# Nonparametric Regression Methods for Causal Mediation Analysis

# Nonparametric Regression Methods for Causal Mediation Analysis

a dissertation presented
by
Katrina Leigh Devick
to
The Department of Biostatistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biostatistics

Harvard University
Cambridge, Massachusetts
May 2018

# Nonparametric Regression Methods for Causal Mediation Analysis

## Abstract

Causal mediation analysis is a popular approach to quantify the mechanisms through which an exposure operates on an outcome. Using causal mediation analysis, one can decompose the total effect (TE) of an exposure on an outcome into the pathway that operates indirectly through an intermediate (mediator) variable and the pathway that is independent of the mediator variable or that operates directly from the exposure to outcome. Researchers' understanding of the pathways operating through an intermediate variable is crucial for policy recommendations to reduce the potentially harmful impact of the exposure(s) and/or uneven burden of disease. With the growing interest in causal mediation analysis, new methods are needed to estimate direct and indirect effects with complex data. In this dissertation, I propose three novel approaches to estimate mediation effects using Bayesian nonparametric regression models. These methods allow for data with the following complexities: (Chapter 1) the exposure is binary and nonmanipulable and a normality assumption for the mediator variable is not suitable; (Chapter 2) the joint effect of multiple exposures (mixture) is of interest; and (Chapter 3) the exposure of interest is a mixture and the mediator and outcome variables are latent constructs composed of multiple measurements. For each method, I discuss how my approach addresses gaps in the literature and demonstrate how my proposed approach preforms compared to current methods via simulation and data application.

# Contents

# List of figures

# List of tables

# Acknowledgments

First and foremost, I want to acknowledge my Lord and Savior, Jesus Christ. Through Christ all things are possible–I have made it to where I am today because of His grace and strength.

I am deeply grateful to my wonderful advisors, Brent Coull and Linda Valeri, for their encouragement, guidance, and mentorship throughout my journey as a graduate student. I would like to give a special thank you to my committee members, Jarvis Chen and Paige Williams. Their insights, both academic and nonacademic, have contributed greatly to this dissertation and to my life.

My journey as a biostatistician would not have started without the encouragement from my mentor and friend, Darcie Delzell. Her love for statistics inspired me to pursue this career, and I will be forever grateful for her advice over the years.

I owe everything to the support of my family and friends. To my parents, Doug and Lynda, none of this would be possible without your love and encouragement. To my siblings, Olivia, Levi, Jake, and Emily Joella, and my honorary sister Alia, you are my best friends and I am inspired by each of you daily. To the incredible friends I have made on this journey, Olivia Maiefski, Gloria Han, Emily Slade, and the Ladies of Biostats, I cherish every memory we have made together and know I would not have made it through the last few years without you by my side.

Lastly, to my number one fan and husband, Lucas. Your love and support have been a constant through my graduate school experience, and for that, I am eternally grateful.

# 1

# The role of body mass index at diagnosis on Black-White disparities in colorectal cancer survival: A density regression mediation approach

Katrina L. Devick[1], Linda Valeri[2,3], Jarvis Chen[4], Alejandro Jara[5], Marie-Abèle Bind[6], and Brent A. Coull[1]

[1]*Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA*

[2]*Psychiatric Biostatistics Laboratory, McLean Hospital, Belmont, MA, USA*

[3]*Department of Psychiatry, Harvard Medical School, Boston, MA, USA*

[4]*Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[5]*Pontificia Universidad Católica de Chile, Santiago, Chile*

[6]*Department of Statistics, Harvard University, Cambridge, MA, USA*

## Abstract

The study of racial/ethnic inequalities in health is important to reduce the uneven burden of disease. In the case of colorectal cancer (CRC), disparities in survival among non-Hispanic Whites and non-Hispanic Blacks are well documented, and mechanisms leading to these disparities need to be studied formally. It has also been established that body mass index (BMI) is a risk factor for developing colorectal cancer, and recent literature shows BMI at diagnosis of colorectal cancer is associated with survival. Since BMI varies by racial/ethnic group, a question that arises is whether disparities in BMI is partially responsible for observed racial/ethnic disparities in CRC survival. This paper presents new methodology to quantify the impact of a hypothetical intervention on BMI on racial/ethnic disparities in survival. We conceptualize the intervention as one that matches the BMI distribution in the Black population to a potentially complex distributional form observed in the White population. We do this by fitting a nonparametric Bayesian density regression

model with a Dirichlet process prior for BMI, coupled with a parametric accelerated failure time model for CRC survival. We perform a simulation that shows our proposed density regression approach performs as well as or better than current methodology allowing for a shift in means only, and that standard practice of categorizing BMI leads to large biases when interactions are present and the distribution of BMI is not normal. When applied to motivating data collected by the Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium, our approach suggests that heterogeneity in the impact of a shift in BMI is present, with the proposed intervention being potentially beneficial for middle-aged Black patients, yet harmful for elderly and high income Black populations. We find allowing for race-covariate interactions is important in determining the result of a hypothetical intervention on BMI on CRC survival.

This work was supported by grants NIH T32ES007142, NIH P01 CA134294, and NIH ES000002.

## 1.1 Introduction

Differences in cancer survival among racial/ethnic groups are well documented [2]. Considering the case of colorectal cancer (CRC), disparities in survival between non-Hispanic Blacks and non-Hispanic-Whites in the United States (U.S.) manifested in 1980 and have widened over time [19, 44]. Since in the U.S. colorectal cancer is the third most commonly diagnosed cancer in men and women, and the third most common cause of cancer-related death [2, 23], it is important to formally study the mechanisms that are causing Black-White disparities in survival to reduce the unfair, unjust, and preventable uneven burden of cancer [32].

It has been well established that body mass index (BMI), as defined by weight in kilograms divided by height in meters squared, is a risk factor for developing colorectal cancer. Recent literature has reported effects of BMI pre-, at-, or post-diagnosis on colorectal cancer survival. It has been hypothesized that higher BMI is associated with worse colorectal cancer prognosis, however, numerous studies have shown the relationship is more complicated [33, 31]. Some studies have observed better survival in overweight or class I obese patients compared to normal weight patients. Since this paradox could result from sample selection bias, reverse causality, and/or collider bias, employing causal methodology to study the relationship between race, BMI, and CRC survival is of utmost importance.

The determinants of disparities in survival for colorectal cancer patients are multifactorial. As differences in the distributions of BMI across racial/ethnic groups are well documented [36, 57], and the effect of

race/ethnicity on BMI has been shown to be complex [57], we propose a causal inference approach based on the theory of counterfactuals [40] to quantify the percentage of disparity in survival that could be reduced had Black-White disparities in BMI been eliminated in colorectal cancer patients. This framework has not been used to implement a distributional shift in a continuous intermediate variable along the race-survival pathway. Using the definition of non-manipulable race as proposed by VanderWeele and Robinson [54], we consider a stochastic intervention on BMI to determine how BMI contributes to disparities in Black-White survival among colorectal cancer patients. The implication for understanding the relationship between race and BMI on survival is fundamental to practical recommendation for physicians to give patients upon diagnosis of colorectal cancer, and ultimately to reduce disparity in colorectal cancer survival.

Since the relationship between race and BMI on colorectal cancer survival is complex in that the Black-White disparity differs across the range of BMI, a simple linear model may not appropriately capture the residual disparity after a hypothetical intervention in BMI. If categories of BMI representing normal-, overweight, class I and class II/III obese individuals are used, then methodology presented by Valeri et al. in 2016 [47] or the parametric equivalent could be applied [48]. This methodology estimates the impact of a hypothetical intervention in each covariate stratum that matches the probability of being in each BMI category for Blacks to that observed in Whites. However, we wish to place no restriction on the distribution of BMI and allow for a shift in the entirety of the Black BMI distribution to match a potentially complex distributional form of BMI observed in the White population, not just allowing for a shift in mean BMI or BMI categories.

To our knowledge, no methods exist to estimate the causal effects of a hypothetical distributional-level intervention of a continuous intermediate variable in the context of estimating racial/ethnic disparities in survival. We propose a nonparametric Bayesian Linear Dependent Dirichlet Process model to estimate densities of BMI for each covariate stratum, coupled with a parametric accelerated failure time model for CRC survival to estimate the effect of this intervention. We will use data from the Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium to answer these questions. This dataset contains numerous individual-level variables not typically found in cancer registries and a high percentage of non-Hispanic Black colorectal cancer patients.

## 1.2 Materials and Methods

### 1.2.1 Study population

We obtained data from the National Cancer Institute's CanCORS Consortium, which contains detailed information from patients and physicians [4]. The consortium collected data on colorectal cancer cases in multiple regions and health care delivery systems across the U.S. The study population consisted of non-Hispanic White and non-Hispanic Black patients enrolled between 2003 and 2005. To be eligible for enrollment as a colorectal cancer case, the study required patients to be at least 21 years old and within 3 months of newly diagnosed invasive adenocarcinoma of the colon or rectum. The CanCORS Consortium oversampled minority groups at several of the sites. Eight out of the eleven CanCORS centers collected survival data, including: Henry Ford Health System (HFHS), Kaiser Permanente Hawaii (KPHI), Kaiser Permanente Northwest (KPNW), Northern California Cancer Center (NCCC), State of Alabama (UAB), Los Angeles County (UCLA), North Carolina (UNC), and Veterans Health Administration (VA). We excluded patients from Kaiser Permanente Hawaii from our analysis due to a small number of non-Hispanic Blacks.

Patients self-reported height and weight at diagnosis of CRC during the CanCORS survey (within 3 months of diagnosis). We used these measurements to calculate BMI, as defined by weight in kilograms divided by height in meters squared. CanCORS collected stage at diagnosis information via medial record abstraction, categorized as stage I-IV according to the American Joint Committee on Cancer (AJCC) staging criteria [22]. For the purpose of the analysis, we excluded patients with unstaged cancer. Approximately 7% of the sample was unstaged.

### 1.2.2 Models

To nonparametrically estimate the density of BMI by covariate pattern, we fit a Linear Dependent Dirichlet Process (LDDP) model:

$$\begin{aligned} m_i|G &\overset{ind.}{\sim} \int N\left(m_i|\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2\right) dG(\boldsymbol{\beta}, \sigma^2), \\ G|\alpha, G_0 &\sim DP(\alpha G_0), \end{aligned} \tag{1.1}$$

where $G_0 \equiv N_p(\boldsymbol{\beta}|\boldsymbol{\mu}_b, \boldsymbol{S}_b)\Gamma(\sigma^{-2}|\tau_1/2, \tau_2/2)$, $m_i$ represents BMI, and $\boldsymbol{x}_i = (R_i, \boldsymbol{C}_i)$ are the covariates in the model [28]. This is a type of dependent Dirichlet Process (DDP), in which an uncountable set of Dirichlet

Processes (DPs) are created and dependence is introduced by modifying the stick-breaking representation of Sethuraman [43]. Let $G$ have a DP prior with base measure $G_0$ and precision parameter $\alpha$, denoted $G \sim DP(\alpha G_0)$. Then, the stick breaking representation of $G$ is:

$$G(B) = \sum_{l=1}^{\infty} \omega_l \delta_{\theta_l}(B), \tag{1.2}$$

where B is a measurable set, $\delta_{\theta_l}(\cdot)$ is the Dirac measure at $\theta_l$, $\theta_l | G_0 \overset{iid}{\sim} G_0$ and $\omega_l = V_l \prod_{j<l}(1 - V_j)$, with $V_l | \alpha \overset{iid}{\sim} Beta(1, \alpha)$. A LDDP model has the specific restriction that the component of the atoms defining the location follows a linear regression model $\theta_l(\boldsymbol{x}) = (\boldsymbol{x}^T, \boldsymbol{\beta}_l, \sigma_l^2)$. We use the following priors for (1.1):

$$\alpha | a_0, b_0 \sim \Gamma(a_0, b_0), \qquad \tau_2 | \tau_{s_1}, \tau_{s_2} \sim \Gamma(\tau_{s_1}/2, \tau_{s_2}/2),$$

$$\boldsymbol{\mu}_b | \mathbf{m}_0, \boldsymbol{S}_0 \sim N_p(\mathbf{m}_0, \boldsymbol{S}_0), \qquad \text{and } \boldsymbol{S}_b | \nu, \boldsymbol{\Psi} \sim IW_p(\nu, \boldsymbol{\Psi}),$$

where $a_0 = 10$, $b_0 = 1$, $\tau_1 = 6.01$, $\tau_{s_1} = 6.01$, $\tau_{s_2} = 2.01$, $\nu = 9$, $\mathbf{m}_0 = (\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{W}^T y$, $\boldsymbol{S}_0 = 1,000(\boldsymbol{W}^T\boldsymbol{W})^{-1}$, $\boldsymbol{\Psi}^{-1} = \boldsymbol{S}_0$, and $\mathbf{W} = [\boldsymbol{J}_n | \mathbf{x}]$

We model the survival time outcome via a Bayesian Weibull accelerated failure time (AFT) model. The joint effects of race and BMI and the total effect of race on CRC survival are given, respectively, by:

$$\log(T) = \theta_0 + \theta_1 R + \theta_{21} M + \theta_{22} M^2 + \theta_{31} RM + \theta_{32} RM^2 + \boldsymbol{\theta}_{\boldsymbol{c}}^T \boldsymbol{c} + \nu\epsilon, \tag{1.3}$$

$$\log(T) = \gamma_0 + \gamma_1 R + \boldsymbol{\gamma}_{\boldsymbol{c}}^T \boldsymbol{c} + \nu\epsilon, \tag{1.4}$$

where $R$ is an indicator for non-Hispanic Black, $M$ represents centered continuous BMI, $\nu$ is the scale parameter, and $\epsilon$ follows an extreme value distribution. We define the disparity in survival between non-Hispanic Blacks and non-Hispanic Whites prior to any intervention as:

$$Disparity = \frac{\mathrm{E}[T | R = 1, \mathbf{c}]}{\mathrm{E}[T | R = 0, \mathbf{c}]} = exp(\gamma_1). \tag{1.5}$$

We define the residual disparity between Black and White individuals after intervention on BMI as the ratio of mean survival time had the distribution of BMI in the Blacks matched that of the Whites, for each covariate pattern:

$$Residual\ disparity(\mathbf{c}) = \frac{\mathrm{E}\left[T_{H_{\mathbf{c}}(0)} | R = 1, \mathbf{c}\right]}{\mathrm{E}\left[T | R = 0, \mathbf{c}\right]} = exp(\theta_1) \frac{\int e^{(\theta_{21} + \theta_{22} + \theta_{31} + \theta_{32})M} d(m | R = 0, \boldsymbol{c})}{\int e^{(\theta_{21} + \theta_{22})M} d(m | R = 0, \boldsymbol{c})}, \tag{1.6}$$

where $H_{\mathbf{c}}(0)$ represents a random draw from the White BMI distribution for the corresponding covariate stratum. The residual disparity (RD) measure assumes 1) no unmeasured mediator-survival confounding and 2) both the mediator and outcome models are correctly specified. The derivation of the residual disparity formula is included in Appendix section A.2 starting on page 45.

To estimate the integrals in (1.6), we use inverse cumulative density sampling to draw 1,000 random samples from the BMI density curves estimated from (1.1). From this model, we estimate a density and CDF for a grid of 200 BMI values for all 96 White covariate strata. We linearly interpolate random draws of BMI from the grid. To implement the shift in distribution of the Black BMI to match the White distribution of BMI, we use the White density and CDF for each covariate pattern to estimate the integral in the numerator. For each of the random samples in a particular covariate stratum, we calculate the necessary integrand and then take the mean of these integrand samples as the estimate of the integral. We do this for the densities and CDFs at each iteration of the MCMC to obtain posterior samples of the residual disparity in each covariate stratum. From these posterior samples, we calculate 95% credible intervals. We acquire posterior samples of the marginal residual disparity by weighting the stratum specific posterior samples of the residual disparity by the empirically estimated probability of being in a given covariate stratum. Lastly, we define the percent reduction in disparity as:

$$Disparity\ Reduction(\mathbf{c}) = \frac{disparity - residual\ disparity(\mathbf{c})}{disparity}. \tag{1.7}$$

### 1.2.3 Simulation

We used simulation to compare the operating characteristics of the residual disparity estimator from our density approach to those obtained by the traditional/difference method [6], the method that assumes a linear regression model for BMI, and the method that categorizes BMI into normal-, over-weight, class I and class II/III obesity and uses a baseline category logit (BCL) regression to model categorical BMI. For the method assuming a linear regression for BMI, we performed nonparametric inference by simulating counterfactuals, as outlined by Imai et al. in 2010 [26]. Model specification and the residual disparity formula for the baseline category logit model can be found in Appendix section A.3 on page 47.

We compared the operating characteristics of these estimators under four scenarios for race-specific distributions of BMI. First, we assumed that the true distribution of BMI was normal and there was only a location

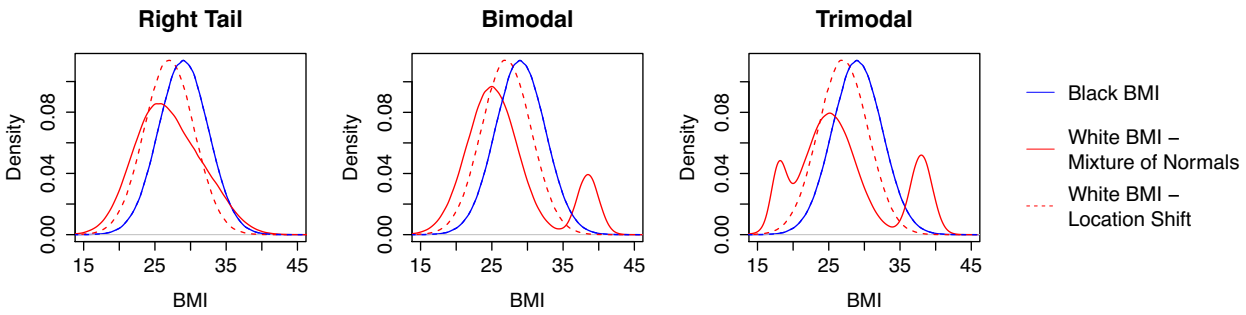


**Figure 1.1** True distributions of BMI under three cases of our simulations. The location shift only case is represented by dashed lines.

shift between the BMI distributions for Blacks and Whites. Then, we considered three additional scenarios where the true target distribution of BMI was a mixture of normals for the White population that cannot be described by a location-scale shift.

We simulated $M$ and $T$ for $K = 1,500$ subjects, letting 750 be Black ($R = 1$) and 750 be White ($R = 0$). We randomly sampled 750 values from the true underlying BMI distribution for the corresponding race. For the location shift only, we generated BMI values for White subjects from a Normal$(27, 3.5^2)$. The true underlying BMI distributions for the three scenarios when the White BMI distribution was a mixture of normals are summarized in Figure 1.1. For each scenario, we fixed the mean of the White BMI distribution at 27.

Since the formula for the residual disparity estimators given in (1.6) only requires estimation of the BMI distribution to which the intervention normalizes this intermediate variable, which we term the target distribution, we generated Black BMI values from a Normal$(29, 3.5^2)$ for all four cases. We then simulated survival times ($T_s$) for each subject from:

$$log(T) = \theta_0 + \theta_1 R + \theta_2 M + \theta_3 RM + \nu\epsilon, \tag{1.8}$$

where $\nu = 0.82$, $\epsilon$ is a random sample from an extreme value distribution, and we set to $\boldsymbol{\theta} = (7.56, -0.88, 0.029, 0)$ or $\boldsymbol{\theta} = (7.56, -0.88, 0.029, 0.022)$ under no interaction and interaction respectively. We generated random censoring times ($T_c$) from a Weibull distribution, such that $\sim 65\%$ of the observed times, $T = \min(T_s, T_c)$, were censored. We truncated the observed times if $T > 1826.25$ days. We simulated 500 datasets for each of the four scenarios of true underlying BMI distributions. For each dataset, we estimated the disparity and residual disparity (RD) for each of the three mediator models and under the difference

method. We compared the performance of these estimators between the different scenarios of underlying distributions of BMI via root mean squared error (rMSE), bias, and standard deviation.

In order to assess how the magnitude of the deviations from standard assumptions would affect the relative performance of the methods under consideration, we repeated our simulation when the effect sizes were doubled, $\boldsymbol{\theta} = (7.56, -1.76, 0.058, 0)$ or $\boldsymbol{\theta} = (7.56, -1.76, 0.058, 0.044)$, under no interaction and interaction respectively.

### 1.2.4 Data analysis

For our CanCORS data analysis, we used survival time in months since diagnosis, and found after restricting to 60-month survival, 62% of the Black and 70% of the White populations were censored. We considered categorized age at diagnosis (age $<$50, 50-65, and $>$65 years), sex (female versus male), income level ($<\$40k$, $\$40\text{-}60k$, $\$60\text{-}80k$, $>\$80k$), and stage at diagnosis (I-IV) to be potential confounders of the BMI-survival relationship. We compared the distribution of baseline characteristics for non-Hispanic Blacks to non-Hispanic Whites via $\chi^2$ statistics.

