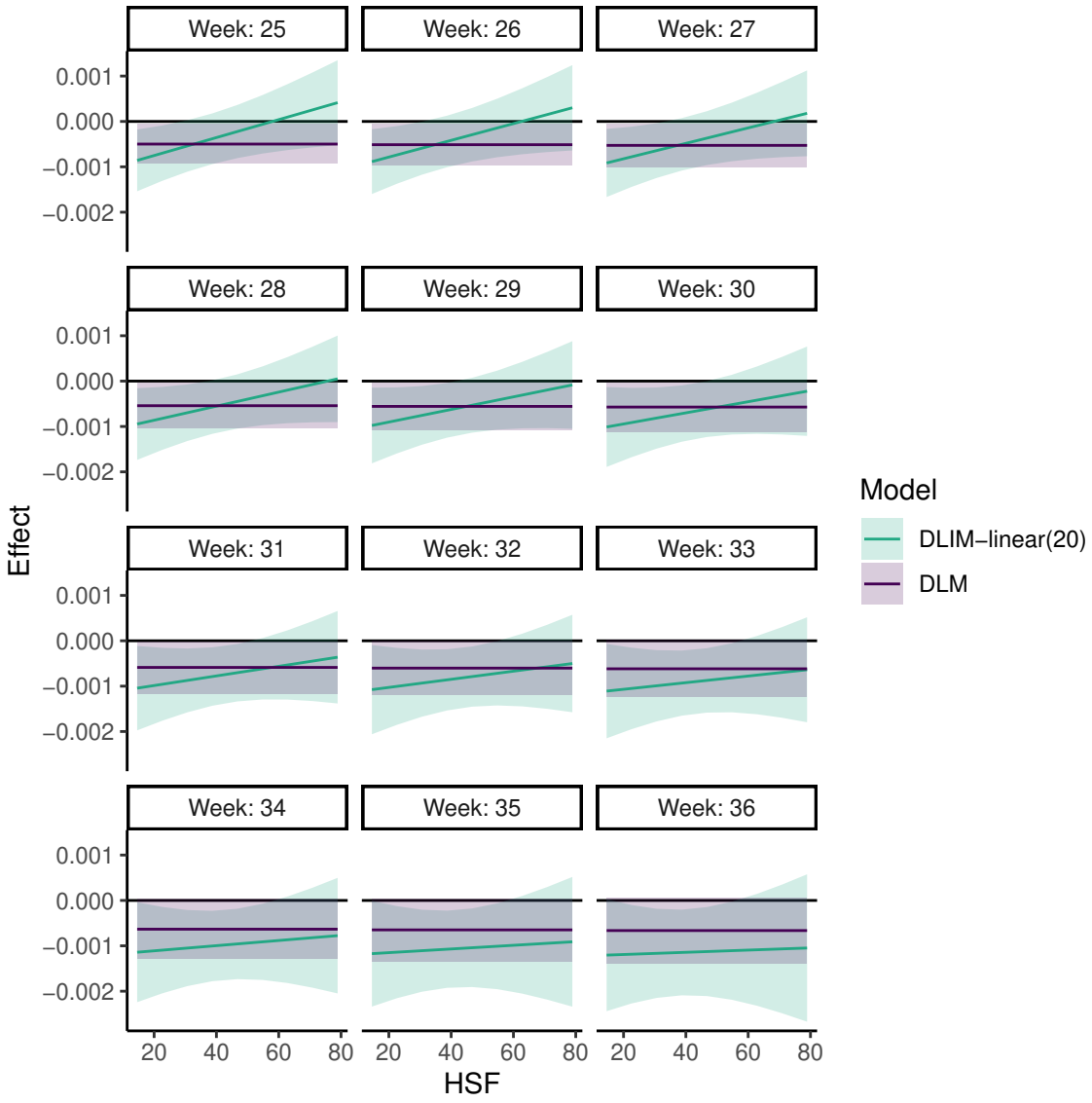


Figure A.4: Estimated exposure-time-response functions for each week of third trimester using both a DLIM-linear with 20 nominal degrees of freedom and standard DLM.

Table A.9: Variable descriptions for the outcome, birth weight gestational age z-score (BWGAZ), and covariates used in the data analysis of this paper.

Covariate	Type	Values
BWGAZ	continuous	-8.29 - 5.95
Age	continuous	11 - 68 yrs
BMI	continuous	10.48 - 83.08 kg/m^2
Height	continuous	38 - 83 in.
Prior weight	continuous	65 - 468 lbs
Race	categorical	Black
		White
		Asian Pacific Islander
		American Indigenous
Ethnicity	categorical	Non-hispanic
		Hispanic
Education	categorical	less than high school
		high school
		advanced
		associate
		college
Income	categorical	<$15,000
		$15,000 - $24,000
		$25,000 - $34,000
		$35,000 - $49,000
		$50,000 - $74,000
		>$75,000
Marital status	categorical	not married
		currently married
		divorced, married/separated, widowed
Self-reported	categorical	yes
prenatal care habits		no
		unknown
Smoking habits	categorical	never smoked
during pregnancy		currently, low amount
		currently, high amount
		former smoker
First trimester avg. ambient temp.	continuous	1.69°C - 35.58°C
Second trimester avg. ambient temp.	continuous	2.15°C - 35.45°C
Third trimester avg. ambient temp.	continuous	1.70°C - 35.66°C
Census-tract of residence elevation	continuous	4577 - 5994 ft.

Appendix B

Distributed Lag Interaction Model with Index Modification

B.1 Additional Methods

We use a Gibbs sampler for the regression coefficients and the error variance. The full conditional distribution for the regression coefficients is

$$\mathbf{\Psi}|\boldsymbol{\rho}, \sigma^2, \mathbf{y} \sim \text{MVN}[\sigma^{-2}\mathbf{V}_\Psi \mathbf{U}'(\boldsymbol{\rho})\mathbf{y}, \mathbf{V}_\Psi], \tag{B.1}$$

where $\mathbf{V}_\Psi = [\mathbf{\Sigma}^{-1} + \sigma^{-2}\mathbf{U}'(\boldsymbol{\rho})\mathbf{U}(\boldsymbol{\rho})]^{-1}$ and Σ is the covariance matrix for the prior distribution on $\mathbf{\Psi}$. The prior on σ^2 is IG(a, b), and the full conditional distribution for σ^2 is

$$\sigma^2|\mathbf{\Psi}, \boldsymbol{\rho}, \mathbf{y} \sim \text{IG}\left[a + \frac{n}{2}, b + \frac{1}{2}\{\mathbf{y} - \mathbf{U}(\boldsymbol{\rho})\mathbf{\Psi}\}'\{\mathbf{y} - \mathbf{U}(\boldsymbol{\rho})\mathbf{\Psi}\}\right]. \tag{B.2}$$

Sampling of other parameters is described in the “Parameter Estimation" section of the main text.

Data: $\mathbf{y}, \mathbf{X}, \mathbf{M}, \mathbf{Z}$

Input: ν_{mod}, ν_{time}, iterations, burn-in, and hyper-parameters ξ^2, τ^2, a, b, $q_1, \ldots, q_L$, and $\nu_1, \ldots, \nu_L$

Initialize parameters $\mathbf{\Psi}, \mathbf{a}, \sigma^2$ using priors

Create exposure-time basis using ν_{time} basis functions

Create initial modifier basis using $m* = 1/L$ and ν_{mod} basis functions

Create initial $\mathbf{U}(\mathbf{a})$ using initial $\mathbf{a}$

for $s = 2, \ldots, N$ **do**

Update $\mathbf{\Psi}$ using Gibbs sampler

Update σ^2 using Gibbs sampler

for $l = 1, \ldots, L$ **do**

Update a_l using algorithm 2

end

Normalize $\mathbf{a}^{(s)}$ to obtain $\boldsymbol{\rho}^{(s)}$

Update cross-basis columns of $\mathbf{U}(\boldsymbol{\rho})$ using $\boldsymbol{\rho}^{(s)}$

end

Result: Posterior samples for $\mathbf{\Psi}, \boldsymbol{\rho}, \sigma^2$

Algorithm 1: MCMC sampler

if $s > 500$ *and* $s \leq$ *burn-in* **then**
| $\delta_l \leftarrow \delta_l$[1-(0.35 - acceptance rate)]

end

$a_l^{(s)} \leftarrow a_l^{(s-1)}$

if *modifier selection* **then**

| **if** $a_l^{(s)} = 0$ **then**
| | $a_l^* \leftarrow \text{Gamma}(q_l, 1)$
| **else**
| | $a_l^* \leftarrow 0$
| **end**
| $a_l^{(s)} \leftarrow a_l^*$ with probability $\min(r_{\text{select}}, 1)$ in (5) in the main text

end

if $a_l^* = 0$ *was just rejected OR no modifier selection* **then**

| $a_l^* \leftarrow \text{FN}(a_l^{(s)}, \delta_l)$
| $a_l^{(s)} \leftarrow a_l^*$ with probability $\min(r, 1)$ in (4) in the main text

end

Result: $a_l^{(s)}$

Algorithm 2: MH step for a_l

B.2 Additional Simulation Table and Figures

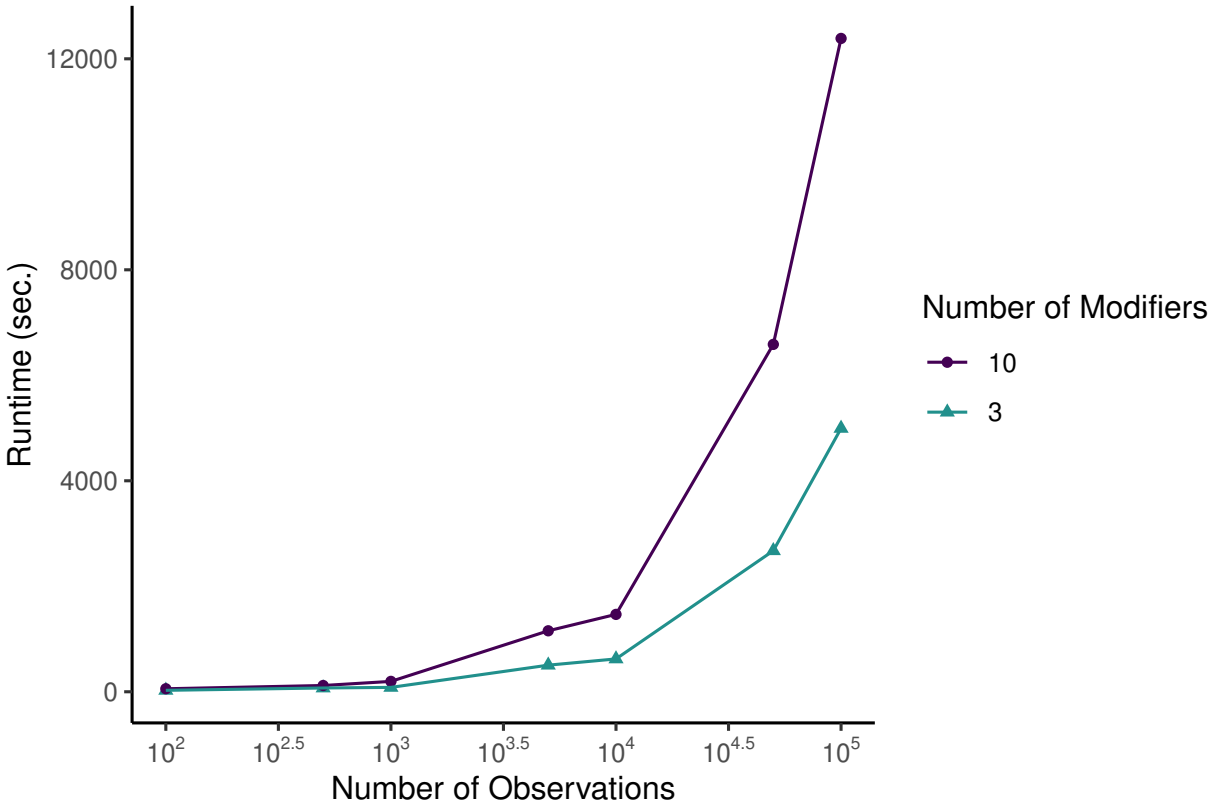


Figure B.1: Runtime averaged across 5 data sets for different number of modifiers. Data sets were simulated using the same generating mechanisms described in the simulation section of the main text.

