Statistical methods for the evaluation of effects of environmental chemical mixtures on adverse pregnancy outcomes

by

Liheng (Harry) Zhuang

M.P.H., Simon Fraser University, 2017

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Master of Science Program
Faculty of Health Sciences

© Liheng (Harry) Zhuang 2023
SIMON FRASER UNIVERSITY
Summer 2023

Copyright in this work is held by the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Declaration of Committee

Name:	**Liheng (Harry) Zhuang**
Degree:	**Master of Science**
Title:	**Statistical methods for the evaluation of effects of environmental chemical mixtures on adverse pregnancy outcomes**
Committee:	**Chair: Sonya Cressman** Assistant Professor, Health Sciences
	Lawrence McCandless Supervisor Professor, Health Sciences
	Aimin Chen Committee Member Professor, Biostatistics Epidemiology and Informatics University of Pennsylvania
	Robert Platt Committee Member Professor, Epidemiology, Biostatistics, and Occupational Health McGill University
	Bruce Lanphear Committee Member Professor, Health Sciences
	Scott Venners Examiner Associate Professor, Health Sciences

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Gestational exposure to environmental chemicals presents an interesting challenge for epidemiological analysis of pregnancy outcomes and fetal development due to concurrent exposure of multiple chemicals. The analysis of chemical biomarkers presents several interesting challenges such as analysis of biomarkers below the limit of detection and handling repeated measures from the same chemical class. Additionally, the issue of co-linearity among the chemicals often results in unreliable effect estimates in traditional analysis approaches such as linear regression. Recent development in statistical methodology demonstrate promising opportunities for novel methods on mixture analysis. We explored the existing methods in the literature and propose to a novel method that combines Bayesian statistics and factor analysis for mixture analysis. We presented two different applications of this method in both regression and mediation analysis of birth weight. We demonstrated that both of the applications show strengths in precision of the estimate and interpretation of the results.

Keywords: Bayesian method; Mixture Analysis; Mediation Analysis; Prenatal outcomes; Environmental Biomarkers; HOME study

Acknowledgement

I would like to start by thanking everyone on my supervisory committee for their guidance and support. Thank you, Lawrence McCandless, Bruce Lanphear, Aimin Chen & Robert Platt, for your constant encouragement and feedback. Biggest thanks to Lawrence for spending much time reading through each of my manuscripts and providing valuable insights and suggestions. I would also like to thank all the collaborators at the HOME study. I am grateful for the opportunity to explore and analyze this data.

Thanks to my grandfather Qiyuan Zhuang for inspiring me to pursue research and thanks to my family for their endless love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1. Overview of methodological challenges in the analysis of environmental chemical mixtures in perinatal epidemiology

Adverse pregnancy outcomes such as preterm birth and low birth weight are one of the leading causes of child mortality and morbidity and important risk factors for developmental disabilities among children [1,2]. According to Tanne [9], preterm birth is the leading cause of newborn death in the US and the rate of preterm birth and low birth weight in the US are rising again in 2016. Numerous risk factors are documented to be associated with preterm birth and low birth weight in the literature [10,11]; among these risk factors, one of the most important risk factors is the exposure to environmental chemicals and their mixtures during pregnancy [3,4].

Environmental chemicals are defined as chemical compounds or chemical elements present in the air, water, food, soil, dust or other environmental media such as consumer products [12]. In the field of environmental and perinatal epidemiology, Exposure to persistent chemicals during pregnancy such as polychlorinated biphenyls, pesticides, and heavy metals are related to pregnancy outcomes such as birth weight reduction and preterm birth [3,4,12–14].

This MSc thesis concerns the challenges of statistics and data analysis of chemical mixtures. According to Braun et al. [15], there are four essential research questions pertaining to the assessment and analysis of environmental chemicals and perinatal outcomes. The first research question is the accurate assessment of environmental chemical exposure; the second research question is the identification of real causal factors in the presences of confounders; the third research question is the identification of periods of vulnerability during the pregnancy; the fourth research question is the evaluation of environmental chemical mixtures (i.e the effects of combined exposure to multiple chemicals). [15].

Among these research questions, the evaluation of environmental chemical mixtures is particularly interesting and important since most of the current knowledge of the health effects of environmental chemicals is informed by a single pollutant statistical model, particularly linear regression and logistic regression models, while chemical mixture exposure profiles resemble the real-life scenarios much closely [13]. Additionally, the real-life exposure profiles of pregnant mothers typically involve low-level exposures of individual chemicals

simultaneously. In particular, the combined effects of multiple chemicals may be large when the effects of individual chemicals seem to be small. Therefore, there is a growing interest in assessing synergistic effects and interactions between individual chemicals and mixture components [16]. As a result, innovations in mixture specific statistical modelling methods are widely supported in multiple articles [5,6].

For models that aim to evaluate the health impact of the environmental chemical mixtures, several statistical challenges are present such as multicollinearity and non-linear exposure-response relationship [5,6]. Among these challenges, multicollinearity present among individual chemicals is the primary challenge because it is widely detected in any environmental chemical mixtures [5,6]. One of the reasons for the presence of multicollinearity is that individual chemicals within mixtures can be traced to a common exposure source. For example, a mixture of PCB congeners is typically present in manufactured commercial products and they are persistent in the environment and a common route of exposure is a diet of fish exposed to such a mixture. According to Vittinghoff et al., [17], multicollinearity between individual chemicals are characterized by the fact that one of the chemicals is almost a linear combination of the other chemicals, resulting in substantial worse precisions of regression coefficients if several chemicals are included in the same predictive model. Therefore, developments of modern statistical methods address this problem by different ways of accounting correlations among individual chemicals.

In order to motivate this MSc thesis, we outline three important areas of background information. First, we review definition of a chemical mixture. Next, we illustrate the challenges of using biomarkers for exposure assessment in perinatal epidemiology. Finally, we describe the challenges of measurement scales in exposure modelling.

1.1.1. Background: What is a chemical mixture?

In order to assess the health effects of environmental mixtures, it is important to outline reasonable definitions and characterizations of these mixtures. Currently, the definition of environmental chemical mixtures can be characterized in three different ways: based on chemical structures, based on exposure sources, or alternatively, based on common biological pathways [18–20]. The chemical structure is an important aspect to consider when classifying different environmental mixtures because the chemical structure is documented to be linked to biological activity and biological persistence in the human body [19,21]. According to Spurgeon et al. [19], chemical structures such as type of bonds between molecules and the length of carbon chains significantly impact the cycle of uptaking, metabolization and excretion

of chemicals. The uptaking of chemicals can be described as a two-stage process that consists of the speciation and adsorption process [21]. Speciation refers to the binding and transportation of chemical molecules in the exposure medium. Adsorption refers to the exposure routes and human body destinations of the chemical molecules. The metabolization of chemicals can also be separated into two stages. The first stage is the distribution of the chemical molecules in the human body and the second stage is the participation of chemical molecules in different metabolic pathways. The excretion can be characterized by the accumulation and elimination of chemical molecules. The biochemistry and physiology of environmental chemicals are extremely complex [19]; therefore, the toxicity of the chemicals at the target biological site is closely related to the chemical structures. Taking Polychlorinated Biphenyl compounds as an example, there is a wide abundance of different chemicals that share similar chemical structures. As described by [20], PCB compounds are able to mimic hormones because of their structural properties, namely the coplanarity of the phenyl rings and the laterality of the chlorine atoms. Both of these two structures participate in specific binding sites with proteins, hence induce toxic responses in biologic systems.

In addition to the chemical structures, the source-based definition of the environmental chemical mixture is also employed in the epidemiological analysis [22,23]. In particular, policy-based intervention studies typically utilize a source-based definition of environmental chemical mixtures to evaluate the impact of the health effects of chemical mixtures specifically related to the exposure source. For example, the total exposure of environmental chemicals for a farmer near an incineration plant can be separated as the exposure from the incineration plant and the exposure from the pesticide use. The Bayesian parametric g-formula method developed by Keil et al., [22] and the multiple sources analysis of pesticides by Petit et al., [23] both employed the source-based definition of environmental chemical mixtures.

Source-based classifications of environmental chemical mixtures have unique values in terms of policy practices to address potential exposures because researchers are able to characterize the health effect of these mixtures attributed to specific sources. It is, therefore, reasonable to develop prevention and intervention strategies related to specific sources identified to be associated with the exposures. Air pollution studies typically employ source-based mixture classifications because the exposure route is predominantly inhalation. In the case of environmental chemical mixtures for pregnancy outcomes, it is more difficult to employ source-based mixture classification because it is difficult to pinpoint exactly the sources of the exposures because of the complex exposure pathways including inhalation, ingestions and dermal contact.

In addition to chemical structures and exposure sources, the biological pathways environmental chemicals participate in can also be used to characterize mixtures. According to Ferguson and Chin [18], inflammation, oxidative stress and endocrine disruption are three major biological pathways that may explain how environmental chemicals induce pregnancy outcomes.

For the inflammatory pathway, environmental chemicals tend to generate tissue-specific inflammation via three different mechanisms [24]. The first pathway is through phagocytosis of chemical molecules such as particulate matter, leading to activation of T helper cells in the human immune system to release cytokines. The second pathway is by binding of chemical molecules to specific cell receptors. In the case of phthalate monoesters, it binds to peroxisome proliferator-activated receptors, which participate in cytokine productions. The third pathway is through inducing epigenetic modification as a result of exposure to environmental chemicals. In this pathway, DNA methylation, histone modification and change in miRNA expressions can all lead to inflammatory responses [18,24]. Although the detailed cellular reactions of the inflammation pathway are still not clear, the systemic inflammatory responses can lead to certain events that are precursors of preterm birth such as cervical ripening and rupture of the amniotic sac [18,24].

For the oxidative stress pathway, environmental chemicals induce oxidative stress by inducing the overproduction of reactive oxygen species, inducing changes in mitochondrial membrane permeability and disrupting antioxidant functions [25]. It is demonstrated that early in gestation, oxidative stress can cause impaired invasion of the spiral arterioles into maternal myometrium, hence resulting in poor placentation that precedes preeclampsia or IUGR. During the pregnancy, an elevated level of oxidative stress could result in premature rupture, resulting in signal changes in the cervix to shorten labour and reduce placental protein synthesis and nutrient support, and eventually lead to a restriction in fetal growth [18,25].

For the endocrine disruption pathway, exposure of environmental chemicals participates in the disruption of essential regulatory hormones during gestation and pregnancy [12]. Hormones are signal molecules responsible for signalling the growth of the fetus as well as the timing of parturition [20]. For example, thyroid and glucocorticoid hormones are two hormones that are involved during pregnancy and polychlorinated biphenyls and perfluorinated compounds can disrupt these two hormones by affecting receptor activity.

However, although these three particular biological pathways are different at the cellular level, it is extremely difficult to attribute the physiological change to any particular

pathways because inflammation is tied closely to hormone regulation as well as oxidation stress may induce inflammation and vice versa [18]. Hence, the grouping of environmental chemical mixtures based on biological pathways requires a further understanding of the toxicological mechanisms.

1.1.2. Background: The challenges of using biomarkers for exposure assessment

For statistical analysis of chemical mixtures, the way researchers measure the exposure of the mixture is as important as careful definitions of mixtures. Biomarkers are widely used in the field of environmental epidemiology to quantify the amount of exposure to toxicological environmental chemicals by measuring chemical concentration in bloodstream or urine [26,27]. One key advantage of biomarkers is that they capture multiple sources of exposure and multiple exposure routes [26]. Biomarkers are less susceptible to measurement error of exposure compared to self-report interview data. In addition, since humans are exposed to hundreds of different environmental chemicals simultaneously, biomarkers allow the researcher to study the association between multiple exposures and the outcomes of interest [26]. However, numerous challenges are present in the epidemiologic analysis of biomarkers. As summarized in the review paper by Hu et al. [28], attributing the cause of a particular health outcome to specific biomarkers is extremely difficult because of the multicollinearity among the biomarkers. For example, in the analysis of the exposure to polychlorinated biphenyls compounds and pregnancy outcomes, over 200 biomarkers are measured in blood plasma and a significant degree of multicollinearity is present among these biomarkers [27], resulting in an extremely complex exposure-response relationship that is challenging to estimate. The curse of dimensionality is another major statistical challenge, which refers to the fact that the human study cohort typically has a smaller sample size compared to the number of biomarkers collected in each sample. The reference scale associated with biomarker measurements is also a challenge in assessing the health effects of chemical exposures because different chemicals may have different dose-response relationships. For example, within a mixture, one chemical can display a linear association with health outcomes while another chemical can display a non-linear association with health outcomes.

Additionally, as noted by Hu et al.[28], chemical concentrations measured in biospecimen are also influenced by multiple factors such as lipid and urine dilution, the persistence of specific chemicals, detection limits of equipment, as well as physiology and metabolic change during pregnancies. These are all important problems, but the focus of this review is on the statistical challenges associated with the analysis of biomarkers.

1.1.3. Background: Measurement scales for biomarkers of chemical exposure

The scale of measurement for biomarkers of chemical exposure is an important challenge in the analysis of biomarkers. Typically, the types of data researchers encounter are labelled as either nominal data or ordinal data [29]. Nominal data have the following features:1) No ordering of the different categories; 2) No measure of the distance between values; 3) Categories can be listed in any order without affecting the relationship between them. Therefore, the causal inference with nominal data is usually straightforward and easy for interpretation. On the other hand, when it comes to ordinal data, whether discrete or continuous, the effect sizes of causal inference are dependent on the scale of measurement [30]. For example, the magnitude of the odds ratio can be drastically different when using different units of measurement on the exposure variables (e.g. consider multivariable modelling of PCB concentrations measured in mg/L versus ng/L).

In the case of multiple biomarkers, conducting statistical tests without adjusting for the scale of measurement in different biomarkers could lead to biased estimates. Therefore, when conducting analysis with multiple exposures of different biomarkers, it is necessary to standardize or account for the variance of the biomarker (i.e. population standard deviation of the biomarker concentration) before inputting them into a statistical model. To account for such a problem, researchers have come up with different techniques to address the scale of measurements.