In CanCORS, we estimated the Black-White disparity in CRC survival prior to intervention on BMI, and the residual disparity after hypothetical intervention on the distribution of Black BMI to match the observed distribution of White BMI using our density approach, linear regression for BMI, BCL regression for categorical BMI, and the traditional/difference method. We allowed for a quadratic effect of BMI on CRC survival in the LDDP, linear regression, and BCL mediator models. We calculated 95% confidence intervals for the disparity, and conditional and marginal residual disparity measures for the linear and BCL models via the bootstrap. We investigated heterogeneity in the effect of the hypothetical intervention on Black BMI in CanCORS across strata.

Motivated by Valeri et al. [47], we considered possible model misspecification by age and sex. We conducted backwards model selection allowing for race by age, race by sex, BMI by age, and BMI by sex interactions. We performed analyses using the best fitting models for BMI and CRC survival in CanCORS, and compared these results to those that did not include any of these interactions.

## 1.3 Results

### 1.3.1 CanCORS

A summary of the CanCORS data can be found in Appendix A on page 45. After we applied exclusion criteria, our sample from CanCORS consisted of 1,585 CRC patients (1,253 non-Hispanic White and 332 non-Hispanic Black individuals). As displayed in Table A.1, Black and White individuals differ significantly on many baseline characteristics. There are more elderly Black individuals than White, Black patients tend to be poorer than White patients, and Black CRC patients are less likely to survive 5 years (log rank test $p$=0.005).

### 1.3.2 Simulation

Table 1.1 shows the rMSE, bias, and standard deviation (SD) for the residual disparity estimators from the LDDP, linear regression, baseline category logit models and the difference method. The rMSE, bias, and SD values exhibit similar patterns when the $\boldsymbol{\theta}$ values are motivated from CanCORS and when they are doubled. When a RM interaction exists ($\theta_3 \neq 0$), our density regression approach performs comparably to or better than the linear regression approach. The bias is smallest when a LDDP mediator model is used, and the variance is only slightly larger than in the linear case. As the effect size increases and the underlying mediator distribution becomes less normal, larger differences among the various residual disparity estimators emerge. Notably, when the mediator is categorized and a baseline category logit (BCL) model is used, we observe large bias and variance. The baseline category logit and the difference method estimators perform equally badly, being about 5% worse than our proposed density regression estimator in terms of bias and rMSE. As we expected, the most efficient method is the difference method, but this leads to large bias when interactions and nonlinearities are present.

When no interaction is present ($\theta_3 = 0$), the four methods perform similarly. This suggest that even when more complicated methods are not needed, our proposed density regression approach performs as well as simpler methods in terms of rMSE.

| | Mediator Model | $\boldsymbol{\theta} = (7.56, -0.88, 0.029, 0.022)$ | | | $\boldsymbol{\theta} = (7.56, -1.76, 0.058, 0.044)$ | | |
|---|---|---|---|---|---|---|---|
| | | RT | Bi | Tri | RT | Bi | Tri |
| **rMSE** | Density | 0.076 | 0.079 | 0.083 | 0.112 | 0.149 | 0.158 |
| | Linear | 0.077 | 0.083 | 0.088 | 0.111 | 0.155 | 0.168 |
| | BCL | 0.078 | 0.094 | 0.095 | 0.122 | 0.194 | 0.209 |
| | Traditional | 0.079 | 0.088 | 0.094 | 0.141 | 0.203 | 0.229 |
| **Bias** | Density | -0.032 | -0.034 | -0.034 | -0.076 | -0.115 | -0.124 |
| | Linear | -0.038 | -0.047 | -0.051 | -0.075 | -0.127 | -0.141 |
| | BCL | -0.026 | -0.059 | -0.056 | -0.069 | -0.165 | -0.175 |
| | Traditional | -0.041 | -0.053 | -0.061 | -0.120 | -0.187 | -0.214 |
| **SD** | Density | 0.068 | 0.072 | 0.076 | 0.082 | 0.095 | 0.098 |
| | Linear | 0.067 | 0.069 | 0.072 | 0.082 | 0.089 | 0.091 |
| | BCL | 0.074 | 0.074 | 0.077 | 0.101 | 0.102 | 0.114 |
| | Traditional | 0.067 | 0.070 | 0.072 | 0.075 | 0.079 | 0.081 |

**Table 1.1** Simulations results when there is an interaction is present ($\theta_3 \neq 0$), when values are motivated from CanCORS, and effect sizes are doubled. Root mean squared error (rMSE), bias, and standard deviation (SD) of the residual disparity estimators under LDDP, linear, baseline category logit (BCL) mediator models and the difference method for 3 different cases of true underlying White BMI distribution is right tailed (RT), binomial (Bi), or trinomial (Tri).

### 1.3.3 Data analysis

Before any hypothetical intervention on BMI at diagnosis, we observe a significant Black-White disparity in survival, adjusting for covariates ($p = 0.023$). Table 1.2 reports marginal results after hypothetical intervention to shift the distribution of BMI in the Blacks to match the BMI distribution of the Whites. We see there is no marginal effect of a shift in BMI on Black-White disparities in survival. This holds across all models considered for the mediator.

Figure 1.2 compares conclusions drawn from the linear regression approach to our density regression approach for each covariate pattern. One plot has points sorted by age categories, and one is color coded by

| | **Estimate (95% CI)** |
|---|---|
| Disparity | 0.833 (0.707, 0.984) |
| RD Density | 0.835 (0.709, 0.984) |
| RD Linear | 0.847 (0.702, 0.990) |
| RD BCL | 0.826 (0.705, 0.995) |
| RD Traditional | 0.809 (0.689, 0.956) |

**Table 1.2** Disparity and residual disparity in CanCORS, without considering potential race×covariate and BMI×covariate interactions. The disparity estimate is from a frequentist AFT model fit, adjusting for covariates. Residual Disparity (RD) measures are estimated from a LDDP, linear, and baseline category logit meditator model as well as the difference method.

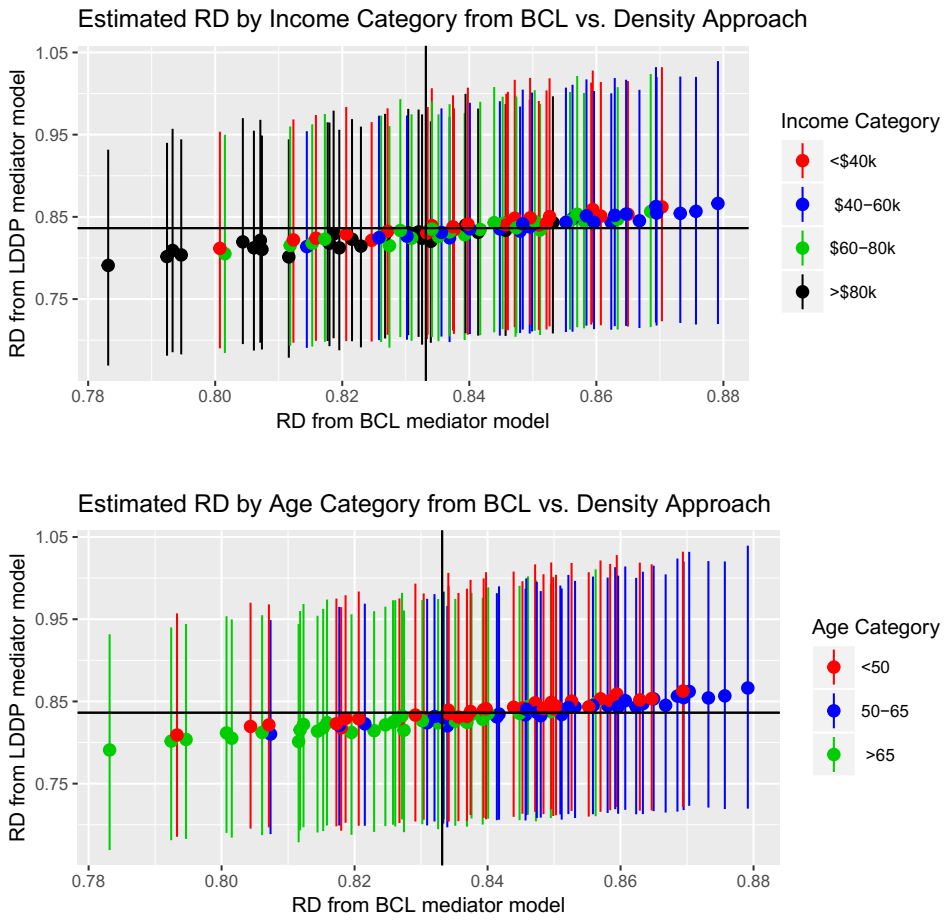


**Figure 1.2** Residual disparity estimates from the BCL regression approach compared to our density regression approach. Each point represents the residual disparity for a particular covariate pattern. Estimated disparity in survival prior to intervention are plotted as vertical and horizontal lines. 95% credible intervals from our density regression approach are depicted in the figure.

income categories. Estimated disparity in survival prior to intervention are plotted as vertical and horizontal lines. Covariate patterns for which both methods estimate the effect of the intervention of BMI to be harmful are in the bottom left quadrant, and points for which both methods estimate the effect to be beneficial are in the upper right quadrant. Points in the upper left and lower right quadrants are covariate patterns for which the two methods lead to different conclusions about the intervention on BMI. 95% Credible Intervals estimated from the density regression approach are included on the plots. We see heterogeneity across subpopulations in CanCORS exists in the impact of intervention on BMI, although no differences are significant. Specifically, there appears to be a trend that people in lower-income categories benefit from an intervention on BMI, whereas, the intervention is harmful for individuals in higher-income categories. Also, the intervention appears to be harmful in the elderly.

| **Estimate (95% CI)** | marginal | age $< 50$ | $50 \leq$ age $< 65$ | $65 \leq$ age |
|---|---|---|---|---|
| Disparity | 0.83 (0.71, 0.98) | 0.64 (0.47, 0.87) | 1.00 (0.81, 1.46) | 0.83 (0.64, 1.09) |
| RD Density | 0.86 (0.73, 1.03) | 0.61 (0.45,0.84) | 1.02 (0.79,1.37) | 0.84 (0.67,1.09) |
| RD Linear | 0.82 (0.70, 0.97) | 0.61 (0.45, 0.83) | 1.08 (0.77. 1.38) | 0.84 (0.65, 1.10) |
| RD BCL | 0.85 (0.72, 1.03) | 0.61 (0.46, 0.83) | 1.00 (0.76, 1.33) | 0.86 (0.67, 1.13) |

**Table 1.3** Marginal and age category specific disparity and residual disparity estimates for best fitting mediator and outcome models. Disparities estimates are from a frequentist AFT model fit, adjusting for covariates and allowing for a race-age interaction. Residual Disparity measures are estimated from a LDDP, linear, and baseline category logit meditator model. The best fitting model for BMI included a race-sex interaction and the best fitting survival model included race-age interactions.

In the CanCORS population, we found race-age interaction terms in the AFT model for the joint effect of race and BMI on CRC survival were statistically significant (LRT $p = 0.050$) and a race-sex interaction term in the linear regression model for BMI was significant (LRT $p < 0.001$). Race-age interaction terms in the AFT model for the total effect of race on CRC survival were marginally significant (LRT $p = 0.098$). This suggests that age category specific disparities should be used when determining the impact of an intervention on BMI. Results from each of the methods after rerunning our analyses with these significant interaction terms are summarized in Table 1.3 and Figure 1.3. As seen in Table 1.3, the marginal residual disparity accounting for race-age and race-sex interactions are almost identical to the estimated residual disparity not accounting for these interactions. However, due to the large effect size of the race-age interactions in the AFT model, we observe gaps in the stratum specific residual disparity and disparity estimates. When comparing the estimated residual disparity in each age category to the corresponding age category specific disparity, we see the impact of the intervention to match the distribution of Black BMI to the distribution of White BMI

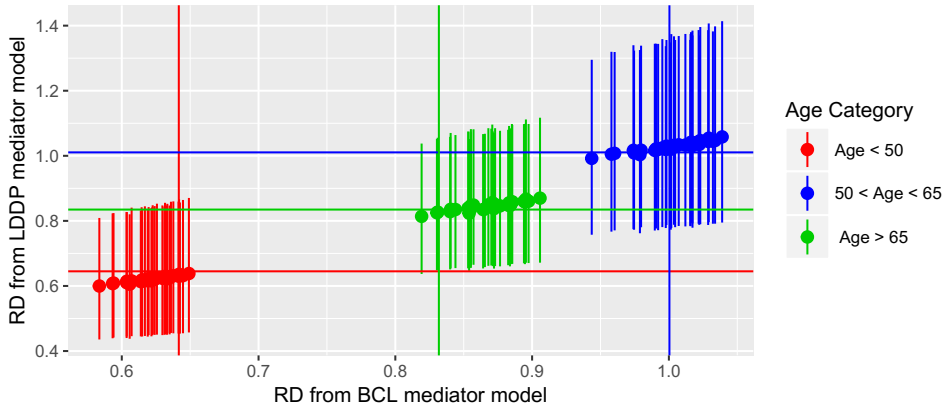


**Figure 1.3** Residual disparity estimates from the BCL regression approach compared to our density regression approach for the best fitting model. Each point represents the residual disparity for a particular covariate pattern. Estimated disparity in survival for each age group prior to intervention are plotted as vertical and horizontal lines.

is null across the age categories. This suggests that the conditional results can be sensitive to race-covariate interactions in the survival outcome and disparity models.

## 1.4 Discussion

To our knowledge, this is the first study to implement a counterfactual causal inference approach to estimate the impact of a distributional-level shift in a continuous intermediate variable along the race-survival pathway in survival disparities. Our methodology relies on the assumptions that there are no unmeasured BMI-survival confounders and that our models are correctly specified.

The main conclusion of our paper is two-fold. First, when race-covariate interactions exist and are not accounted for, incorrect conclusions can be drawn about the impact of intervention on the intermediate variable in reducing race-survival disparities. Second, we show it is not only inefficient to categorize a mediator variable, but categorization results in more bias and variance than if the continuous analog was used, even if the distribution of the mediator is not normal. This is particularly important when the intermediate variable is BMI, since categorizing BMI into well-established categories of normal-, over-weight, and obese is standard in the biomedical literature. One way to avoid the potential for large biases in such settings is to use the nonparametric regression methods proposed here.

Our approach had several advantages. The density mediator model in our approach is robust to model misspecification. This is a consequence of the BMI densities being estimated nonparametrically for each covariate pattern. Also, even with the loss of efficiency from a nonparametric mediator model, our simulations show the reduction in bias can outweigh any increase in efficiency, resulting in decreases in rMSE. Our density regression approach is the only method that allows for arbitrary shifts in a distribution, as opposed to a simple location shift in the distributions, of a continuous intermediate variable. In situations in which there are unmeasured variables, which when marginalized over lead to complex, multi-model distributional forms for the mediator, or when the sample size is insufficient to stratify on all observed variables that would reduce the potential for such complex distributions, our proposed methodology allows the distribution of the mediator to be flexible.

There are limitations to our analysis. Since the data are sparse, race by income interactions were not considered and additional covariates were not added to our models. Thus, estimates could be biased due to

residual confounding, although, our methods should be robust to the exclusion of additional covariates. In the case of our data application to CanCORS, we only have a slightly right tailed distribution of BMI, which does not deviate enough from a normal distribution to see significant differences between RD estimates from the various approaches in our results. Even though CanCORS contained individual level income, due to low numbers of Black individuals in higher income categories, we needed to the use crude income categories. Ideally, we would have data on BMI one year after diagnosis of CRC in addition to BMI at diagnosis of CRC, so individual trajectories of weight loss/gain could be considered. The major limitation in our analysis is the small number of non-Hispanic Blacks included in CanCORS after applying exclusion criteria (n=332).

Further study is needed to more thoroughly understand the relationship between race-survival and moderating factors. Methods that consider weighting approaches will allow for more covariates and interactions to be included in the analysis. Extensions to our approach to allow for dependence between the mediator and outcome models will be considered in future research.

# 2

# Bayesian kernel machine causal mediation analysis

Katrina L. Devick[1], Jennifer F. Bobb[2], Maitreyi M. Mazumdar[3,4], Birgit Claus Henn[5], David C. Bellinger[3,4], David C. Christiani[4], Robert O. Wright[6], Paige L. Williams[1], Brent A. Coull[1,4], and Linda Valeri[7,8]

[1]*Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA*

[2]*Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA*

[3]*Department of Neurology, Boston Children's Hospital, Boston, MA, USA*

[4]*Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[5]*Department of Environmental Health, Boston University School of Public Health, Boston, MA, USA*

[6]*Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

[7]*Psychiatric Biostatistics Laboratory, McLean Hospital, Belmont, MA, USA*

[8]*Department of Psychiatry, Harvard Medical School, Boston, MA, USA*

## Abstract

Exposure to complex mixtures is a real-world scenario. As such, it is important to understand the mechanisms through which a mixture operates in order to reduce the burden of disease. Currently, there are few methods in the causal mediation analysis literature to estimate the direct and indirect effects of a exposure mixture on an outcome operating through a intermediate (mediator) variable. This paper presents new statistical methodology to estimate the natural direct effect (NDE), natural indirect effect (NIE), and controlled direct effects (CDEs) of a potentially complex mixture exposure on an outcome through a mediator variable. We implement Bayesian kernel machine regression (BKMR) to allow for all possible interactions and nonlinear effects of the co-exposures on the mediator, and the co-exposures and mediator on the outcome. From the posterior predictive distributions of the mediator and the outcome, we simulate counterfactual outcomes

to obtain posterior samples, estimates, and credible intervals (CI) of the NDE, NIE, and CDE. We perform a simulation study that shows when the exposure-mediator and exposure-mediator-outcome relationships are complex, our proposed Bayesian kernel machine regression – causal mediation analysis (BKMR–CMA) preforms better than current mediation methods. We apply our methodology to quantify the contribution of birth length as a mediator between in utero co-exposure of arsenic, manganese and lead, and children's neurodevelopment, in a prospective birth cohort in rural Bangladesh. We found a negative association of co-exposure to lead, arsenic, and manganese and neurodevelopment, a negative association of exposure to this metal mixture and birth length, and evidence that birth length mediates the effect of co-exposure to lead, arsenic, and manganese on children's neurodevelopment. If birth length were fixed to its $75^{th}$ percentile value of 48cm, the effect of the metal mixture on neurodevelopment decreases, suggesting that nutritional interventions to help increase birth length could potentially block some of the harmful effects of the metal mixture on neurodevelopment.

This work was supported by grants NIH T32ES007142, NIH P01 CA134294, and NIH ES000002.

**Keywords**: Children's neurodevelopment; Environmental mixture; Mixture; Multi-pollutant exposure.

## 2.1 Introduction

The ability to identify mechanisms through which a complex exposure profile operates is critical for the development of policy protective of public health. As such, the National Institute of Environmental Health Sciences (NIEHS) has prioritized the development of statistical methods that quantify the effect of environmental mixtures on health outcomes [12]. With this increased priority, new methodology is needed to measure these effects and minimize the burden of disease.

Exposure to complex mixtures are representative of real-world scenarios. Since elements of a mixture can exhibit complex interactions, it is important to consider the whole mixture when evaluating the nature of the relationship of a mixture on a health outcome [62, 16, 14]. Once a relationship between a mixture and outcome is established, questions regarding the pathways through which the mixture operates arise.

One approach to quantify operating mechanisms is the use of causal mediation analysis [37, 55, 56, 50]. Causal mediation analysis allows for the decomposition of a total effect (TE) of an exposure on an outcome into the pathway that operates indirectly through an intermediate (mediator) variable and the pathway that is

independent of the intermediate variable or that operates directly from the exposure to outcome. Researchers' understanding of the pathways operating through an intermediate variable is crucial for policy recommendations to reduce the harmful impact of environmental mixtures on health outcomes. Few methods exist to estimate mediation effects when the exposure of interest is a mixture. If the mediator variable has a linear effect on the outcome, a direct extension of the closed form solutions derived by VanderWeele and Vandsteelandt (2009, 2010) and Valeri and VanderWeele (2013) is applicable [55, 56, 50]. In the presence of a nonlinear effect of the mediator on the outcome, the algorithm presented by Imai et al. in 2010 can be used to estimate the natural direct effect (NDE), natural indirect effect (NIE), and controlled direct effects (CDE) of a mediator on the relationship of a mixture on a outcome, through prediction of counterfactuals [26]. However, both of these methods assume no model misspecification. Thus, all interactions between the individual elements of the mixture, the elements of the exposure mixture and mediator, and any nonlinearities need to be included in the models for the mediator and outcome to obtain valid inference. As the dimensions of a multi-dimensional exposure increase, it becomes exponentially difficult to use current methods to obtain unbiased estimates of the mediated effects. To our knowledge, no other methods currently exist to estimate the NDE, NIE, and CDE of a potentially complex exposure mixture on an outcome through a mediator variable.