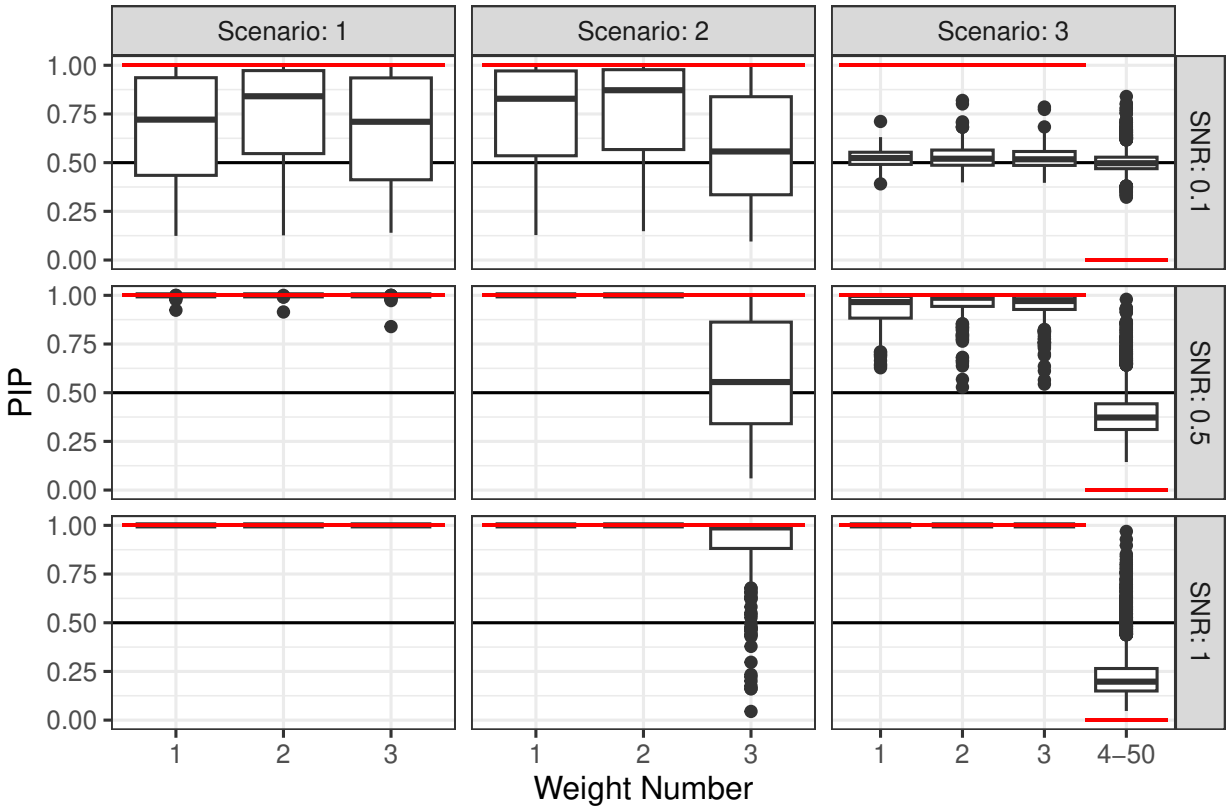


Figure B.2: Posterior inclusion probabilities for each weight for DLIM-IM with modifier selection for all 200 simulated data sets with for all weight scenarios and signal-to-noise (SNR) settings. The red lines are fixed at 1 for non-zero weights and 0 for zero weights. The black line is fixed at the inclusion threshold 0.5.

Table B.1: Extending Table 3.1 in the main text to include other weight scenarios. The different 10 scenario is the following weights: 0.3, 0.2, 0.1, 0.1, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05. The equal 10 scenario is 10 weights that are all 0.1. The sparse 10 scenario is the following weights: 0.5, 0.4, 0.1, 0, 0, 0, 0, 0, 0, 0.

			Cumulative			Point-wise		
Weight	SNR	Model	RMSE	Coverage	Width	RMSE	Coverage	Width
different 10	0.1	DLIM-IM	1.06	0.97	5.24	0.25	0.95	0.94
different 10	0.1	DLIM-IM sel.	1.05	0.97	5.18	0.24	0.96	0.96
different 10	0.1	Fixed-index	1.82	0.95	7.51	0.12	0.92	0.43
different 10	0.5	DLIM-IM	0.31	0.97	1.41	0.06	0.89	0.20
different 10	0.5	DLIM-IM sel.	0.30	0.97	1.38	0.06	0.89	0.20
different 10	0.5	Fixed-index	0.38	0.95	1.57	0.04	0.71	0.11
different 10	1	DLIM-IM	0.19	0.95	0.78	0.04	0.78	0.10
different 10	1	DLIM-IM sel.	0.19	0.95	0.78	0.04	0.78	0.10
different 10	1	Fixed-index	0.21	0.95	0.84	0.03	0.81	0.08
equal 10	0.1	DLIM-IM	1.06	0.97	5.21	0.25	0.95	0.93
equal 10	0.1	DLIM-IM sel.	1.04	0.97	5.15	0.24	0.96	0.95
equal 10	0.1	Fixed-index	1.80	0.95	7.46	0.12	0.92	0.43
equal 10	0.5	DLIM-IM	0.32	0.97	1.43	0.06	0.89	0.20
equal 10	0.5	DLIM-IM sel.	0.31	0.97	1.41	0.06	0.89	0.20
equal 10	0.5	Fixed-index	0.38	0.95	1.55	0.04	0.73	0.11
equal 10	1	DLIM-IM	0.20	0.95	0.79	0.04	0.78	0.10
equal 10	1	DLIM-IM sel.	0.20	0.95	0.79	0.04	0.78	0.10
dqual 10	1	Fixed-index	0.20	0.95	0.81	0.03	0.88	0.09
sparse 10	0.1	DLIM-IM	1.08	0.97	5.32	0.25	0.95	0.95
sparse 10	0.1	DLIM-IM sel.	1.06	0.97	5.27	0.24	0.96	0.97
sparse 10	0.1	Fixed-index	1.85	0.95	7.63	0.12	0.92	0.44
sparse 10	0.5	DLIM-IM	0.29	0.97	1.36	0.06	0.89	0.19
sparse 10	0.5	DLIM-IM sel.	0.28	0.97	1.30	0.06	0.89	0.19
sparse 10	0.5	Fixed-index	0.40	0.94	1.62	0.05	0.68	0.10
sparse 10	1	DLIM-IM	0.18	0.95	0.75	0.04	0.78	0.10
sparse 10	1	DLIM-IM sel.	0.17	0.96	0.74	0.04	0.77	0.10
sparse 10	1	Fixed-index	0.24	0.93	0.91	0.04	0.63	0.07

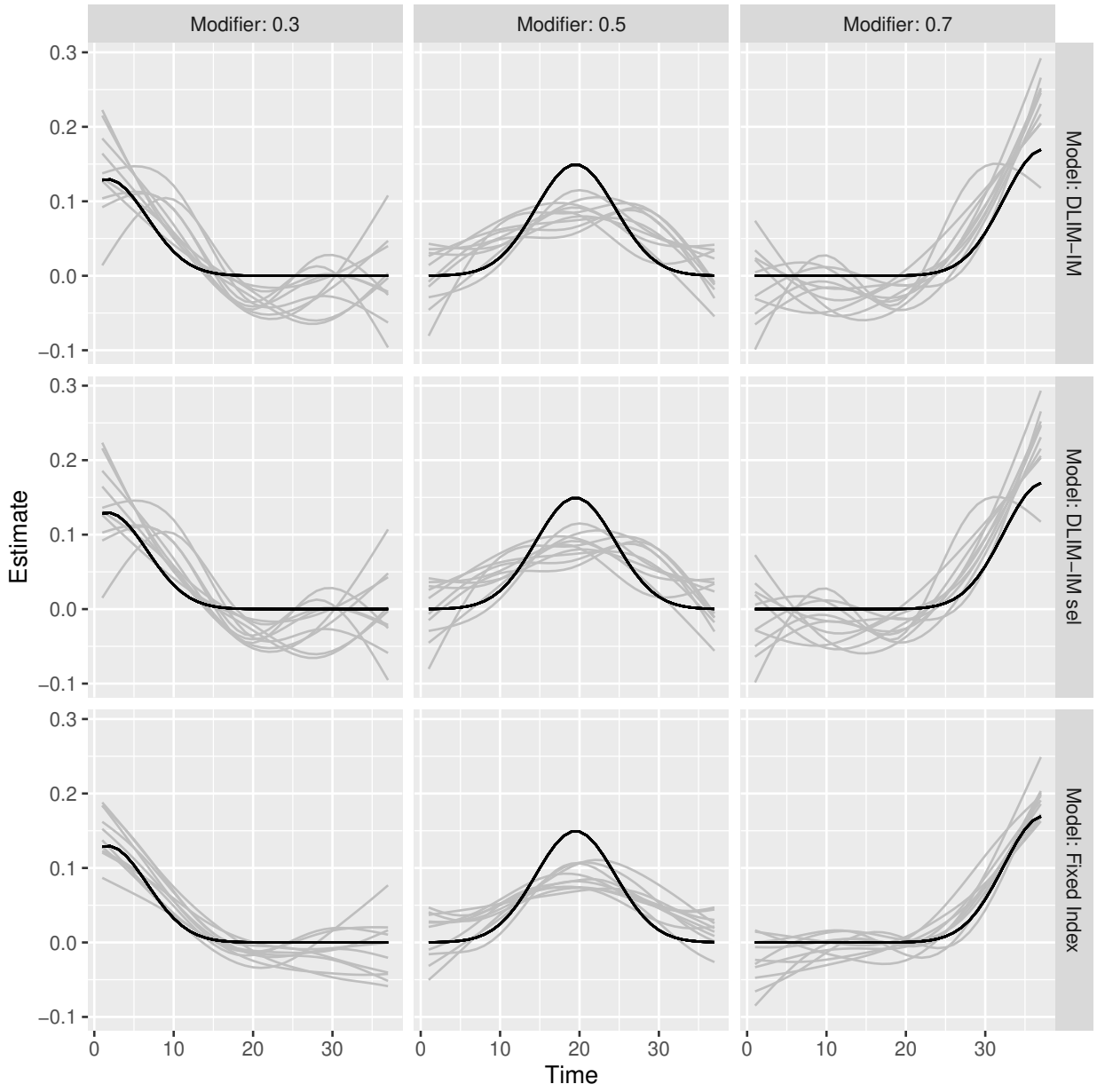


Figure B.3: Estimated exposure-time-response functions (gray lines) from the three models considered in the main text (DLIM-IM, DLIM-IM with selection, fixed index model) at 3 modifier index values (0.3, 0.5, 0.7) for 10 individual simulated data sets. The true exposure-time-response function is presented in black. The data sets were simulated using a signal-to-noise ratio of 1 in the equal weights scenario (scenario 1).