The most commonly used class of technique is feature scaling [31]. It is a normalization technique to rescale the existing data onto some arbitrary range such as [0,1] to ensure uniform comparison across data. An example of feature scaling adjustment is by subtracting the mean of the biomarker from the particular measurement and dividing it by the standard deviation of the biomarkers to create “Z-score” of exposure. For example, +2 would imply that the exposure of a participant was 2 standard deviations above the mean of the sample. However, the difficulty with this approach is that the biomarker concentrations are typically right-skewed, and therefore, the standard deviation is not a meaningful measure of exposure variation. The second class of technique is quantile approaches [32]. In this type of adjustment, the data are grouped into multiple equal parts. An example of a quantile approach would be quartile adjustment which the subject is divided into four different exposure categories [32] (e.g 0-25th percentile, 25-50th percentile, 50-75th percentile, and 75-100th percentile). Continuous quantile can also be used by calculating using the formula: $q=k(n+1)/100$ [32]. The third approach, recommended by Greenland [8], is to calculate a standard reference such as interquartile range and report the effect size in relation to the

standard reference point. This approach provides an easier interpretation of the results compare to the previous two approaches since it is easier to talk about the health effects with a point of comparison [11].

1.2. Study Objectives

In this MSc thesis, I set out to achieve three primary objectives. Firstly, I aimed to utilize Bayesian factor analysis to examine the association between exposure to biomarkers of multiple chemical mixtures and birth weight. My second objective involved investigating the association between exposure to chemical mixtures and birth weight considering potential mediation by thyroid hormones while employing a latent variable approach. My third objective centered around comparison of the analysis results to other existing methods such as multiple linear regression to provide insights and suggestions for mixture analysis.

The study population of all my analyses came from the HOME (Health Outcomes and Measures of Environment) study, The HOME study is a prospective birth cohort of pregnant mothers and their infants established in 2003 at the Cincinnati Children's Environmental Health Center, Ohio. To examine the impact of environmental chemicals on child health, pregnant mothers who were >18 years old and at 13-19 weeks of gestation were recruited from seven prenatal clinics and hospitals [15].

It is important to note that this thesis follows a manuscript-based format. Chapter 2 will encompass the detailed literature review of existing methods on mixture analysis. Chapter 3 is a manuscript already published: *Zhuang LH, Chen A, Braun JM, Lanphear BP, Hu JMY, Yolton K, McCandless LC. Effects of gestational exposures to chemical mixtures on birth weight using Bayesian factor analysis in the Health Outcome and Measures of Environment (HOME) Study. Environ Epidemiology. 2021 Jun 8;5(3):e159.* Chapter 4 is a manuscript planned for submission for publication to an epidemiology or biostatistics journal. Additionally, Chapter 5 will include my overall conclusions, suggestions for future research, and the implications of my findings for researchers and policymakers.

1.3. References

1. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet. Elsevier; 2015;385: 430–440.

2. Schieve LA, Tian LH, Rankin K, Kogan MD, Yeargin-Allsopp M, Visser S, et al. Population impact of preterm birth and low birth weight on developmental disabilities in US children. Ann Epidemiol. Elsevier; 2016;26: 267–274.

3. Windham G, Fenster L. Environmental contaminants and pregnancy outcomes. Fertil Steril. Elsevier; 2008;89: e111–6; discussion e117.

4. Wigle DT, Arbuckle TE, Turner MC, Bérubé A, Yang Q, Liu S, et al. Epidemiologic evidence of relationships between reproductive and child health outcomes and environmental chemical contaminants. J Toxicol Environ Health B Crit Rev. Taylor & Francis; 2008;11: 373–517.

5. Billionnet C, Sherrill D, Annesi-Maesano I, GERIE study. Estimating the health effects of exposure to multi-pollutant mixture. Ann Epidemiol. 2012;22: 126–141.

6. Sun Z, Tao Y, Li S, Ferguson KK, Meeker JD, Park SK, et al. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. Environ Health. 2013;12: 85.

7. Taylor KW, Joubert BR, Braun JM, Dilworth C, Gennings C, Hauser R, et al. Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology: Lessons from an Innovative Workshop. Environ Health Perspect. 2016;124: A227–A229.

8. Greenland S, Pearce N. Statistical foundations for model-based adjustments. Annu Rev Public Health. 2015;36: 89–108.

9. Tanne JH. Preterm and low weight births rise again in the US. BMJ. 2017;358: j3311.

10. Evens A, Hryhorczuk D, Lanphear BP, Rankin KM, Lewis DA, Forst L, et al. The impact of low-level lead toxicity on school performance among children in the Chicago Public Schools: a population-based retrospective cohort study. Environ Health. 2015;14: 21.

11. Kingsley SL, Eliot MN, Glazer K, Awad YA, Schwartz JD, Savitz DA, et al. Maternal ambient air pollution, preterm birth and markers of fetal growth in Rhode Island: results of a hospital-based linkage study. J Epidemiol Community Health. 2017;71: 1131–1136.

12. Bergman A, Heindel JJ, Kasten T, Kidd KA, Jobling S, Neira M, et al. The impact of endocrine disruption: a consensus statement on the state of the science. Environ Health Perspect. 2013;121: A104–6.

13. Carlin DJ, Rider CV, Woychik R, Birnbaum LS. Unraveling the Health Effects of Environmental Mixtures: An NIEHS Priority. Environ Health Perspect. 2013;121: a6–a8.

14. Burns JS, Williams PL, Sergeyev O, Korrick S, Lee MM, Revich B, et al. Serum dioxins and polychlorinated biphenyls are associated with growth among Russian boys. Pediatrics. 2011;127: e59–68.

15. Braun JM, Gray K. Challenges to studying the health effects of early life environmental chemical exposures on children's health. PLoS Biol. 2017;15: e2002800.

16. Sexton K. Cumulative risk assessment: an overview of methodological approaches for evaluating combined health effects from exposure to multiple environmental stressors. Int J Environ Res Public Health. 2012;9: 370–390.

17. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics:

Linear, Logistic, Survival, and Repeated Measures Models. Springer Science & Business Media; 2011.

18. Ferguson KK, Chin HB. Environmental chemicals and preterm birth: Biological mechanisms and the state of the science. Curr Epidemiol Rep. 2017;4: 56–71.

19. Spurgeon DJ, Jones OAH, Dorne J-LCM, Svendsen C, Swain S, Stürzenbaum SR. Systems toxicology approaches for understanding the joint effects of environmental chemical mixtures. Sci Total Environ. 2010;408: 3725–3734.

20. McKinney JD, Waller CL. Polychlorinated biphenyls as hormonally active structural analogues. Environ Health Perspect. ehp.niehs.nih.gov; 1994;102: 290–297.

21. Cassee FR, Groten JP, van Bladeren PJ, Feron VJ. Toxicological evaluation and risk assessment of chemical mixtures. Crit Rev Toxicol. 1998;28: 73–101.

22. Keil AP, Daza EJ, Engel SM, Buckley JP, Edwards JK. A Bayesian approach to the g-formula. Stat Methods Med Res. 2018;27: 3183–3204.

23. Petit C, Blangiardo M, Richardson S, Coquet F, Chevrier C, Cordier S. Association of environmental insecticide exposure and fetal growth with a Bayesian model including multiple exposure sources: the PELAGIE mother-child cohort. Am J Epidemiol. 2012;175: 1182–1190.

24. Vadillo-Ortega F, Osornio-Vargas A, Buxton MA, Sánchez BN, Rojas-Bracho L, Viveros-Alcaráz M, et al. Air pollution, inflammation and preterm birth: a potential mechanistic link. Med Hypotheses. 2014;82: 219–224.

25. Longini M, Perrone S, Vezzosi P, Marzocchi B, Kenanidis A, Centini G, et al. Association between oxidative stress in pregnancy and preterm premature rupture of membranes. Clin Biochem. 2007;40: 793–797.

26. Savitz DA, Wellenius GA. Invited Commentary: Exposure Biomarkers Indicate More Than Just Exposure. Am J Epidemiol. 2018;187: 803–805.

27. Savitz DA. Invited commentary: interpreting associations between exposure biomarkers and pregnancy outcome. Am J Epidemiol. 2014;179: 545–547.

28. Hu JMY, Zhuang LH, Bernardo BA, McCandless LC. Statistical Challenges in the Analysis of Biomarkers of Environmental Chemical Exposures for Perinatal Epidemiology. Current Epidemiology Reports. 2018;5: 284–292.

29. Marateb HR, Mansourian M, Adibi P, Farina D. Manipulating measurement scales in medical statistical analysis and data mining: A review of methodologies. J Res Med Sci. 2014;19: 47–56.

30. Solomon JD, Vallero D, Benson K. Evaluating risk: A revisit of the scales, measurement theory, and statistical analysis controversy. 2017 Annual Reliability and Maintainability Symposium (RAMS). 2017. pp. 1–6.

31. Youn E, Jeong MK. Class dependent feature scaling method using naive Bayes classifier for text datamining. Pattern Recognit Lett. 2009;30: 477–485.

32. Altman DG, Bland JM. Quartiles, Quintiles, Centiles, And Other Quantiles. BMJ: British Medical Journal. BMJ; 1994;309: 996–996.

Chapter 2. Literature review

Adverse pregnancy outcomes are one of the most important causes of child mortality, morbidity, and developmental disabilities [1,2]. Among various risk factors associated with these adverse pregnancy outcomes, environmental chemicals and their mixtures are deemed to be one of the most important risk factors [3,4]. However, most of the current knowledge of the health effects of environmental chemicals is informed by a single pollutant statistical model due to specific challenges in multi-pollutant modelling. Therefore, a growing interest in mixture specific modelling is well merited in the field of environmental and perinatal epidemiology.

Billionnet et al.[5] conducted a comprehensive review of statistical methods exploring the health impact of air pollutant mixtures. Sun et al. [6] used simulation studies to simulate different profiles of correlated exposures to evaluate the performances of five different statistical methods. Taylor et al. [7] conducted a more comprehensive review of currently available statistical methods for mixture assessment and categorized different methods based on their functions. However, as argued by Greenland and Pearce [8], the discussion of statistical methodologies cannot be independent of the contextual details of the epidemiologic questions. Therefore, more comprehensive reviews of statistical methods in mixture assessment of different epidemiological outcomes are well merited.

The aim of this literature review chapter is to review the current literature that employs modern statistical methods for the assessment of mixtures effects to identify potential gaps and hypothesize innovations to address them. Due to the extremely complex nature of the mixture assessment, we aim to focus our review to specifically address the assessment of environmental chemical mixture effects on perinatal outcomes using biomarkers. In Section 2.1 of this paper, the theoretical basis of statistical model selection pertaining to the assessment of mixtures is discussed. In Section 2.2 of this paper, a detailed comprehensive overview of existing methods that can be used for mixture modelling is included. In Section 2.3 of this paper, literature reviews of research focus on environmental chemical mixtures and perinatal outcomes are provided. In section 2.4 of this paper, potential innovation and future directions are hypothesized based on findings from previous sections.

2.1 Statistical method selection to estimate the health effects of environmental chemical mixtures: how do researchers select one method instead of another?

Billionnet et al. [5] argue that the selection of statistical methods for evaluating the effects of mixtures must be built upon an understanding of the exposure pathways of

environmental chemicals. Taking polychlorinated biphenyl compounds as an example, within this particular chemical group, there is a wide abundance of different chemicals that share similar chemical structures. The major exposure pathways for this group of chemical is through ingestion, inhalation and dermal contact. Within the ingestion pathway, bio-accumulation of the chemical is possible because animals such as fish are also exposed to these chemicals [13]. It is expected that humans accumulate more variety of PCB congeners as well as higher concentrations of these chemicals in the bloodstreams. This results in the presence of collinearity and interaction effects between individual chemicals within the mixtures. Due to the presence of collinearity among chemical variables, traditional linear regression methods which include all variables in a single model will generate extremely unstable parameter estimates. In extreme cases, drastically different estimation of regression coefficients can occur even with a single data point change in the available data. Hence, handling multiple biomarkers in different chemical classes is an important and challenging task. With recent developments in statistical methodologies and the advancement of machine learning algorithms, there are more available statistical methods that aim to tackle the problem [13]. However, the current consensus of the scientific community agrees that there is no perfect method that performs consistently better than other methods [33]. Hence, it is essential for researchers to understand how to select particular methods and why such selections work. Three specific criteria (1. The research objectives 2. The data and variables 3. Causal inference or prediction) for researchers to consider are provided when selecting statistical procedures [33].

2.1.1. Criteria #1 for choosing a statistical method: The research objectives

The first and most important criterion for choosing a statistical method is the scientific question of the study. The selection of methods is dependent on the primary causal relationship the researcher wishes to explore. If the researcher is interested in the health effects of several individual biomarkers within a chemical class (e.g. the particular PCB congeners that are most strongly associated with infant birth weight), then the primary goal of the statistical procedure is to identify the most important biomarkers that are associated with the health outcome [6]. Therefore, variable selection methods are more suitable for the task (also known as “subset selection” or “feature selection” in the computer science literature). Common variable selection methods include automatic model selection such as stepwise selection [5], least absolute shrinkage and selection operator (LASSO) [34], and elastic net [35]. LASSO and elastic net also do shrinkage regression which tends to improve the model predictions (e.g. reduce prediction error when predicting a health outcome such as birth weight). The above statistical methods will perform algorithms to detect dominating

biomarkers associated with the health outcome and output the measure of this association. Other machine learning methods such as Random forest [36] can also be used to determine the relative importance of biomarkers within the data, however, the magnitude and direction of the associations with the health outcome are not summarized like variable selection methods.

If the researcher is interested in the interactions and causal framework of biomarkers within a chemical class associated with the health effects, the primary goal of the statistical procedure is to either explore potential causal pathways or confirming a proposed causal pathway. Therefore, a graph representation of the causal relationship is essential for answering this question. Direct acyclic graph (DAG) [37] is often constructed during analysis. The methods that tend to work well for this scientific question is called latent variable model or structural equation modelling (SEM) [38]. SEM enables the researcher to perform exploratory factor analysis [39,40] to identify the potential latent variable in the model from individual biomarker data to construct a causal diagram. SEM also enables researchers to perform confirmatory factor analysis [39] to estimate the health effects of both individual biomarker and latent variable (mixture) given a specific causal diagram.