In this paper, we present a novel method to estimate the NDE, NIE, and CDE for a potentially complex mixture of exposures on an outcome operating through an intermediate variable. We allow for highly complex exposure-mediator and exposure-response functions using Bayesian kernel machine regression (BKMR). BKMR has been shown to perform well compared to other kernel machine approaches [8]. We use BKMR to model the mediator and outcome since BKMR allows for all possible nonlinearities and interactions between the mixture elements, and between the mixture and the mediator, without *a priori* specification. We predict counterfactuals using the posterior predictive distributions of the mediator and the outcome and present an algorithm for estimation of mediation effects.

We apply this method to model data from a prospective birth cohort in Bangladesh. Arsenic, manganese, and lead are known neurotoxicants [9, 13, 38, 46, 64] that are abundant in the Bangladeshi environment, including in drinking water [29]. The relationship between arsenic, manganese and lead on child neurodevelopment is shown to be complex [60, 61, 59, 58, 62, 15, 16, 14, 25, 49]. In our data application, we estimate the NDE and NIE to bring light to the relationship of this metal mixture on child neurodevelopment operating through *in utero* growth, specifically birth length. Also, we estimate the CDE of the metal mixture

on neurodevelopment at different quantiles of birth length to assess if birth growth measures can potentially block some of the harmful effects of the metal mixture on neurodevelopment.

## 2.2 Materials and Methods

### 2.2.1 Bayesian kernel machine regression

We first review BKMR presented by Bobb et al. (2015) as a framework to estimate the effect of a complex mixture on a health outcome [8]. For each subject $i = 1, \ldots, n$, we assume:

$$Y_i = h\left(\boldsymbol{z}_i\right) + \boldsymbol{C}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{2.1}$$

where $Y_i$ is a continuous health outcome, $\boldsymbol{z}_i = \left(z_{1i}, \ldots, z_{Li}\right)^T$ is a vector of $L$ exposure variables (e.g. metals), $\boldsymbol{C}_i = \left(C_{1i}, \ldots, C_{Pi}\right)^T$ is a vector of potential confounders, and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Model (2.1) relates the outcome to the exposure mixture through a flexible function, $h(\cdot)$, which accommodates for nonlinearity and/or interaction among the mixture components. Due to this complexity, identifying a set of basis functions to represent $h(\cdot)$ can be difficult, thus, we employ a kernel machine representation [18].

The unknown function $h(\cdot)$ can be specified in two ways. One can either use basis functions or a positive-definite kernel function $K(\cdot, \cdot)$ to identify $h(\cdot)$. Mercer's theorem [18] established that the kernel function $K(\cdot, \cdot)$ implicitly specifies a unique function space spanned by a particular set of orthogonal basis functions, under regularity conditions. Therefore, any $h(\cdot)$ in this function space can be represented by a set of basis functions or by the dual representation kernel function $K(\cdot, \cdot)$. Liu et al. (2007) showed that model (2.1) can be expressed as the mixed model (2.2) [35]:

$$\begin{aligned} Y_i &\sim N(h_i + \boldsymbol{C}_i^T \boldsymbol{\beta}, \sigma^2) \quad \text{independent}, i = 1, \ldots n, \\ \boldsymbol{h} &= \left(h_i, \ldots, h_n\right)^T \sim N(0, \tau \boldsymbol{K}), \end{aligned} \tag{2.2}$$

where $\boldsymbol{K}$, the kernel matrix, has $(i, j)$-th element $K(\boldsymbol{z}_i, \boldsymbol{z}_j)$.

The kernel function $K(\cdot, \cdot)$ uses a metric of similarity to establish how close exposure profiles $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ are for subjects $i$ and $j$. We will focus on the Gaussian kernel, which uses Euclidean distance as a means to quan-

tify this similarity. Under the Gaussian kernel, we assume $cor(h_i, h_j) = \exp\left\{-(1/\rho)\sum_{\ell=1}^{L}(z_{\ell i} - z_{\ell j})^2\right\}$, where $\rho$ is a tuning parameter that regulates the smoothness of the dose-response function. Intuitively, this assumption means subjects with similar exposure profiles ($\boldsymbol{z}_i$ close to $\boldsymbol{z}_j$) will have more similar risks ($h_i$ will be close to $h_j$).

To fit (2.1), we assume a flat prior on the coefficients for the confounding variables, $\boldsymbol{\beta} \sim 1$, and assume $\sigma^{-2} \sim Gamma(a_\sigma, b_\sigma)$, were we set both the shape parameter $a_\sigma$ and the scale parameter $b_\sigma$ to 0.001. For convenience, we parameterize BKMR model (2.1) with $\lambda = \tau\sigma^{-2}$, where we assume a Gamma prior distribution for $\lambda$ with mean $\mu_\lambda = 10$ and variance $\sigma^2_\lambda = 100$. We assume a uniform distribution $\rho \sim Unif(a, b)$ with $a = 0$ and $b = 100$ for the smoothness parameter $\rho$. For additional details regarding BKMR and prior specification, see Bobb et al. [8].

When the exposure mixture is comprised of numerous elements, it may be of interest to fit (2.1) with component-wise variable selection. To allow for variable selection, the kernel function $K(\cdot, \cdot)$ is augmented. In the case of the Gaussian kernel, the kernel function is expanded as:

$$K(\boldsymbol{z}, \boldsymbol{z}^\top; \boldsymbol{r}) = \exp\left\{-\sum_{\ell=1}^{L} r_\ell (z_\ell - z_\ell^\top)^2\right\}, \tag{2.3}$$

where $\boldsymbol{r} = (r_1, \ldots, r_\ell)^\top$, and we assume a "slab-and-spike" prior for the auxiliary parameters,

$$\begin{aligned} r_\ell \,|\, \delta_\ell &\sim \delta_\ell f_1(r_\ell) + (1 - \delta_\ell) P_0, \quad m = 1, \ldots, M, \\ \delta_\ell &\sim \text{Bernoulli}(\pi), \end{aligned} \tag{2.4}$$

where $\delta_\ell$ is an indicator that element $\ell$ is included in the kernel, $f_1(\cdot)$ denotes a pdf with support on $\mathbb{R}^+$, and $P_0$ is the density with a point mass at 0.

### 2.2.2 Causal mediation analysis

In order to define causal contrasts in a mediation context, we first define our notation. Let $Y_{am}$ denote the counterfactual outcome $Y$ if the exposure level $A$ was set to $a$ and mediator level $M$ was set to $m$. Let $M_a$ be the counterfactual mediator level $M$ that would have been observed if the exposure $A$ was set to $a$. Accordingly, $Y_{aM_{a^*}}$ represents the counterfactual outcome $Y$ if the exposure level $A$ was set to $a$ and the mediator $M$ was set to the level it would have taken if the exposure level $A$ was $a^*$.

The mediated effects of interest, the natural direct effect (NDE), the natural indirect effect (NIE), and the controlled direct effects (CDEs), are formally defined as:

$$NDE = \mathrm{E}\left[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}\right], \tag{2.5}$$

$$NIE = \mathrm{E}\left[Y_{aM_a} - Y_{aM_{a^*}}\right], \tag{2.6}$$

$$CDE(m) = \mathrm{E}\left[Y_{am} - Y_{a^*m}\right]. \tag{2.7}$$

The NDE captures the average difference in the counterfactual outcomes for a change in exposure level $a^*$ to $a$, while fixing the mediator to the level it would have taken if unexposed, $a^*$. The NIE measures the average difference in counterfactual outcomes when fixing the exposure to level $a$, while the mediator varies from the level it would have taken if exposed to $a$ to the level it would have taken if unexposed, $a^*$. The CDE quantifies the average difference in the counterfactual outcomes for a change in exposure level from $a^*$ to $a$, while intervening to fix the mediator to a specified level, $m$.

### 2.2.3 Bayesian kernel machine regression – causal mediation analysis

We consider a single health outcome $Y$, single mediator variable $M$, exposure mixture $\boldsymbol{A}$ comprised of $L$ components, and confounder matrix $\boldsymbol{C}$. To allow for potentially complex relationships between the mixture elements, we model the mediator variable using BMKR model (2.8):

$$M_i = h_M(\boldsymbol{A}_i) + \boldsymbol{C}_i^T\boldsymbol{\beta} + \epsilon_{Mi}, \tag{2.8}$$

where $\epsilon_{Mi} \overset{iid}{\sim} N(0, \sigma_M^2)$. Since accounting for exposure-mediator interactions is important to obtain unbiased effect estimates, we include the mediator variable along with the exposure mixture in the kernel function when modeling the health outcome in (2.9):

$$Y_i = h_Y(\boldsymbol{A}_i, M_i) + \boldsymbol{C}_i^T\boldsymbol{\theta} + \epsilon_{Yi}, \tag{2.9}$$

where $\epsilon_{Yi} \overset{iid}{\sim} N(0, \sigma_Y^2)$. By fitting the models separately, we assume $\epsilon_{Mi}$ and $\epsilon_{Yi}$ are independent. To model

the total effect of the exposure mixture on the outcome, we consider BKMR model (2.10):

$$Y_i = g(\boldsymbol{A}_i) + \boldsymbol{C}_i^T\boldsymbol{\gamma} + \xi_i, \tag{2.10}$$

where $\xi_i \overset{iid}{\sim} N(0, \sigma_\xi^2)$.

We estimate the NDE, NIE, and TE for a change in exposure profile from $\boldsymbol{a}^*$ to $\boldsymbol{a}$ via the following algorithm.

1. Fit BKMR mediator, outcome, and total effect models (2.8), (2.9), and (2.10), respectively.
2. For each Markov chain Monte Carlo (MCMC) iteration, $j = 1, \ldots, J$:
   (a) Sample $k = 1, \ldots, K$ replicates of the mediator for the mean level of covariates under exposure level $\boldsymbol{a}^*$ from mediator model (2.8):

$$\begin{aligned} M_{\boldsymbol{a}^*}^{(jk)}(\bar{\boldsymbol{c}}) &= \mathrm{E}^{(j)}(M|\boldsymbol{A} = \boldsymbol{a}^*, \boldsymbol{C} = \bar{\boldsymbol{c}}) + \sigma_M^{(j)} N(0,1) \\ &= h_M^{(j)}(\boldsymbol{A} = \boldsymbol{a}^*) + \bar{\boldsymbol{c}}^T\boldsymbol{\beta}^{(j)} + \sigma_M^{(j)} N(0,1). \end{aligned}$$

   (b) For each of the $j = 1, \ldots, J$ and $k = 1, \ldots, K$ samples of $M_{\boldsymbol{a}^*}$, estimate the average outcome value for the mean level of covariates for $Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}$ from outcome model (2.9):

$$\begin{aligned} Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}^{(jk)}(\bar{\boldsymbol{c}}) &= \mathrm{E}^{(j)}(Y|\boldsymbol{A} = \boldsymbol{a}, M = M_{\boldsymbol{a}^*}^{(jk)}(\bar{\boldsymbol{c}}), \boldsymbol{C} = \bar{\boldsymbol{c}}) \\ &= h_Y^{(j)}(\boldsymbol{A} = \boldsymbol{a}, M = M_{\boldsymbol{a}^*}^{(jk)}(\bar{\boldsymbol{c}})) + \bar{\boldsymbol{c}}^T\boldsymbol{\theta}^{(j)}. \end{aligned}$$

   (c) Let the $j^{th}$ posterior sample of $Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}$ be $Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}^{(j)}(\bar{\boldsymbol{c}}) = \frac{1}{K}\sum_{k=1}^K Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}^{(jk)}(\bar{\boldsymbol{c}})$
   (d) Since $Y_{\boldsymbol{a}^*} = Y_{\boldsymbol{a}^*M_{\boldsymbol{a}^*}}$ and $Y_{\boldsymbol{a}} = Y_{\boldsymbol{a}M_{\boldsymbol{a}}}$, we sample the $j^{th}$ posterior sample of $Y_{\boldsymbol{a}^*}$ and $Y_{\boldsymbol{a}}$ from the total effect model instead of sampling $Y_{\boldsymbol{a}^*M_{\boldsymbol{a}^*}}$ and $Y_{\boldsymbol{a}M_{\boldsymbol{a}}}$ for ease of computation. Calculate the average outcome value for the mean level of covariates at $\boldsymbol{a}_{\text{new}} = \begin{pmatrix} \boldsymbol{a} & \boldsymbol{a}^* \end{pmatrix}^{\mathsf{T}}$ from (2.10):

$$\begin{aligned} Y_{\boldsymbol{a}_{\text{new}}}^{(j)}(\bar{\boldsymbol{c}}) &= \mathrm{E}^{(j)}(Y|\boldsymbol{A} = \boldsymbol{a}_{\text{new}}, \boldsymbol{C} = \bar{\boldsymbol{c}}) \\ &= g^{(j)}(\boldsymbol{A} = \boldsymbol{a}_{\text{new}}) + \bar{\boldsymbol{c}}^T\boldsymbol{\gamma}^{(j)} \end{aligned}$$

3. Obtain the $j^{th}$ posterior sample of the NDE, NIE, and TE by:

$$\begin{aligned} NDE^{(j)} &= Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}^{(j)}(\bar{\boldsymbol{c}}) - Y_{\boldsymbol{a}^*}^{(j)}(\bar{\boldsymbol{c}}), \\ NIE^{(j)} &= Y_{\boldsymbol{a}}^{(j)}(\bar{\boldsymbol{c}}) - Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}^{(j)}(\bar{\boldsymbol{c}}), \\ TE^{(j)} &= Y_{\boldsymbol{a}}^{(j)}(\bar{\boldsymbol{c}}) - Y_{\boldsymbol{a}^*}^{(j)}(\bar{\boldsymbol{c}}). \end{aligned}$$

4. Estimate the NDE, NIE and 95% credible intervals from these posterior samples.

Four no unmeasured confounding assumptions are required for the NDE and NIE to have a causal interpretation: (i) $Y_{\boldsymbol{a}m} \amalg \boldsymbol{A}|\boldsymbol{C}$, (ii) $Y_{\boldsymbol{a}m} \amalg M|\boldsymbol{C}, \boldsymbol{A}$, (iii) $M_{\boldsymbol{a}} \amalg \boldsymbol{A}|\boldsymbol{C}$, and (iv) $Y_{\boldsymbol{a}m} \amalg M_{\boldsymbol{a}^*}|\boldsymbol{C}$. Namely, there are no unmeasured exposure-outcome confounders, there are no unmeasured mediator-outcome confounders, there are no unmeasured exposure-mediator confounders, and the exposure does not affect any mediator-outcome confounders.

To estimate the CDE, only two no unmeasured confounding assumptions are required: (i) and (ii). The algorithm to estimate the CDE is similar to the algorithm presented above. We include explicit steps to estimate the CDE in Appendix B.2 on page 49.

## 2.3 Simulation

We evaluated the ability of Bayesian kernel machine regression – causal mediation analysis (BKMR-CMA) to estimate the joint mediated effects of a mixture compared to traditional mediation methods [6] and causal mediation analysis [55, 56, 50] under numerous signal-to-noise ratios motivated from our Bangladesh application (Section 2.4).

### 2.3.1 Setup

We first generated a true underlying dataset for each simulation scenario. The dataset consisted of a health outcome $Y_i$, a mediator value $M_i$, and an exposure mixture $\boldsymbol{A}_i = (A_{1i}, A_{2i}, A_{3i})^T$ for $i = 1, \ldots, 50,000$ subjects. The exposure mixture was generated as $\boldsymbol{A}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, the mediator as $M_i \sim N(h_M(\boldsymbol{A}_i), \sigma^2_M)$, and the health outcome as $y_i \sim N(h_Y(\boldsymbol{A}_i), \sigma^2_Y)$, where $\boldsymbol{\Sigma}$ is the covariance structure of *in utero* exposure to manganese (Mn), arsenic (As), and lead (Pb) after log transform and standardization from our Bangladesh application (Section 2.4), and $h_M(\cdot)$ and $h_Y(\cdot)$ are the dose-response surfaces observed in Bangladesh. The correlation structure of Mn, As, and Pb and graphical summaries of $h_M(\cdot)$ and $h_Y(\cdot)$ are included in Appendix section B.1 on page 48. We set $\sigma^2_M$ and $\sigma^2_Y$ to various signal-to-noise ratios motivated from Bangladesh where we took the estimated residual variances after fitting BKMR models (2.8) and (2.9), $\hat{\sigma}^2_M$ and $\hat{\sigma}^2_Y$ respectively, and divided them by $x = 1, 2, 5, 10, 20$, and 30. We took this approach instead of increasing the sample size to decrease the computational burden of our simulations. For each simulation scenario, we randomly sampled 500 datasets of 300 observations each, $\{Y_i, \boldsymbol{A}_i, M_i\}_{i=1}^{300}$, from the true underlying dataset

of 50,000 subjects.

For each simulation dataset, we first fit BKMR models (2.11) - (2.13) with and without component-wise variable selection:

$$M_i = \beta_0 + h_M(\boldsymbol{A}_i) + \epsilon_{Mi}, \tag{2.11}$$

$$Y_i = \theta_0 + h_Y(\boldsymbol{A}_i, M_i) + \epsilon_{Yi}, \tag{2.12}$$

$$Y_i = \gamma_0 + g(\boldsymbol{A}_i) + \xi_i, \tag{2.13}$$

where $\epsilon_{Mi} \overset{iid}{\sim} N(0, \sigma^2_M)$, $\epsilon_{Yi} \overset{iid}{\sim} N(0, \sigma^2_Y)$, and $\xi_i \overset{iid}{\sim} N(0, \sigma^2_\xi)$. Using our proposed BKMR–CMA approach, we estimated the NDE, NIE, and TE for a change of the exposure mixture from $\boldsymbol{a}^*$, the exposures set equal to their $25^{th}$ percentile in the true underlying dataset, to the $\boldsymbol{a}$, the exposures set equal to their $50^{th}$ percentile in the true underlying dataset. Since the exposures were generated from a multivariate normal with mean 0 and variance 1, all elements in the $\boldsymbol{a}^*$ vectors were extremely close to $\Phi^{-1}(0.25) = Z_{.25} = -0.674$ and all elements in the $\boldsymbol{a}$ vectors were extremely close to $\Phi^{-1}(0.5) = Z_{.5} = 0$. We estimated the CDE for the same change in the exposures, from $\boldsymbol{a}^*$ to $\boldsymbol{a}$, fixing the mediator at the observed $25^{th}$, $50^{th}$, and $75^{th}$ percentiles of the mediator in the corresponding true underling dataset.

Second, we conducted meditation analyses for the joint effect of the metal mixture using both causal mediation methods and traditional approaches. For these analyses, we considered the same change in exposures as with BKMR–CMA, a change in the metals from their $25^{th}$ to $50^{th}$ percentiles in the underlying data for the NDE, NIE, TE, and CDE. For the CDE, we fixed the mediator to its $25^{th}$, $50^{th}$, and $75^{th}$ percentiles in the corresponding true underling dataset. We modeled the mediator and outcome using linear regression allowing for exposure-mediator interactions in the outcome model. Derivations of the closed form solutions to estimate causal mediation effects with multiple exposures are included in Appendix B.3. We refer to the approach that uses linear regression models for both the mediator and outcome and allows for exposure-mediator interactions in the outcome model as the "linear approach" [55, 56, 50]. We also estimated the NDE and NIE without considering exposure-mediator interactions using the product method extended for multiple exposures, we refer to this as the "traditional approach" [6].

| | | rMSE | | | | | | Coverage Probability | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Effect | Models | $\hat{\sigma}^2/1$ | $\hat{\sigma}^2/2$ | $\hat{\sigma}^2/5$ | $\hat{\sigma}^2/10$ | $\hat{\sigma}^2/20$ | $\hat{\sigma}^2/50$ | $\hat{\sigma}^2/1$ | $\hat{\sigma}^2/2$ | $\hat{\sigma}^2/5$ | $\hat{\sigma}^2/10$ | $\hat{\sigma}^2/20$ | $\hat{\sigma}^2/50$ |
| TE | BKMR | 0.043 | 0.033 | 0.023 | 0.017 | 0.013 | 0.009 | 0.964 | 0.972 | 0.968 | 0.962 | 0.958 | 0.960 |
| | BKMR-VS | 0.048 | 0.039 | 0.023 | 0.018 | 0.014 | 0.009 | 0.922 | 0.898 | 0.928 | 0.930 | 0.932 | 0.952 |
| | linear | 0.042 | 0.032 | 0.024 | 0.021 | 0.020 | 0.019 | 0.930 | 0.912 | 0.834 | 0.716 | 0.518 | 0.222 |
| NDE | BKMR | 0.038 | 0.029 | 0.021 | 0.017 | 0.015 | 0.013 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.990 |
| | BKMR-VS | 0.038 | 0.031 | 0.021 | 0.017 | 0.014 | 0.013 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.990 |
| | linear | 0.040 | 0.030 | 0.021 | 0.018 | 0.015 | 0.014 | 0.948 | 0.928 | 0.904 | 0.888 | 0.854 | 0.782 |
| | traditional | 0.039 | 0.029 | 0.021 | 0.019 | 0.018 | 0.019 | 0.946 | 0.926 | 0.882 | 0.812 | 0.720 | 0.498 |
| NIE | BKMR | 0.026 | 0.020 | 0.018 | 0.017 | 0.016 | 0.014 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 |
| | BKMR-VS | 0.026 | 0.021 | 0.015 | 0.015 | 0.015 | 0.014 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 |
| | linear | 0.014 | 0.013 | 0.012 | 0.011 | 0.011 | 0.010 | 0.954 | 0.926 | 0.910 | 0.898 | 0.884 | 0.874 |
| | traditional | 0.013 | 0.012 | 0.010 | 0.009 | 0.008 | 0.006 | 0.944 | 0.932 | 0.934 | 0.928 | 0.934 | 0.956 |
| CDE25 | BKMR | 0.044 | 0.032 | 0.022 | 0.017 | 0.013 | 0.011 | 0.984 | 0.980 | 0.980 | 0.978 | 0.972 | 0.980 |
| | BKMR-VS | 0.042 | 0.038 | 0.031 | 0.023 | 0.016 | 0.012 | 0.926 | 0.876 | 0.866 | 0.910 | 0.932 | 0.964 |
| | linear | 0.050 | 0.038 | 0.029 | 0.026 | 0.025 | 0.025 | 0.936 | 0.918 | 0.812 | 0.706 | 0.552 | 0.322 |
| CDE50 | BKMR | 0.042 | 0.031 | 0.021 | 0.016 | 0.013 | 0.010 | 0.976 | 0.976 | 0.970 | 0.966 | 0.958 | 0.960 |
| | BKMR-VS | 0.039 | 0.036 | 0.030 | 0.022 | 0.015 | 0.011 | 0.902 | 0.848 | 0.850 | 0.900 | 0.928 | 0.932 |
| | linear | 0.042 | 0.032 | 0.024 | 0.020 | 0.019 | 0.019 | 0.918 | 0.898 | 0.850 | 0.782 | 0.686 | 0.468 |
| CDE75 | BKMR | 0.044 | 0.033 | 0.022 | 0.017 | 0.013 | 0.010 | 0.980 | 0.976 | 0.972 | 0.966 | 0.966 | 0.956 |
| | BKMR-VS | 0.033 | 0.031 | 0.028 | 0.021 | 0.015 | 0.012 | 0.936 | 0.892 | 0.878 | 0.900 | 0.938 | 0.936 |
| | linear | 0.048 | 0.034 | 0.023 | 0.018 | 0.015 | 0.014 | 0.952 | 0.938 | 0.934 | 0.916 | 0.904 | 0.802 |

**Table 2.1** Root mean squared error (rMSE) and coverage probabilities from our simulations under varying signal-to-noise ratios comparing our BKMR–CMA approach using BKMR models fit with and without variable selection (BKMR-VS and BKMR respectively), the linear approach, and the traditional approach. CDE25 represents the CDE when the mediator is set to its $25^{th}$ percentile, CDE50 represents the CDE when the mediator is set to its $50^{th}$ percentile, and CDE75 represents the CDE when the mediator is set to it $75^{th}$ percentile in the true underlying dataset.