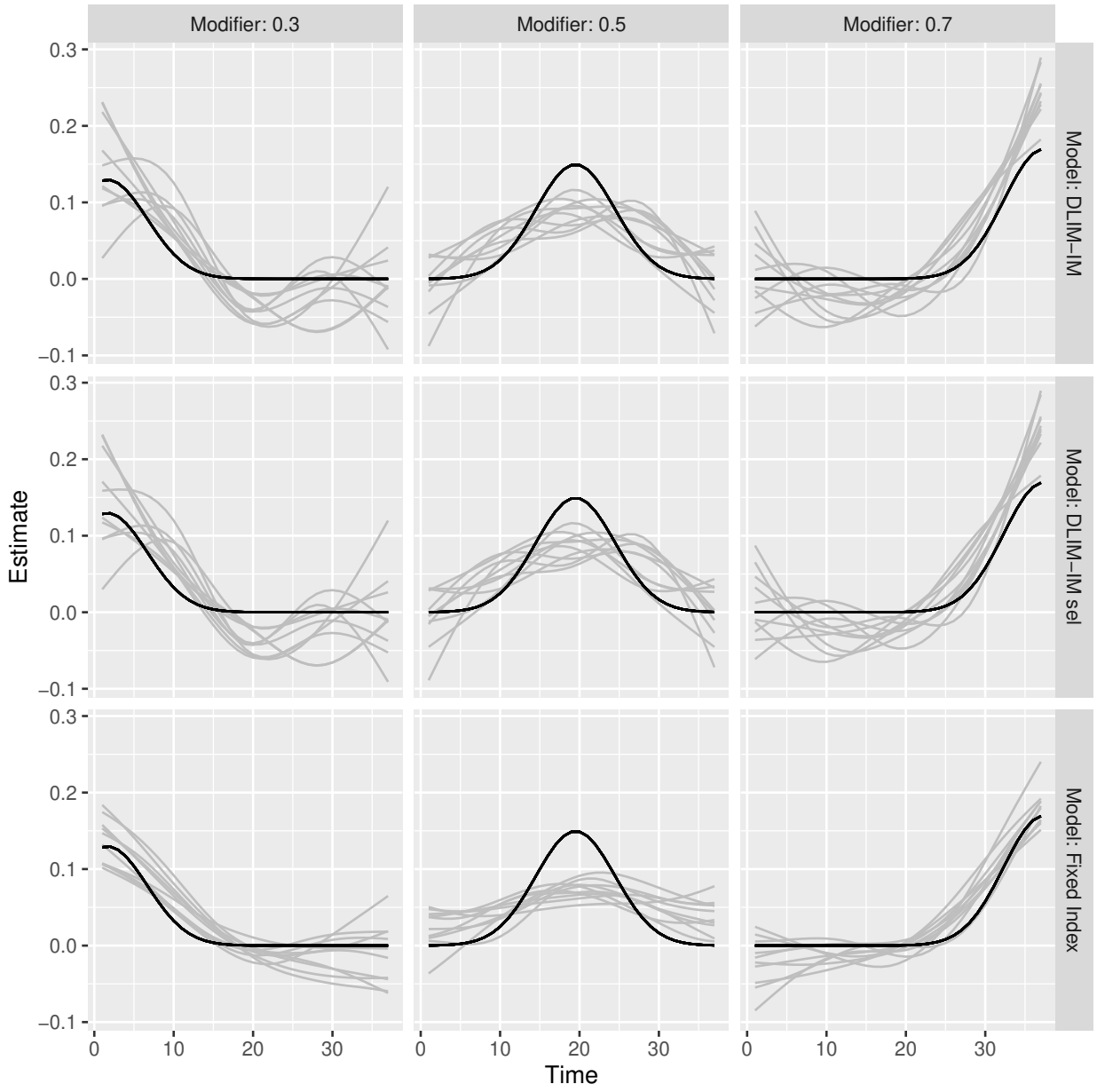


Figure B.4: Estimated exposure-time-response functions (gray lines) from the three models considered in the main text (DLIM-IM, DLIM-IM with selection, fixed index model) at 3 modifier index values (0.3, 0.5, 0.7) for 10 individual simulated data sets. The true exposure-time-response function is presented in black. The data sets were simulated using a signal-to-noise ratio of 1 in the different weights scenario (scenario 2).

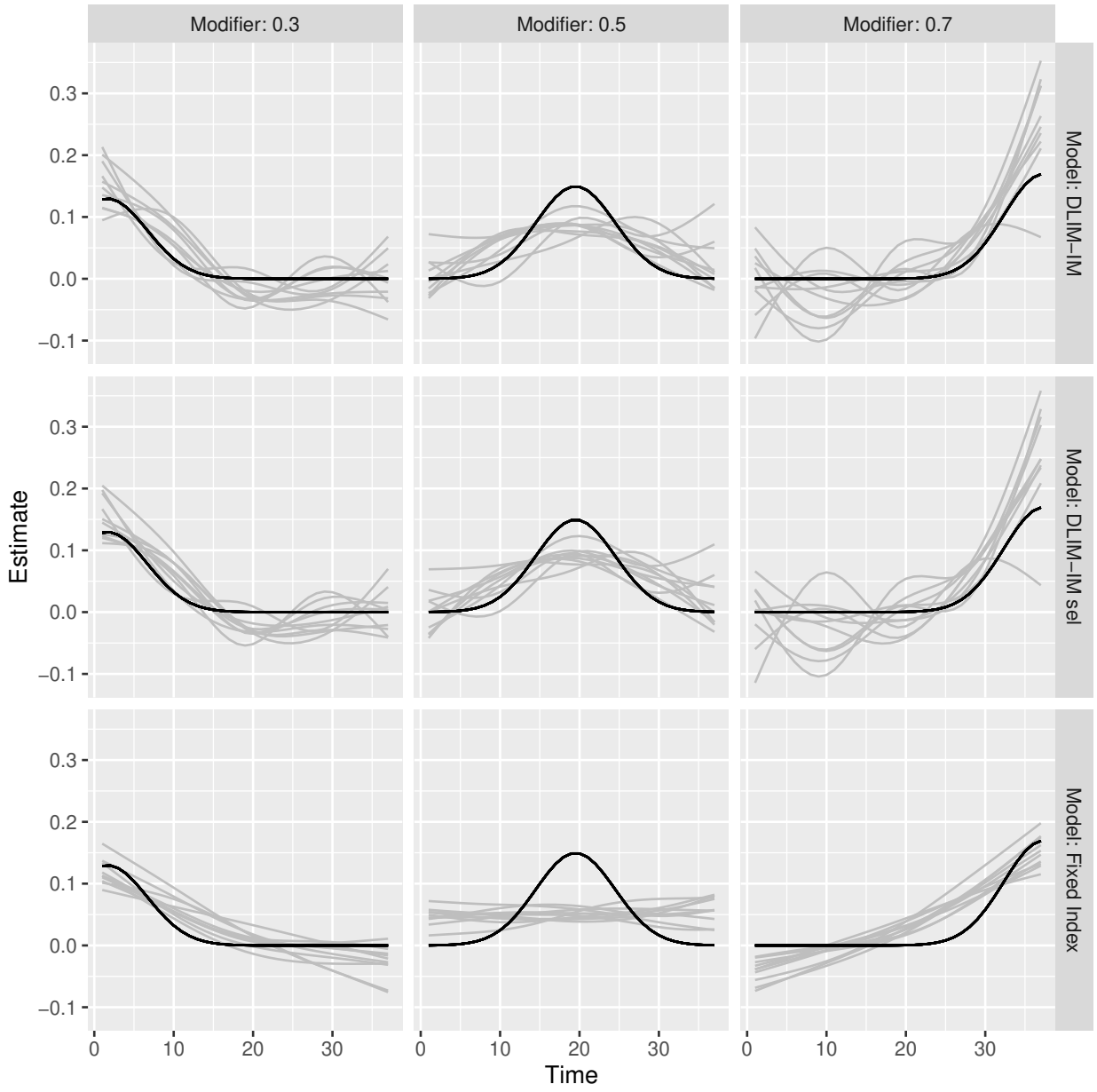


Figure B.5: Estimated exposure-time-response functions (gray lines) from the three models considered in the main text (DLIM-IM, DLIM-IM with selection, fixed index model) at 3 modifier index values (0.3, 0.5, 0.7) for 10 individual simulated data sets. The true exposure-time-response function is presented in black. The data sets were simulated using a signal-to-noise ratio of 1 in the sparse weights scenario (scenario 3).

B.3 Additional Analysis Table and Figures

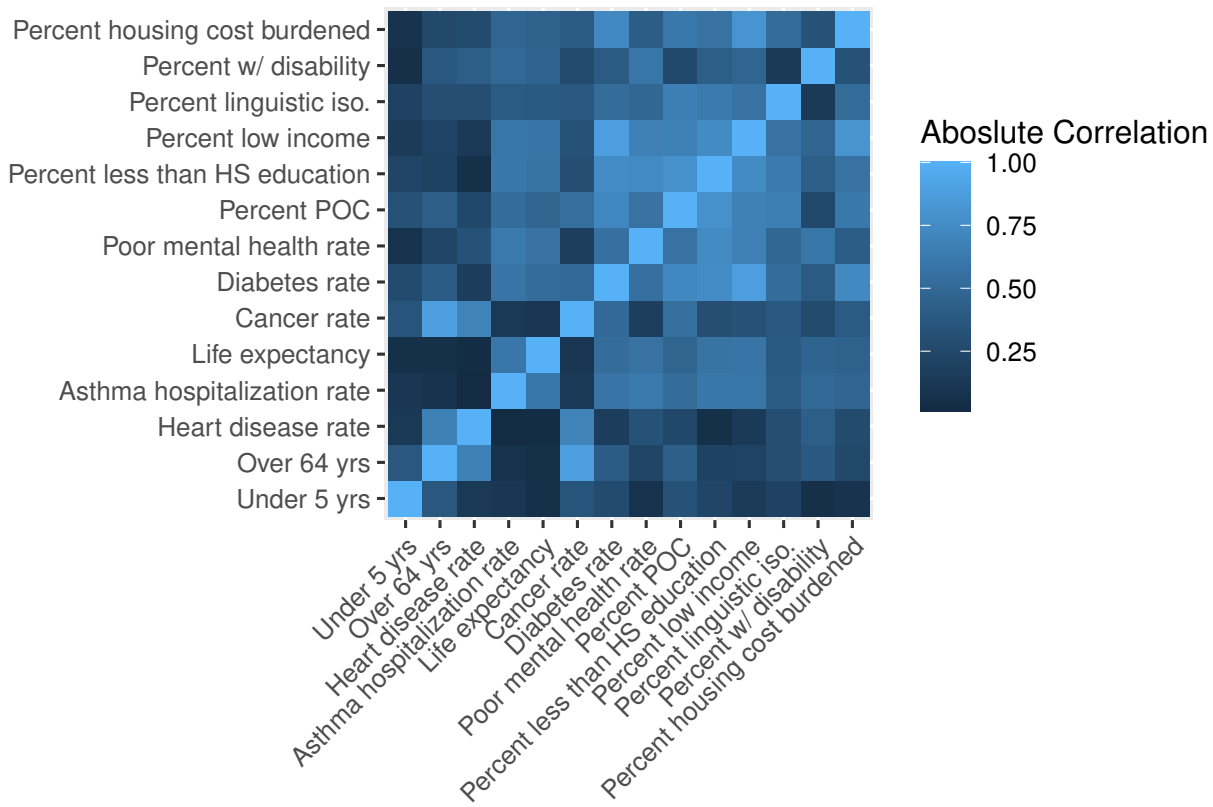


Figure B.6: Heat map of the absolute correlations for the 14 candidate modifiers used in the Colorado EnviroScreen analysis.