If the researcher is interested in the combined/synergistic health effects of all biomarkers within a chemical class on a health outcome, then the primary goal of the statistical procedure is to provide an estimation of health effects with respect to the total mixture of biomarkers, or the total health effects pertaining to a specific chemical class. For this type of analysis, the underlying assumption is that different biomarkers within a specific class act similarly on the biological pathway [33], thus making it reasonable to use different mixture approaches. This can include linear regression and extensions, LASSO based method and extensions, Latent variable modelling approaches, Non-linear exposure-response surface methods, Bayesian methods as well as modern machine learning approaches. Both variable shrinkage method and latent variable modelling can be used to estimate the health effects of the total mixtures; however, to determine which method works better will depend on the intrinsic correlation and interaction of biomarkers within the data. This brings us to the second criteria.

2.1.2. Criteria #2 for choosing a statistical method: The data and variables

The second criterion is the understanding of variables in the data. This refers to the nature of the dataset. Important factors to consider for these criteria are types of the variable (continuous or categorical), the sparsity of the dataset (the extent of how many biomarkers in the data have strong effects on the outcome), colinearity of biomarkers (high or low correlation

between biomarkers), and sample size of the data. Taking types of variable as an example, if most variables in the dataset are categorical, clustering approaches such as k-means clustering and recursive partitioning [41] performs better compared to a dataset with more continuous variables. If the researcher employs these methods without considering the type of variables, it could result in significant loss of data because the continuous variables are forced to be categorized during the analysis [36]. The sparsity of the dataset refers to the number of important biomarkers in a chemical class. If there are only a few biomarkers in the data that are particularly associated with the health outcome, variable shrinkage methods such as LASSO would perform very well because it can improve the precision of the estimates by discarding biomarkers not associated with the health outcome [6]. If most of the biomarkers in the data are associated with the health outcome and the magnitudes tend to be quite similar among different biomarkers, then it is better to employ latent variable analysis for more reliable estimates because it retains all variables within the data and produce effect estimates of latent variables specified by the researchers while the variable selection method may eliminate potential important biomarkers due to the collinearity in the data in this context [42]. Therefore, the domain expert knowledge such as the biological pathway and relative importance of multiple biomarkers are very important in the selection of statistical methods.

Moreover, the assumptions needed for statistical methods also need to be taken in consideration. Different statistical methods carry different assumptions and violations of one or more assumptions could make the analysis results unreliable. The most common mistake is to assume certain statistical distribution for variables when the variables do not meet the criteria. Therefore, in order to prevent such mistakes, data transformation procedures such as logarithm transformation are often performed before feeding the data into statistical models [29]. Another example would be assuming a linear relationship among variables when variables are actually related to each other non-linearly. In the case of the non-linear relationships among biomarkers, exposure smooth surfacing techniques such as Bayesian Kernel Machine Regression (BKMR) would perform much better compare to models that assume linear operations between biomarkers such as SEM [36].

2.1.3. Criteria #3 for choosing a statistical method: Causal inference or prediction

In addition to the above criterions, statistical methods assessing the health effects of environmental chemical mixtures can also be generally divided into two major categories based on whether the model is designed to explain effects or predict outcomes [43]. The first category entails parameter estimation, whereas the second category entails outcome prediction. The first category focuses on addressing the following question: given the available

data, how much of the change in the outcome can be attributed to the change in the exposure [43]. The second category focuses on addressing the following question: given a potential new candidate with specific observed variables, what outcome would the prediction model predicts. However, it is noted by Savitz in the context of biomarkers and pregnancy outcomes [27] that although outcome prediction studies involve a potential change in the exposure level of biomarkers, it is not addressing whether the exposure of biomarker can be directly altered to affect the outcome since biomarker levels cannot be artificially increased or decreased in this context. Instead, it is addressing whether strategies such as environmental regulations or behavioural change could potentially alter the exposure level of the biomarkers and whether such strategies can alter the targeted health outcomes.

Regarding the first category (causal inference rather than prediction), parameter estimation methods usually require detailed procedures of confounder selections and adjustments. Traditionally, confounder selections were usually informed by expert opinions and previous literature [44]. The Directed Acyclic Graph (DAG) approach is gaining more popularity since it provides a specific graphic representation of the causal structures hypothesized by the researchers [44]. This improves the dialogue between researchers by establishing common grounds and identifying exactly what parameters the researchers are estimating based on the proposed framework.

DAG is characterized by several factors or variables connected with arrows where the arrows represent the direction of the causal relationship. The acyclic nature ensures that no path in the diagram can form a closed loop because a factor cannot be the causal factor for itself. Based on the causal framework, the researcher then selects potential variables believed to be included in the statistical models and fits the observational data to estimate the regression coefficients of the variables. Suttorp et al. [45] provided a detailed description of the procedures for the identification of potential confounders in the model through the identification of colliders and backdoor paths. Collider is defined as a common factor in which two arrows collide. The backdoor path is defined as a path from exposure to an outcome that has segments of arrows going towards the exposure and ends with an arrow to the outcome. Confounders are identified on each open backdoor path of the DAG. [46] demonstrated that the traditional approach of the selection of confounders can potentially introduce systematic bias into the model; therefore, they developed a six-step process aiming to eliminate systematic bias using a DAG framework. Traditional epidemiology relies heavily on association studies that carefully address bias and confounding. With the introduction of DAG, the procedures to select potential confounders and the procedures to eliminate bias can be automated using computer software.

However, the limitation of DAG are still apparent [47]. When constructing a DAG, there are several ways to introduce potential bias and confounding into the model. The first one is the decision to not draw an arrow between two factors. This indicates that absolute certainty that the two factors are not causal. The second one is the direction of the arrow. Sometimes it is extremely difficult to decide which arrow direction to put, especially when the number of exposure variables and covariates is very large. Even in the case of a small number of potential variables on the DAG, the possibility of arrow directions increases exponentially with the increase of the number of variables. Although the construction of the DAG cannot solve the potential of bias and confounding, it made the assumption explicit in the model framework. The epidemiological analysis carried on DAG is based upon the assumption of a reasonable causal framework and it offers space to test and debate assumptions on model explicitly.

The second category (outcome prediction rather than causal inference) is becoming more and more prevalent in the literature recently inspired by modern machine learning methods [48]. In this category of methods, the inherent structures of the model are undetermined and data-driven. Therefore, it does not suffer from disadvantages pertaining to parameter estimation methods discussed earlier. Outcome prediction methods usually employ cross-validation techniques to create separate datasets for training models and performance testings. The researcher then selects specific criteria to evaluate the outcome prediction accuracy and precision of the models. One common criterion used for continuous outcomes is the mean-squared prediction error (MSPE). Multiple statistical models are tested against each other and compete for the best prediction performances.

Criticisms for both approaches are documented and it is usually guided by what is the definition of causality and what types of questions the researchers are trying to answer [43]. In particular, statistical analysis assessing human health effects is often critiqued more than statistical analysis assessing market trends or movie preferences, especially for the category of prediction methods. For example, the generalizability of the data-driven prediction model is severely impacted by the sample sizes of the model. It is apparent that researchers have the perception that the causality issue with human health effects need to be much more robust because policymakers would implement interventions or treatment based on the assessment of statistical models.

Hernan [49] stated that the concept of causality is argued to be impossible to prove dated back in the 18th century by David Hume. Within the field of epidemiology, the concept of causality is currently guided by quantitative counterfactual theory, also referred to as potential outcomes approach. Hernan[49] explicitly stated that the quantitative counterfactual theory

does not aim to determine whether a risk factor is a cause of the outcome but to predict what would happen if an intervention is performed on the risk factor. This theory, therefore, is consistent with the predictive modelling approach. However, Vandenbroucke [50] argues that restricting the concept of causality by the quantitative theoretical framework is problematic because it restricts the type of research questions and hypotheses one can take on. In addition, the quantitative counterfactual theory cannot deal with the problem of unfeasible human interventions. For example, the inferences about variable Sex and Race in such analysis is meaningless since no feasible intervention is available. Vandenbroucke [50] argues for a pragmatic pluralistic approach of causality in epidemiology and advocates for traditional theoretical frameworks such as the Bradford Hill criteria of causality.

2.2 Modelling approaches for effects estimation or outcome prediction

Regardless of whether the model aims to explain effects or predict outcomes, the statistical methods in the analysis of environmental chemical mixtures need to operate on the principle of accounting correlations among variables in the analysis. As a result, There are several strategies that can be used to account for these correlations: 1) variable selection methods (also known as “feature selection” in the computer science literature, 2) non-linear exposure-response methods, 3) Bayesian methods, and 4) machine learning techniques.

Variable selection methods such as LASSO (least absolute shrinkage and selection operator) and ridge regression remove the collinearity in the data by selecting a subset of variables from the total data. The criteria of the importance of these variables are usually based on how much variability within the total data (e.g. variability in the birthweight outcome variable) can be explained by the selected variables [6,34]. This approach has two major limitations, one is it discards a significant amount of data because it needs to reduce the dimensions to avoid collinearity. The other is that the variables selected are not guaranteed to be the actual important variables that participated in the exposure-disease pathway (e.g. variables that genuinely predict the outcome within the population of patients under investigation).

In contrast, non-linear exposure-response methods such as BKMR (Bayesian Kernel Machine Regression), GAM (Generative Additive Models) and MARS (Multivariate Adaptive Regression Splines) are the second major group of the method for tackling high dimensional correlated exposure variables [6,36,51]. For this group of methods, all variables are included in the model. However, linear and non-linear transformations of original variables are introduced to mitigate the effect of collinearity and interactions. This group of the method may

suffer from the “curse of dimensionality” more often compared to variable selection methods in the context of epidemiological studies which usually have relatively small sample sizes. Therefore, the methods may suffer from not having enough data points to allow the non-linear transformations and hence result in estimation with low precision due to the additional variances introduced by the transformations [6,51]. Moreover, a further limitation is that the final estimated parameters of the model lose interpretability because so many transformations have been performed on the original variables. In other words, it is difficult to communicate the analysis results to knowledge users in a simple way.

More generally, many methods involve the use of Bayesian statistics, which is a modern brand of statistics that is constantly evolving. The fundamental conceptual differences between Bayesian statistics and traditional frequentist statistics are commonly depicted as the different interpretations of probability under both frameworks [52–54]. In Bayesian statistics, the probability is defined as the degree of belief of a certain event while in frequentist statistics, the probability is defined as the relative frequency of a certain event under a large number of trials. This conceptual difference indicates that Bayesian statistical methods models the uncertainty quantified by probability [54]. In terms of epidemiology, the conceptual differences between Bayesian statistics and frequentist statistics can be illustrated by the different underlying assumptions of the estimating parameters in the model. For Bayesian statistics, the assumption is that the parameters estimated are probability distributions rather than a single fixed value for frequentist statistics. This is very appealing to epidemiologists because it gives a clearer understanding of the analysis results and uncertainty compared to frequentist methods such as 95% confidence intervals and p-values [53]. This is particularly relevant in environmental epidemiology where effect sizes are usually small and not statistically significant. In this case, Bayesian methods can give a more nuanced quantification of uncertainty.

A related collection of methods is machine learning techniques. Machine learning approaches are becoming more and more prevalent in the field of epidemiology and focus primarily on prediction tasks [55]. This approach typically splits available data on hand into training data and testing data. Training data is used to build statistical models and testing data is used to evaluate the built statistical models. By employing statistical techniques such as cross-validation and bootstrap sampling, the built statistical model is evaluated by minimizing the residual sum of squared errors of the predicted outcomes. Machine learning approaches can be extremely powerful and produce very good prediction results when the available data is rich and abundant[48].

In the following sections, a detailed comprehensive overview of all statistical methods in the analysis of the mixture is conducted to compare and contrast the strengths and limitations of different approaches.

2.2.1 Linear regression and extensions

Before reviewing the numerous methodologies for multiple correlated exposures, we give a brief overview of linear regression analysis and it's extensions. Linear regression is the most commonly used statistical tool for epidemiological assessment [56]. In the context of environmental chemical mixture exposure and pregnancy outcomes, linear regression approaches can be separated into single pollutant analysis and multiple pollutant analysis. For example, in the study of gestational PCB exposure and infant birth weight, we consider models that incorporate a single PCB, or alternatively, several PCBs simultaneously.

For both linear regression approaches, the statistical model presumes that the outcome variable (e.g. birth weight) follows the following equation:

$$Y = \alpha + \beta_x X + \beta_c C + \epsilon$$

where α is the Y-intercept, β_x is the vector of regression coefficients of risk factors, X is the vector of risk factors such as environmental chemicals, β_c is the vector of regression coefficients of confounders, C is the vector of confounders such as age and socioeconomic status and ϵ is the random effect in the model.

For single pollutant analysis, the researcher usually performs variable selection to select the most important variable before fitting the models. This can be done statistically by using methods such as stepwise regression and LASSO or it can also be determined by empirical evidence from the literature. Regression in parallel is also a popular approach where single pollutant analysis is done for multiple chemical variables separately and the results are aggregated together afterwards.

For multiple pollutant analysis, several approaches can be used to assess the effects of environmental chemical mixtures. The first one is multiple linear regression with all variables present which can be described by the equation: $Y = \alpha + \beta_x X + \beta_c C + \epsilon$. The second one is multiple linear regression using the sum-of-chemical approach. This method models the mixture variable by summing up the individual chemicals within the mixtures. The regression model fits using the following equation:

$$Y = \alpha + \beta_x \sum_{j=1}^{k} X_j + \beta_c C + \epsilon$$

where β_x is the mixture-specific regression coefficients (in this approach, the mixture is defined as the total amount of exposure of all individual chemicals) and X_j are the jth environmental chemicals. The third approach is multiple linear regression with averaging. This method models the mixture variable by taking the average of the individual chemical concentrations within the mixtures (e.g. the average of the PCB concentrations for a dozen+ PCB biomarkers, adjusted for lipid dilution). The regression model fits using the following equation:

$$Y = \alpha + \beta_x \sum_{j=1}^{k} X_j/k + \beta_c C + \epsilon$$

where β_x is the mixture-specific regression coefficients (in this approach, the mixture is defined as the average amount of exposure of individual chemicals within a mixture) and X_j are the jth environmental chemicals.