### 2.3.2 Simulation results

The root mean squared error (rMSE) and coverage probabilities for the TE, NDE, NIE, and CDEs in our simulation are summarized in Table 2.1. We observe similar results for the TE as we do for the CDEs when we intervene to fix the mediator at its $25^{th}$, $50^{th}$, and $75^{th}$ percentiles in the true underlying datasets. As the variance in the data generation decreased, we see a decrease in the rMSE across all effects and approaches considered. Although the rMSE is about the same for the TE and CDEs using our proposed BKMR–CMA approach compared to the linear approach at lower signal-to-noise ratios, there is a noticeable gap in the rMSE at higher signal-to-noise ratios. Specifically for the TE, we see that the rMSE has approached 0.019 for the linear approach, but the rMSE for our BKMR–CMA approach both when the models are fit with and without variable selection is still decreasing at the higher signal-to-noise ratios we considered. We also observe significant differences in the coverage probabilities for the TE and CDEs comparing the linear approach to our BKMR–CMA approach. While our BKMR–CMA approach is conservative, the coverage for the linear approach is very bad, most notable for the TE when the signal-to-noise ratio is high the coverage probability is as low as 0.22.

For the NDE, the rMSE is similar for the linear and our BKMR–CMA approach, and slightly higher for the traditional approach. However, the coverage probabilities of the NDE for both the linear and traditional approaches are lower than acceptable. The rMSE of the NIE is lower for the linear and traditional approaches than our BKMR–CMA approach. This results from the fact that the TE decomposes into the NDE and NIE, thus, the bias of the NIE is function of the magnitude and direction of the biases for the TE and NDE. Although the biases for the TE and NDE are greater in magnitude for the linear approach compared to our BKMR–CMA approach, the are of similar magnitude and operate in the same direction for the linear approach opposed to our BKMR–CMA approach. Therefore, the biases cancel each other out for linear and traditional approaches and the rMSE for the NIE is observed to be lower.

## 2.4 Application to Bangladeshi cohort

### 2.4.1 Study population

We apply our BKMR-CMA methodology to quantify the contribution of birth length (BL) as a mediator between *in utero* co-exposure to arsenic, manganese and lead, and children's neurodevelopment, in a prospective birth cohort in rural Bangladesh. This cohort has previously been described [24, 30, 49]. For the purpose of illustrating our method, we only include mother-infant pairs enrolled at the Pabna clinic and exclude 5 pairs where the infant had outlying birth lengths (BL > 3.7 standard deviations (SD) from the mean), for a total sample of 382. Researchers measured *in utero* metal exposure to arsenic, manganese, and lead from umbilical cord venous blood samples. Collaborators in Bangladesh administered the Bayley Scales of Infant and Toddler Development$^{TM}$, Third Edition (BSID-III$^{TM}$) to children 20-40 months after birth and neurodevelopment was measured as the raw cognitive development score (CS) [7]. We control for child sex, child's age at the time of the Bayley Scale administration, maternal IQ, maternal education (less than high school vs. at least high school), maternal protein intake (low vs. medium vs. high tertiles), secondhand smoke exposure at baseline (smoking environment vs. non-smoking environment), HOME score, and maternal age at delivery in all analyses. When conducting our analyses, we log transformed, centered, and scaled metal concentrations, and centered and scaled CS, BL, and continuous confounder variables.

### 2.4.2 Models

We model the effect of co-exposure to arsenic, manganese, and lead on birth length via a BKMR mediator model (2.8). To model the joint effect of the metal mixture and birth length on neurodevelopment, we fit a BKMR outcome model (2.9) with all three metals and birth length in the kernel function. We graphically examine the relationship between the metal mixture and birth length and the relationship between the metal mixture and birth length on the neurodevelopment.

We simulate counterfactuals to estimate the NDE, NIE, and TE for a change in the raw exposures from $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL), all metals set at their corresponding $25^{th}$ percentile, to $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL), all metals set at their $75^{th}$ percentile. We also calculate the CDE for a change in exposure from $\boldsymbol{a}^*$ to $\boldsymbol{a}$ when birth length is set to its $25^{th}$ percentile of 46cm, median value of 47cm, and $75^{th}$ percentile of 48cm.

To further examine which metals mediate the effect on neurodevelopment through birth length, we calculate the NDE and NIE for a change of a single metal from its $25^{th}$ to $75^{th}$ percentiles, while fixing the other metals at their $25^{th}$, $50^{th}$, or $75^{th}$ percentile values.

### 2.4.3 Results

We found nonlinear associations of the metals and birth length, and of manganese with neurodevelopment, and an interaction between arsenic and manganese on neurodevelopment. There also appears to be small interactions between the metals and birth length on neurodevelopment. Graphical summaries of BKMR mediator, outcome, and total effect models are included in Figure 2.1 and in Appendix section B.1.

Figure 2.1D suggests that the metal mixture has a harmful effect on child neurodevelopment when comparing higher percentiles to lower percentiles of metal exposure. A change in the metal mixture from the median to higher quantiles operates by significantly reducing birth length as seen in Figure 2.1E. We see in Figure 2.1A that manganese is responsible for the majority of the harmful effect of the metal mixture on neurodevelopment and in Figure 2.1B the significant inverse effect of the metal mixture on birth length also is driven by manganese. After adjustment for birth length and intervening to fix birth length at its median value of 47cm, the effect of the metal mixture on neurodevelopment is reduced (Figure 2.1F compared to Figure 2.1D).

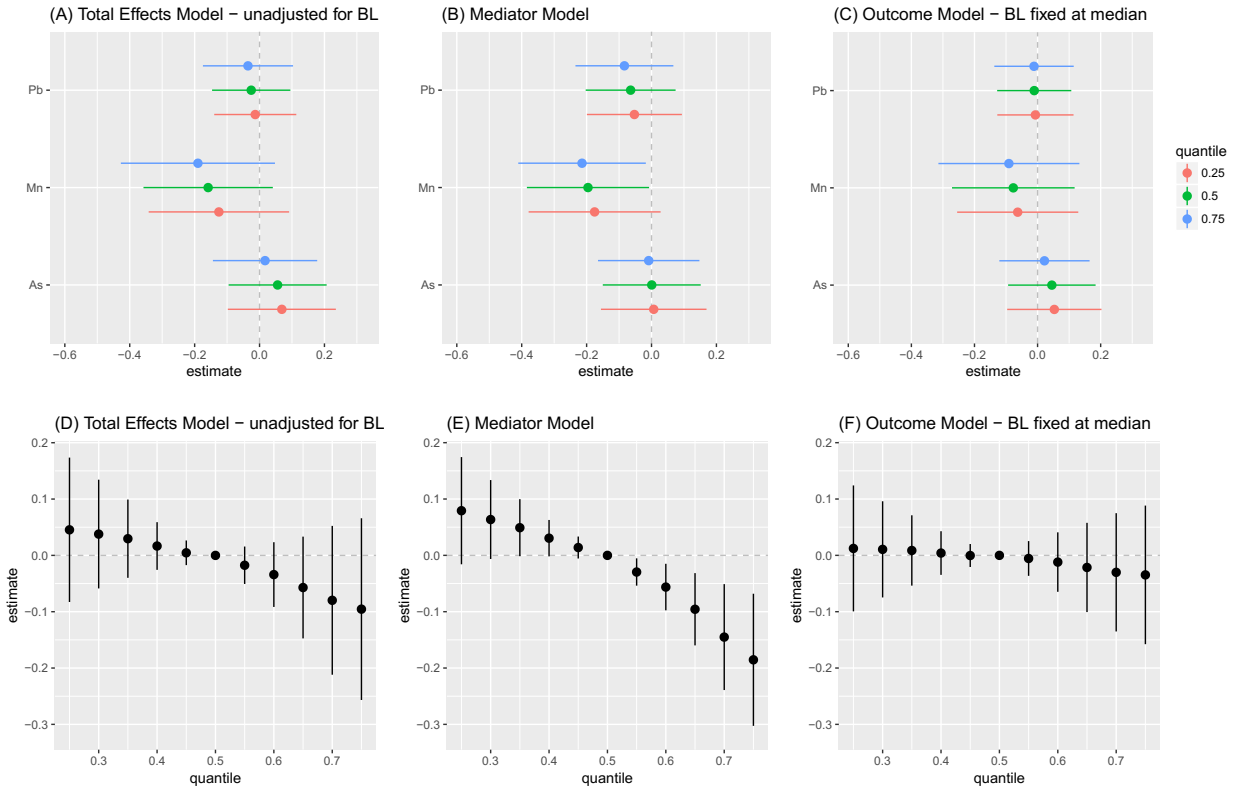


**Figure 2.1** The joint effect of the metal mixture on cognitive score (CS) in our Bangladeshi cohort, adjusted and unadjusted for birth length, and the joint effect of the metal mixture on birth length estimated by Bayesian kernel machine regression (BKMR). (A, C) Overall effect of the metal mixture on cognitive score adjusted and unadjusted for birth length (estimates and 95% credible intervals). These figures shows the average change in neurodevelopment for a change in each element of the metal mixture from a particular percentile to the their median value, from the total effects model (unadjusted for birth length) and the outcome model (fixing birth length at its median value of 47cm). (B) This plot shows the expected change in birth length for a change in the metal mixture described for A/C. (D, F) Single metal associations with neurodevelopment (estimates and 95% CI, gray dashed line a the null). These figures show the average change in neurodevelopment (adjusted and unadjusted for birth length) for a change in a single metal from is $25^{th}$ to $75^{th}$ percentile values, fixing the other metals at their $25^{th}$, $50^{th}$, or $75^{th}$ percentiles. (E) Single metal associations with birth length.

Table 2.2 summarizes the mediation effects for a change of the raw metal mixture from $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL) to $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL). We see a negative association between the metal mixture and neurodevelopment, independent of birth length, comparing the co-exposure of metals at their 75th percentiles to their 25th percentiles, although not significant (NDE: -0.04, 95% CI: (-0.27, 0.18)). The percent mediated through birth length was estimated to be 73%, and a negative indirect effect of birth length was observed (NIE: -0.10, 95% CI: (-0.34, 0.14)). Upon hypothetical intervention to fix birth length at the 75th percentile value of 48cm, the direct effect was reduced, suggesting, targeted interventions on fetal growth can block part of the adverse effect

| Effect | Estimate (95% CI) |
|---|---|
| TE | -0.14 (-0.36, 0.03) |
| NDE | -0.04 (-0.27, 0.18) |
| NIE | -0.10 (-0.34, 0.14) |
| CDE($M_{.25}$=46cm) | -0.06 (-0.28, 0.12) |
| CDE($M_{.50}$=47cm) | -0.04 (-0.25, 0.13) |
| CDE($M_{.75}$=48cm) | -0.02 (-0.22, 0.17) |

**Table 2.2** Mediation effects estimated in Bangladesh using BKMR–CMA. All effects are estimated for a change of the mixture $\boldsymbol{A}$ =(As, Mn, Pb) from its raw $25^{th}$ percentile $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL) to its raw $75^{th}$ percentile $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL). The CDE are calculated as the direct effect from $\boldsymbol{a}^*$ to $\boldsymbol{a}$ intervening to fix the mediator at its $25^{th}$, $50^{th}$, $75^{th}$ percentiles values of 46, 47, and 48cm respectively.

of metals on neurodevelopment (CDE: -0.02, 95% CI: (-0.22, 0.17)). Due to the flexibility of our BKMR–CMA, we require greater sample sizes to detect mediation effects. In our Bangladeshi cohort, we are not even powered to estimate the TE using BKMR, so although the mediation effects are not significant, the trends observed in the data are informative.

In Figure 2.2, we observe the negative association between the metal mixture and neurodevelopment is driven by manganese, both indirectly through birth length and direct. When manganese and arsenic are set to higher levels, lead is mediating some of effect of the metal mixture on neurodevelopment.

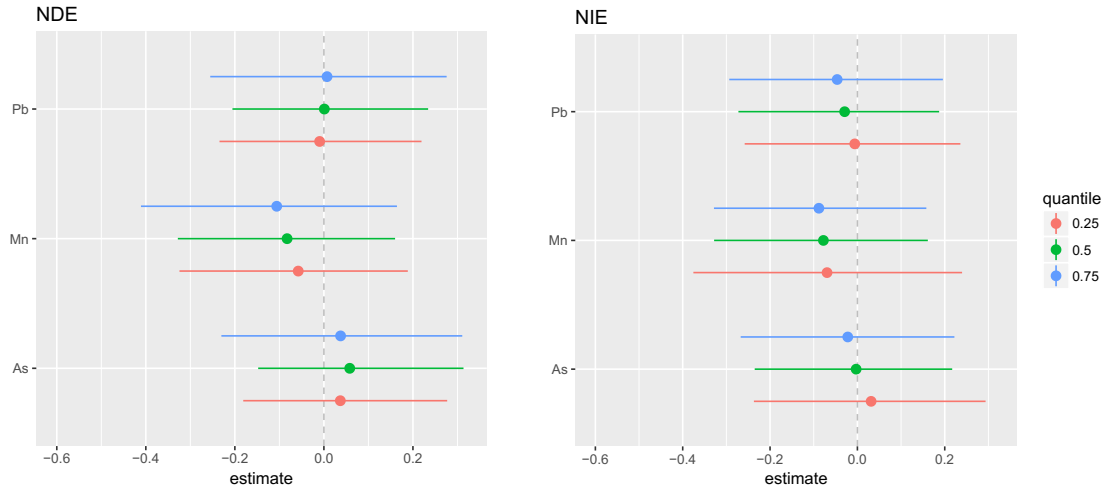


**Figure 2.2** Single metal NDE and NIE effects on neurodevelopment (estimates and 95% CI). These figures show the NDE and NIE for a change in one metal from its $25^{th}$ to $75^{th}$ percentile values, fixing the other metals at their $25^{th}$, $50^{th}$, or $75^{th}$ percentiles.

## 2.5 Discussion

We have proposed Bayesian kernel machine regression – causal mediation analysis (BKMR–CMA) as a way to estimate the direct and indirect effects of an environmental mixture on a outcome through an intermediate variable. To our knowledge, this is the first method presented in the causal inference literature to estimate these effects when the exposure of interest is a potentially complex mixture, without *a priori* knowledge of the exposure-mediator or exposure-mediator-outcome relationship. This method allows for complex relationships between the elements of the mixture and the mediator variable through the joint kernel specification in the outcome model. Our extension of causal mediation methodology that allows for a mixture of exposures is important for many environmental health applications.

We estimate the TE, NDE, NIE, and CDEs through simulation of counterfactuals from the posterior predictive distribution for each MCMC iteration and make inference from these posterior samples of the mediation effects. Our simulation showed our proposed BKMR–CMA approach performs better than current methods to estimate the TE, NDE, and CDEs. We observe noticeable differences in the coverage probability for the linear and traditional approaches compared to our BKMR–CMA approach. In the presence of complex data generation scenarios, we advise to use our approach over other methods.

Applying these methods to a prospective Bangladeshi birth cohort, we found a negative association of co-exposure to lead, arsenic, and manganese on neurodevelopment, a negative association of exposure to this metal mixture on birth length, and some evidence that birth length mediates the effect of co-exposure to lead, arsenic, and manganese on children's neurodevelopment. If birth length were fixed to its $75^{th}$ percentile value of 48cm, the effect of the metal mixture on neurodevelopment is smaller, suggesting that nutritional interventions to help increase birth length could potential block the harmful effect of the metal mixture.

Our BKMR–CMA algorithm easily extends beyond linear models for the mediator and outcome. If the outcome is binary, the logistic regression option in the `bkmr` R package can be used and our code be implemented to estimate the mediation effects. While one can also consider variable selection with BKMR–CMA for high dimensional exposures, we did not apply it to our Bangladeshi cohort because we were interested in the effect of three metals, so high dimensionality is not a primary concern here. The general approach we present can be used to estimate mediation effects for any Bayesian mediator and outcome models.

Many limitations of our method are due to the exponentially increasing computation time required to fit

BKMR and predict counterfactuals as the number of exposures, covariates, and sample size increase. In many applications, exposure to mixtures with more than three elements is common. In the presence of a high dimensional exposure, simulations studies would need be conducted to see how our method preforms. In the current formulation of our algorithm, we assume the mediator and outcome models are independent. Although this is a common assumption in causal mediation literature, this is a limitation of our methods. In our data application, our results are limited by potential residual confounding by malnutrition and low power due to a small sample size.

In future work, we plan to consider joint specification of the mediator and outcome models to reduce the assumptions need for BKMR–CMA to be interpreted causally. We also hope to extend these methods to allow for multiple mediators and/or multiple outcomes.

# 3

# Integrative causal mediation analysis for multiple exposure, mediator, and outcome measures

Katrina L. Devick[1], Jennifer F. Bobb[2], Maitreyi M. Mazumdar[3,4], Birgit Claus Henn[5], David C. Bellinger[3,4], David C. Christiani[4], Robert O. Wright[6], Paige L. Williams[1], Brent A. Coull[1,4], and Linda Valeri[7,8]

[1]*Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA*

[2]*Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA*

[3]*Department of Neurology, Boston Children's Hospital, Boston, MA, USA*

[4]*Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[5]*Department of Environmental Health, Boston University School of Public Health, Boston, MA, USA*

[6]*Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

[7]*Psychiatric Biostatistics Laboratory, McLean Hospital, Belmont, MA, USA*

[8]*Department of Psychiatry, Harvard Medical School, Boston, MA, USA*

## Abstract

More powerful methodology is needed to estimate mediation effects when flexible models are used to capture the relationship between multiple exposure, mediator, and outcome measures. We consider the case when both the mediator and outcome variables of interest cannot be fully captured by one measurement, but rather are constructs reflected by multiple measures. In this paper, we present a framework to estimate the total effect of a potentially complex mixture jointly on the multiple outcome measurements and a decomposition of this effect into the direct effect of the exposures on the outcome measures and the indirect effect that operates through multiple mediator instruments. We do this by jointly modeling the mediator measurements

and jointly modeling the outcome measurements using Bayesian kernel machine regression (BKMR) with subject-specific intercepts. In a case study of a prospective Bangladeshi birth cohort, we observe a harmful total effect of co-exposure to manganese, arsenic, and lead on children's neurodevelopment, with about 50% of this effect operating through fetal growth. When first combining information from three mediator measures, we observe a 11-29% decrease in the width of the credible intervals for the mediated effects, and an additional 13-20% decrease when also pooling information from two outcome measurements.