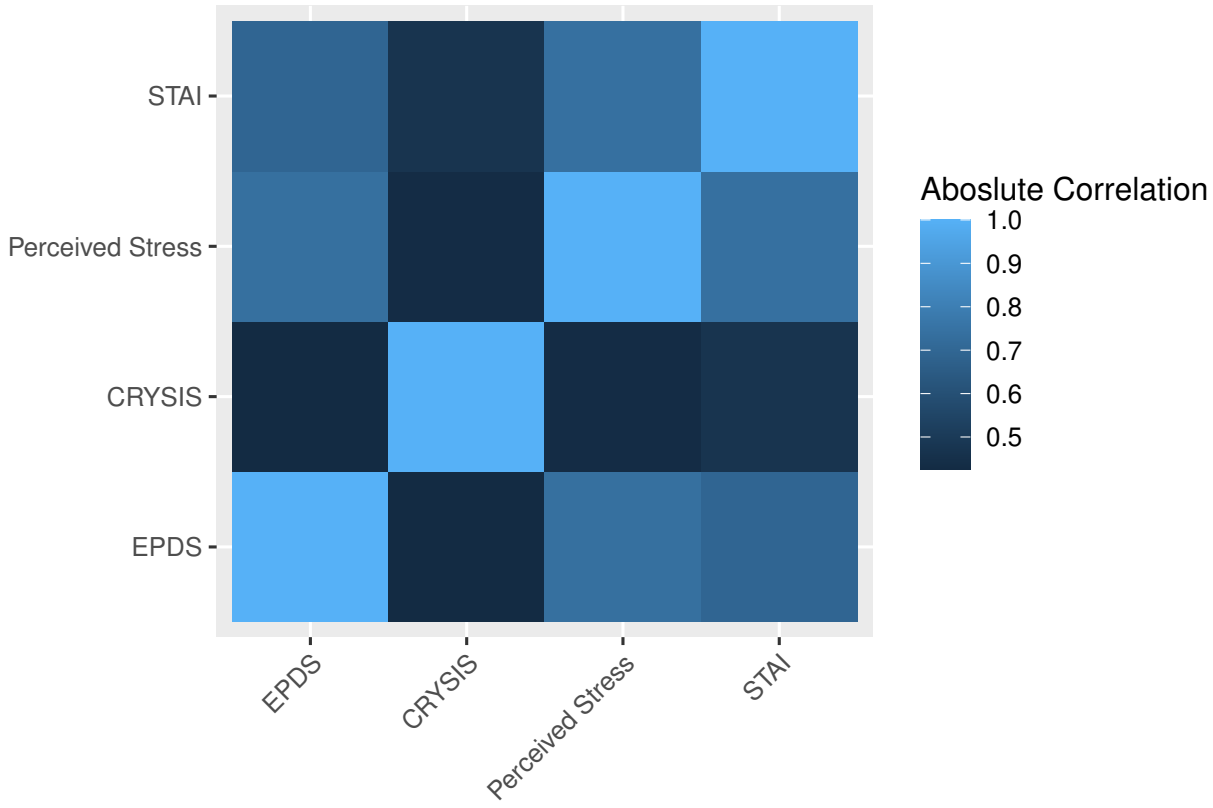


Figure B.7: Heat map of the absolute correlations for the 4 stress modifiers used in the PROGRESS analysis.

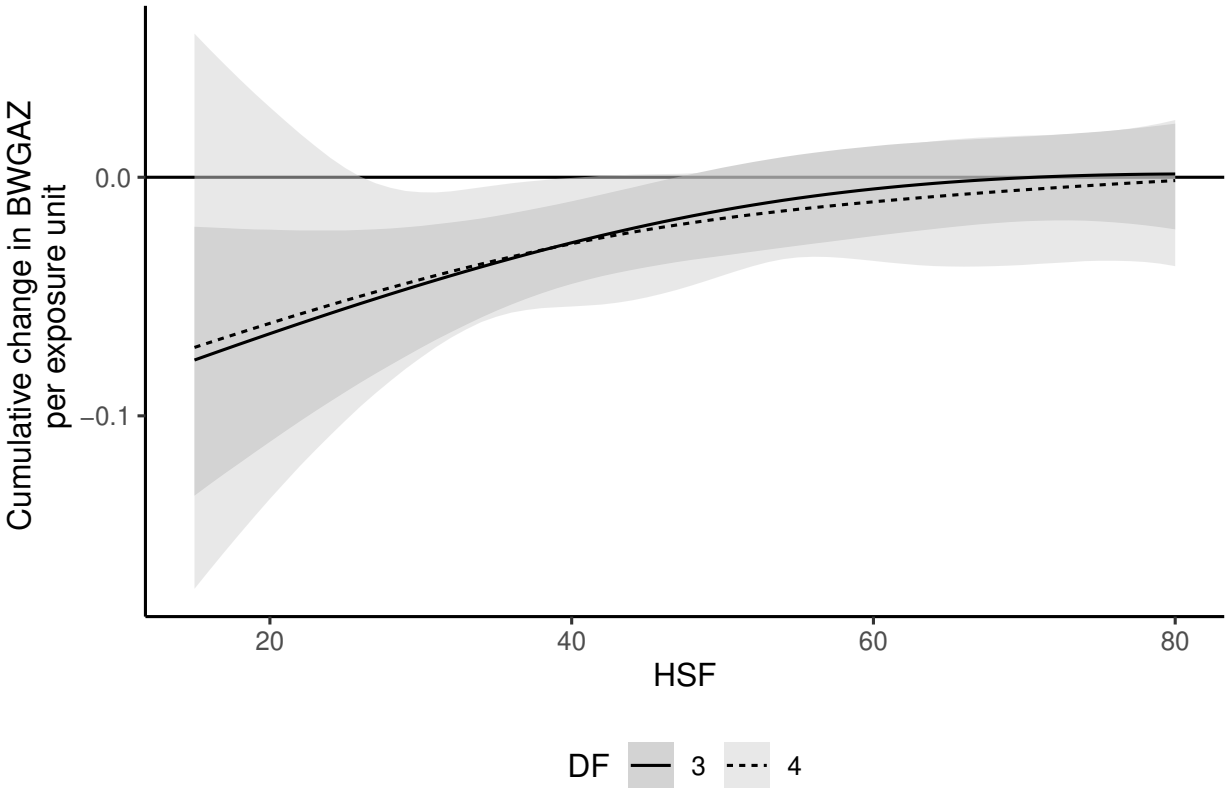


Figure B.8: Estimated cumulative effect from DLIM-IMs with selection using the Colorado birth cohort data set with 3 and 4 degrees of freedom in both the modifier and exposure-time dimensions.

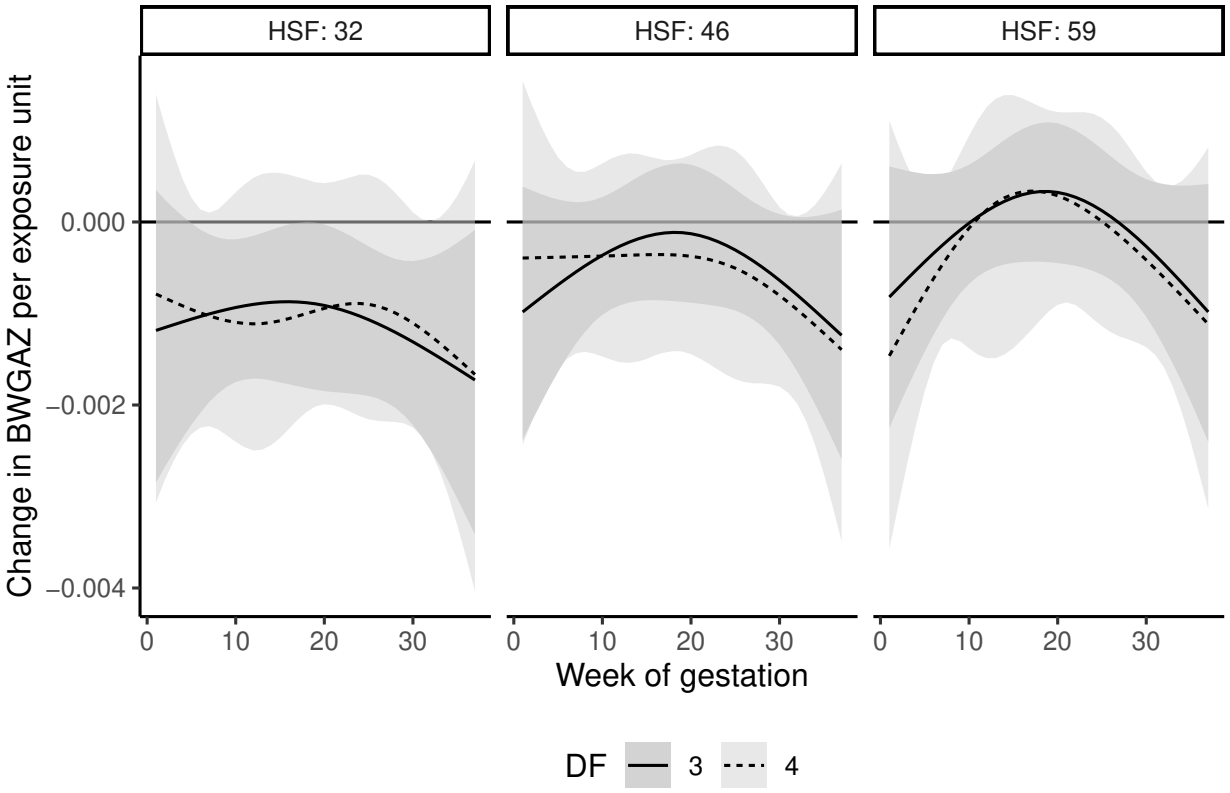
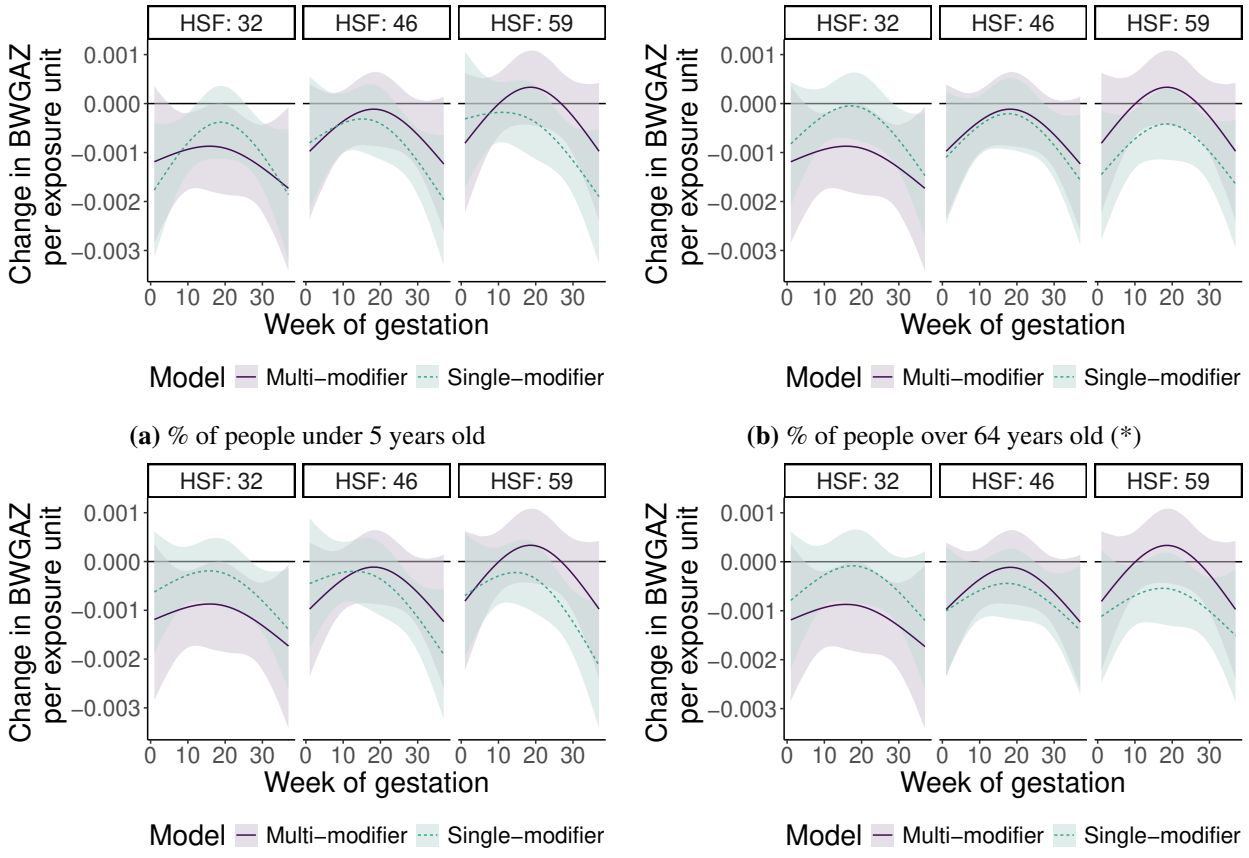