Stepwise regression also belongs to the group of linear regression methods [57]. It modifies the multiple linear regression models by performing automatic addition/substitution/deletion of variables using standards such as Akaike information criterion or Bayesian information criterion [5]. However, stepwise selection has been criticized for the analysis of correlated exposures because it tends to give analysis results that are highly unstable and susceptible to random variations in the data. For example, in the analysis of PCBs, We anticipate that stepwise regression will select PCB molecules in a way that is random and haphazard depending on the quirks of the data.

2.2.2 LASSO and LASSO based extensions

Efron et al. [58] developed an improved version of a stepwise procedure known as least angle regression which is proved to be mathematically equivalent to the LASSO method that will be discussed in detail. As demonstrated by Tibshirani, et al. [59], the multiple linear regression model can be modified with a penalty term λ for shrinkage of the regression coefficient. This modification is known as ridge regression illustrated by the following formula [59].

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2.$$

As λ takes values from 0 to ∞, the coefficient takes an increasing amount of shrinkage to regression coefficients from the original ordinary least square estimate to 0. Therefore, ridge regression performs shrinkage of regression coefficients without a selection of variables [59]. LASSO (Least Absolute Shrinkage and selection operator) is a modification of ridge regression which estimates coefficients of the model that minimizes the following criterion where λ is a tuning parameter for shrinkage [59]:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|.$$

It is different from Ridge regression in the penalty term where L2-norm is replaced with L1-norm. This favours shrinkage on parameters so that most of the regression coefficients tend to 0, achieving the purpose of variable selections. Therefore, if there are predominant chemicals within the chemical mixture, LASSO is able to select the chemical variables out of the mixture that explains the greatest percentage of variation in the outcomes and produces the estimates of the corresponding regression coefficients.

Although LASSO performs really well in many applications, Zou and Hastie [60] demonstrated two key limitations of the LASSO algorithm. The first one is that when the number of variables p is greater than the sample size n, LASSO can only select at most n variables for the model. The second one is that when correlated variables are present in data, LASSO tends to select one or only a few of them in a half-hazard manner while shrinking the regression coefficients of other correlated variables to 0.

To address the first limitation, Zou and Hastie [60] proposed a combination approach of ridge regression and LASSO, known as Elastic net to enable variable sections greater than n when p >n, in which the penalty term can be described by the following criterion, where α is a tuning parameter between 0 and 1 to dictate the proportion of L1-norm penalty versus L2-norm penalty [60] :

$$\lambda\sum_{j=1}^{p}[\alpha\beta_j^2 + (1-\alpha)|\beta|]$$

Due to the combining of L1-norm penalty and L2-norm penalty, correlated variables are forced to be together, meaning that they are either all selected or all removed by the model. However, this approach loses its ability to accurately estimate the effects of correlated variables that have opposite effects on the outcome because of the ridge penalty forces similar regression coefficients estimations (i.e same direction). For example, fish consumption is usually correlated with both mercury concentration and omega-fatty acids concentration in blood. While the true health effect of these two variables on child neurodevelopment (e.g IQ) is anticipated to be positive for omega fatty acids and negative for mercury concentrations, elastic net tends to generate unstable estimates of regression coefficients for these variables.

In order to achieve more accurate estimation of regression coefficients to better characterize the true underlying relationships in the population, Zou [3] developed a procedure called adaptive LASSO where he introduces a weighted penalty term in LASSO illustrated by the following equation[60,61]]:

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|,$$

$$w_j = 1/|\hat{\beta}_j^{\text{ols}}|^r, \qquad w_j = 1/|\hat{\beta}_j^{\text{ridge}}|^r$$

Essentially, adaptive LASSO is a two-step procedure where the first step is to identify the importance of all variables by fitting an ordinary least square regression model or ridge regression model and estimate regression coefficients for all variables. The variables with a bigger magnitude of regression coefficients are considered to be more important. The second step is applying LASSO with a weighted penalty where the more important variables identified in the previous step receive less amount of shrinkage. Zou[61] has demonstrated mathematically that adaptive LASSO is able to approximate the true underlying model asymptotically. This property is known as the oracle property[3]. However, in actual practice, we do not have an infinite amount of observations in our data. Therefore, the parameter estimated by adaptive LASSO can still be inaccurate. As a matter of fact, when collinearity is high among variables since the estimated regression coefficients in the first step becomes more unstable, adaptive LASSO tends to generate poorer estimates compared to LASSO. This echoes with the concept of bias-variance tradeoff. When the overall causal relationship is high-dimensional and complex and when the amount of data is relatively low resulting in larger variances, the precision of a more complex model is often worse than a simpler model.

2.2.3 Latent variable methods

A quite different modelling approach to multiple correlated variables is to use latent variable methods. The primary motivation to use latent variable methods is to reduce the dimensions of exposure data [62]. Thus the approach is similar in spirit to variable selection, which also seeks to discard predictor variables. The dimensions of exposure data are characterized by the number of existing exposure variables in the data. In order to estimate regression coefficients for all the exposure variables, the amount of data required increases exponentially [62]. Mathematically, when performing regression analysis, the number of exposure variables (p) and the number of data points (n) must satisfy criteria “n greater or equal to p” to achieve a solvable solution. This phenomenon is known as “the curse of dimensionality” [63]. However, only achieving a solvable solution is not sufficient to generate reliable estimates (small standard errors) due to variations from different datasets. Therefore, it is sufficient to say that for the purpose of estimating regression coefficients of multiple exposure variables, the number of data points (n) should be much larger than the number of exposure variables (p) and this is why dimensions reduction techniques are warranted in the analysis of high-dimensional data [62].

In the field of epidemiology, latent variable methods have two predominant contexts [62]. The first context is exploring available data to identify potential latent exposure variables, typically much fewer than observed exposure variables, to capture the most amount of variances in the outcome variable. The second context is performing parameter estimations of hypothetical latent variables, usually calculated from regression coefficients, given a pre-proposed hypothetical causal framework.

Underneath the first context, two statistical techniques are commonly used known as exploratory factor analysis (EFA) and principal component analysis (PCA) [64]. Both of these two techniques are data driven and are able to identify latent variables. In EFA, the latent variables identified are called common factors; in PCA, the latent variables identified are called principal components. Both EFA and PCA share similar statistical assumptions such as the linear relationship between outcome variables (e.g birth weight) and latent variables, normal distribution for each observed variables and bivariate normal distributions for each pair of observed variables. However, EFA and PCA are different with respect to the correlation matrix involved in the computation. In PCA, variance from each observed variable, variance common among variables and error variances are all included in the matrix while in EFA, only variances shared among observed variables are included. Mathematically, the PCA model can be expressed as:

$$X = \beta Z_0$$

where X is a matrix of observed variables, Z_0 is a matrix of principal components and β is a matrix of weights. In contrast, the EFA model can be expressed as:

$$X = \beta_0 + \gamma Z + \varepsilon$$

where β_0 is a matrix of intercepts, X is a matrix of observed variables, Z is a matrix of common factors, γ is a matrix of weight and ε is a matrix of error associated with factors.

Orthogonal rotation and oblique rotation are common rotation methods used to determine the number of common factors and the number of principal components [64]. It is noted that the latent variables in both of these two methods are essentially linear combinations of the observed variables. However, since there is no measurement error term in the PCA model, it does not assume the existence of a model framework while in the EFA model, the existence of a model framework is assumed.

The second context in which latent variable methods are used in epidemiology is when using the statistical technique called confirmatory factor analysis(CFA) [62]. In this approach, the researcher hypothesizes a causal framework before performing any transformation of the data. The researchers decide on the exact number of latent variables and how they are related to the observed variables in the data. Similar to EFA and PCA, the assumptions are linear relationships between observed variables and latent variables, normal distribution for each observed variables and bivariate normal distributions for each pair of observed variables [65]. Typically, multiple plausible hypothetical causal frameworks would be evaluated during the model development process to generate different estimates.

EFA and CFA are also sometimes referred to as the path analysis and the measurement model components of structured equation modelling (SEM) [66]. Path analysis depicts the procedure to uncover the underlying model framework using information from available data and the measurement model depicts the procedure to assess the impact given a particular model framework. Both EFA and CFA assume the existence of a particular causal structure, therefore, the approach that combines both EFA and CFA is also used in some fields such as psychology [39] where EFA and CFA are performed sequentially.

PCA is different from EFA and CFA in the sense that it does not assume a causal structure [64]. The latent variables or the principal components in the PCA do not have a

theoretical meaning. They are strictly linear combinations of observed variables. Therefore, the regression coefficients on the latent variable in PCA analysis is not exactly interpretable [67]. It is a simple technique to reduce the amount of correlation and dimension within the data. PCA analysis is also usually used in prediction modelling where the researcher is only interested in predict outcome more accurately with available data, but not interested in the specific causal structure and pathways between the observed variables and the outcome of interest.

An extension of PCA known as sparse principal component analysis (SPCA) has gained more attention in the study of the environmental chemical mixture because of its ability to produce sparsity after identification of principal components [68]. Sparsity means that the regression coefficients of most of the variables that are not identified to be included in the principal components are reduced to zero. Using techniques originated from the LASSO-based method, SPCA differs from PCA by introducing a specific criterion for the penalization of each principal component determined [68]. The principal component that has minimal influence on the residual sum of squared error is forced to be close to zero due to this penalization introduced. Therefore, SPCA can further reduce the collinearity present in the data by limiting the number of principal components compared to ordinary PCA. However, when working with a large number of variables that all contribute small effects to the outcome, SPCA tends to perform poorly due to the elimination of potentially important variables or principal components specifically for the purpose of creating sparsity [68].

2.2.4 Nonlinear exposure-response surface methods

Due to the extremely complex interactions and unidentifiable causal structures among chemicals in mixtures, researchers developed approaches that use non-linear exposure-response surface estimation techniques [55,69,70]. It is built upon linear methods through expansion and transformation. In a review of epidemiology analysis of endocrine disrupting chemical mixtures conducted by Lazarevic et al. [71], the most notable non-linear methods include the Generalized Additive Model (GAM), the Multivariate Adaptive Regression Splines (MARS) and finally weighted quantile sum regression (WQSR). We describe each of these three non-linear methods in detail.

Generalized Additive Model is developed through two-step modifications of the multiple linear regression models[69,70]. Recall the equation for multiple linear regression is given by:

$$Y = \alpha + \beta_x X + \beta_c C + \epsilon$$

where α is the Y-intercept, β_x is the vector of regression coefficients of risk factors, X is the vector of risk factors, β_c is the vector of regression coefficients of confounders, C is the vector of confounders and ϵ is the random effect in the model. By performing different types of transformation on the outcome variable Y, different generalized linear models (GLM) can be expressed by the following equation:

$$g(\mu(X)) = \alpha + \beta_1 X_1 + \ldots + \beta_p X_p$$

where g(u(X)) is the transformed outcome variable. For example, if g() is the logit function, then the generalized linear model after transformation is the logistic regression model. Generalized Additive Models build upon the generalized linear model framework by expanding the regression coefficient vector β. It can be described by the following equation:

$$g(\mu(X)) = \alpha + f_1(X_1) + f_2(X_2) + \ldots + f_p(X_p)$$

where f(x) are different possible mathematical functions instead of regression coefficients. The typical f(x) includes, but is not limited to polynomial functions, logarithm-based functions, and exponential functions [70]. The fitting of the model is by a back-fitting algorithm that iteratively fits individual f(x) while calculating residuals based on the fit. Through numerous iterative cycles, the minimized residual model converged is considered as the final model. Published examples of epidemiology studies using GAMs include the work of Zanobetti et al., Ramsay et al., and Dominici et al. which all focus on the assessment of air pollutants. [72–74]

Weighted quantile sum regression [75] is an interesting nonlinear method that combines the concept of percentiles and mathematical transformations to estimate the effect of a mixture of correlated variables on the outcome variable. In this model, correlated variables are combined into an index where each variable is scored into quantiles. The basic index model takes the following form [75]:

$$g(\mu) = \beta_0 + \beta_1 \left(\sum_{i=1}^{c} w_i q_i \right) + \mathbf{z}'\boldsymbol{\varphi},$$

where g(u) denote any types of link functions in the generalized linear model (e.g. logit), w is the unknown weight for the ith variable, β_1 is the regression coefficient for the weighted

quantile sum, z is a vector of covariates and phi is a vector of regression coefficients for covariates.

After the formulation of the above basic model, the total data is split into a training dataset and validation dataset [75]. Using bootstrap sampling techniques, maximum likelihood estimators are applied for each bootstrap sample to optimized the weights in the basic model. The total sums of weights are constrained to one in order to potentially remove some variables that is not as significant or influential on the data. The final estimated model is determined by averaging over all bootstrap samples for the regression coefficients.

One unique advantage of WQSR is the use of quantile to remove extreme outliers in the data, which is typical for environmental chemical mixtures (e.g. extreme levels of chemical exposures). This procedure reduces the amount of variance since it transforms continuous variables into categorical variables in the model. However, this procedure also produces a disadvantage since chemical exposure is drastically different among individuals. By using quantile approaches, it is impossible to distinguish between people at 90 percentile from people at 76 percentile since they are both grouped into the same quantile [76]. In addition, chemical exposures in the population are observed to have long-tail distributions, which result in uneven differences across quantiles. For example, people at 80 percentile may have exposure concentrations five times more than people at 70 percentile and twenty times more than people at 60 percentiles. Therefore, the interpretation of the model needs to be carefully articulated and it is difficult to interpret the regression coefficients since the weight assigned to each variable is different and the quantile assigned to each variable has different distributions as well.

Classification And Regression Trees (CART) are a unique method that operates under the recursive partitioning algorithm [6]. Given a specific data sample, this algorithm selects a single point in one variable to separate the whole data into two splits to reduce the residual sum of squares for multiple cycles until it reaches some specific stopping criteria set by researchers. Consider the following two equations:

$$RSS(\text{Full Data}) = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$RSS(\text{Split}) = \sum_{i \in R_1}(y_i - \bar{y}_1)^2 + \sum_{i \in R_2}(y_i - \bar{y}_2)^2$$

where RSS stands for the residual sum of squares, the algorithm aims to minimize the RSS of each cycle by selecting a unique point in an exposure variable to split the outcome variable into two sets.