This work was supported by grants NIH T32ES007142, NIH P01 CA134294, and NIH ES000002.

**Keywords**: Bayesian kernel machine regression; Children's neurodevelopment; Metal mixture; Mixed models; Multiple mediators; Multiple outcome model; Multi-pollutant exposure.

## 3.1 Introduction

More powerful methodologies are needed to quantify complex relationships between exposure mixtures and observed phenotypes, and determine if the relationship is operating through intermediate (mediator) variables. In many life science contexts, interest lies in understanding the mechanisms through which complicated outcomes arise. These phenotypes are complex in that more than one measure is needed to fully quantify them. In this setting, information needs to be combined across multiple correlated outcome measurements. It is possible that the exposure mixture of interest operates through an intermediate phenotype that is also composed of more than one measure. For example, it is of interest to quantify the effect of *in utero* co-exposure to manganese, arsenic, and lead on children's neurodevelopment, and how much of this effect is operating through growth measures at birth. There are many aspects to consider for children's neurodevelopment, such as thinking, reasoning, motor development, and language, and multiple measures to quantify fetal growth, such as birth weight, birth length, and head circumference.

There are a number of papers that address the problem of estimating direct and indirect effects in the presence of multiple mediators [3, 1, 63, 27, 53, 52, 51]. Most of these papers use one of two general approaches. They either estimate the joint effect of the mediators or they estimate path-specific effects. To our knowledge, no methods are available to estimate the direct and indirect effects of multiple mediators when the exposure of interest is a mixture. The main distinction between the multiple mediator setting and our setting is that the multiple mediator measures we consider are thought to all reflect a single underlying mediator, e.g.

fetal growth, rather than separate mediator variables. Accordingly, we are interested in estimating the joint indirect effect through this underlying mediator rather than separate effects for each of the multiple mediator measures.

Numerous papers have addressed the question of combining information across several outcome measures in a single model. There are two general approaches to address this issue. The first approach introduces one or more latent variables for the outcome and views the multiple outcome measurements as different manifestations of these latent variables [41, 21, 11, 10, 42]. The second approach directly models the effect of the exposure on the outcomes using generalized linear mixed models (GLMMs) to account for the correlation among each subject's measurements through subject-specific effects [41, 34, 17, 39, 45]. It is known that combining information across multiple outcome measurements improves power. However, to our knowledge, no one has combined information from multiple instruments for both the mediator and outcome in mediation analysis.

Typically, when data on multiple mediators and outcomes are available, studies conduct separate mediation analyses for each mediator-outcome pair. Statistical power is a concern when conducting any mediation analysis, as the power to detect direct and indirect effects of an exposure on an outcome through a mediator variable is less than the power to detect a total effect of an exposure on an outcome [5]. This problem of low power is a compounded when conducting separate mediation analyses since researchers need to adjust for multiple comparisons. When the effect of the exposure of interest on the multiple mediators and the effect of the exposure and mediators on the multiple outcomes are similar, power can be gained by estimating global direct and indirect effects across the multiple mediators and outcomes. Additional complexity arises when considering an environmental mixture as the exposure. It is desirable to allow for complex dose-response relationships; however, a more flexible model is generally less powerful. To boost our power to detect nonlinear and non-additive effects of a mixture within a causal mediation framework, we propose a joint mediator and joint outcome model approach.

In this paper, we extend Bayesian kernel machine regression – causal mediation analysis (BMKR–CMA) to allow the mediator and/or outcome to be complex phenotypes best characterized by multiple measurements [20]. To do this, we implement Bayesian kernel machine regression (BKMR) to model both the mediator measurements and the outcome measurements with subject-specific and outcome-specific intercepts [17, 45, 8]. We account for the multiple mediator measurements when modeling the outcomes through group variable

selection within the kernel function [8]. We compare the efficiency gains in a two staged fashion. First, we combine information across the mediator measurements and compare the efficiency to separate mediation analyses for each mediator-outcome pair. Second, we additionally combine information across the multiple outcome measurements and compare the efficiency gain to case where only information is combined across the mediator measurements. To our knowledge, this is the first time ideas behind the mixed model framework for multiple outcome models has been incorporated within a causal mediation analysis.

We apply our methodology to a case study in rural Bangladesh to examine the direct and indirect effects of *in utero* co-exposure to manganese, arsenic, and lead through fetal growth on neurodevelopment. We quantify efficiency gains by first combining information across multiple growth measures at birth, and then, from further combining information from multiple neurodevelopment scores at 20-40 months.

## 3.2 Materials and Methods

### 3.2.1 Bayesian kernel machine regression

We first briefly review BKMR presented by Bobb et al. (2015) as a framework to estimate the effect of a complex mixture on a health outcome [8]. For each subject $i = 1, \ldots, n$, we assume:

$$Y_i = h\left(\boldsymbol{z}_i\right) + \boldsymbol{C}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{3.1}$$

where $Y_i$ is a continuous health outcome, $\boldsymbol{z}_i = \left(z_{1i}, \ldots, z_{Li}\right)^T$ is a vector of $L$ exposure variables (e.g. metals), $\boldsymbol{C}_i = \left(C_{1i}, \ldots, C_{Pi}\right)^T$ is a vector of potential confounders, and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Model (3.1) relates the outcome to the exposure mixture through a flexible function, $h(\cdot)$, which accommodates for nonlinearity and/or interaction among the mixture components. We specify a positive-definite kernel function $K(\cdot, \cdot)$ to identify $h(\cdot)$ and focus on the Gaussian kernel.

When the exposure mixture is comprised of numerous elements, it may be of interest to fit (3.1) with component-wise variable selection. To allow for variable selection, the kernel function $K(\cdot, \cdot)$ is augmented.

In the case of the Gaussian kernel, the kernel function is expanded as:

$$K(\boldsymbol{z}, \boldsymbol{z}^{\mathsf{T}}; \boldsymbol{r}) = \exp\left\{-\sum_{\ell=1}^{L} r_\ell (z_\ell - z_\ell^{\mathsf{T}})^2\right\}, \tag{3.2}$$

where $\boldsymbol{r} = (r_1, \ldots, r_\ell)^{\mathsf{T}}$, and we assume a "slab-and-spike" prior for the auxiliary parameters,

$$\begin{aligned} r_\ell \,|\, \delta_\ell &\sim \delta_\ell f_1(r_\ell) + (1 - \delta_\ell) P_0, \quad m = 1, \ldots, M, \\ \delta_\ell &\sim \text{Bernoulli}(\pi), \end{aligned} \tag{3.3}$$

where $\delta_\ell$ is an indicator that element $z_\ell$ is included in the kernel, $f_1(\cdot)$ denotes a pdf with support on $\mathbb{R}^+$ and $P_0$ is the density with a point mass at 0.

When elements included in the kernel function are highly correlated, an additional step of variable selection may be implemented. We can partition the mixture components $z_1, \ldots, z_\ell$ into groups $\mathcal{S}_g$ $(g = 1, \ldots, G)$, such that the correlation among elements in a group is high while the correlation across groups is low. To execute group variable selection, we assume the indicator variables from the slab-and-spike prior in (3.3) have the following distributions:

$$\begin{aligned} \boldsymbol{\delta}_{\mathcal{S}_g} \,|\, \omega_g &\sim \text{Multinomial}(\omega_g, \boldsymbol{\pi}_{\mathcal{S}_g}), \quad g = 1, \ldots, G, \\ \omega_g &\sim \text{Bernoulli}(\pi), \end{aligned} \tag{3.4}$$

where $\boldsymbol{\delta}_{\mathcal{S}_g} = (\delta_\ell)_{z_\ell \in \mathcal{S}_g}$ is a vector of indicator variables representing whether element $z_\ell$ in group $\mathcal{S}_g$ is included in the kernel and $\boldsymbol{\pi}_{\mathcal{S}_g}$ is the corresponding vector of prior probabilities. With this specification, at most one element from each group is included in the kernel at a time. For additional details regarding BKMR and prior specification, see Bobb et al. [8].

### 3.2.2 Bayesian kernel machine regression – causal mediation analysis

Next, we will review Bayesian kernel machine regression – causal mediation analysis (BMKR–CMA) as a means to decompose the total effect of an exposure mixture on an outcome into the effect that operates directly from the mixture to the outcome and the effect that operates indirectly through a mediator variable [20]. Let $Y_{\boldsymbol{a}m}$ denote the counterfactual outcome $Y$ if the levels of the exposure mixture $\boldsymbol{A}$ were set to

$\boldsymbol{a}$ and mediator level $M$ was set to $m$. Let $M_{\boldsymbol{a}}$ be the counterfactual mediator level $M$ that would have been observed if the levels of the exposure mixture $\boldsymbol{A}$ were set to $\boldsymbol{a}$. Accordingly, $Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}$ represents the counterfactual outcome $Y$ if the levels of the exposure mixture $\boldsymbol{A}$ were set to $\boldsymbol{a}$ and the mediator $M$ was set to the level it would have taken if the levels of the exposure mixture $\boldsymbol{A}$ were $\boldsymbol{a}^*$. The effects of interest, the natural direct effect (NDE), the natural indirect effect (NIE), and the controlled direct effects (CDEs), are formally defined as:

$$\begin{aligned} NDE &= \mathrm{E}\left[Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}} - Y_{\boldsymbol{a}^*M_{\boldsymbol{a}^*}}\right], && (3.5)\\ NIE &= \mathrm{E}\left[Y_{\boldsymbol{a}M_{\boldsymbol{a}}} - Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}\right], && (3.6)\\ CDE(m) &= \mathrm{E}\left[Y_{\boldsymbol{a}m} - Y_{\boldsymbol{a}^*m}\right]. && (3.7) \end{aligned}$$

Devick et al. (2018) considered a single health outcome $Y$, single mediator variable $M$, exposure mixture $\boldsymbol{A}$ comprised of $L$ components, and confounder matrix $\boldsymbol{C}$ [20]. To estimate the TE, NDE, NIE, and CDEs allowing for potential complex dose-response relationships between the exposure mixture, mediator, and outcome, first fit the following BKMR models [8]:

$$\begin{aligned} M_i &= h_M(\boldsymbol{A}_i) + \boldsymbol{C}_i^T\boldsymbol{\beta} + \epsilon_i^M, && (3.8)\\ Y_i &= h_Y(\boldsymbol{A}_i, M_i) + \boldsymbol{C}_i^T\boldsymbol{\theta} + \epsilon_i^Y, && (3.9)\\ Y_i &= g(\boldsymbol{A}_i) + \boldsymbol{C}_i^T\boldsymbol{\gamma} + \xi_i, && (3.10) \end{aligned}$$

where $\epsilon_i^M \overset{iid}{\sim} N(0, \sigma_M^2)$, $\epsilon_i^Y \overset{iid}{\sim} N(0, \sigma_Y^2)$, and $\xi_i \overset{iid}{\sim} N(0, \sigma_\xi^2)$. Then, simulate posterior samples of $Y_{\boldsymbol{a}^*M_{\boldsymbol{a}^*}}$, $Y_{\boldsymbol{a}M_{\boldsymbol{a}}}$, $Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}$, $Y_{\boldsymbol{a}^*m}$, and $Y_{\boldsymbol{a}m}$ from the posterior predictive distributions. From these chains, posterior samples of the NDE, NIE, and CDEs are calculated and inference is made from these samples. For details about the algorithm to estimate the mediation effects using BKMR–CMA, please refer to Devick et al. (2018) [20].

### 3.2.3 BKMR–CMA for a latent mediator and latent outcome

Now, we consider the extension to BMKR–CMA when both the mediator and outcome are latent variables best captured by multiple measurements. We proposed a single mediator, outcome, and total effects model

to combine information from these multiple measurements. Let $\boldsymbol{M}_i = (M_{i1}, M_{i2}, \ldots, M_{iR})$ be the $R$ mediator measurements for subject $i$ and $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iS})$ be the $S$ outcome measurements for subject $i$. To flexibly model the exposure-mediator relationship and account for the correlation among the multiple mediator measurements, we extend (3.8) to include subject-specific intercepts:

$$M_{ir} = \beta_i + \beta_r + h_M(\boldsymbol{A}_i) + \boldsymbol{\beta_c^\top} \boldsymbol{C}_i + \epsilon_{ir}^M, \tag{3.11}$$

where $i = 1, \ldots, N$, $\beta_i \overset{iid}{\sim} N(0, \sigma^2_{b_M})$, $\epsilon_{ir}^M \overset{iid}{\sim} N(0, \sigma^2_M)$, and $\beta_r$ represents the fixed effect for each mediator measure where $r = 1, \ldots, R$. Here, we consider $\beta_r$ to be fixed effects since we only consider three mediator measurements in our analysis. We also include subject-specific intercepts when modeling the outcome measurements. To allow for all possible exposure-mediator interactions, we include the multiple mediator measurements in the kernel function with the exposure mixture when modeling the outcome:

$$Y_{is} = \theta_i + \theta_s + h_Y(\boldsymbol{A}_i, \boldsymbol{M}_i) + \boldsymbol{\theta_c^\top} \boldsymbol{C}_i + \epsilon_{is}^Y, \tag{3.12}$$

where $i = 1, \ldots, N$, $\theta_i \overset{iid}{\sim} N(0, \sigma^2_{b_Y})$, $\epsilon_{is}^Y \overset{iid}{\sim} N(0, \sigma^2_Y)$, and $\theta_s$ represents the fixed effect for each outcome measure where $s = 1, \ldots, S$. Here, we consider $\theta_s$ to be fixed effects since we are only considering two outcome measurements in our analysis. When fitting model (3.12), we perform group variable selection for the elements in the kernel. Since the mediator measurements can be highly correlated, our variable selection first decides if the entire group of the mediators are to be included in the model at each Markov chain Monte Carlo (MCMC) iteration. Then, if the mediator measurements are selected for a particular MCMC iteration, an additional variable selection procedure is implemented to determine which of the $r$ mediator measurements is to be included in the model. The exposure elements are selected component-wise, but they can be partitioned into groups for the group variable selection if they are highly correlated. To model the total effect of the exposure mixture on the outcome measurements, we consider BKMR model (3.13):

$$Y_{is} = \gamma_i + \gamma_s + g(\boldsymbol{A}_i) + \boldsymbol{\gamma_c^\top} \boldsymbol{C}_i + \xi_{is}, \tag{3.13}$$

where $i = 1, \ldots, N$, $\gamma_i \overset{iid}{\sim} N(0, \sigma^2_b)$, $\xi_{is} \overset{iid}{\sim} N(0, \sigma^2_\xi)$, and $\gamma_s$ represents the fixed effect for each outcome measure where $s = 1, \ldots, S$. For model details and prior specification, see Bobb et al. (2015) [8].

We estimate the global NDE, NIE, and TE for a change in exposure profile from $\boldsymbol{a}^*$ to $\boldsymbol{a}$ via the following algorithm.

1. Fit joint BKMR mediator, outcome, and total effect models (3.11), (3.12), and (3.13), respectively.
2. For each MCMC iteration, $j = 1, \ldots, J$:
   (a) Sample $k = 1, \ldots, K$ replicates of each mediator $r = 1, \ldots, R$ for the mean level of covariates under exposure level $\boldsymbol{a}^*$ from mediator model (3.11):

$$\begin{aligned} M_{r\boldsymbol{a}^*}^{(jk)}(\bar{\boldsymbol{c}}) &= \mathrm{E}^{(j)}(M_r|\boldsymbol{A} = \boldsymbol{a}^*, \boldsymbol{C} = \bar{\boldsymbol{c}}) + \sigma_M^{(j)} N(0,1) \\ &= \beta_r^{(j)} + h_M^{(j)}(\boldsymbol{A}_i = \boldsymbol{a}^*) + \bar{\boldsymbol{c}}^\mathsf{T}\boldsymbol{\beta}_{\boldsymbol{c}}^{(j)} + \sigma_M^{(j)} N(0,1). \end{aligned}$$

   (b) For each of the $j = 1, \ldots, J$ and $k = 1, \ldots, K$ samples of $\boldsymbol{M_a} = (M_{1\boldsymbol{a}}, \ldots, M_{R\boldsymbol{a}})$ and $\boldsymbol{M}_{\boldsymbol{a}^*} = (M_{1\boldsymbol{a}^*}, \ldots, M_{R\boldsymbol{a}^*})$, estimate the average outcome value for the reference group ($s = 1$) at the mean level of covariates for $Y_{\boldsymbol{aM}_{\boldsymbol{a}^*}}$ from outcome model (3.12):

$$\begin{aligned} Y_{\boldsymbol{aM}_{\boldsymbol{a}^*}}^{(jk)}(\bar{\boldsymbol{c}}) &= \mathrm{E}^{(j)}(Y_{s=1}|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{M} = \boldsymbol{M}_{\boldsymbol{a}^*}^{(jk)}(\bar{\boldsymbol{c}}), \boldsymbol{C} = \bar{\boldsymbol{c}}) \\ &= \theta_{s=1}^{(j)} + h_Y^{(j)}(\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{M} = \boldsymbol{M}_{\boldsymbol{a}^*}^{(jk)}(\bar{\boldsymbol{c}})) + \bar{\boldsymbol{c}}^\mathsf{T}\boldsymbol{\theta}_{\boldsymbol{c}}^{(j)}. \end{aligned}$$

   (c) Let the $j^{th}$ posterior sample of $Y_{\boldsymbol{aM}_{\boldsymbol{a}^*}}$ be $Y_{\boldsymbol{aM}_{\boldsymbol{a}^*}}^{(j)}(\bar{\boldsymbol{c}}) = \frac{1}{K}\sum_{k=1}^{K} Y_{\boldsymbol{aM}_{\boldsymbol{a}^*}}^{(jk)}(\bar{\boldsymbol{c}})$
   (d) Since $Y_{\boldsymbol{a}^*} = Y_{\boldsymbol{a}^*\boldsymbol{M}_{\boldsymbol{a}^*}}$ and $Y_{\boldsymbol{a}} = Y_{\boldsymbol{aM_a}}$, we sample the $j^{th}$ posterior sample of $Y_{\boldsymbol{a}^*}$ and $Y_{\boldsymbol{a}}$ from the total effect model instead of sampling $Y_{\boldsymbol{a}^*\boldsymbol{M}_{\boldsymbol{a}^*}}$ and $Y_{\boldsymbol{aM_a}}$ for ease of computation. Calculate the average outcome value for the reference group (this needs to be the same reference group as in (b), $s = 1$) for the mean level of covariates at $\boldsymbol{a}_{\text{new}} = \begin{pmatrix} \boldsymbol{a} & \boldsymbol{a}^* \end{pmatrix}^\mathsf{T}$ from (3.13):

$$\begin{aligned} Y_{\boldsymbol{a}_{\text{new}}}^{(j)}(\bar{\boldsymbol{c}}) &= \mathrm{E}^{(j)}(Y_{s=1}|\boldsymbol{A} = \boldsymbol{a}_{\text{new}}, \boldsymbol{C} = \bar{\boldsymbol{c}}) \\ &= \gamma_{s=1}^{(j)} + g^{(j)}(\boldsymbol{A} = \boldsymbol{a}_{\text{new}}) + \bar{\boldsymbol{c}}^\mathsf{T}\boldsymbol{\gamma}_{\boldsymbol{c}}^{(j)} \end{aligned}$$

3. Obtain the $j^{th}$ posterior sample of the global NDE, NIE, and TE by:

$$\begin{aligned} NDE_{\text{global}}^{(j)} &= Y_{\boldsymbol{aM}_{\boldsymbol{a}^*}}^{(j)}(\bar{\boldsymbol{c}}) - Y_{\boldsymbol{a}^*}^{(j)}(\bar{\boldsymbol{c}}), \\ NIE_{\text{global}}^{(j)} &= Y_{\boldsymbol{a}}^{(j)}(\bar{\boldsymbol{c}}) - Y_{\boldsymbol{aM}_{\boldsymbol{a}^*}}^{(j)}(\bar{\boldsymbol{c}}), \\ TE_{\text{global}}^{(j)} &= Y_{\boldsymbol{a}}^{(j)}(\bar{\boldsymbol{c}}) - Y_{\boldsymbol{a}^*}^{(j)}(\bar{\boldsymbol{c}}). \end{aligned}$$

4. Estimate the global NDE, NIE, TE, and their 95% credible intervals from these posterior samples.

Four no unmeasured confounding assumptions are required for the global NDE and NIE to have a causal interpretation: (i) $\boldsymbol{Y_{am}} \amalg \boldsymbol{A}|\boldsymbol{C}$, (ii) $\boldsymbol{Y_{am}} \amalg \boldsymbol{M}|\boldsymbol{C}, \boldsymbol{A}$, (iii) $\boldsymbol{M_a} \amalg \boldsymbol{A}|\boldsymbol{C}$, and (iv) $\boldsymbol{Y_{am}} \amalg \boldsymbol{M}_{\boldsymbol{a}^*}|\boldsymbol{C}$. Namely, there

are no unmeasured exposure-outcome confounders, there are no unmeasured mediator-outcome confounders, there are no unmeasured exposure-mediator confounders, and the exposure does not affect any mediator-outcome confounders.