Figure B.9: Estimated exposure-time-response curves for DLIM-IMs with selection using the Colorado birth cohort data set with 3 and 4 degrees of freedom in both the modifier and exposure-time dimensions.



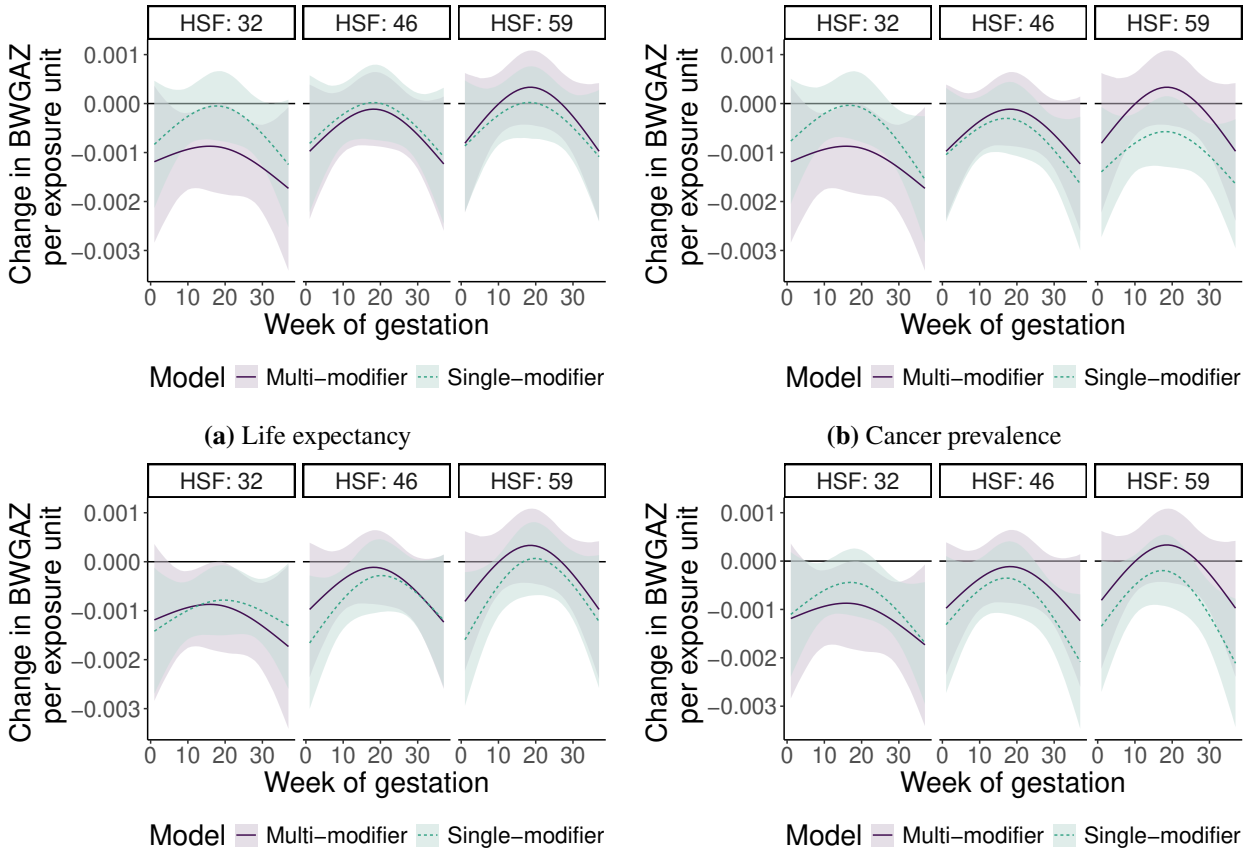
(a) % of people under 5 years old

(b) % of people over 64 years old (*)

(c) Rate of heart disease

(d) Asthma hospitalization rate (*)

Figure B.10: Estimated exposure-time-response functions across a grid of modifier index values (HSF) from the DLIM-IM with selection on the 14 modifiers (Multi-modifier model) presented in the main text and a DLIM-IM with one modifier (single-modifier model). The asterisk denotes modifiers that were found to contribute to modification in the multi-modifier model in the main text analysis.



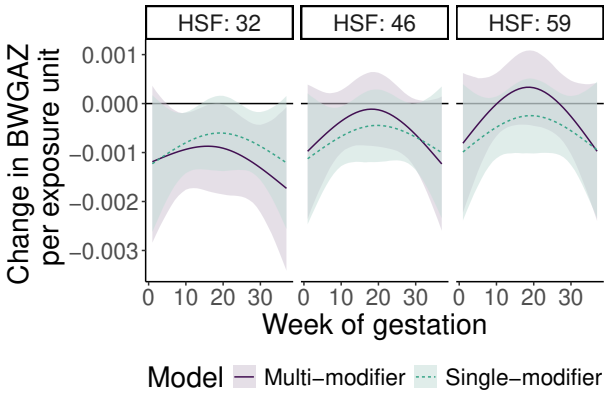
(a) Life expectancy

(b) Cancer prevalence

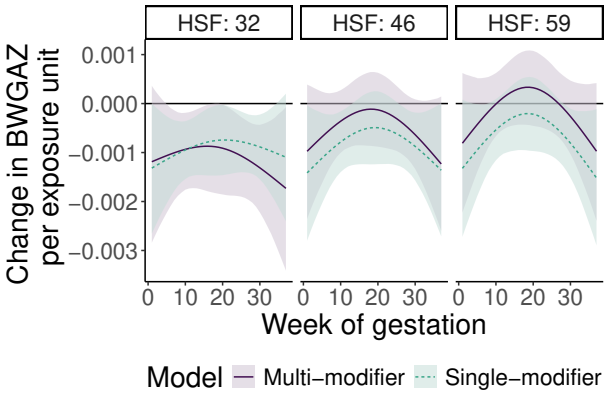
(c) Diabetes prevalence

(d) Rate of poor mental health

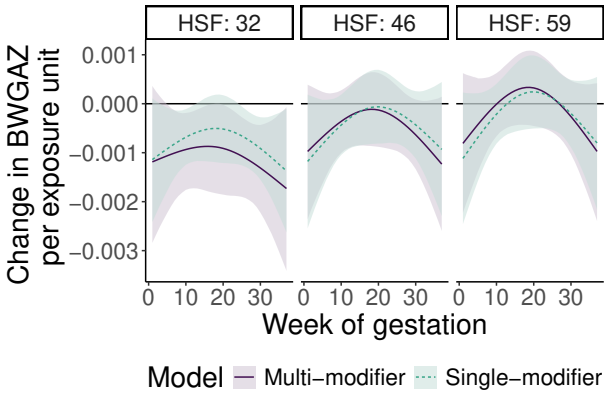
Figure B.11: Estimated exposure-time-response functions across a grid of modifier index values (HSF) from the DLIM-IM with selection on the 14 modifiers (Multi-modifier model) presented in the main text and a DLIM-IM with one modifier (single-modifier model). The asterisk denotes modifiers that were found to contribute to modification in the multi-modifier model in the main text analysis.