The criteria of the stopping rule are decided by the researcher [55]; One common criterion is the number of observations in the terminal node, another common criterion is when the RSS is not improved by additional splits. After the classification/regression tree is constructed, pruning is usually performed to reduce the unnecessary branches in the tree structures. Sometimes, an additional node on the tree may improve RSS by a very little amount, yet introducing a larger amount of variances. The pruning of the trees can be done by limiting the size of the tree (the number of splits) or by cross-validation methods when the sample size is relatively large.

Tree-based methods are usually very variable and unstable because they are extremely discrete [77]. They usually have poor performances on smooth relationships including linear relationships. By splitting the data, each time it is automatically assuming interactions among the exposure variables. Consequently, a tree-based method on its own is not a very good modelling method. However, when it is combined with other techniques, it is able to produce much better prediction results [77]. Published epidemiology studies that employ CART methods include the work of Stel et al. which aim to study recurrent falling of the elderly in community-dwelling, the work of Mueser et al. which aim to study the risk factors of substance use disorders in hospitalized psychiatric patients and Nishida et al. which aim to study the association between smoking and obesity and the risk of developing periodontitis [78–80].

Multivariate adaptive regression splines (MARS) is built upon the previously described generalized additive model and tree-based model [81]. As demonstrated in the previous section, the tree-based regression model handles the splitting of the data discontinuously. Therefore, it lacks the ability to model smooth monotone patterns such as linearity. On the other hand, the generalized additive model, when operating with high dimensional data, face the problem with too many parameters and basis functions to select from [77]. MARS works on the following regression models:

$$\hat{Y} = \hat{\alpha} + \hat{\beta} h_{j1}(X)$$

where h(x) represents the basis function selected by the model. Then the MARS model will consider each single observation point as potential knots for splitting data into half, meaning

that for each possible knots, a pair of basis function is available to select from represented by the following equations:

$$(x-t)_+ = x-t \quad \text{if } x > t \qquad (t-x)_+ = t-x \quad \text{if } x < t$$
$$= 0 \quad \text{if } x \leq t \qquad = 0 \quad \text{if } x \geq t$$

,

where t is the knot selected for splitting the data. Then a linear combination of these paired basis functions is constructed for the final model expressed in the following equation:

$$\hat{Y} = \sum \beta_k h_x(X)$$

where β is a different regression coefficient assigned for each paired basis functions. Similar to the CART model, the MARS model also requires a stopping criterion selected by the researcher [77]. The commonly used stopping criterion is the number of knots used for building the models. Typically, the fully developed MARS model tend to overfit the data given the numerous number of basis functions selected. Therefore, the pruning of the model works in a similar manner to the CART model. For each knot identified by the model, the backward elimination method is used to determine whether the elimination of the left side basis function or the right side basis function is able to significantly improve the mean squared prediction error (MSPE) by doing cross-validations. In general, MARS combines the advantages of GAM and CART by using regression spline models to replace tree-based split to allow modelling of linearity and other monotone relationship with the use of knots to avoid excessive parameterization and interactions that can be potentially introduced by the GAM approach [82].

2.2.5. Bayesian methods

As discussed above, Bayesian statistics is a modern branch of statistics that is constantly evolving. Different fields including epidemiology started to adopt the Bayesian statistical framework [51]. Before going into details of different types of statistical models, a brief introduction of Bayesian statistics is necessary to understand the importance and motivation behind this type of approach.

The general procedure for any Bayesian methods follows the following equation known as Bayes' theorem:

$$P(parameter/data) = \frac{P(data/parameter)P(prior)}{P(data)}$$

where P(prior) is defined as the prior probability distribution, P(data/parameter) is defined as the likelihood given sampling data, P(data) is defined as the marginal likelihood given sampling data and P(parameter/data) is defined as the posterior probability distribution [53]. Therefore, the typical Bayesian approaches can be summarized as combining information obtained from sampled data with prior beliefs about data in order to generate a final estimate which is depicted as the posterior probability distribution. For example, we can calculate the posterior distribution of the odds ratio for epidemiologic studies.

As the sample sizes increase, the likelihood part of the Bayesian estimate (posterior distribution) receives higher weight in the final estimation. In the field of epidemiology, since it is costly and time-consuming to collect human data, the sample sizes of the study are typically limited. Therefore, the selection of priors becomes very important. Typically, priors are selected based either on a review of existing research in the literature or informed by expert opinions [52]. However, according to a systematic review conducted by Rietbergen et al. [83], the process of prior elicitation, specification and evaluation are often lacking in Bayesian epidemiological analysis and a more transparent reporting of such process is critical. Conducting sensitivity analysis with different priors including non-informative prior to illustrate how much influences prior have on the final posterior probability estimation is encouraged [83].

With recent developments in computer algorithms (e.g. Markov chain Monte Carlo computation), the uses of Bayesian methods have become more and more popular because the flexibility of Bayesian methods is able to model complex relationships without introducing a significant amount of computation time. The use of Bayesian methods also allows the researcher to greatly reduce the amount of variability among the data by introducing prior information. Published epidemiology studies using Bayesian methods include the work of Papathomas et al. which aims to study the risk factors of lung cancer among non-smokers and the work of Valeri et al. which aims to study the relationship between exposure to metal mixtures and neurodevelopmental outcomes in children. [84,85].

With regard to the Bayesian analysis of multiple correlated exposures (chemical mixtures), we now describe two different Bayesian methods: Bayesian Kernel Machine Regression (BKMR), and Bayesian hierarchical linear models (BHLM).

Bobb et al. [36] proposed the Bayesian approach to kernel machine regression. The method is extremely popular and combines Bayesian inference with modern machine learning methods. The kernel machine regression models independent variables (exposures) with dependent variables (outcome) using the following mathematical formula:

$$g(\mu_i) = h(z_{i1}, \ldots, z_{iM}) + \beta \mathbf{x}_i, \quad i = 1, \ldots, n$$

where g is a monotonic function that denotes some form of the transformation of the outcome variable, h is a flexible exposure-response function, z is the vector of predictors, x is the vector of covariates and B is the coefficients for covariates. Since the kernel machine regression allows a great degree of flexibility, the non-additive and non-linear relationships can be modelled using the proposed framework. One of the common kernel function h used is the Gaussian kernel, which can be represented by:

$$K(\mathbf{z}, \mathbf{z}') = \exp\left\{ - \sum_{m=1}^{M} r_m (z_m - z'_m)^2 \right\}.$$

where z and z' denote the predictor of two individuals in the sample, and rm denote the tuning parameter that controls the smoothness of the kernel function. Essentially, the effect on the outcome is shrunk for individuals with closer predictor values to smooth the exposure-response curve.

Since BKMR combines the advantage of Bayesian approaches and kernel machine regressions approaches, it allows the researchers the flexibility of the exposure-response relationship while restricting the amount of variance. Therefore, it is possible to model multiple pollutant mixture in addition to the modelling of individual pollutants. Bobb et al. also published a statistical package in R for the implementation of BKMR in the epidemiological analysis [86]. Some notable published work utilized BKMR in epidemiology include the work of Buckley et al. which aims to study the window of susceptibility to environmental exposures [87] and the work of Harley et al. which aims to study the associations between perinatal phthalate exposures and childhood obesity [88].

The Bayesian hierarchical linear model [51,89–91] is a simpler form of Bayesian methods originally proposed by Greenland to analyze multiple correlated exposures [92]. The hierarchical component of the method is from the prior distributions introduced on the

regression coefficients in the model. By allowing a different type of prior distributions determined from available literature or existing evidence, Hamilton Monte Carlo sampling method can be applied to generate a posterior credible interval of estimated regression coefficients [51].

The main advantage of this method is that it can be implemented relatively easier than BKMR as well as a significant amount of flexibility because of the number of different prior distributions that can be chosen for any particular data. However, since this method assume a linear relationship between exposure variables and outcome variables, the fit according to this model is usually not as accurate as BKMR when a potential non-linear relationship is present in the data [83]. Examples of published environmental epidemiology literature that utilizes BHLM include the work of Braun et al. and Hamra et al. which both aim to study the association between exposure to endocrine disrupting chemicals and autistic behaviour [90,91].

2.2.6. Modern machine learning approaches

Machine learning approaches of model selection for epidemiology is based on the quantitative counterfactual framework and focus on prediction tasks [70]. When building models for prediction purposes, machine learning approaches typically distinguish between training data and testing data where training data is used to build statistical models and testing data is used to evaluate the built statistical models. In this type of approach, the available data on hand dictate the modelling process and it is not uncommon to build different optimal models when using different selection criteria of training data and test data. One general approach of training data and test data selection is splitting by a certain percentage such as 80% training data with 20% test data. A more complicated approach includes general cross-validation and bootstrap sampling. General cross-validation is accomplished by splitting data multiple times by a certain percentage point and each split result in a different sub-sample of training data and test data [70]. Bootstrap sampling uses the idea of resampling with replacement from original data to construct test data. Both of these complicated approaches aim to reduce the possibility of overfitting due to specific choices of training and test data [77].

Most machine learning methods generate the mean estimates of the outcome Y by minimizing the residual sum of squared errors. The resulting mean estimates of Y have variances associated with it in the following form:

$$Var(\bar{Z}) = \frac{1}{B^2}\left[\sum_{b=1}^{B} Var(Z_b) + \sum_{b=1}^{B}\sum_{k\neq b} Cov(Z_b, Z_k)\right]$$

where Z is representing any outcome variables in data, B representing the sample size [77]. By transforming the above equation in terms of the pairwise correlation between individual Z, the following equation is deduced:

$$Var(\bar{Z}) = \sigma^2(\rho + [1-\rho]/B)$$

where σ^2 is the variance of the mean estimates and ρ is the pairwise correlation between Z.

From this equation, if the pairwise correlation between Z is equal to zero, then $Var(Z) = \sigma^2/B$, whereas if the pairwise correlation between Z is equal to one, then $Var(Z) = \sigma^2$. By making B as large as possible, the total variances approach $Var(Z) = \sigma^2 * \rho$. Therefore, this demonstrates that by effectively making sample size bigger, it is possible to reduce variances associated with the mean estimates only to a certain amount and it is highly dependent on the pairwise correlations. The idea of the ensemble method is therefore proposed to construct individual weak predictors that are uncorrelated with high individual variances to replace strong predictors that are more correlated with low individual variances to achieve better prediction performances. Out of the ensemble method family, the Random forest and Gradient Boosting Machine are the most well-known methods.

Random forest [55], as suggested by its name, is an advanced modification of the tree-based methods. It also utilizes resampling techniques such as bootstrap to generate a test set for evaluation of prediction performances. Similar to regression trees, the model randomly selects split points for each independent variable to recursively partition data into splits to predict the outcome variables. The unique feature of random forest is to create as many regression trees as possible while selecting a random subset of the total independent variables in the data. As a result, multiple regression trees will be created. This group of regression trees would have high individual variances, but low correlation among each other, generally refer to as weak learners. Afterwards, all these weak predictors are averaged across different regression trees to produce the final ensemble random forest.

Random forest is a significant improvement in modern machine learning approaches,

however, it requires specific criteria for it to perform well. As noted by [55], the two most important criteria are the selection of hyperparameter m, which represents the number of independent variables to select for each regression tree and the sample size of available data.

The selection of hyperparameter m is dependent on the available data and prediction task. Ideally, the researcher wants to select m that is small enough to have a low correlation among the individual independent variables, but not too small to result in extreme increases of individual variances of each independent variable. For prediction tasks that deal with continuous outcome variables, the recommended hyperparameter m is one-third of the total amount of independent variables. For prediction tasks that deal with categorical outcome variables, the recommended hyperparameter m is the square root of the total amount of independent variables. Regarding the sample size for using Random forest for prediction tasks, a minimum of 500 data points is needed and more data points would generate better results because it yields greater reduction of variances by averaging the results.

Gradient boosting machine(GBM) [82] is another type of ensemble method that was built upon the previously described LASSO method as well as tree-based method. It creates ensemble predictors differently from the random forest by adding predictors iteratively instead of generating large pools of weak predictors simultaneously.

The general approach of GBM is to approximate LASSO estimates by increments. Consider the following equation:

$$f(x) = \sum_{k=1}^{K} \beta_k T(x, \Theta_k)$$

where B represents the estimates of the regression coefficient and T(x) represents the tree structures in the current iteration. Starting from the following condition:

$$f^{(0)}(x_i) = \bar{y},\ e_i^{(0)} = y_i - f^{(0)}(x_i)\ i = 1, \ldots, n,\ \boldsymbol{\beta}^{(0)} = 0$$

The algorithm searches for a potential tree structure with regression estimates that are most correlated with the current residual. Then the equation is updated as follows:

$$f^{(m)}(x_i) = f^{(m-1)}(x_i) + \nu T(x_i, \Theta_{k^*})$$

where m represents the current iteration and v represents a small value in terms of change updated for the tree structures. A new residual is calculated for the next cycle of boosting by the following equation:

$$e_i^{(m)} = y_i - f^{(m)}(x_i) = e_i^{(m-1)} - \nu T(x_i, \Theta_{k^*})$$

Similar to the random forest, the hyperparameter m, v and tree sizes are all tuning parameters for the GBM model and for each unique sample data, it is selected by some resampling techniques such as bootstrap or cross-validation. It is noted by Trevor et al. that when m approaches infinity, the updated residual approaches zero, indicating the extreme case of complete data overfitting [55]. The recommended tree sizes for GBM is 6 which represent 5-way interactions of independent variables. The quantity v is the learning rate of the boosting machine which indicates how fast each update of increment is computed. The smaller v is, the more smooth the iteration residual is computed and the longer it takes to converge to the final boosted model. Gradient boosting can be generalized by using other arbitrary loss functions instead of the residual sum squared error [55]. This allows a more flexible framework such as Poisson regression, logistic regression and quantile regression model to use gradient boosting algorithm to approximate the final model.

GBM and random forest are the most commonly used machine learning approaches in prediction tasks [70]. Both are extremely powerful in its ability to predict outcomes with significant lower errors compared to traditional approaches. However, these methods generally require a very large sample size to achieve greater performances since the model complexity is high for these ensemble methods. In the case of epidemiology, these methods are usually not ideal due to the limitation of the sample size bounded by the cost of recruiting human subjects. In addition, the number of variables is typically high in the case of environmental chemical mixtures, and the potential benefits gained by averaging weak predictors may not outweigh the significant amount of individual variance associated with each independent variable. Therefore, careful considerations need to be taken before employing an ensemble method for causal inference of environmental and perinatal epidemiology.