To estimate the global CDE, only two no unmeasured confounding assumptions are required: (i) and (ii). The algorithm to estimate the global CDE is similar to the algorithm presented above. We include explicit steps to estimate the global CDE in Appendix C.2.

## 3.3 Case study of a Bangladeshi cohort

### 3.3.1 Study population

We apply our joint BKMR–CMA to estimate the global total effect of *in utero* co-exposure to manganese, arsenic, and lead across two neurodevelopment scores and how much of this global effect operates through three fetal growth measures in a prospective birth cohort in rural Bangladesh. This cohort has previously been described [24, 30, 49]. For the purpose of illustrating our method, we only include mother-infant pairs enrolled at the Pabna clinic and exclude five pairs where the infant had outlying birth lengths (BL > 3.7 standard deviations (SD) from the mean), for a total sample of 382. Researchers measured *in utero* metal exposure to arsenic, manganese, and lead from umbilical cord venous blood samples. Collaborators in Bangladesh administered the Bayley Scales of Infant and Toddler Development$^{TM}$, Third Edition (BSID–III$^{TM}$) to children 20-40 months after birth [7]. We consider two domains of neurodevelopment as measured by BSID–III: the raw cognitive development score (cogn) and the raw fine motor score (fmotor). To assess fetal growth, we consider measurements for birth length (BL; cm), birth weight (BW; kg), and head circumference (HC; cm). We control for child sex, child's age at the time of the Bayley Scale administration, maternal IQ, maternal education (less than high school vs. at least high school), maternal protein intake (low vs. medium vs. high tertiles), secondhand smoke exposure at baseline (smoking environment vs. non-smoking environment), HOME score, and maternal age at delivery in all analyses. When conducting our analyses, we log transform, center, and scale metal concentrations, and center and scale the neurodevelopment and fetal growth measures, and continuous confounder variables.

### 3.3.2 Models

To quantify the efficiency gain of combining information from multiple mediator measurements and multiple outcome measurements, we first conduct individual BKMR–CMA for each mediator-outcome combination. To be consistent throughout our comparisons, we fit each BKMR model with component-wise variable selection for the elements in the kernel. We simulate counterfactuals to estimate the NDE, NIE, and TE for a change in the raw exposures from $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL), all metals set at their corresponding $25^{th}$ percentile, to $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL), all metals set at their $75^{th}$ percentile. To fix the overall type 1 error rate to 0.05 for the 6 mediator-outcome BKMR–CMA analyses, we use a Bonferroni correction when calculating the credible intervals. Thus, for each of the mediation effects, we take the $\frac{0.05}{6\times2}$ and $1 - \frac{0.05}{6\times2}$ quantiles of the posterior chains as the credible intervals.

Since we desire to isolate the efficiency gain by first combining information across the multiple mediator measurements, we perform BKMR–CMA for a change in exposure from $\boldsymbol{a}^*$ to $\boldsymbol{a}$ with a joint mediator model, but separate outcome models:

$$M_{ir} = \beta_i + \beta_r + h_M(\boldsymbol{A}_i) + \boldsymbol{\beta}_{\boldsymbol{c}}^{\mathsf{T}}\boldsymbol{C}_i + \epsilon_{ir}^{M}, \tag{3.14}$$

$$Y_i = h_Y(\boldsymbol{A}_i, \boldsymbol{M}_i) + \boldsymbol{\theta}_{\boldsymbol{c}}^{\mathsf{T}}\boldsymbol{C}_i + \epsilon_i^{Y}. \tag{3.15}$$

We fit (3.14) with component-wise variable selection and (3.15) with group variable selection in the kernel. For (3.15), the elements of $\boldsymbol{A}$ are chosen component-wise and the mediator measurements $\boldsymbol{M}$ are first selected (or not selected) as a group. If $\boldsymbol{M}$ is selected, then component-wise variable selection decides which of the three mediator measurements is included in the kernel, for each MCMC iteration. To control the type 1 error rate for these two BKMR–CMA analyses, we take the $\frac{0.05}{2\times2}$ and $1 - \frac{0.05}{2\times2}$ quantiles of the posterior chains as the credible intervals for the NDE, and NIE. Note that the TE when combining information across the mediators is the same as the TE when considering separate mediator-outcome BKMR–CMA.

Next, we fit our proposed joint BKMR–CMA approach to estimate the global TE, NDE, and NIE, for a change in exposure from $\boldsymbol{a}^*$ to $\boldsymbol{a}$, across the two neurodevelopment scores and three fetal growth measurements. To quantify the efficiency gain from fitting six separate mediator-outcome BKMR–CMA analyses to combining information from the three mediator measurements, we divide the width of the credible interval

of a given effect (NDE or NIE) for a given outcome using the joint mediator model by the width of the corresponding credible interval when each of the three mediators are modeled separately. To further evaluate the efficiency gain when combining information from the two outcome measurements, we divide the width of the credible interval for the global NDE or NIE by the width of the credible interval for the NDE or NIE at each outcome when the mediators are modeled jointly. Since the TE is the same when combining information across the mediators compared to when the mediator-outcome pairs are considered separately, we can only quantify the efficiency gain from combining information across outcomes. To examine the efficiency gain for the TE when combining information across the two outcome measurements, we divide the width of the credible interval for the global TE by the width of the credible interval for the TE when the outcomes are modeled separately.

### 3.3.3 Results

Estimates and 95% credible intervals for the TE, NDE, NIE, and CDEs from the global, joint mediator, and separate mediator-outcome BKMR–CMA for a change in the raw exposures from $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL), all metals set at their corresponding $25^{th}$ percentile, to $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL), all metals set at their $75^{th}$ percentile, are in Appendix section C.1 on page 53. We include visual summaries of the posterior distributions of these effects in Figure 3.1. After combining information across the two outcome measurements, we observe stronger evidence of a harmful total effect of co-exposure to manganese, arsenic, and lead on children's neurodevelopment (TE:-0.11, 95% CI: (-0.26, 0.04)). The global NIE is driven by birth length, as birth length mediates about 90% of the harmful effect metal exposure on cognitive score and about 70% on fine motor score. Head circumference and birth weight mediate less than 10% of the harmful effect of the metal mixture on both neurodevelopment scores.

We summarize the efficiency gain from first combining information across our three mediator measurements, then from further pooling information across our two outcome measurements in Table 3.1. For the TE, we see a 31% decrease in the width of the credible interval for the TE of the metal mixture on cognitive score and a 16% decrease for fine motor score. There is notable efficiency gain in both stages of our analysis. We observe a 14% to 29% decrease in the width of the credible interval for the NDE and a 11% to 25% decrease for the NIE when combining information across our three mediator measurements. When further combining

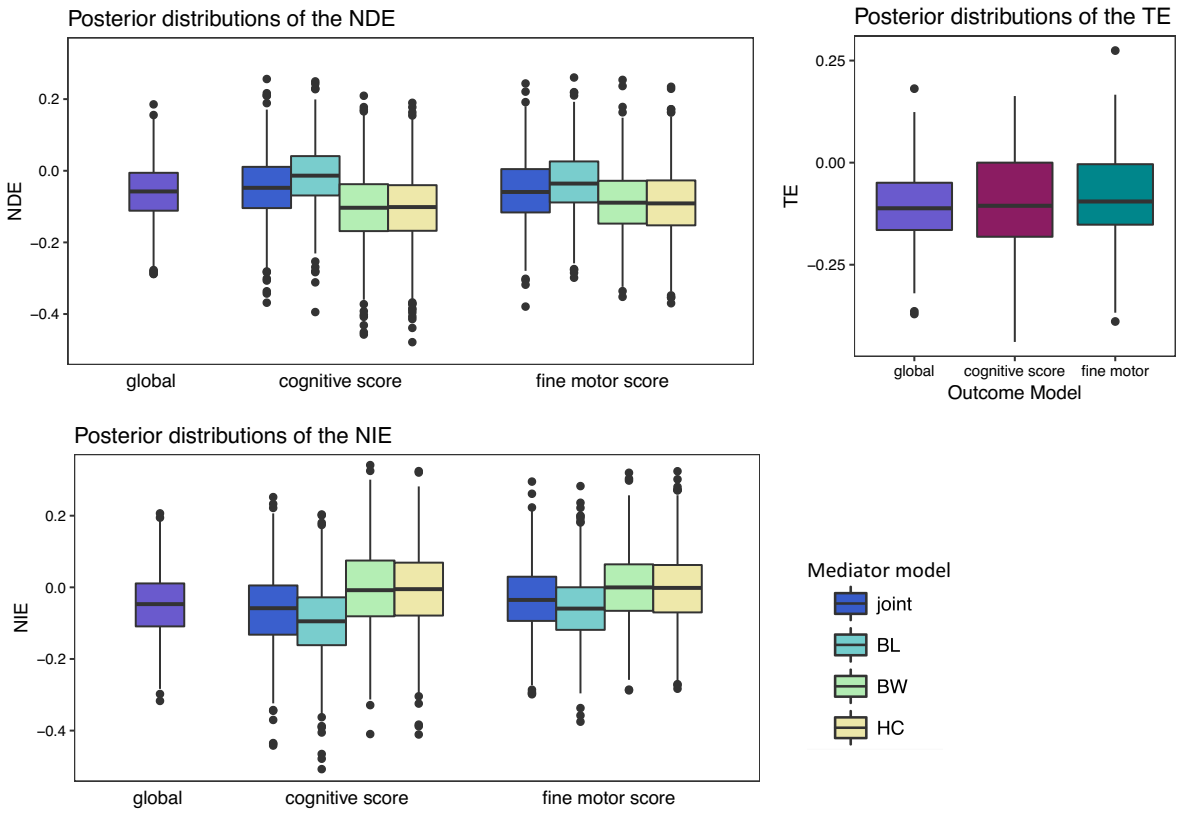


**Figure 3.1** Box plots displaying the posterior distributions of the TE, NDE, and NIE in a Bangladeshi cohort from separate mediator-outcome models, a joint mediator model, and joint mediator and outcome models. All effects are for a change in the raw exposures from $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL), all metals set at their corresponding $25^{th}$ percentile, to $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL), all metals set at their $75^{th}$ percentile.

information from our two outcome measurements, we see an additional 20% decrease in the width of the credible interval for the NDE on cognitive score and an additional 14% decrease for the corresponding NIE. For fine motor score, we observe an added 20% decrease in the width of the credible interval for the NDE and an added 13% decrease for the NIE. We see similar patterns in efficiency gain for the CDEs as the NIE and NDE. The efficiency gains for the CDE intervening to set the mediators to their $25^{th}$, $50^{th}$, and $75^{th}$ percentiles are in Appendix section C.1 on page 53.

| | cognitive score | fine motor score |
|---|---|---|
| BL[1] | NDE: 0.85<br>NIE: 0.81 | NDE: 0.86<br>NIE: 0.89 |
| HC[1] | NDE: 0.73<br>NIE: 0.79 | NDE: 0.78<br>NIE: 0.79 |
| BW[1] | NDE: 0.71<br>NIE: 0.75 | NDE: 0.83<br>NIE: 0.86 |
| joint[2] | TE: 0.69<br>NDE: 0.80<br>NIE: 0.86 | TE: 0.84<br>NDE: 0.80<br>NIE: 0.87 |

**Table 3.1** Efficiency gains of the mediation effects at each stage of combining information across the multiple mediator and outcome measurements in a Bangladeshi cohort, when considering a change in the raw exposures from $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL), all metals set at their corresponding $25^{th}$ percentile, to $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL), all metals set at their $75^{th}$ percentile. [1]Width of the credible interval for a given effect and outcome estimated from the joint mediator model divided by the width of the corresponding credible interval estimated from separate mediator-outcome BKMR–CMA. [2]Width of the credible interval for the global mediated effect, combining information from all mediator and outcome measurements, divided by the width of the credible interval for the corresponding effect estimated from a joint mediator, but separate outcome, model. We calculated these ratios using credible intervals constructed with the appropriate Bonferroni correction for multiple comparisons.

## 3.4 Discussion

We propose a Bayesian kernel machine regression (BKMR) approach to causal mediation analysis (BKMR–CMA) to integrate information from multiple continuous exposure, mediator, and outcomes measures. We define the global TE, NDE, and NIE and present an algorithm to estimate these effects for an exposure mixture on a complex outcome phenotype operating through a complex mediator phenotype. We combine information by jointly modeling the multiple mediator measurements and jointly modeling the multiple outcome measurements allowing for subject-specific intercepts. We account for the potential high correlation of the mediator measures in the outcome model through group variable selection. We quantify the efficiency gain when first combining information across the mediator measurements, then further quantify the efficiency gain by combining information across the outcome measurements.

Although the global TE of co-exposure to manganese, arsenic, and lead on neurodevelopment is not significantly different from zero in our Bangladeshi cohort, we found notable efficiency gains at each stage of our analysis. This suggests that if we were to consider the global effect with additional mediator and

outcome measurements of similar effect magnitude, we could obtain significant results. We did not consider additional neurodevelopment scores from the BSID–III$^{TM}$ in our analysis since the other scores, receptive communication, expressive communication, and gross motor, were not associated with the metal mixture.

Our analysis shows that power will increase even if one applies the methodology across only one complex phenotype–either the mediator or the outcome. However, we observe additional gain in efficiency when pooling information from both. For our case study, BKMR models with subject-specific intercepts seem to adequately fit the data. We empirically verify this assumption in Figure 3.1 since the distributions of the effects are not significantly different across the multiple mediator and outcome measurements.

As with any flexible model, there are trade-offs that come with relaxing model assumptions. We choose to flexibly model the relationships between the exposure, mediator, and outcome measurements, but at the expense of power. Our solution to this problem is to use random subject-specific intercepts through a mixed model framework to gain efficiency by pooling information across the multiple mediator and outcome measurements. By considering only subject-specific intercepts and not also subject-specific slopes, our model is making a strong assumption. However, this assumption can be empirically verified by checking that the effects of interest are approximately the same magnitude and direction across the multiple mediator and outcome measures.

Our current model specification does not allow for time-varying confounding, which is plausible in many data applications. If a time-dependent confounder is also a part of the latent construct one considers for the mediator variable, it can be included in the set of the mediator measures to avoid possible bias. By fitting the mediator and outcome models separately, we assume independence between the models. Although this is a common assumption in the mediation literature, it is a limitation to our approach presented.

In this paper, we provide a flexible model to increase power when evaluating the relationship between an exposure mixture, and complex mediator and outcome phenotypes. There are many directions for future work in this setting. First, we hope to extend our approach to also allow for subject-specific random slopes. Second, we want to allow for weighting in our models to account for time-dependent confounding. Lastly, we want to develop a framework to model the mediator and outcome models jointly to minimize potential biases due to model incompatibility.

# A

# Supplementary material to accompany Chapter 1

## A.1 Summary of baseline characteristics in CanCORS

| | **Total** ($n = 1,585$) | **Black** ($n = 332$) | **White** ($n = 1,253$) | ***p*-value** |
|---|---|---|---|---|
| **5-year survival** | 1,077 (68%) | 206 (62%) | 871 (70%) | 0.010 |
| **BMI** (kg/m$^2$) | | | | |
| <25 | 612 (39%) | 103 (31%) | 509 (41%) | 0.001 |
| 25-30 | 564 (36%) | 118 (36%) | 446 (36%) | 0.986 |
| 30-35 | 259 (16%) | 69 (21%) | 190 (15%) | 0.014 |
| $\geq$ 35 | 150 (9%) | 42 (13%) | 108 (9%) | 0.026 |
| **Stage** | | | | |
| I | 362 (23%) | 71 (21%) | 291 (23%) | 0.478 |
| II | 363 (23%) | 69 (20%) | 294 (23%) | 0.250 |
| III | 532 (33%) | 119 (36%) | 412 (33%) | 0.361 |
| IV | 331 (21%) | 75 (22%) | 256 (20%) | 0.389 |
| **Female** | 678 (43%) | 162 (49%) | 516 (41%) | 0.013 |
| **Age at diagnosis** | | | | |
| <50 | 340 (21%) | 92 (28%) | 248 (20%) | 0.002 |
| 50-65 | 542 (34%) | 136 (41%) | 406 (32%) | 0.003 |
| >65 | 703 (44%) | 104 (31%) | 599 (48%) | <0.001 |
| **Income** | | | | |
| <\$40$k$ | 850 (54%) | 236 (71%) | 614 (49%) | <0.001 |
| \$40-60$k$ | 278 (18%) | 45 (14%) | 233 (19%) | 0.032 |
| \$60-80$k$ | 158 (10%) | 22 (7%) | 136 (11%) | 0.022 |
| >\$80$k$ | 299 (19%) | 29 (9%) | 270 (22%) | <0.001 |

**Table A.1** A summary of baseline characteristics for colorectal cancer patients from CanCORS centers that collected survival data 2003-2005 [n (%)], excluding KPHI. BMI is calculated from self-reported height and weight within 3 months of diagnosis. All $p$-values were calculated from $\chi^2$ tests.

## A.2 General residual disparity formula derivation

The general formula to estimate the residual disparity in covariate stratum $\boldsymbol{c}$ for any mediator model, assuming (A.1) holds, is (A.2). This formula allows for all possible race-covariate interactions, mediator-covariate

interactions, and a possible n-degree polynomial effect of the mediator (M) on survival.

$$log(T) = \theta_0 + \theta_1 R + \sum_{j=1}^{n} \theta_{2j} M^j + \sum_{j=1}^{n} \theta_{3j} R M^j + \boldsymbol{\theta}_4^T \boldsymbol{C} + \boldsymbol{\theta}_5^T R \boldsymbol{C} + \sum_{j=1}^{n} \boldsymbol{\theta}_{6j}^T M^j \boldsymbol{C} + \nu\epsilon \tag{A.1}$$

$\mathrm{E}_T(T_{H_x(0)}|R=1, \boldsymbol{C}=\boldsymbol{c})$ represents the expected survival time for a Black individual with covariate levels $\boldsymbol{c}$ if their BMI at diagnosis of colorectal cancer had been set to a random draw from that of the white BMI distribution with covariate levels $\boldsymbol{c}$.

$$\begin{aligned}
&\mathrm{E}_T(T_{H_x(0)}|R=1, \boldsymbol{C}=\boldsymbol{c}) \\
&= \int_m \mathrm{E}_T(T_m|R=1, H_x(0)=m, \boldsymbol{C}=\boldsymbol{c}) dP_{H_x(0)}(m|R=1, \boldsymbol{c}) \\
&= \int_m \mathrm{E}_T(T_m|R=1, \boldsymbol{C}=\boldsymbol{c}) dP_{H_x(0)}(m|R=1, \boldsymbol{c}) \\
&= \int_m \mathrm{E}_T(T|R=1, M=m, \boldsymbol{C}=\boldsymbol{c}) dP_M(m|R=0, \boldsymbol{c}) \\
&= \int_m \mathrm{E}_T\left[e^{\theta_0+\theta_1+\sum_{j=1}^{n}(\theta_{2j}+\theta_{3j})m^j+\left(\boldsymbol{\theta}_4^T+\boldsymbol{\theta}_5^T\right)\boldsymbol{c}+\sum_{j=1}^{n}\boldsymbol{\theta}_{6j}^T m^j \boldsymbol{c}+\nu\epsilon}\right] dP_M(m|R=0, \boldsymbol{c}) \\
&= \int_m e^{\theta_0+\theta_1+\sum_{j=1}^{n}(\theta_{2j}+\theta_{3j})m^j+\left(\boldsymbol{\theta}_4^T+\boldsymbol{\theta}_5^T\right)\boldsymbol{c}+\sum_{j=1}^{n}\boldsymbol{\theta}_{6j}^T m^j \boldsymbol{c}} \mathrm{E}_T\left[\nu\epsilon\right] dP_M(m|R=0, \boldsymbol{c}) \\
&= e^{\theta_0+\theta_1+\left(\boldsymbol{\theta}_4^T+\boldsymbol{\theta}_5^T\right)\boldsymbol{c}} \int_m e^{\sum_{j=1}^{n}(\theta_{2j}+\theta_{3j})m^j+\sum_{j=1}^{n}\boldsymbol{\theta}_{6j}^T m^j \boldsymbol{c}} dP_M(m|R=0, \boldsymbol{c})
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}_T(T_m|R=0, \boldsymbol{C}=\boldsymbol{c}) \\
&= \int_m \mathrm{E}_T(T_m|R=0, M=m, \boldsymbol{C}=\boldsymbol{c}) dP_M(m|R=0, \boldsymbol{c}) \\
&= \int_m \mathrm{E}_T(T_m|R=0, \boldsymbol{C}=\boldsymbol{c}) dP_M(m|R=0, \boldsymbol{c}) \\
&= \int_m \mathrm{E}_T(T|R=0, M=m, \boldsymbol{C}=\boldsymbol{c}) dP_M(m|R=0, \boldsymbol{c}) \\
&= \int_m \mathrm{E}_T\left[e^{\theta_0+\sum_{j=1}^{n}\theta_{2j}m^j+\boldsymbol{\theta}_4^T\boldsymbol{c}+\sum_{j=1}^{n}\boldsymbol{\theta}_{6j}^T m^j \boldsymbol{c}+\nu\epsilon}\right] dP_M(m|R=0, \boldsymbol{c}) \\
&= \int_m e^{\theta_0+\sum_{j=1}^{n}\theta_{2j}m^j+\boldsymbol{\theta}_4^T\boldsymbol{c}+\sum_{j=1}^{n}\boldsymbol{\theta}_{6j}^T m^j \boldsymbol{c}} \mathrm{E}_T\left[\nu\epsilon\right] dP_M(m|R=0, \boldsymbol{c}) \\
&= e^{\theta_0+\boldsymbol{\theta}_4^T\boldsymbol{c}} \mathrm{E}_T\left[e^{\nu\epsilon}\right] \int_m e^{\sum_{j=1}^{n}\theta_{2j}m^j+\sum_{j=1}^{n}\boldsymbol{\theta}_{6j}^T m^j \boldsymbol{c}} dP_M(m|R=0, \boldsymbol{c})
\end{aligned}$$

$$\frac{\mathrm{E}\left[T_{H_{\mathbf{c}}(0)}|R=1, \mathbf{c}\right]}{\mathrm{E}\left[T|R=0, \mathbf{c}\right]} = e^{\theta_1+\boldsymbol{\theta}_5^T\boldsymbol{c}} \frac{\int_m e^{\sum_{j=1}^{n}(\theta_{2j}+\theta_{3j})m^j+\sum_{j=1}^{n}\boldsymbol{\theta}_{6j}^T m^j \boldsymbol{c}} dP_M(m|R=0, \boldsymbol{c})}{\int_m e^{\sum_{j=1}^{n}\theta_{2j}m^j+\sum_{j=1}^{n}\boldsymbol{\theta}_{6j}^T m^j \boldsymbol{c}} dP_M(m|R=0, \boldsymbol{c})} \tag{A.2}$$