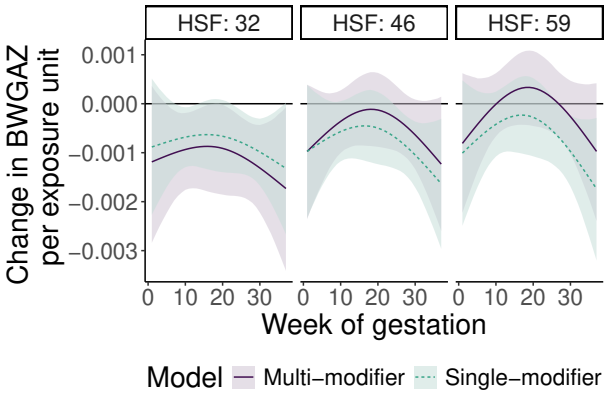
(a) % of people of color (*)



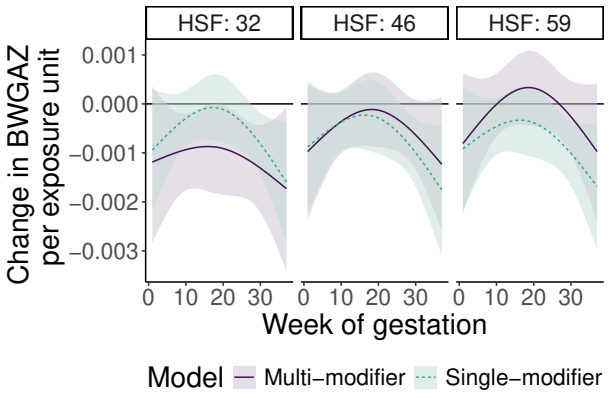
(b) % of people less than high school ed. (*)



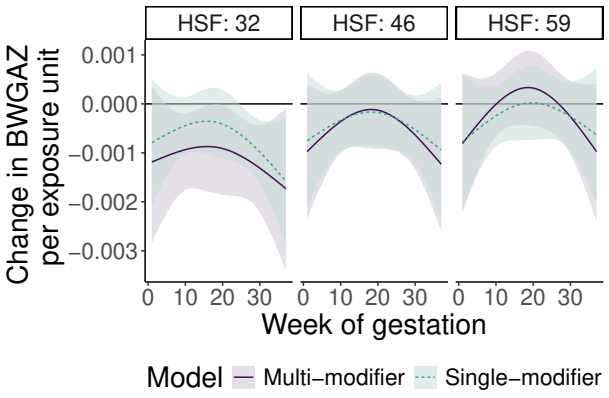
(c) % of people with low income



(d) % of households linguistically isolated (*)



(e) % of people with disabilities



(f) % of households burdened by housing costs

Figure B.12: Estimated exposure-time-response functions across a grid of modifier index values (HSF) from the DLIM-IM with selection on the 14 modifiers (Multi-modifier model) presented in the main text and a DLIM-IM with one modifier (single-modifier model). The asterisk denotes modifiers that were found to contribute to modification in the multi-modifier model in the main text analysis.

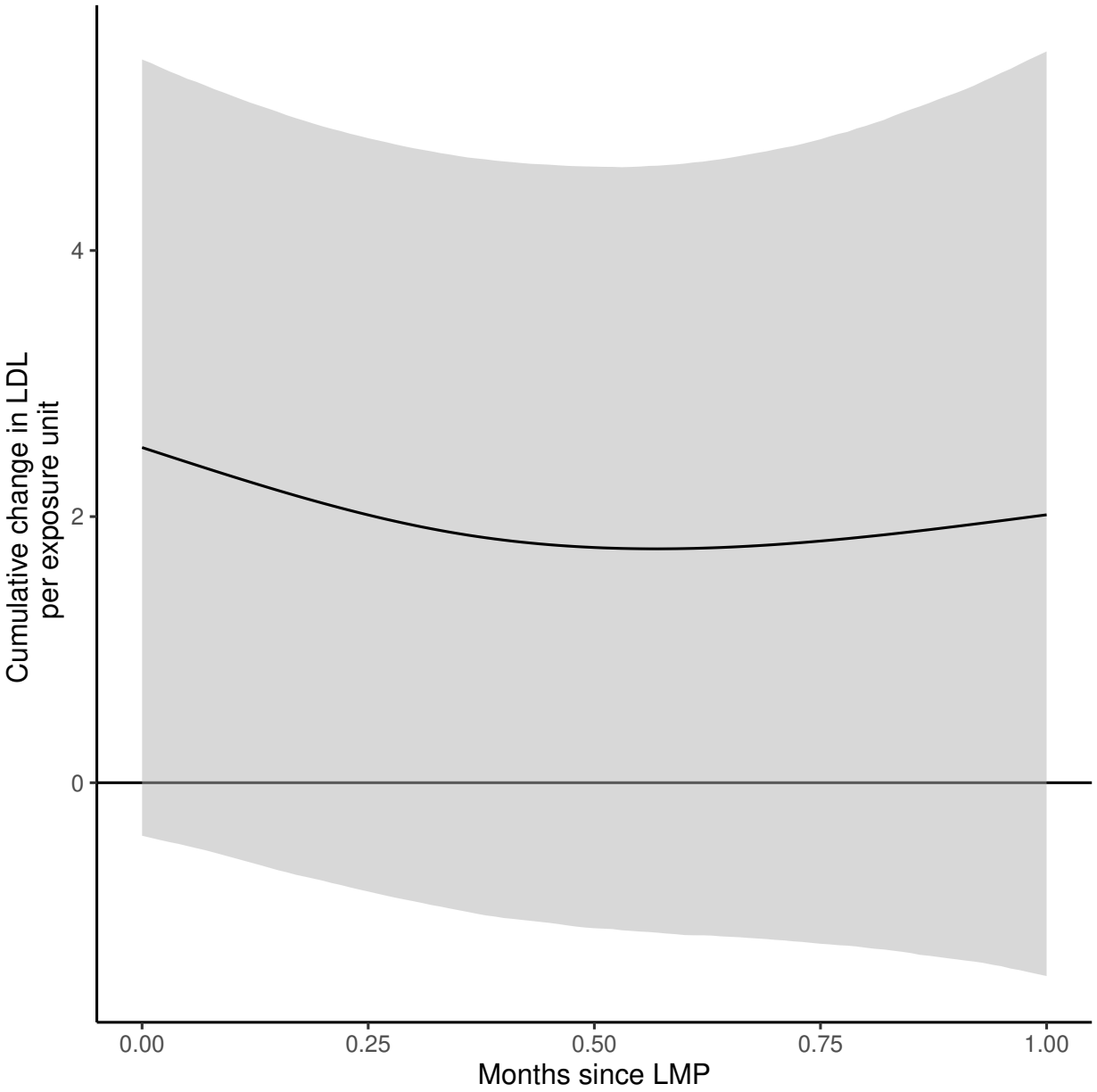


Figure B.13: Cumulative effect estimates for LDL 48 months after parturition from a DLIM-IM without selection using the PROGRESS cohort data set fit with 3 degrees of freedom in both the modifier and exposure-time dimensions.

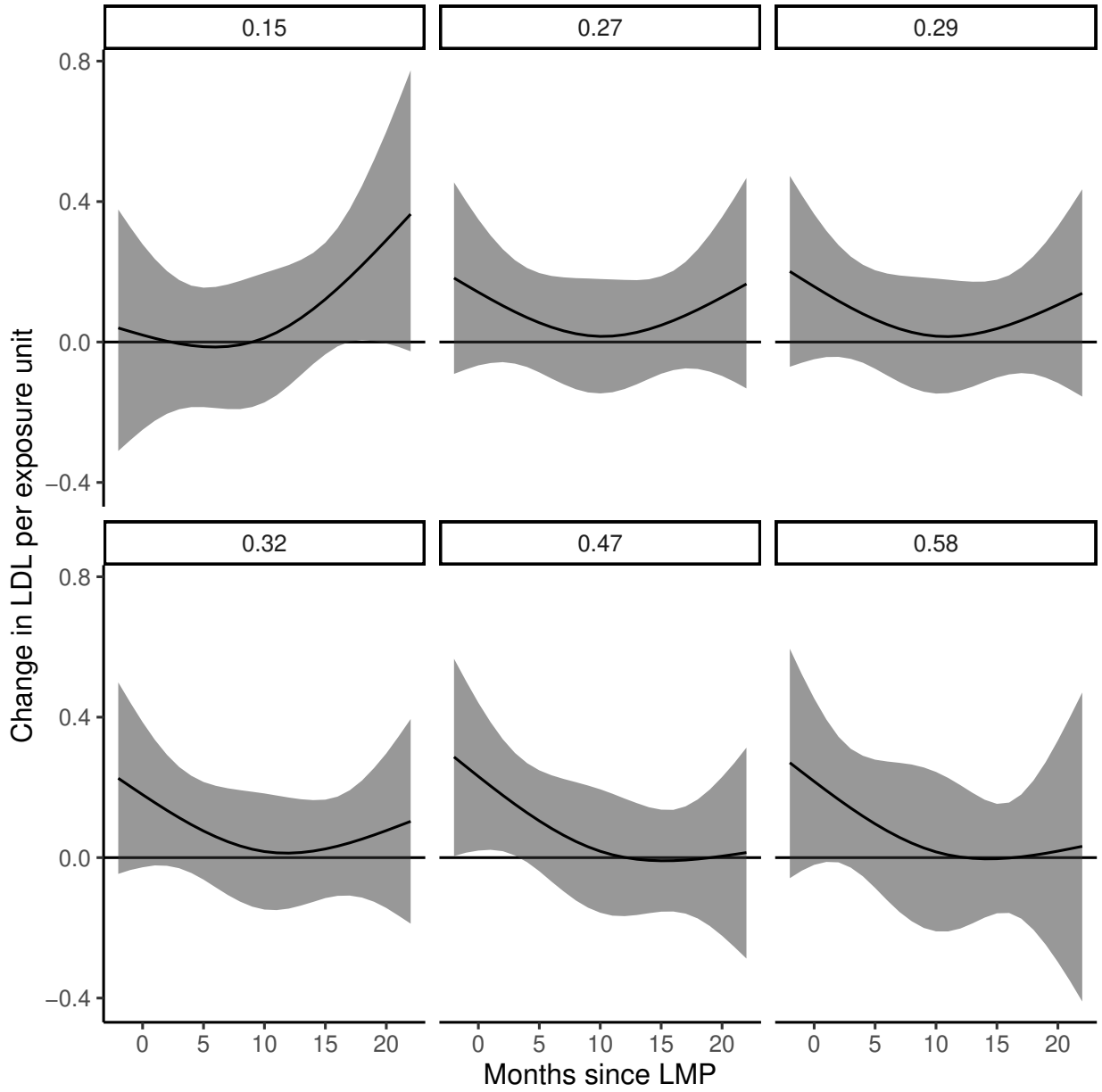


Figure B.14: Estimated exposure-time-response curves for LDL 48 months after parturition from a DLIM-IM without selection using the PROGRESS cohort data set fit with 3 degrees of freedom in both the modifier and exposure-time dimensions, evaluated at the 10^{th} ($\hat{m}^* = 0.15$), 25^{th} ($\hat{m}^* = 0.27$), 30^{th} ($\hat{m}^* = 0.29$), 40^{th} ($\hat{m}^* = 0.32$), 75^{th} ($\hat{m}^* = 0.47$), and 90^{th} ($\hat{m}^* = 0.58$) percentiles of the estimated modifier index.