2.3 Examples of epidemiology studies using innovative mixture methods in the analysis of pregnancy outcomes

In this section, we review seven examples of papers that employ innovative methods to examine the effects of environmental chemical mixtures in relation to pregnancy outcomes.

In particular, birth weight (five papers) and preterm birth (two papers) are considered as the outcome while any biomarkers of gestational environmental chemical exposure that are potentially involved in any biological pathway in the earlier sections are considered. Papers with traditional statistical approaches are excluded. We also excluded papers examining other child health outcomes (e.g neurodevelopment), as well as papers examining environmental pollutants other than toxic chemicals (e.g air pollution).

Crucially, the selected papers are by no means an attempt to be a comprehensive review of the literature (e.g. a systematic review [5]). Instead, this Section is written in the spirit of the biostatistics literature, where the hallmark is a selection of real-data examples, hand-picked by the authors, which demonstrate opportunities and pitfalls in statistical science. By examining the state of the art methods on the estimation of the effects of environmental chemical mixtures on adverse pregnancy outcomes, it is possible to identify the strengths and weaknesses of each method. In addition, a comparison of the effect estimates generated by traditional methods and the effect estimates generated by modern methods provides potential insights and challenges on the causal inferences of the relationship between environmental chemical exposures and pregnancy outcomes.

2.3.1. Innovative mixture methods for birthweight outcome

Woods et al. [89]: Gestational exposure to endocrine disrupting chemicals in relation to infant birth weight: a Bayesian analysis of the HOME study (2017)

Woods et al. [89] examined the relationship between environmental chemicals and birth weight using the Bayesian hierarchical linear model (BHLM). 272 pregnant women enrolled between 2003-2006 from Health Outcomes and Measures of Environment (HOME) Study was used for the statistical analysis. The HOME study is a prospective birth cohort study in Cincinnati, Ohio. Biomarkers in the blood and urine sample collected at 16 weeks and 26 weeks gestation were used to quantify the concentration of chemicals in each individual human sample.

Woods et al. [89] utilized a direct acyclic graph to construct a causal diagram that informed the selection of potential covariates in the models. A total of 53 exposure variables span across five endocrine disrupting chemical classes, two heavy metals and two organophosphate pesticides were included in the model for birth weight, along with 9 potential confounders such as maternal age and household income. Gestation age was included in the model for birth weight using spline techniques to account for the non-linear relationship. The

Bayesian hierarchical linear model (BHLM) takes the following form of the regression model:

$$Y = \alpha + \beta_x X + \beta_c C + spline(GA) + \epsilon$$

For the BHLM approach, the author used uninformative prior distributions (mean=0, variance=10^5) on the group based regression coefficients (e.g. for the collection of PCBs that were included in the model) for the purpose of shrinkage. Because the chemicals within a given class are highly correlated, the shrinkage prior served to bias the regression coefficients towards the group mean. Hamilton Monte Carlo sampling with 20,000 iterations coupled with 500 burn-in iterations was used to generate the posterior credible interval estimates for model parameters. The final estimates from BHLM were compared with LASSO and elastic net method.

According to the results obtained by the BHLM model, a ten-fold increase in gestational exposure towards bisphenol A (BPA), polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs) and organochlorine pesticides (OCPs) had a small or near-zero association with birth weight based on both 95% and 50% credible interval. Whereas ten-fold increase in gestation exposure towards perfluoroalkyl substances (PFAS), lead and organophosphate pesticides (OPPs) were shown to have an imprecise association with lower birth weight based on 50% credible intervals. It is also documented that LASSO and elastic net regression coefficient estimates were shown to have a larger magnitude with a wider confidence interval.

The major strength of the BHLM employed in this paper is the ability to shrink the variances of regression coefficients to generate more precise results as compared with LASSO and elastic net [89]. However, the major limitation of BHLM is that it models each exposure to individual chemicals such as PCB 128 and PCB 153 separately (single pollutant assessment). Since environmental chemical mixtures such as endocrine disrupting chemicals are consisting of numerous individual chemicals with similar structures, the effect estimates (regression coefficients) produced by BHLM do not represent the effects of mixtures. Instead, it produces individual chemical-specific regression coefficients. Since exposure towards environmental chemicals is usually simultaneous and correlation among these individual chemicals are observed, the individual chemical-specific regression coefficients estimates lose meaning because they represent the effect of one individual chemical independent of the other which is never the case for environmental chemical exposures.

Petit et al.[23]: Association of environmental insecticide exposure and fetal

growth with a Bayesian model including multiple exposure sources: the PELAGIE mother-child cohort (2012)

Petit et al. [23] examined the relationship between environmental insecticide exposure and birth weight as well as head circumferences using an integrated Bayesian latent variable model. The Bayesian latent variable model is an extension to the confirmatory factor analysis model described earlier in section 4.2. It adds the Bayesian component by introducing prior information on the regression coefficients. Using data from the PELAGIE mother and child prospective cohort located in France established in 2002, a total of 1,213 samples were included for the statistical analysis. The environmental insecticide exposures were classified by four different potential sources: 1. Insecticide exposure from the non-organic diet, 2. Insecticide exposure from the household use of products applied to plants, 3. Insecticide exposure from the household use of products against insects, 4. Insecticide exposure from agricultural activities. The exposure data were characterized by a survey questionnaire with categorical ordinal responses span across two years.

The Bayesian latent variable model characterized four different sources as mixtures of different possible routes of exposure to insecticides. Different prior distributions were implemented for regression coefficients associated with different sources defined according to existing evidence about the impact on birth weight. The multinomial distribution was used to model non-organic diets; Bernoulli distribution was used for household use and agricultural activities. The link between the outcome and the latent variable was characterized by multinomial linear models.

The major results generated by the Bayesian latent variable model indicated that insecticide exposure from agricultural activities was associated with a 0.10cm decrease in head circumference with a 95% credible interval [-0.22cm, 0.01cm] and the exposure from household insecticide use to treat plants were associated with a 27g decrease in average birth weight with a 95% credible interval [-59.2g, 6.4g] and 0.12cm decrease in head circumference with a 95% credible interval [-0.26cm, 0.01cm].

The major strength of the Bayesian latent variable model is its ability to allow researchers to characterize mixture exposures by sources and generate effect estimates associated with each source [23]. Therefore, potential interventions can be implemented according to the sources to mitigate the health effects. However, the limitation of this particular study resides in the survey questionnaire design of the study. Potential bias associated with survey questionnaires such as recall bias when answer questions and selection bias resulted

from non-response to questionnaire due to specific confounders can be potentially magnified when using a Bayesian framework, especially when the researchers employed informative prior on the source-specific regression coefficients. In addition, the definition of insecticide exposure strictly by the source is potentially problematic since different insecticides have different chemical structures and may participate in different biological pathways. The response from one individual regarding insecticide can be non-analogous to the response from another individual. In conjunction with different perceptions on the ordinal scale in the questionnaire, additional sources of bias can be introduced into the modelling process.

Chiu et al. [93] : Evaluating effects of prenatal exposure to phthalate mixtures on birth weight: A comparison of three statistical approaches (2018)

Chiu et al. [93] examined three different statistical approaches to evaluate the association between phthalates mixture exposure during pregnancy and birth weight. They were interested to see whether different statistical approaches can lead to different estimates of the effects. Using data from the Environmental and Reproductive Health Study (EARTH) that consists about 300 mother-infant pairs located in Massachusetts, Boston from 2005 to 2016. Urine samples were used to quantify the chemical exposures and birth weight outcomes were collected from hospital medical records.

Regarding the statistical approaches, Chiu et al. [93] first identified potential covariates and confounders using the direct acyclic graph (DAG) conceptualized by prior knowledge informed by literature. Additional covariates such as parity and infant sex were also included in some models for the purpose of reducing random variability. Afterwards, single pollutant linear regression, latent variable methods including principal component analysis (PCA) and structural equation modelling (SEM), and Bayesian kernel machine regression (BKMR) were all used to compare and contrast the effect estimates. For single pollutant linear regression, each phthalate metabolite was analyzed separately. Two-way interactions were also incorporated in a linear regression approach to account for correlation among phthalate metabolites. For PCA, the varimax rotation method was used to produce the principal components from all phthalate metabolites based on the correlation matrix among these metabolites independent of outcome variables. Exploratory factor analysis was used to reduce the metabolites mixtures into two latent constructs for the fitting of SEM. For BKMR, the Gaussian kernel was used for each phthalate metabolite to allow non-linear smoothing of exposure-response surfaces.

The results obtained from the single pollutant linear regression model generated a

negative association between phthalate metabolite exposure and birth weight. However, all effect estimates had wide confidence intervals representing imprecise estimates. When potential metabolite interaction terms were included in the model, the effect estimates became unstable with some metabolite showing positive effects and some showing negative effects. This result effectively demonstrated the inability to model complex mixtures using simple linear regression models. The results obtained from both PCA and SEM indicate that phthalates metabolites mixtures can be characterized into two significant latent variables labelled as the DEHP component and non-DEHP component. The mixture specific regression coefficients for PCA and SEM also demonstrated a negative association between phthalates exposure and birth weight with non-statistical significances. It is noted that the effect estimates from latent variable model is smaller in magnitude compared to linear regression models and the confidence intervals associated with the estimates are much narrower compared to linear regression models. BKMR produced similar results to the latent variable method except for the fact that it identified two individual metabolites (MEP and MEHP) instead of latent mixture constructs.

Although all methods employed in this study generated imprecise estimates with very wide confidence intervals, latent variable model and BKMR effectively demonstrate their ability to model complex mixture exposure better than multiple linear regression models because of their ability to effectively reduce correlations and collinearity among individual phthalates metabolites by either reducing the number of variables using latent constructs (i.e. dimension reduction) or allowing non-linear exposure-response curve using kernel functions. In addition, latent variable models and BKMR were able to distinguish within the phthalates metabolites between major contributors and minor contributors. Therefore, although the overall effect estimates are statistically non-significant, the relative importance of the individual phthalate metabolites can be evaluated. On the other hand, the limitation of using latent variable models and BKMR is related to how collinearity is reduced in each approach. For latent variable models, the latent constructs were determined by correlation within the chemical mixtures independent of the outcome variable. Hence it is difficult to determine whether the latent constructs selected are actually important constructs associated with the outcome. Therefore, such constructs should be informed by external evidence for validation. For BKMR, since nonlinear smoothing techniques are being used, it should be noted that with limiting sample sizes, the number of variances introduced by more complex kernel functions might outweigh the benefit of collinearity reduced by the approach. Therefore, careful selection of priors and hyperparameters are needed to optimize the effect estimations.

Lenters et al.[94]: Prenatal Phthalate, Perfluoroalkyl Acid, and Organochlorine

Exposure and Term Birth Weight in Three birth cohorts: Multi-pollutant models Based on Elastic Net Regression (2016)

Lenters et al. [94] examined the association between exposure to individual chemicals within complex mixtures and birth weight reduction. The study cohort consists of 1,321 mother-infant pairs span from 19 different municipalities in Greenland, Ukraine and Poland. The chemical exposures were measured using biomarkers in blood and the birth weight outcome was collected from hospital records. Covariates were established according to the literature and ascertained by interviews and self-reported survey data. A total of 16 biomarkers span across perfluoroalkyl substances, phthalates metabolites and organochlorine pesticides were analyzed using elastic-net regression.

As described by Zou & Hastie [60], elastic-net is an extension of the LASSO-based method that combines the benefit of LASSO and ridge regression. The major benefit achieved by such a method is to effectively reduce correlation and collinearity among chemicals by the selection of a subset of variables. In this study, Lenters et al. [94] used 10-fold cross-validation techniques to determine the optimal hyper-parameter selected for the penalization term in the elastic-net models. 8 biomarkers were selected out of the 16 biomarkers for the unadjusted model. Upon the further adjustment of different sets of covariates and confounders, different numbers of biomarkers were selected. However, four biomarkers were consistently selected regardless of covariates adjusted. They are MEHHP, p,p-DDE, PFOA and MOiNP. The regression coefficients estimated for these four biomarkers were characterized by three different levels: 1. negative with statistical significance across all adjusted models (MEHHP, p,p-DDE); 2. negative with statistical significance across some adjusted models (PFOA). 3. positive with no statistical significance across all adjusted models (MOiNP). p,p-DDE and MEHHP were identified to be the most significant variables within the mixtures that associate with birth weight reduction. Interestingly, the magnitude of effect estimates generally decreases with more adjustments applied to the covariates. This indicates that a certain degree of unmeasured confounding or mediation effect is present.

The strength of employing the elastic-net regression model in this study is that it effectively reduced collinearity among the individual chemical variables by identifying four primary chemicals out of 16 individual chemicals. However, the limitation is apparent since the statistical interpretation of regression coefficients for these primary chemicals becomes ambiguous. It is difficult to state whether the effect estimates obtained is from the identified primary chemicals or a mixture of correlated individual chemicals because the regression coefficients produced to represent the hypothetical effects when all chemical variables not

selected are assumed to be absent. In addition, if several chemicals within a mixture actually have a health effect, but it is relatively smaller compared to other chemicals in the analysis. These chemicals would then be very likely removed from the model. Therefore, the selection of variables will be dependent on the comprehensiveness of biomarkers being measured in the study.

Additionally, for this particular study, three different cohorts from three different countries were combined together. Therefore, the amount of exposure within one community might be quite different from the amount of exposure within another community. By employing the elastic-net approach, the particular community with the highest amount of exposure might influence more on the final selection of variables in the model. Therefore, normalization techniques might be important in data preparation to ensure the scale of measurement is comparable across different population groups being selected by the study.

Govarts et al. [95] : Combined effects of Prenatal Exposures to Environmental Chemicals on Birth Weight (2016)

Govarts et al. [95] explored the association between birth weight and environmental chemical mixtures in a total of 16 individual chemicals spanning across heavy metals, polychlorinated biphenyls (PCBs), phthalates and perfluorinated compounds. Using data from the Flemish environmental and health studies (FLEHS II) mother-child cohort, a total of 248 samples were included in the statistical analysis. Principal component regression was performed on a subset of the total sample (n=157) due to missing data issues. Chemical concentrations were measured using biomarkers obtained from blood serum and birth outcomes were collected from medical records. Exposure measurements below the limit of detection were imputed with half of the LOD value. All biomarkers concentrations were log transformed and Z-scores were calculated to ensure the normalized comparison between different individual chemicals. Covariates were selected based on existing evidence from the literature.