## A.3 Residual disparity formula for BCL mediator model

Now, let us categorize our mediator ($M$) into $K+1$ categories. Consider the following baseline category logit (BCL) mediator model and AFT Weibull outcome model:

$$\log\{P(M=k|R,\boldsymbol{C})/P(M=0|R,\boldsymbol{C})\} = \beta_{0k} + \beta_{1k}R + \boldsymbol{\beta}_{2k}^T\boldsymbol{C} + \boldsymbol{\theta}_{3k}^T R\boldsymbol{C} \tag{A.3}$$

$$\log(T) = \theta_0 + \theta_1 R + \sum_{k=1}^{K}\theta_{2k}M_k + \sum_{k=1}^{K}\theta_{3k}RM_k + \boldsymbol{\theta}_4^T\boldsymbol{C} + \boldsymbol{\theta}_5^T R\boldsymbol{C} + \sum_{k=1}^{K}\boldsymbol{\theta}_{6k}^T M_k\boldsymbol{C} + \nu\epsilon, \tag{A.4}$$

where $M_k$ is an indicator that the mediator value is in category $k$. Under these models, the residual disparity estimator for a particular covariate pattern $\boldsymbol{c}$ is:

$$\frac{\mathrm{E}\left[T_{H_{\mathbf{c}}(0)}|R=1,\mathbf{c}\right]}{\mathrm{E}\left[T|R=0,\mathbf{c}\right]} = e^{\theta_1+\boldsymbol{\theta}_5^T\boldsymbol{c}}\left\{\frac{1+e^{\beta_{01}+\boldsymbol{\beta}_{21}^T\boldsymbol{c}+\theta_{21}+\theta_{31}+\boldsymbol{\theta}_{61}^T\boldsymbol{c}}+\ldots+e^{\beta_{0K}+\boldsymbol{\beta}_{2K}^T\boldsymbol{c}+\theta_{2K}+\theta_{3K}+\boldsymbol{\theta}_{6K}^T\boldsymbol{c}}}{1+e^{\beta_{01}+\boldsymbol{\beta}_{21}^T\boldsymbol{c}+\theta_{21}+\boldsymbol{\theta}_{61}^T\boldsymbol{c}}+\ldots+e^{\beta_{0K}+\boldsymbol{\beta}_{2K}^T\boldsymbol{c}+\theta_{2K}+\boldsymbol{\theta}_{6K}^T\boldsymbol{c}}}\right\}. \tag{A.5}$$

# B

# Supplementary material to accompany Chapter 2

## B.1 Correlation and kernel functions used in our simulations

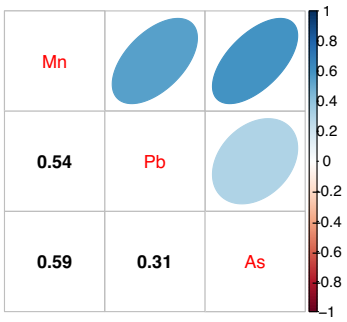


**Figure B.1** Covariance structure $\boldsymbol{\Sigma}$ considered in our simulation. The covariance for manganese (Mn), arsenic (As), and lead (Pb) from Bangladesh after log transform and standardization.

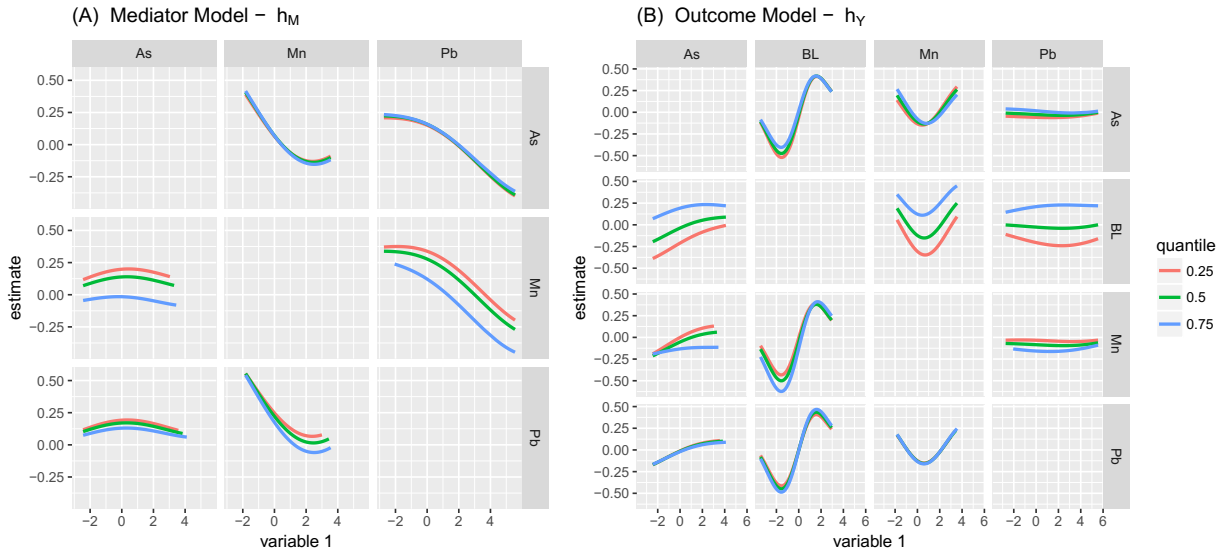


**Figure B.2** Kernel functions $h_M(\cdot)$ and $h_Y(\cdot)$ considered in our simulation. These functions are obtain by fitting BKMR mediator model (2.8) and BKMR outcome model (2.9) in our Bangladeshi cohort. Bivariate (A) exposure-mediator and (B) exposure-mediator-response functions for the each element listed on the top when the element listed on the right is fixed at its $25^{th}$, $50^{th}$ or $75^{th}$ percentiles.

## B.2 Algorithm to estimate CDEs using BKMR

1. Fit BKMR outcome model (2.9).

2. For each MCMC iteration, $j = 1, \ldots, J$:

   (a) Estimate the average outcome value for the mean level of covariates at the specific mediator value of interest for $\boldsymbol{a}_{\text{new}} = \begin{pmatrix} \boldsymbol{a} & \boldsymbol{a}^* \end{pmatrix}^{\mathsf{T}}$ from (2.9). (I.e. estimate $Y_{\boldsymbol{a}^* m}$ and $Y_{\boldsymbol{a} m}$ for each MCMC iteration).

$$\begin{aligned} \boldsymbol{Y}^{(j)}_{\boldsymbol{a}_{\text{new}} m}(\bar{\boldsymbol{c}}) &= \mathrm{E}^{(j)}(Y|\boldsymbol{A} = \boldsymbol{a}_{\text{new}}, M = m, \boldsymbol{C} = \bar{\boldsymbol{c}}) \\ &= h^{(j)}_Y(\boldsymbol{A} = \boldsymbol{a}_{\text{new}}, M = m) + \bar{\boldsymbol{c}}^T \boldsymbol{\theta}^{(j)} \end{aligned}$$

   (b) Obtain the $j^{th}$ posterior sample of the CDE for a change of exposure from $\boldsymbol{a}^*$ to $\boldsymbol{a}$ intervening to fix the mediator at $m$ by:

$$CDE^{(j)} = Y^{(j)}_{\boldsymbol{a} m}(\bar{\boldsymbol{c}}) - Y^{(j)}_{\boldsymbol{a}^* m}(\bar{\boldsymbol{c}}).$$

3. Estimate the CDE and 95% credible intervals from these posterior samples.

Only two no unmeasured confounding assumptions are required for the CDE to have a causal interpretation: $Y_{\boldsymbol{a} m} \amalg \boldsymbol{A}|\boldsymbol{C}$, $Y_{\boldsymbol{a} m} \amalg M|\boldsymbol{C}, \boldsymbol{A}$. Namely, there are no unmeasured exposure-outcome confounders and there are no unmeasured mediator-outcome confounders.

## B.3 Formulas to estimate causal mediation effects when the exposure is a mixture

Consider the following linear regression models for the mediator and outcome:

$$\mathrm{E}[M] = \beta_0 + \boldsymbol{\beta}_1^{\mathsf{T}}\boldsymbol{A} + \boldsymbol{\beta}_2^{\mathsf{T}}\boldsymbol{C}, \tag{B.1}$$

$$\mathrm{E}[Y] = \theta_0 + \boldsymbol{\theta}_1^{\mathsf{T}}\boldsymbol{A} + \theta_2 M + \boldsymbol{\theta}_3^{\mathsf{T}}\boldsymbol{A}M + \boldsymbol{\theta}_4^{\mathsf{T}}\boldsymbol{C}, \tag{B.2}$$

where $\boldsymbol{A} = (A_1, \ldots, A_L)^{\mathsf{T}}$ is a exposure mixture of $L$ components, $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots \beta_{1L})^{\mathsf{T}}$, $\boldsymbol{\theta}_1 = (\theta_{11}, \ldots \theta_{1L})^{\mathsf{T}}$, and $\boldsymbol{\theta}_3 = (\theta_{31}, \ldots \theta_{3L})^{\mathsf{T}}$.

$\mathrm{E}_Y(Y_{\boldsymbol{a}m}|\boldsymbol{C} = \boldsymbol{c})$ represents the expected outcome value had everyone been exposed to level $\boldsymbol{a}$ and had their mediator been set to level $m$, fixing covariates to level $\boldsymbol{c}$. $\mathrm{E}_Y(Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}|\boldsymbol{C} = \boldsymbol{c})$ represents the expected outcome value had everyone been exposed to level $\boldsymbol{a}$ and had their mediator been set to the level it would have taken if exposure is set to $\boldsymbol{a}^*$, fixing covariates to level $\boldsymbol{c}$. Then, considering models (B.1) and (B.2), and assuming (i) $Y_{\boldsymbol{a}m} \amalg \boldsymbol{A}|\boldsymbol{C}$, (ii) $Y_{\boldsymbol{a}m} \amalg M|\boldsymbol{C}, \boldsymbol{A}$, (iii) $M_{\boldsymbol{a}} \amalg \boldsymbol{A}|\boldsymbol{C}$, and (iv) $Y_{\boldsymbol{a}m} \amalg M_{\boldsymbol{a}^*}|\boldsymbol{C}$, we can estimate these effects as:

$$\begin{aligned}
\mathrm{E}_Y(Y_{\boldsymbol{a}m}|\boldsymbol{C}) &\overset{(i)-(ii)}{=} \mathrm{E}_Y(Y|\boldsymbol{A} = \boldsymbol{a}, M = m, \boldsymbol{C} = \boldsymbol{c}) && \text{by consistency} \\
&= \theta_0 + \boldsymbol{\theta}_1^{\mathsf{T}}\boldsymbol{a} + \theta_2 m + \boldsymbol{\theta}_3^{\mathsf{T}}\boldsymbol{a}m + \boldsymbol{\theta}_4^{\mathsf{T}}\boldsymbol{c} \\
\mathrm{E}_Y(Y_{\boldsymbol{a}^*m}|\boldsymbol{C}) &\overset{(i)-(ii)}{=} \mathrm{E}_Y(Y|\boldsymbol{A} = \boldsymbol{a}^*, M = m, \boldsymbol{C} = \boldsymbol{c}) && \text{by consistency} \\
&= \theta_0 + \boldsymbol{\theta}_1^{\mathsf{T}}\boldsymbol{a}^* + \theta_2 m + \boldsymbol{\theta}_3^{\mathsf{T}}\boldsymbol{a}^* m + \boldsymbol{\theta}_4^{\mathsf{T}}\boldsymbol{c}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}_Y(Y_{\boldsymbol{a}M_{\boldsymbol{a}^*}}|\boldsymbol{C}) &= \int_m \mathrm{E}_Y(Y_{\boldsymbol{a}m}|M_{\boldsymbol{a}^*} = m, \boldsymbol{C} = \boldsymbol{c}) dP_{M_{\boldsymbol{a}^*}}(m|\boldsymbol{C} = \boldsymbol{c}) \\
&\overset{(iii)}{=} \int_m \mathrm{E}_Y(Y_{\boldsymbol{a}m}|M_{\boldsymbol{a}^*} = m, \boldsymbol{C} = \boldsymbol{c}) dP_{M_{\boldsymbol{a}^*}}(m|\boldsymbol{A} = \boldsymbol{a}^*, \boldsymbol{C} = \boldsymbol{c}) \\
&\overset{(iv)}{=} \int_m \mathrm{E}_Y(Y_{\boldsymbol{a}m}|\boldsymbol{C} = \boldsymbol{c}) dP_M(m|\boldsymbol{A} = \boldsymbol{a}^*, \boldsymbol{C} = \boldsymbol{c}) && \text{by consistency} \\
&\overset{(i)-(ii)}{=} \int_m \mathrm{E}_Y(Y|\boldsymbol{A} = \boldsymbol{a}, M = m, \boldsymbol{C} = \boldsymbol{c}) dP_M(m|\boldsymbol{A} = \boldsymbol{a}^*, \boldsymbol{C} = \boldsymbol{c}) && \text{by consistency}
\end{aligned}$$

$$
\begin{aligned}
&= \int_m \theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a} + \theta_2 m + \boldsymbol{\theta}_3^\intercal \boldsymbol{a} m + \boldsymbol{\theta}_4^\intercal \boldsymbol{c}\, dP_M(m|\boldsymbol{A}=\boldsymbol{a}^*, \boldsymbol{C}=\boldsymbol{c}) \\
&= \theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a} + \boldsymbol{\theta}_4^\intercal \boldsymbol{c} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a}) \int_m m\, dP_M(m|\boldsymbol{A}=\boldsymbol{a}^*, \boldsymbol{C}=\boldsymbol{c}) \\
&= \theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a} + \boldsymbol{\theta}_4^\intercal \boldsymbol{c} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a})\, \mathrm{E}_M(M|\boldsymbol{A}=\boldsymbol{a}^*, \boldsymbol{C}=\boldsymbol{c}) \\
&= \theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a} + \boldsymbol{\theta}_4^\intercal \boldsymbol{c} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a})\, [\beta_0 + \boldsymbol{\beta}_1^\intercal \boldsymbol{a}^* + \boldsymbol{\beta}_2^\intercal \boldsymbol{c}]
\end{aligned}
$$

By similar logic,

$$
\begin{aligned}
\mathrm{E}_Y(Y_{\boldsymbol{a}M_{\boldsymbol{a}}}|\boldsymbol{C}) &= \theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a} + \boldsymbol{\theta}_4^\intercal \boldsymbol{c} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a})\, [\beta_0 + \boldsymbol{\beta}_1^\intercal \boldsymbol{a} + \boldsymbol{\beta}_2^\intercal \boldsymbol{c}] \\
\mathrm{E}_Y(Y_{\boldsymbol{a}^*M_{\boldsymbol{a}^*}}|\boldsymbol{C}) &= \theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a}^* + \boldsymbol{\theta}_4^\intercal \boldsymbol{c} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a}^*)\, [\beta_0 + \boldsymbol{\beta}_1^\intercal \boldsymbol{a}^* + \boldsymbol{\beta}_2^\intercal \boldsymbol{c}]
\end{aligned}
$$

Thus,

$$
\begin{aligned}
CDE(m) &= \mathrm{E}_Y(Y|\boldsymbol{A}=\boldsymbol{a}, M=m, \boldsymbol{C}=\boldsymbol{c}) - \mathrm{E}_Y(Y|\boldsymbol{A}=\boldsymbol{a}^*, M=m, \boldsymbol{C}=\boldsymbol{c}) \\
&= (\boldsymbol{\theta}_1^\intercal + \boldsymbol{\theta}_3^\intercal m)\, (\boldsymbol{a} - \boldsymbol{a}^*) \\
NDE &= \int_{\boldsymbol{c}} \mathrm{E}_Y(Y_{aM_{a^*}}|\boldsymbol{C}) - \mathrm{E}_Y(Y_{a^*M_{a^*}}|\boldsymbol{C}) dP_{\boldsymbol{C}}(\boldsymbol{c}) \\
&\approx \mathrm{E}_Y(Y_{aM_{a^*}}|\bar{\boldsymbol{C}}) - \mathrm{E}_Y(Y_{a^*M_{a^*}}|\bar{\boldsymbol{C}}) \\
&= \theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a} + \boldsymbol{\theta}_4^\intercal \bar{\boldsymbol{c}} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a})\, [\beta_0 + \boldsymbol{\beta}_1^\intercal \boldsymbol{a}^* + \boldsymbol{\beta}_2^\intercal \bar{\boldsymbol{c}}] - \\
&\quad (\theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a}^* + \boldsymbol{\theta}_4^\intercal \bar{\boldsymbol{c}} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a}^*)\, [\beta_0 + \boldsymbol{\beta}_1^\intercal \boldsymbol{a}^* + \boldsymbol{\beta}_2^\intercal \bar{\boldsymbol{c}}]) \\
&= \boldsymbol{\theta}_1^\intercal\, (\boldsymbol{a} - \boldsymbol{a}^*) + \boldsymbol{\theta}_3^T\, (\boldsymbol{a} - \boldsymbol{a}^*)\, [\beta_0 + \boldsymbol{\beta}_1^\intercal \boldsymbol{a}^* + \boldsymbol{\beta}_2^\intercal \bar{\boldsymbol{c}}] \\
NIE &= \int_{\boldsymbol{c}} \mathrm{E}_Y(Y_{aM_a}|\boldsymbol{C}) - \mathrm{E}_Y(Y_{aM_{a^*}}|\boldsymbol{C}) dP_{\boldsymbol{C}}(\boldsymbol{c}) \\
&\approx \mathrm{E}_Y(Y_{aM_a}|\bar{\boldsymbol{C}}) - \mathrm{E}_Y(Y_{aM_{a^*}}|\bar{\boldsymbol{C}}) \\
&= \theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a} + \boldsymbol{\theta}_4^\intercal \bar{\boldsymbol{c}} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a})\, [\beta_0 + \boldsymbol{\beta}_1^\intercal \boldsymbol{a} + \boldsymbol{\beta}_2^\intercal \bar{\boldsymbol{c}}] - \\
&\quad (\theta_0 + \boldsymbol{\theta}_1^\intercal \boldsymbol{a} + \boldsymbol{\theta}_4^\intercal \bar{\boldsymbol{c}} + (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a})\, [\beta_0 + \boldsymbol{\beta}_1^\intercal \boldsymbol{a}^* + \boldsymbol{\beta}_2^\intercal \bar{\boldsymbol{c}}]) \\
&= (\theta_2 + \boldsymbol{\theta}_3^\intercal \boldsymbol{a})\, [\boldsymbol{\beta}_1^\intercal\, (\boldsymbol{a} - \boldsymbol{a}^*)]
\end{aligned}
$$

When considering traditional approaches to model the outcome, we do not include exposure-mediator interactions in (B.2). We therefore model the outcome as:

$$\mathrm{E}[Y] \quad = \quad \gamma_0 + \boldsymbol{\gamma}_1^{\mathsf{T}} \boldsymbol{A} + \gamma_2 M + \boldsymbol{\gamma}_3^{\mathsf{T}} \boldsymbol{C}. \tag{B.3}$$

We estimate the traditional mediation effects for an exposure mixture as:

$$\begin{aligned} NDE &= \boldsymbol{\gamma}_1^{\mathsf{T}} \left( \boldsymbol{a} - \boldsymbol{a}^* \right), \\ NIE &= \theta_2 \boldsymbol{\beta}_1^{\mathsf{T}} \left( \boldsymbol{a} - \boldsymbol{a}^* \right). \end{aligned}$$