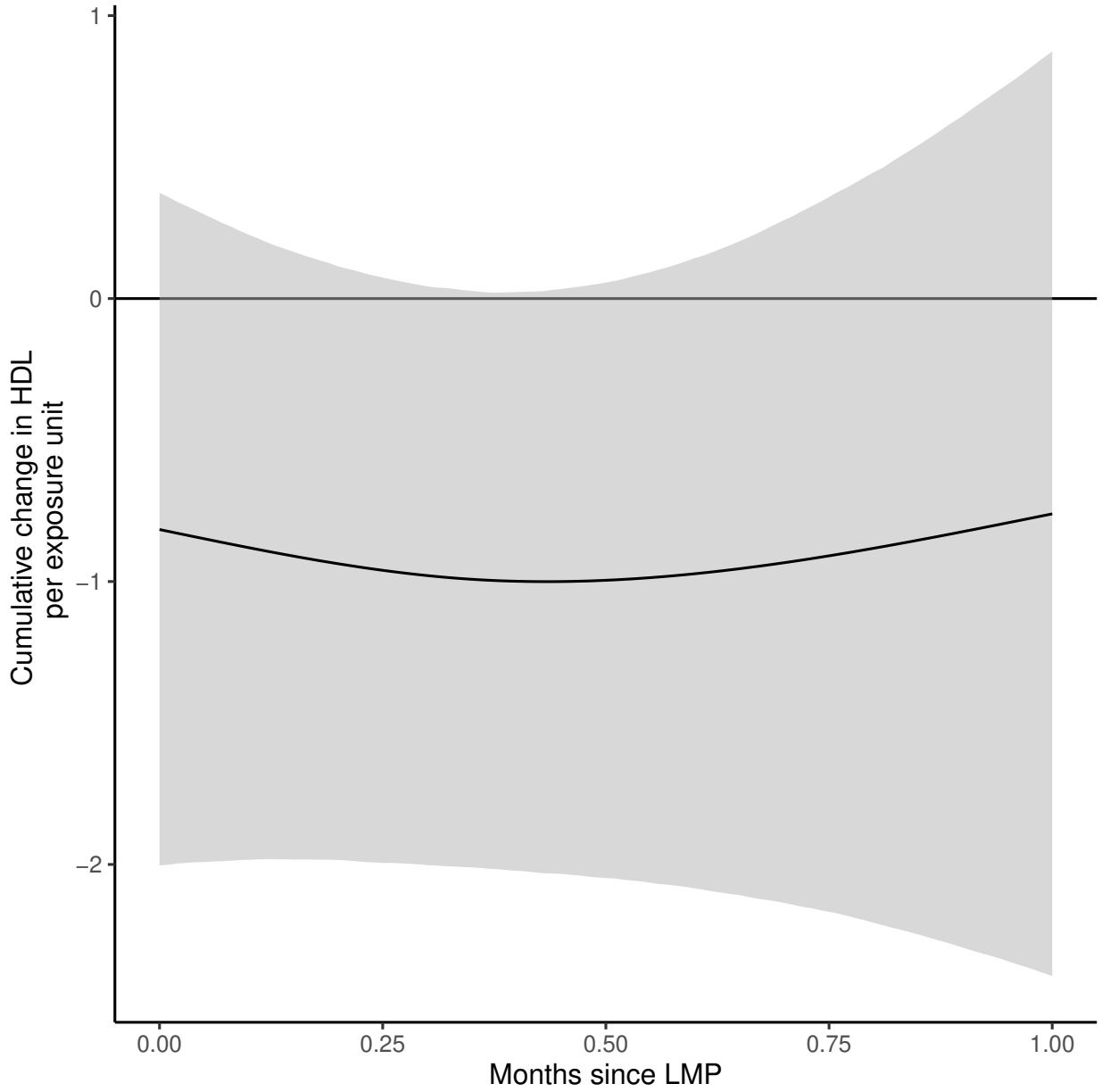
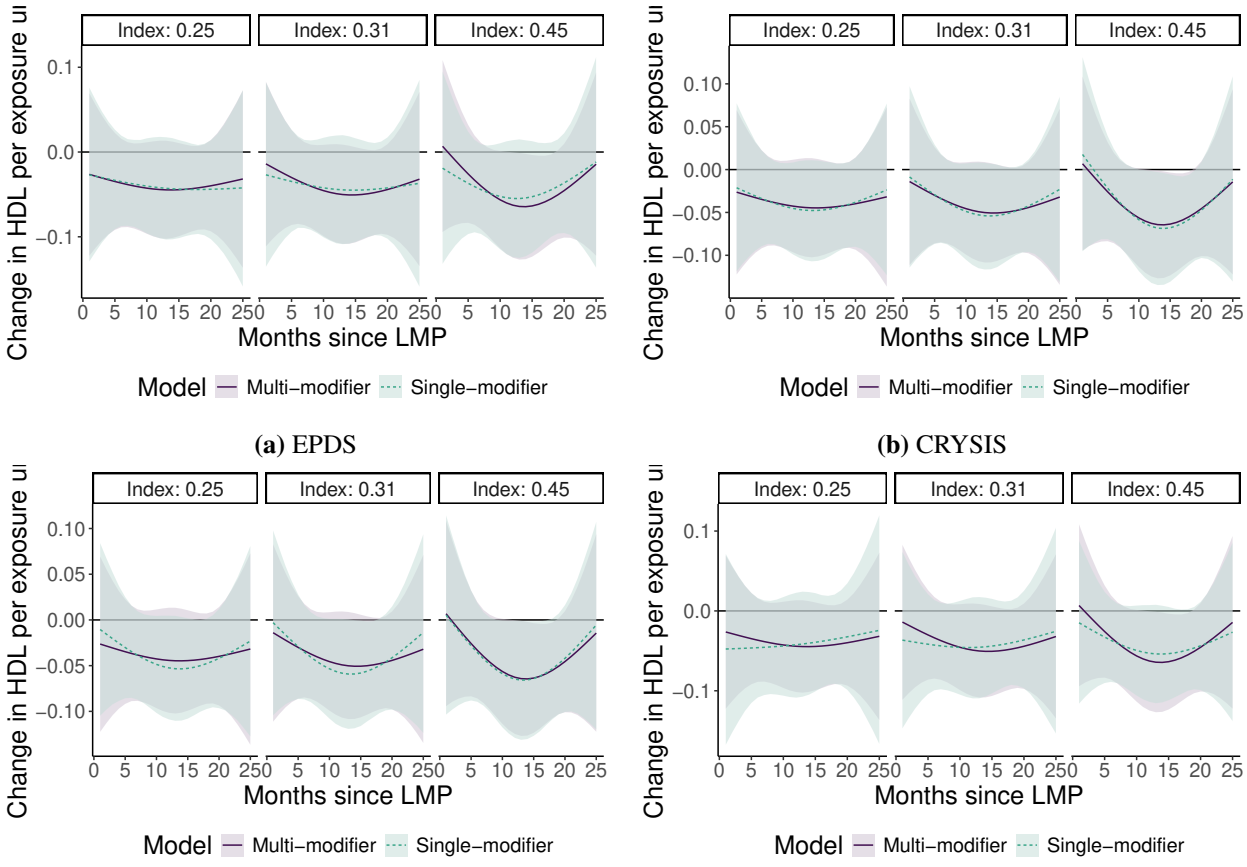


Figure B.15: Cumulative effect estimates for HDL 48 months after parturition from a DLIM-IM without selection using the PROGRESS cohort data set fit with 3 degrees of freedom in both the modifier and exposure-time dimensions.



(a) EPDS

(b) CRYSIS

(c) Perceived Stress Score

(d) STAI

Figure B.16: Estimated exposure-time-response functions from a DLIM-IM using the 4 stress modifiers (multi-modifier model in solid black) presented in the main text and a DLIM-IM with one of the stress modifiers (single-modifier model in dashed blue) from the PROGRESS data set. The estimated exposure-time-response curve is presented at the 25th, 40th, and 75th percentiles of the estimated modifier index from the multi-modifier model presented in the main text.

Table B.2: Weight estimates and posterior inclusion probabilities (PIPs) for a DLIM-IM with modifier selection using ν_{mod} degrees of freedom in the modifier dimension and ν_{time} degrees of freedom in the exposure-time dimension, denoted (ν_{mod}, ν_{time}). The flag (*) indicates that the PIPs disagree in terms of selection, and the flag (**) indicates that the estimated weights are not similar.

	Estimate		SD		PIP	
Indicator	(3, 3)	(4, 4)	(3, 3)	(4, 4)	(3, 3)	(4, 4)
Asthma	0.14	0.14	0.12	0.10	0.74	0.82
Cancer	0.03	0.06	0.06	0.12	0.41	0.45
Diabetes (*)	0.05	0.08	0.07	0.09	0.47	0.61
Heart disease	0.01	0.03	0.05	0.07	0.24	0.35
Life expectancy	0.01	0.02	0.03	0.04	0.25	0.32
Mental health	0.03	0.05	0.06	0.08	0.37	0.46
Percent over 64 (*)	0.10	0.06	0.13	0.09	0.58	0.49
Percent under 5	0.02	0.02	0.05	0.05	0.31	0.34
Housing cost	0.06	0.06	0.10	0.10	0.49	0.49
Disability (*)	0.06	0.11	0.10	0.16	0.46	0.61
Less than high school education (**)	0.19	0.09	0.19	0.13	0.75	0.57
Linguistic isolation	0.07	0.12	0.10	0.12	0.53	0.71
Low income	0.01	0.01	0.03	0.04	0.26	0.29
POC (**)	0.22	0.14	0.18	0.14	0.81	0.69

Appendix C

Subgroup Analyses and Effect Modification with Bayesian Kernel Machine Regression

C.1 Additional Methods

C.1.1 Prior specification for models

We used the same prior distributions and hyperparameters for all models considered. We used a flat prior on the regression coefficients $\boldsymbol{\beta}$ and an inverse Gamma prior on the error variance, $\sigma^2 \sim \text{IG}(0.001, 0.001)$. Following Bobb et al. (2015), we parameterized $\lambda = \tau\sigma^{-2}$ and assigned λ a Gamma prior with mean 100 and variance 100. We used the same priors for the group-specific λ parameters in the second group-separable model. Finally, the smoothing parameters ρ_m for each exposure $m = 1, \ldots, M$ has prior $\rho_m \sim \text{Uniform}(0, 100)$. We used the same priors for ρ_p in the modifier-in-kernel model.

C.1.2 Simulation Methods

Simulation surfaces were generated using the following equations: $h_1(\mathbf{z}) = 4f[\frac{1}{4}(\Sigma_{m=1}^2 z_m + \frac{1}{2}\prod_{m=1}^2 z_m), 0, 0.3]$, $h_2 = \frac{1}{4}h_1$, and $h_3(\mathbf{z}) = 0$. The function f is the logistic distribution density function $f(x, \mu, \sigma) = \frac{1}{\sigma}\text{exp}[(x-\mu)/\sigma]/[1 + \text{exp}\{(x-\mu)/\sigma\}]^2$.

We estimated the exposure-response surface across a grid of quantiles ranging from the first to third quartiles. We estimated the group-specific total mixture effects across that same grid of quantiles for q_2, with a change in effect compared to the median ($q_1 = 0.5$). We estimated the group-specific single-exposure effects for an IQR change ($q_1 = 0.25, q_2 = 0.75$) in each exposure while the others were fixed at each quartile ($q = 0.25, 0.5, 0.75$). We summarized these group-specific effects across each modifier group. We estimated interaction effects, or the difference in effect between groups, for the same quantiles as the group-specific effects.

We consider different exposure ranges for each effect estimate to prevent extrapolation. For group-specific effect estimates, we evaluate performance in each group only over the range of exposures in that group. For between-group differences, we evaluate performance using the intersection of the range of exposures observed in both groups of the difference.