For the single pollutant model, only arsenic was found to have a statistically significant association with lower birth weight. For one interquartile range increase in arsenic concentration in blood was associated with a 91g decrease of birth weight. For the principal component analysis, four and six principal components were identified using different samples consist of different numbers of individual chemicals. For the 16 individual chemicals sample population, 157 individual data points were included in the analysis and six principal components were identified. For the 12 individual chemicals sample population, 217

individual data points were included in the analysis and four principal components were identified. None of the principal components identified in the 16 individual chemicals sample population were associated with a significant change in birth weight and only one principal component (mainly consist of cadmium and arsenic) was identified with a significant association with birth weight reduction.

The strength of this study was that it used normalization techniques to deal with the scale of measurements combined with principal component analysis to assess the effect of mixtures. This ensured a meaningful comparison between different individual chemicals that may display drastically different scales in terms of blood chemical concentrations. This is crucial when comparing effect estimates across different chemicals since chemical concentrations without normalization can artificially inflate or deflate the estimates based on the scale used [28]. However, not all individual chemicals share the same exposure-response curves. For example, an IQR increase in heavy metal such as arsenic and cadmium are not necessarily comparable to an IQR increase in endocrine disrupting chemicals such as polychlorinated biphenyls [95]. Therefore, this study employed an assumption that all individual chemicals share the same exposure-response relationship. Additionally, due to the limitation of the sample sizes in this study, the statistical power of the model was greatly reduced as more chemicals are included in the model. With significant missing data issues, when more chemicals are included in the analysis model, sample sizes become smaller and smaller. This resulted in high variances associated with effect estimates in principal component analysis and yield confidence interval wide enough to conclude non-statistical significances.

2.3.2. Innovative mixture methods for preterm birth outcomes

The preterm birth outcome is another important pregnancy outcome in perinatal epidemiology, and that is possibly affected by environmental chemical exposures during pregnancy. Preterm birth is typically categorically defined by using cutoff points such as 37 weeks, or alternatively, the investigation can model the length of gestational duration directly. Typically, this birth outcome is not exclusively researched independent of birth weight and there are far fewer articles focus only on preterm birth compared to articles focus only on birth weight [96].

Chen et al.[96] Statistical methods for modelling repeated measures of maternal environmental exposure biomarkers during pregnancy in association with preterm birth (2015)

Chen et al. [96] examined nine different statistical approaches concerning the association of phthalates metabolites mixtures and preterm birth. In this study, preterm birth is defined as binary non-time varying the categorical outcome. All live birth with gestational age less than 37 weeks were considered as cases of preterm birth. The data were from a nested case-control study that consists of a prospective cohort recruited at Brigham and Women's Hospital in Boston. A total of 130 cases of preterm birth mothers with 352 random controls were included in the statistical analysis. Phthalate exposures were measured at four different time points during pregnancy using urine samples. Therefore, these measurements display two types of collinearity. The first type of collinearity is from collinearity among different individual phthalate metabolites and the second type of collinearity is from collinearity among different timed-measurements of the same individual phthalate metabolites.

For the statistical analysis, multiple logistic regression models, parallel cross-section logistic regression models, models using mean exposure as a summary, models using maximum exposure as a summary, hierarchical mixed effects models, Gaussian mixture clustering methods, functional clustering model, and function logistic regression models were compared and contrasted. To further elaborate on the methods employed, the models can be described as logistic regression with increasing complexities in terms of exposure characterizations.

For the multiple logistic regression models, all exposure measurements of phthalate metabolites at different time-points were included in the same model. For a parallel cross-section logistic regression model, only exposure measurements of phthalate metabolites at the same time-point were included in the model and different models were generated for different time-points. For models using mean exposure and models using maximum exposure, the phthalate metabolite concentrations either averaged across different time-points and the maximum phthalate metabolite concentrations of all different time-points were included in the model respectively. For hierarchical mixed effects model, time is included as an additional weighted term for determination of the mean values of exposures. Gaussian mixture clustering utilizes a splitting algorithm similar to tree-based recursive partitioning to separate the sample population into different clusters with Gaussian distribution. Functional clustering further removes the assumption of Gaussian distribution and employs non-parametric criteria such as K-means clustering which cluster the sample population based on the Euclidean distances. The functional logistic regression model takes time as an additional predictor for exposure measurements of individual phthalate metabolites.

Although the results generated from all different models display statistically non-

significant odds ratios between phthalates metabolites and preterm birth, the magnitude of these odds ratios were really different. This is expected because the exposure variables are characterized differently in different models. The use of these models should be determined by the research questions as noted by the author. In particular, for examining sensitive windows of exposure in association with an outcome, parallel cross-sectional logistic regression is recommended. For summarizing repeated measures to estimate the association, logistic regression with mean exposure or hierarchical mixed effects model are recommended. For identifying acute exposures and contribution of temporal patterns in exposure levels, functional clustering and functional logistic regression are recommended.

Kalloo et al. [97]: Chemical mixture Exposures during pregnancy and the Birth Outcomes (2018)

Kalloo et al. [97] examined the associations between exposure to 35 organic pollutants, cotinine and four metals and preterm birth defined by gestational duration. Birth weight, birth length and head circumferences were also examined. Using data from Health Outcomes and Measurement of Environment study (HOME) which is a prospective birth cohort study in Cincinnati, Ohio, 380 pregnant mother-infant pairs were included in the prospective cohort. Urine and blood biomarkers collected at 16 weeks and 26 weeks gestation were used to quantify the concentrations of chemicals.

The goals of the Kalloo et al. were to identify potential clusters within the study population representing different degrees of exposure towards different types of environmental chemicals. The statistical model used is selected to specifically model mixture effects of environmental chemical exposures instead of single pollutant analysis.

K-means clustering and non-negative principal component analysis were used to identify latent clusters and principal components representing chemical mixtures. Three clusters were generated representing three different levels of exposure to all chemicals in general. Six principal components were then identified within each cluster to further remove the collinearity present in the data space. Multivariate linear regression was then employed to evaluate the associations between gestational duration, birth weight, birth length and head circumferences and the produced principal component.

Although Kalloo et al. and Woods et al. (discussed in Section 5.1.1) used the same dataset, the research goals were different and the effect estimates obtained were different. Kalloo et al. aim to study the impact of mixtures while Woods et al. aim to assess the effects of

individual chemicals independently within a mixture. The results generated from this study showed that chemical biomarkers that share similar chemical structures tend to load onto the same principal components. Birth length reduction was associated with exposure to two principal components. One principal component was loaded by organochlorine pesticides, cadmium, and lead while the other principal component was loaded by mercury and monoethyl phthalate. It is also found that clusters representing higher general chemical concentrations had higher loadings on these two principal components, indicating that the majority difference of exposures among the study population is from the individual chemicals loaded on these two principal components. However, all of the associations observed for the gestational duration, birth weight and head circumferences had a wider confidence interval representing imprecise estimates. This is consistent with the results obtained by Woods et al.[89], which indicates that even after using different statistical techniques to account for correlations present in the data (Bayesian statistics in Woods et al. and PCA/clustering in Kalloo et al.), the obtained estimates remain imprecise. Potential innovation combining both approaches can be implemented to further account for the correlations present in the study.

The strength of this study is the flexibility in the modelling approaches. K-means clustering is a non-parametric statistical method and does not require any assumptions of probability distributions in the model. Therefore, it is extreme flexibility in terms of the number of clusters and the degree of differences among the clusters. The principal component analysis can reduce a significant amount of collinearity by combining multiple individual chemical biomarkers into latent constructs known as a principal component. However, due to the fact that the clustering procedure is extremely flexible and can change drastically from different sample populations, it is difficult to evaluate the effect estimates horizontally across different populations. Additionally, after clustering and principal component analysis, the interpretability of the model becomes more difficult for general audiences since different types of transformation have been performed on the variables in the model. If the researcher is able to identify the clusters in specific contexts, then the results generated might be more useful. For example, one cluster represents exposures of individual chemicals from an agricultural farm while another cluster represents exposures of individual chemicals near a mining site. Then the effect estimates can be attributed to these sources, giving exclusive meaning to these latent clusters or principal components.

2.4 Potential innovations and future directions

Through a careful review of the available statistical tools and a careful review of existing literature exploring the relationship between environmental chemical mixtures and

pregnancy outcomes, it is apparent that different methods have unique advantages and limitations. However, two common limitations are identified among multiple statistical methods. The first limitation is how to provide meaningful parameter estimation in the public health context, and the second limitation is addressing two specific forms of trade-offs during modelling procedures, particularly the bias-variance tradeoff and the flexibility-interpretability tradeoff. This section aims to address these limitations and propose potential innovations addressing the identified issues.

2.4.1. Meaningful parameter estimation in the public health context

The first commonly identified problem in methods is how to provide meaningful parameter estimation in the public health context. According to Greenland [8], one limitation of current modern methods in the field of epidemiology is that the methods do not adequately select confounders and interactions among exposure variables in a meaningful and interpretable manner. Currently, modern statistical methods often use data pre-processing techniques to transform exposure variables based on assumptions from specific statistical models. Greenland argued that meaningful centring and scaling of the exposure variables should be considered before the selection of specific modelling strategies because this is essential for generating parameter estimation that is meaningful and interpretable within the public health context [8].

It is also important to document specific confounders and interactions identified in the causal pathway of the exposure-response relationship. Particularly, clinical explanations of the relationship between exposure variables and the outcome variables are ideal for providing information on the selection of modelling strategies. For example, if we have clinical evidences on how PCB congeners induce inflammation biologically, this piece of information needs to be incorporated in our modelling procedures. This shifts the focus of passive prediction tasks towards outcome-based prediction tasks that are meaningful from a public health perspective rather than exclusively meaningful from a statistical perspective. The interpretation of the model selected using this strategy becomes more transparent and understandable. This indicates that the biological pathway of chemicals in human bodies should be carefully examined. Meaningful biological pathways generated from toxicology studies and biochemistry research can be extremely helpful for epidemiological assessment since it allows clearer explanations of the estimated effects measures.

2.4.2. Trade-off: bias vs variances and flexibility vs interpretability

The second commonly identified limitation in all methods concerns the concept of trade-off. Most statistical models illustrated in the previous sections face two forms of trade-off during the modelling process. The first one is the bias-variance trade-off and the second one is the flexibility-interpretability trade-off. The first form of trade-off can be quantified and researchers have developed specific mathematical techniques to evaluate the trade-off. The second form of trade-off is more tricky since there is no quantifiable definition of model interpretability. We will examine these two forms of trade-off more closely in the following paragraphs.

The bias-variance tradeoff as entailed by the name is the tradeoff between statistical bias and variance. Statistical bias is defined as the difference between the average predicted value of the parameter in the model and the true value of the parameter. Variance is defined as the variability of predicted values depending on the sampling data supplied for the modelling process. In the field of epidemiology, bias is usually associated with the term accuracy and variance is usually associated with the term precision and both of these two concepts are used to describe errors. Typically, in a given data set, the more complex the statistical model is, the more errors associated with variances there are. Therefore, to deal with the bias-variance tradeoff problem, the researcher wants to minimize the sum of errors for both the bias and variance. For example, in variable selection methods such as LASSO [59], some variables are removed based on the relative importance determined by the method. By doing this, we are purposefully introducing statistical bias into the modelling procedure since we assume the regression coefficients of the removed variables are zero. This reduces the variances associated with the regression coefficients of these variables simultaneously since they are constantly estimated as zero. This is a bias-variance tradeoff where we introduce statistical bias to reduce variance. On the other hand, when we use non-linear methods such as GAM [77], we are allowing the model to become more complex by allowing more transformation of the variables. This procedure then increases the error associated with variances but reduces the amount of bias by allowing more flexibility in the model.

It is important to note that the term bias described in this section is referring to the statistical definition of bias, not bias usually described in the context of epidemiology such as selection bias, recall bias, or confounding bias. Although the statistical term bias can be influenced by the epidemiological biases (e.g. confounding), it is necessary to point out that bias in the epidemiology context usually deals with problems involving sample representation and model generalization, but not problems involving bias-variance trade-off.

According to Fortmann-Roe [98], one common misconception is that researchers

should always favour approaches that minimize bias at the expense of variances because the accuracy of the model is usually valued higher than the precision of the model. This misconception is rooted in the assumption that by doing multiple studies and averaging results, it is possible to mitigate the error due to variances, but not error due to biases. Although statistically, this is true, in practice, each individual data set can only be analyzed once and the repetition of analysis is meaningless when confounding and interaction variables in each dataset are not defined universally. Therefore, always favouring approaches that minimize bias at the expense of variances is problematic. Kaufman [99] also discussed the phenomenon of epidemiologists being biased against statistical bias over variances that can be potentially resulted from confusion between statistical bias and epidemiological bias discussed earlier. Kaufman [99] argues that researchers should treat both bias and variances as sources of errors and both are equally bad.

The flexibility-interpretability trade-off is another interesting trade-off that is not discussed as often as the bias-variance trade-off [43]. It is particularly relevant in epidemiological analyses of biomarkers where the interpretation of results can be challenging. In any statistical model concerning human epidemiological investigations, the interpretation of the results is extremely important since it can be used to inform policy and practices. For example, PCB exposures are typically measured via a biomarker of concentration. The dose-response curve to a particular health outcome can be extremely complex. Nonetheless, policy-makers require results that are easily understood. Therefore, the interpretability of the model is an important aspect. Interestingly, through a review of the current methods, models that typically have good interpretability usually have less amount of flexibility. This is understandable since it is difficult to provide clear interpretation statements when the exposure-response relationship is more complex, especially when additional mathematical transformation including linear and non-linear transformation is performed during the data cleaning processes before analysis. As a result, potential innovations that increase interpretability while retaining model flexibility are merited.