# C

# Supplementary material to accompany Chapter 3

## C.1 Supplemental tables

| | cognitive score | fine motor score |
|---|---|---|
| BL[1] | CDE25: 0.78<br>CDE50: 0.75<br>CDE75: 0.69 | CDE25: 0.85<br>CDE50: 0.79<br>CDE75: 0.81 |
| HC[1] | CDE25: 0.64<br>CDE50: 0.58<br>CDE75: 0.54 | CDE25: 0.65<br>CDE50: 0.7<br>CDE75: 0.74 |
| BW[1] | CDE25: 0.58<br>CDE50: 0.58<br>CDE75: 0.56 | CDE25: 0.72<br>CDE50: 0.71<br>CDE75: 0.75 |
| joint[2] | CDE25: 0.65<br>CDE50: 0.71<br>CDE75: 0.75 | CDE25: 0.71<br>CDE50: 0.67<br>CDE75: 0.65 |

**Table C.1** Efficiency gains of the CDEs at each stage of combining information across the multiple mediator and outcome measurements in a Bangladeshi cohort, when considering a change in the raw exposures from $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL), all metals set at their corresponding $25^{th}$ percentile, to $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL), all metals set at their $75^{th}$ percentile. [1]Width of the credible interval for a given effect and outcome estimated from the joint mediator model divided by the width of the corresponding credible interval estimated from separate mediator-outcome BKMR–CMA. [2]Width of the credible interval for the global mediated effect, combining information from all mediator and outcome measurements, divided by the width of the credible interval for the corresponding effect estimated from a joint mediator, but separate outcome, model. We calculated these ratios using credible intervals constructed with the appropriate Bonferroni correction for multiple comparisons. CDE25 represents the CDE when the mediator(s) are set to their $25^{th}$ percentile(s), CDE50 represents the CDE when the mediator(s) are set to their $50^{th}$ percentile(s), and CDE75 represents the CDE when the mediator(s) are set to their $75^{th}$ percentile(s).

| | cognitive score | fine motor score |
|---|---|---|
| BL[1] | NDE: -0.01 (-0.27, 0.20)<br>NIE: -0.10 (-0.39, 0.17)<br>CDE25: -0.02 (-0.28, 0.14)<br>CDE50: -0.01 (-0.25, 0.14)<br>CDE75: 0.00 (-0.22, 0.17) | NDE: -0.03 (-0.26, 0.19)<br>NIE: -0.06 (-0.28, 0.20)<br>CDE25: -0.02 (-0.23, 0.13)<br>CDE50: -0.02 (-0.26, 0.13)<br>CDE75: -0.03 (-0.27, 0.12) |
| HC[1] | NDE: -0.10 (-0.39, 0.15)<br>NIE: -0.01 (-0.30, 0.27)<br>CDE25: -0.10 (-0.40, 0.12)<br>CDE50: -0.11 (-0.40, 0.11)<br>CDE75: -0.11 (-0.40, 0.11) | NDE: -0.09 (-0.33, 0.16)<br>NIE: 0.00 (-0.26, 0.27)<br>CDE25: -0.08 (-0.32, 0.15)<br>CDE50: -0.09 (-0.33, 0.10)<br>CDE75: -0.09 (-0.33, 0.10) |
| BL[1] | NDE: -0.10 (-0.40, 0.16)<br>NIE: -0.01 (-0.31, 0.29)<br>CDE25: -0.11 (-0.47, 0.11)<br>CDE50: -0.10 (-0.40, 0.11)<br>CDE75: -0.09 (-0.37, 0.12) | NDE: -0.09 (-0.32, 0.15)<br>NIE: 0.00 (-0.25, 0.24)<br>CDE25: -0.09 (-0.32, 0.11)<br>CDE50: -0.09 (-0.31, 0.11)<br>CDE75: -0.08 (-0.31, 0.11) |
| joint[2] | TE: -0.11 (-0.40, 0.12)<br>NDE: -0.05 (-0.30, 0.19)<br>NIE: -0.06 (-0.34, 0.20)<br>CDE25: -0.01 (-0.25, 0.15)<br>CDE50: -0.01 (-0.22, 0.15)<br>CDE75: 0.00 (-0.20, 0.16) | TE: -0.09 (-0.31, 0.12)<br>NDE: -0.06 (-0.28, 0.16)<br>NIE: -0.03 (-0.27, 0.21)<br>CDE25: -0.02 (-0.24, 0.17)<br>CDE50: -0.03 (-0.25, 0.13)<br>CDE75: -0.03 (-0.27, 0.11) |
| global[3] | TE: -0.11 (-0.26, 0.04)<br>NDE: -0.06 (-0.22, 0.10)<br>NIE: -0.05 (-0.22, 0.12)<br>CDE25: -0.01 (-0.15, 0.07)<br>CDE50: -0.01 (-0.15, 0.06)<br>CDE75: -0.01 (-0.15, 0.06) | |

**Table C.2** Estimated mediation effects in a Bangladeshi cohort for a change in the raw exposures from $\boldsymbol{a}^* = (As_{.25} = 0.56\mu$g/dL, $Mn_{.25} = 4.72\mu$g/dL, $Pb_{.25} = 1.15\mu$g/dL), all metals set at their corresponding $25^{th}$ percentile, to $\boldsymbol{a} = (As_{.75} = 1.58\mu$g/dL, $Mn_{.75} = 17.80\mu$g/dL, $Pb_{.75} = 2.42\mu$g/dL), all metals set at their $75^{th}$ percentile. [1]Estimates from separate mediator-outcome BKMR–CMA with a Bonferroni correction for six analyses, [2]effect estimates after combining information across the three fetal growth measurements with a Bonferroni correction for two analyses, and [3]global mediated effects pooling information from all mediator and outcome measurements. CDE25 represents the CDE when the mediator(s) are set to their $25^{th}$ percentile(s), CDE50 represents the CDE when the mediator(s) are set to their $50^{th}$ percentile(s), and CDE75 represents the CDE when the mediator(s) are set to their $75^{th}$ percentile(s).

## C.2 Algorithm to estimate global CDEs using BKMR:

1. Fit BKMR outcome model (3.12).

2. For each MCMC iteration, $j = 1, \ldots, J$:

   (a) Estimate the average outcome value for the reference group ($s = 1$), the mean level of covariates, the specific mediator vector of interest $\mathbf{m}$, and $\boldsymbol{a}_{\text{new}} = \begin{pmatrix} \boldsymbol{a} & \boldsymbol{a}^* \end{pmatrix}^{\mathsf{T}}$ from (3.12). (I.e. estimate $Y_{\boldsymbol{a}^*\mathbf{m}}$ and $Y_{\boldsymbol{a}\mathbf{m}}$ for each MCMC iteration).

$$\begin{aligned} \boldsymbol{Y}^{(j)}_{\boldsymbol{a}_{\text{new}}\mathbf{m}}(\bar{\boldsymbol{c}}) &= \mathrm{E}^{(j)}(Y_{s=1}|\boldsymbol{A} = \boldsymbol{a}_{\text{new}}, \boldsymbol{M} = \mathbf{m}, \boldsymbol{C} = \bar{\boldsymbol{c}}) \\ &= \theta^{(j)}_{s=1} + h^{(j)}_{Y}(\boldsymbol{A} = \boldsymbol{a}_{\text{new}}, \boldsymbol{M} = \mathbf{m}) + \bar{\boldsymbol{c}}^T\boldsymbol{\theta}^{(j)} \end{aligned}$$

   (b) Obtain the $j^{th}$ posterior sample of the global CDE for a change of exposure from $\boldsymbol{a}^*$ to $\boldsymbol{a}$ intervening to fix the mediator measurements at $\mathbf{m}$ by:

$$CDE^{(j)}_{\text{global}} = Y^{(j)}_{\boldsymbol{a}\mathbf{m}}(\bar{\boldsymbol{c}}) - Y^{(j)}_{\boldsymbol{a}^*\mathbf{m}}(\bar{\boldsymbol{c}}).$$

3. Estimate the global CDE and 95% credible intervals from these posterior samples.

Only two no unmeasured confounding assumptions are required for the global CDE to have a causal interpretation: $\boldsymbol{Y}_{\boldsymbol{a}\mathbf{m}} \amalg \boldsymbol{A}|\boldsymbol{C}, \boldsymbol{Y}_{\boldsymbol{a}\mathbf{m}} \amalg \boldsymbol{M}|\boldsymbol{C}, \boldsymbol{A}$. Namely, there are no unmeasured exposure-outcome confounders and there are no unmeasured mediator-outcome confounders.

# References


[1] Albert, J. M. and Nelson, S. (2011). Generalized causal mediation analysis. *Biometrics*, 67(3):1028–1038.

[2] American Cancer Society (2016). Cancer facts & figures for African Americans 2016–2018. *Atlanta, GE: American Cancer Society*.

[3] Avin, C., Shpitser, I., and Pearl, J. (2005). Identifiability of path-specific effects. *In: Proceedings of the International Joint Conferences on Artificial Intelligence*, pages 357–363.

[4] Ayanian, J. Z., Chrischilles, E. A., Wallace, R. B., Fletcher, R. H., Fouad, M. N., Kiefe, C. I., Harrington, D. P., Weeks, J. C., Kahn, K. L., Malin, J. L., et al. (2004). Understanding cancer treatment and outcomes: the cancer care outcomes research and surveillance consortium. *Journal of Clinical Oncology*, 22(15):2992–2996.

[5] Barfield, R., Shen, J., Just, A. C., Vokonas, P. S., Schwartz, J., Baccarelli, A. A., VanderWeele, T. J., and Lin, X. (2017). Testing for the indirect effect under the null for genome-wide mediation analyses. *Genetic Epidemiology*, 41(8):824–833.

[6] Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182.

[7] Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development*. Harcourt Assessment Inc., San Antonio, TX, 3rd edition.

[8] 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.

[9] Bressler, J., Kim, K.-a., Chakraborti, T., and Goldstein, G. (1999). Molecular mechanisms of lead neurotoxicity. *Neurochemical Research*, 24(4):595–600.

[10] Budtz-Jørgensen, E., Grandjean, P., and Weihe, P. (2007). Separation of risks and benefits of seafood intake. *Environmental Health Perspectives*, 115(3):323.

[11] Budtz-Jørgensen, E., Keiding, N., Grandjean, P., and Weihe, P. (2002). Estimation of health effects of prenatal methylmercury exposure using structural equation models. *Environmental Health*, 1(1):2.

[12] Carlin, D. J., Rider, C. V., Woychik, R., and Birnbaum, L. S. (2013). Unraveling the health effects of environmental mixtures: an NIEHS priority. *Environmental Health Perspectives*, 121(1):A6–A8.

[13] Clarkson, T. W. (1987). Metal toxicity in the central nervous system. *Environmental Health Perspectives*, 75:59–64.

[14] Claus Henn, B., Coull, B. A., and Wright, R. O. (2014). Chemical mixtures and children's health. *Current Opinion in Pediatrics*, 26(2):223–229.

[15] Claus Henn, B., Ettinger, A. S., Schwartz, J., Téllez-Rojo, M. M., Lamadrid-Figueroa, H., Hernández-Avila, M., Schnaas, L., Amarasiriwardena, C., Bellinger, D. C., Hu, H., et al. (2010). Early postnatal blood manganese levels and children's neurodevelopment. *Epidemiology (Cambridge, Mass.)*, 21(4):433–439.

[16] Claus Henn, B., Schnaas, L., Ettinger, A. S., Schwartz, J., Lamadrid-Figueroa, H., Hernández-Avila, M., Amarasiriwardena, C., Hu, H., Bellinger, D. C., Wright, R. O., et al. (2012). Associations of early childhood manganese and lead coexposure with neurodevelopment. *Environmental Health Perspectives*, 120(1):126–132.

[17] Coull, B. A., Hobert, J. P., Ryan, L. M., and Holmes, L. B. (2001). Crossed random effect models for multiple outcomes in a study of teratogenesis. *Journal of the American Statistical Association*, 96(456):1194–1204.

[18] Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.

[19] DeLancey, J. O. L., Thun, M. J., Jemal, A., and Ward, E. M. (2008). Recent trends in black-white disparities in cancer mortality. *Cancer Epidemiology Biomarkers & Prevention*, 17(11):2908–2912.

[20] Devick, K. L., Bobb, J. F., Mazumdar, M. M., Claus Henn, B., Bellinger, D. C., Christiani, D. C., Wright, R. O., Coull, B. A., and Valeri, L. (2018). Bayesian kernel machine causal mediation analysis. *Manuscript in progress*.

[21] Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):355–366.

[22] Edge, S., Byrd, D. R., Compton, C. C., Fritz, A. G., Greene, F., and Trotti, A., editors (2010). *AJCC Cancer Staging Handbook*. Springer-Verlag New York.

[23] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2013). GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: IARC CancerBase No. 11. Available from: http://globocan.iarc.fr, accessed on 14/04/2018.

[24] Gleason, K., Shine, J. P., Shobnam, N., Rokoff, L. B., Suchanda, H. S., Hasan, I., Sharif, M. O., Mostofa, G., Amarasiriwardena, C., Quamruzzaman, Q., et al. (2014). Contaminated turmeric is a potential source of lead exposure for children in rural Bangladesh. *Journal of Environmental and Public Health*, 2014.

[25] Hamadani, J., Tofail, F., Nermell, B., Gardner, R., Shiraji, S., Bottai, M., Arifeen, S., Huda, S. N., and Vahter, M. (2011). Critical windows of exposure for arsenic-associated impairment of cognitive function in pre-school girls and boys: a population-based cohort study. *International Journal of Epidemiology*, 40(6):1593–1604.

[26] Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309–334.

[27] Imai, K. and Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis*, 21(2):141–171.

[28] Jara, A., Hanson, T. E., Quintana, F. A., Müller, P., and Rosner, G. L. (2011). Dppackage: Bayesian semi-and nonparametric modeling in r. *Journal of Statistical Software*, 40(5):1–30.

[29] Kile, M., Wright, R., Amarasiriwardena, C., Quamruzzaman, Q., Rahman, M., Mahiuddin, G., and Christiani, D. (2009). Maternal and umbilical cord blood levels of arsenic, cadmium, manganese, and lead in rural Bangladesh. *Epidemiology*, 20(6):S149–S150.

[30] Kile, M. L., Rodrigues, E. G., Mazumdar, M., Dobson, C. B., Diao, N., Golam, M., Quamruzzaman, Q., Rahman, M., and Christiani, D. C. (2014). A prospective cohort study of the association between drinking water arsenic exposure and self-reported maternal health symptoms during pregnancy in Bangladesh. *Environmental Health*, 13(1):29.

[31] Kocarnik, J. M., Chan, A. T., Slattery, M. L., Potter, J. D., Meyerhardt, J., Phipps, A., Nan, H., Harrison, T., Rohan, T. E., Qi, L., et al. (2016). Relationship of prediagnostic body mass index with survival after colorectal cancer: Stage-specific associations. *International Journal of Cancer*, 139(5):1065–1072.

[32] Krieger, N. (2005). Defining and investigating social disparities in cancer: critical issues. *Cancer Causes & Control*, 16(1):5–14.

[33] Kroenke, C. H., Neugebauer, R., Meyerhardt, J., Prado, C. M., Weltzien, E., Kwan, M. L., Xiao, J., and Caan, B. J. (2016). Analysis of body mass index and mortality in patients with colorectal cancer using causal diagrams. *JAMA Oncology*, 2(9):1137–1145.

[34] Lin, X., Ryan, L., Sammel, M., Zhang, D., Padungtod, C., and Xu, X. (2000). A scaled linear mixed model for multiple outcomes. *Biometrics*, 56(2):593–601.

[35] Liu, D., Lin, X., and Ghosh, D. (2007). Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics*, 63(4):1079–1088.

[36] Ogden, C. L., Carroll, M. D., Curtin, L. R., McDowell, M. A., Tabak, C. J., and Flegal, K. M. (2006). Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA*, 295(13):1549–1555.

[37] Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, pages 411–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[38] Polańska, K., Jurewicz, J., and Hanke, W. (2013). Review of current evidence on the impact of pesticides, polychlorinated biphenyls and selected metals on attention deficit/hyperactivity disorder in children. *International Journal of Occupational Medicine and Environmental Health*, 26(1):16–38.

[39] Roy, J., Lin, X., and Ryan, L. M. (2003). Scaled marginal models for multiple continuous outcomes. *Biostatistics*, 4(3):371–383.

[40] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

[41] Sammel, M., Lin, X., and Ryan, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statistics in Medicine*, 18(1718):2479–2492.

[42] Sánchez, B. N., Budtz-Jørgensen, E., Ryan, L. M., and Hu, H. (2005). Structural equation models: a review with applications to environmental epidemiology. *Journal of the American Statistical Association*, 100(472):1443–1455.

[43] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, pages 639–650.

[44] Siegel, R., DeSantis, C., and Jemal, A. (2014). Colorectal cancer statistics. *A Cancer Journal for Clinicians*, 64(2):104–117.

[45] Thurston, S. W., Ruppert, D., and Davidson, P. W. (2009). Bayesian models for multiple outcomes nested in domains. *Biometrics*, 65(4):1078–1086.

[46] Vahter, M. (2008). Health effects of early life exposure to arsenic. *Basic & Clinical Pharmacology & Toxicology*, 102(2):204–211.

[47] Valeri, L., Chen, J. T., Garcia-Albeniz, X., Krieger, N., VanderWeele, T. J., and Coull, B. A. (2016). The role of stage at diagnosis in colorectal cancer black–white survival disparities: a counterfactual causal inference approach. *Cancer Epidemiology, Biomarkers and Prevention*, 25(1):83–89.

[48] Valeri, L. and Coull, B. A. (2016). Estimating causal contrasts involving intermediate variables in the presence of selection bias. *Statistics in Medicine*, 35(26):4779–4793.

[49] Valeri, L., Mazumdar, M. M., Bobb, J. F., Claus Henn, B., Rodrigues, E., Sharif, O. I., Kile, M. L., Quamruzzaman, Q., Afroz, S., Golam, M., et al. (2017). The joint effect of prenatal exposure to metal mixtures on neurodevelopmental outcomes at 20–40 months of age: Evidence from rural Bangladesh. *Environmental Health Perspectives*, 125(6).

[50] Valeri, L. and VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, 18(2):137–150.

[51] VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.

[52] VanderWeele, T. and Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiologic Methods*, 2(1):95–115.

[53] VanderWeele, T. J. (2012). Mediation analysis with multiple versions of the mediator. *Epidemiology (Cambridge, Mass.)*, 23(3):454.

[54] VanderWeele, T. J. and Robinson, W. R. (2014). On causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology (Cambridge, Mass.)*, 25(4):473–485.

[55] VanderWeele, T. J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, 2(4):457–468.

[56] VanderWeele, T. J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, 172(12):1339–1348.

[57] Wang, Y. and Beydoun, M. A. (2007). The obesity epidemic in the united states—gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. *Epidemiologic Reviews*, 29(1):6–28.

[58] Wasserman, G. A., Liu, X., Factor-Litvak, P., Gardner, J. M., and Graziano, J. H. (2008). Developmental impacts of heavy metals and undernutrition. *Basic & Clinical Pharmacology & Toxicology*, 102(2):212–217.

[59] Wasserman, G. A., Liu, X., Parvez, F., Ahsan, H., Factor-Litvak, P., Kline, J., Van Geen, A., Slavkovich, V., LoIacono, N. J., Levy, D., et al. (2007). Water arsenic exposure and intellectual function in 6-year-old children in Araihazar, Bangladesh. *Environmental Health Perspectives*, 115(2):285–289.

[60] Wasserman, G. A., Liu, X., Parvez, F., Ahsan, H., Factor-Litvak, P., van Geen, A., Slavkovich, V., Lolacono, N. J., Cheng, Z., Hussain, I., et al. (2004). Water arsenic exposure and children's intellectual function in Araihazar, Bangladesh. *Environmental Health Perspectives*, 112(13):1329–1333.

[61] Wasserman, G. A., Liu, X., Parvez, F., Ahsan, H., Levy, D., Factor-Litvak, P., Kline, J., van Geen, A., Slavkovich, V., LoIacono, N. J., et al. (2006). Water manganese exposure and children's intellectual function in Araihazar, Bangladesh. *Environmental Health Perspectives*, 114(1):124–129.

[62] Wright, R. O., Amarasiriwardena, C., Woolf, A. D., Jim, R., and Bellinger, D. C. (2006). Neuropsychological correlates of hair arsenic, manganese, and cadmium levels in school-age children residing near a hazardous waste site. *Neurotoxicology*, 27(2):210–216.

[63] Zheng, W. and van der Laan, M. J. (2012). Targeted maximum likelihood estimation of natural direct effects. *The International Journal of Biostatistics*, 8(1):1–40.

[64] Zoni, S. and Lucchini, R. G. (2013). Manganese exposure: cognitive, motor and behavioral effects on children: a review of recent findings. *Current Opinion in Pediatrics*, 25(2):255–260.