C.2 Additional Figures

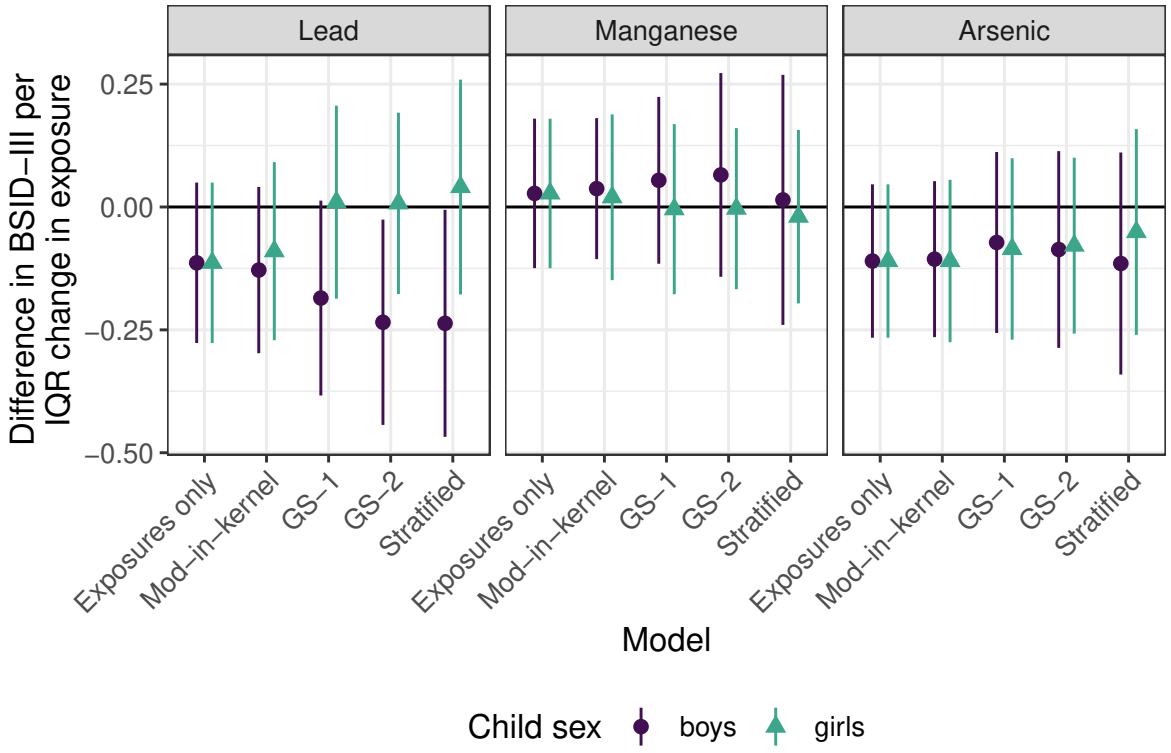


Figure C.1: Estimated sex-specific single-exposure effects of lead, manganese, and arsenic on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. This is the estimated average change in BSID-III for an IQR change in exposure, while other exposures are fixed at their medians, for both girls (green triangles) and boys (purple circles) separately.

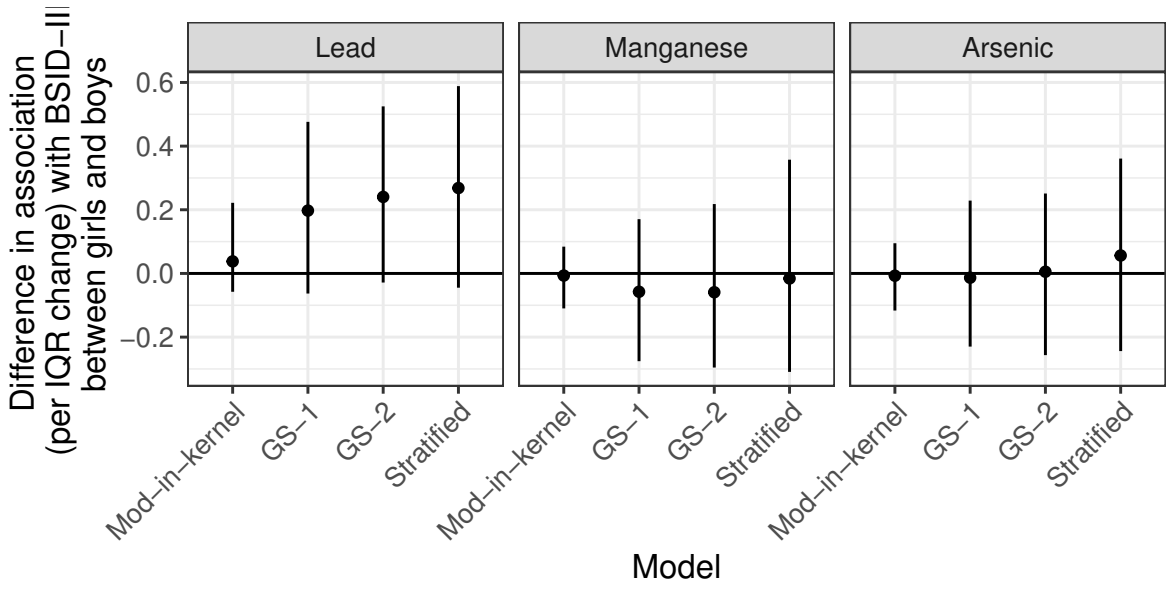


Figure C.2: Estimated between-group differences (girls effect minus boys effect) in single-exposure effects of lead, manganese, and arsenic on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. This is the estimated difference between girls and boys average change in BSID-III for an IQR change in exposure, while other exposures are fixed at their medians.

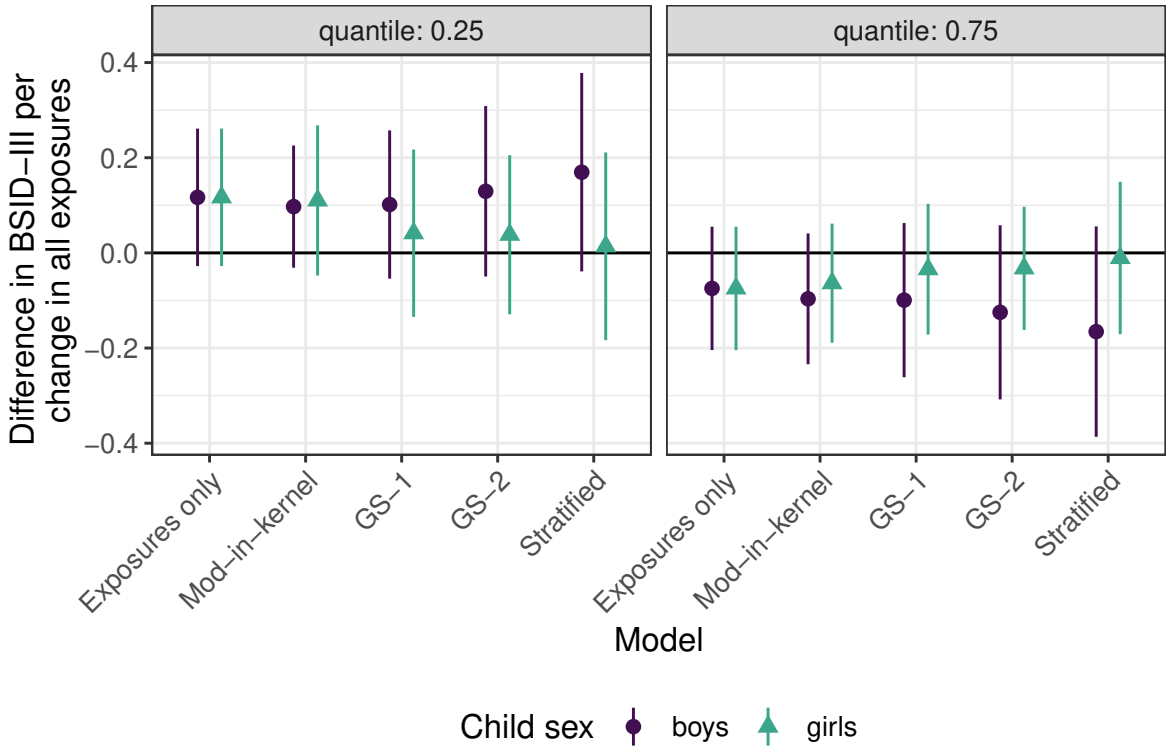


Figure C.3: Estimated sex-specific total effects of the mixture on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. This is the estimated average change in neurodevelopment score for a change in all exposures between their first quartile and median (facet 1) and between their third quartile and median (facet 2) for girls (green triangles) and boys (purple circles).

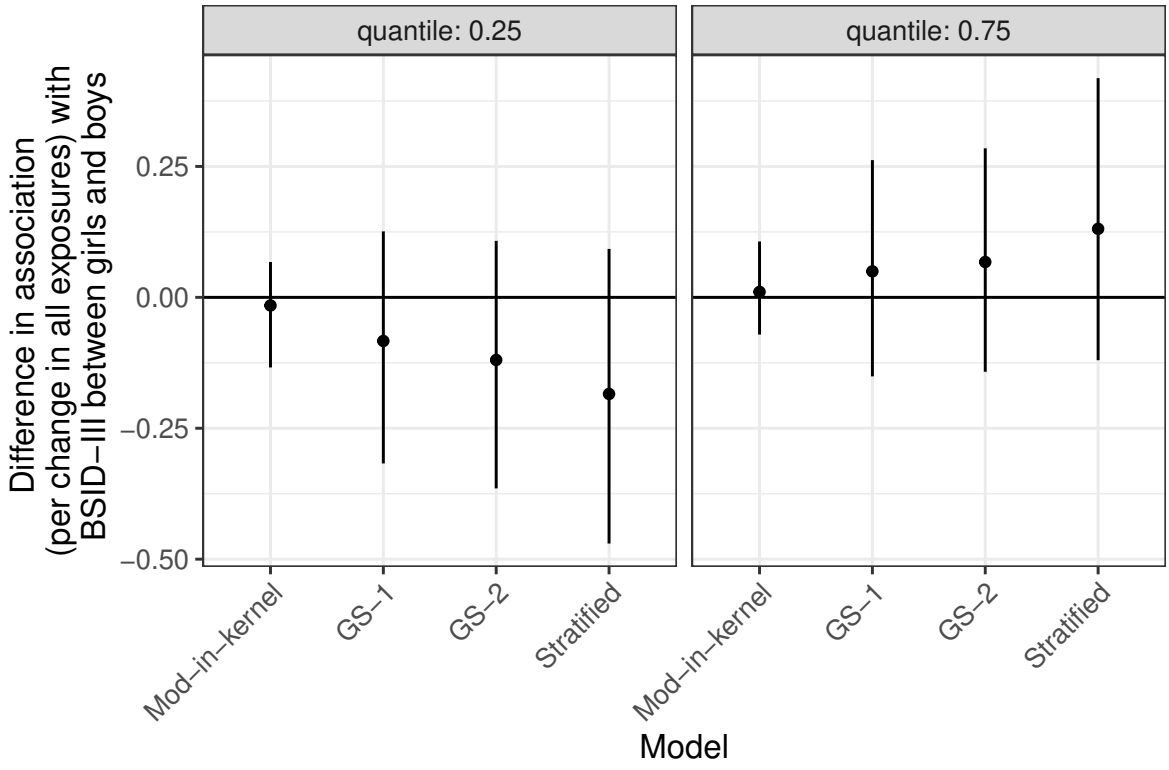


Figure C.4: Estimated differences (girls effect minus boys effect) between girls and boys in total effects of the mixture on BSID-III from each model (x-axis) in $n = 349$ children in the Bangladesh cohort. This is the estimated difference girls and boys average change in neurodevelopment score for a change in all exposures between their first quartile and median (facet 1) and between their third quartile and median (facet 2).