2.4.3. Future directions

With two essential forms of trade-off in mind, future innovations in statistical methods should not only focus on improving the accuracy and precision associated with parameter estimations, but also improving model interpretability [43]. As discussed in the previous sections, Bayesian statistics is a branch of statistics that conceptualize probability as a degree of belief which allows modelling of both the parameter estimates and the uncertainty as probability distributions [54]. Through the incorporation of existing knowledge into the data

analysis, the precision of estimation can be improved due to specific forms of shrinkage on parameters. This is an important advantage over ensemble-based methods since it has the ability to perform well even when the sample size is relatively small. The latent variable model discussed in previous sections is used to measure latent variables that are associated with observable variables to estimate the overall effect [66]. This type of modelling procedure has unique benefits in model interpretability because it can be used to conceptualize an actual causal framework. Therefore, the use of latent variables can provide contextual details in analysis to improve model interpretations. In the context of environmental chemical mixtures analysis, simply doing latent variable analysis may suffer from poor estimates due to the nature of high correlations among different chemical variables [66]. However, combining Bayesian statistics and latent variable analysis is a promising way to mitigate the problem of high collinearity among chemicals in chemical mixtures while retaining good model interpretability.

The advantages of combined approaches using both Bayesian statistics and latent variable analysis can be further demonstrated in the analysis of environmental chemical mixtures. The first advantage is its ability to use latent variables to model complex chemical mixtures and return mixture specific regression coefficients as the estimation of the overall effect of the mixtures on a health outcome (e.g. birth weight). The second advantage is its ability to identify the individual chemical contribution to the overall mixtures to display a relative array of important chemicals within particular chemical mixtures. The third advantage is it retains the interpretability of the model without discarding any observation from the data. These advantages of this model make it a model that compromise between the advantages and disadvantages of variable selection methods and high dimensional non-linear estimation methods.

However, the novel Bayesian latent variable analysis also has some potential limitations. The limitation mainly lies in the assumptions made in the model which are the prior distribution of the parameters and the hypothetical causal diagram. If these assumptions are deviated from the actual scenario by a significant degree, the estimation generated can be unreliable. Therefore, sensitivity analysis with different causal diagram and prior assumptions need to be explored in to validate how reliable the assumptions are during the modelling process.

2.5. Conclusion

Environmental chemicals exposures in pregnant women may have an impact on

pregnancy outcomes through multiple biological mechanisms [1,2]. The epidemiological assessment of the health effects of environmental chemical mixtures is an important and difficult problem to solve [21]. The challenges of such assessment result mainly from multicollinearity present among different individual chemicals [5,15,28,71]

A comprehensive review of statistical methods was conducted on mixtures problems. These statistical methods deal with collinearity differently. Dimensional reduction methods [39,62,66] which consist of variable selection methods and latent variable modelling remove correlation by using a smaller number of predictor variables in the final models for health outcome (e.g. birth weight). The non-linear exposure-response method [55,70,77] allows the non-linear transformation of variables to remove correlation among individual variables. Ensemble methods utilize random sampling techniques to select a subset of variables to produce multiple weak learning models, then average across all weak learners to generate the final models. Finally, Bayesian-based methods [51,53,54] use specific priors for the purpose of variable shrinkage to reduce the total amount of variances in the model. All these methods have their own strengths and limitations, and innovative approaches that combine two or more of these methods may be promising in the assessment of pregnancy health impacts of environmental chemical mixtures.

Literature reviews of epidemiological studies employing modern statistical methods in exploring associations between environmental chemical mixtures and pregnancy outcomes were conducted. The strengths and limitations of each study are discussed. Common gaps were identified from these studies, particularly the difficulties in balancing model interpretability and model complexity as well as the difficulties in balancing bias and variances associated with parameter estimates in complex models. Potential innovation approaches can be developed by taking up strengths from multiple types of different statistical models to optimize the challenges depicted.

2.6. References

1. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet. Elsevier; 2015;385: 430–440.

2. Schieve LA, Tian LH, Rankin K, Kogan MD, Yeargin-Allsopp M, Visser S, et al. Population impact of preterm birth and low birth weight on developmental disabilities in US children. Ann Epidemiol. Elsevier; 2016;26: 267–274.

3. Windham G, Fenster L. Environmental contaminants and pregnancy outcomes. Fertil Steril. Elsevier; 2008;89: e111–6; discussion e117.

4. Wigle DT, Arbuckle TE, Turner MC, Bérubé A, Yang Q, Liu S, et al. Epidemiologic evidence of relationships between reproductive and child health outcomes and environmental chemical contaminants. J Toxicol Environ Health B Crit Rev. Taylor & Francis; 2008;11: 373–517.

5. Billionnet C, Sherrill D, Annesi-Maesano I, GERIE study. Estimating the health effects of exposure to multi-pollutant mixture. Ann Epidemiol. 2012;22: 126–141.

6. Sun Z, Tao Y, Li S, Ferguson KK, Meeker JD, Park SK, et al. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. Environ Health. 2013;12: 85.

7. Taylor KW, Joubert BR, Braun JM, Dilworth C, Gennings C, Hauser R, et al. Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology: Lessons from an Innovative Workshop. Environ Health Perspect. 2016;124: A227–A229.

8. Greenland S, Pearce N. Statistical foundations for model-based adjustments. Annu Rev Public Health. 2015;36: 89–108.

9. Tanne JH. Preterm and low weight births rise again in the US. BMJ. 2017;358: j3311.

10. Evens A, Hryhorczuk D, Lanphear BP, Rankin KM, Lewis DA, Forst L, et al. The impact of low-level lead toxicity on school performance among children in the Chicago Public Schools: a population-based retrospective cohort study. Environ Health. 2015;14: 21.

11. Kingsley SL, Eliot MN, Glazer K, Awad YA, Schwartz JD, Savitz DA, et al. Maternal ambient air pollution, preterm birth and markers of fetal growth in Rhode Island: results of a hospital-based linkage study. J Epidemiol Community Health. 2017;71: 1131–1136.

12. Bergman A, Heindel JJ, Kasten T, Kidd KA, Jobling S, Neira M, et al. The impact of endocrine disruption: a consensus statement on the state of the science. Environ Health Perspect. 2013;121: A104–6.

13. Carlin DJ, Rider CV, Woychik R, Birnbaum LS. Unraveling the Health Effects of Environmental Mixtures: An NIEHS Priority. Environ Health Perspect. 2013;121: a6–a8.

14. Burns JS, Williams PL, Sergeyev O, Korrick S, Lee MM, Revich B, et al. Serum dioxins and polychlorinated biphenyls are associated with growth among Russian boys. Pediatrics. 2011;127: e59–68.

15. Braun JM, Gray K. Challenges to studying the health effects of early life environmental chemical exposures on children's health. PLoS Biol. 2017;15: e2002800.

16. Sexton K. Cumulative risk assessment: an overview of methodological approaches for evaluating combined health effects from exposure to multiple environmental stressors. Int J Environ Res Public Health. 2012;9: 370–390.

17. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. Springer Science & Business Media; 2011.

18. Ferguson KK, Chin HB. Environmental chemicals and preterm birth: Biological mechanisms and the state of the science. Curr Epidemiol Rep. 2017;4: 56–71.

19. Spurgeon DJ, Jones OAH, Dorne J-LCM, Svendsen C, Swain S, Stürzenbaum SR. Systems

toxicology approaches for understanding the joint effects of environmental chemical mixtures. Sci Total Environ. 2010;408: 3725–3734.

20. McKinney JD, Waller CL. Polychlorinated biphenyls as hormonally active structural analogues. Environ Health Perspect. ehp.niehs.nih.gov; 1994;102: 290–297.

21. Cassee FR, Groten JP, van Bladeren PJ, Feron VJ. Toxicological evaluation and risk assessment of chemical mixtures. Crit Rev Toxicol. 1998;28: 73–101.

22. Keil AP, Daza EJ, Engel SM, Buckley JP, Edwards JK. A Bayesian approach to the g-formula. Stat Methods Med Res. 2018;27: 3183–3204.

23. Petit C, Blangiardo M, Richardson S, Coquet F, Chevrier C, Cordier S. Association of environmental insecticide exposure and fetal growth with a Bayesian model including multiple exposure sources: the PELAGIE mother-child cohort. Am J Epidemiol. 2012;175: 1182–1190.

24. Vadillo-Ortega F, Osornio-Vargas A, Buxton MA, Sánchez BN, Rojas-Bracho L, Viveros-Alcaráz M, et al. Air pollution, inflammation and preterm birth: a potential mechanistic link. Med Hypotheses. 2014;82: 219–224.

25. Longini M, Perrone S, Vezzosi P, Marzocchi B, Kenanidis A, Centini G, et al. Association between oxidative stress in pregnancy and preterm premature rupture of membranes. Clin Biochem. 2007;40: 793–797.

26. Savitz DA, Wellenius GA. Invited Commentary: Exposure Biomarkers Indicate More Than Just Exposure. Am J Epidemiol. 2018;187: 803–805.

27. Savitz DA. Invited commentary: interpreting associations between exposure biomarkers and pregnancy outcome. Am J Epidemiol. 2014;179: 545–547.

28. Hu JMY, Zhuang LH, Bernardo BA, McCandless LC. Statistical Challenges in the Analysis of Biomarkers of Environmental Chemical Exposures for Perinatal Epidemiology. Current Epidemiology Reports. 2018;5: 284–292.

29. Marateb HR, Mansourian M, Adibi P, Farina D. Manipulating measurement scales in medical statistical analysis and data mining: A review of methodologies. J Res Med Sci. 2014;19: 47–56.

30. Solomon JD, Vallero D, Benson K. Evaluating risk: A revisit of the scales, measurement theory, and statistical analysis controversy. 2017 Annual Reliability and Maintainability Symposium (RAMS). 2017. pp. 1–6.

31. Youn E, Jeong MK. Class dependent feature scaling method using naive Bayes classifier for text datamining. Pattern Recognit Lett. 2009;30: 477–485.

32. Altman DG, Bland JM. Quartiles, Quintiles, Centiles, And Other Quantiles. BMJ: British Medical Journal. BMJ; 1994;309: 996–996.

33. Braun JM, Gennings C, Hauser R, Webster TF. What Can Epidemiological Studies Tell Us about the Impact of Chemical Mixtures on Human Health? Environ Health Perspect. 2016;124: A6–9.

34. Bien J, Taylor J, Tibshirani R. A LASSO FOR HIERARCHICAL INTERACTIONS. Ann Stat. 2013;41: 1111–1141.

35. Agier L, Portengen L, Chadeau-Hyam M, Basagaña X, Giorgis-Allemand L, Siroux V, et al. A Systematic Comparison of Linear Regression-Based Statistical Methods to Assess Exposome-Health Associations. Environ Health Perspect. 2016;124: 1848–1856.

36. Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics. 2015;16: 493–508.

37. Kaplan D, Lee C. Bayesian Model Averaging Over Directed Acyclic Graphs With Implications for the Predictive Performance of Structural Equation Models. Struct Equ Modeling. Routledge; 2016;23: 343–353.

38. Nikolov MC, Coull BA, Catalano PJ, Godleski JJ. An informative Bayesian structural equation model to assess source-specific health effects of air pollution. Biostatistics. 2007;8: 609–624.

39. Marsh HW, Morin AJS, Parker PD, Kaur G. Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. Annu Rev Clin Psychol. 2014;10: 85–110.

40. Conti G, Frühwirth-Schnatter S, Heckman JJ, Piatek R. Bayesian Exploratory Factor Analysis. J Econom. 2014;183: 31–57.

41. Gass K, Klein M, Chang HH, Flanders WD, Strickland MJ. Classification and regression trees for epidemiologic research: an air pollution example. Environ Health. 2014;13: 17.

42. Tomarken AJ, Waller NG. Structural equation modeling: strengths, limitations, and misconceptions. Annu Rev Clin Psychol. 2005;1: 31–65.

43. Shmueli G. To Explain or to Predict? Stat Sci. Institute of Mathematical Statistics; 2010;25: 289–310.

44. Krieger N, Davey Smith G. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. Int J Epidemiol. 2016;45: 1787–1808.

45. Suttorp MM, Siegerink B, Jager KJ, Zoccali C, Dekker FW. Graphical presentation of confounding in directed acyclic graphs. Nephrol Dial Transplant. 2015;30: 1418–1423.

46. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Med Res Methodol. 2008;8: 70.

47. Dawid AP. Beware of the DAG! In: Guyon I, Janzing D, Schölkopf B, editors. Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008. Whistler, Canada: PMLR; 2010. pp. 59–86.

48. Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. Annu Rev Public Health. 2018;39: 95–112.

49. Hernán MA. Does water kill? A call for less casual causal inferences. Ann Epidemiol. 2016;26: 674–680.

50. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. Int J Epidemiol. 2016;45: 1776–1786.

51. MacLehose RF, Dunson DB, Herring AH, Hoppin JA. Bayesian methods for highly correlated

exposure data. Epidemiology. 2007;18: 199–207.

52. Dunson DB. Commentary: practical advantages of Bayesian analysis of epidemiologic data. Am J Epidemiol. 2001;153: 1222–1226.

53. Eddy SR. What is Bayesian statistics? Nat Biotechnol. Nature Publishing Group; 2004;22: 1177.

54. Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics. Br J Math Stat Psychol. Wiley Online Library; 2013;66: 8–38.

55. Trevor H, Robert T, Jh F. The elements of statistical learning: data mining, inference, and prediction [Internet]. New York, NY: Springer; 2009. Available: http://thuvien.thanglong.edu.vn:8081/dspace/handle/DHTL_123456789/4053

56. Bender R. Introduction to the use of regression models in epidemiology. Methods Mol Biol. 2009;471: 179–195.

57. Myers RH, Myers RH. Classical and modern regression with applications. Duxbury press Belmont, CA; 1990.

58. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat. Institute of Mathematical Statistics; 2004;32: 407–499.

59. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Series B Stat Methodol. [Royal Statistical Society, Wiley]; 1996;58: 267–288.

60. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series. Wiley Online Library; 2005; Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00503.x%4010.1111/%28ISSN%291467-9868.TOP_SERIES_B_RESEARCH

61. Zou H. The Adaptive Lasso and Its Oracle Properties. J Am Stat Assoc. Taylor & Francis; 2006;101: 1418–1429.

62. Gallagher MW, Brown TA. Introduction to Confirmatory Factor Analysis and Structural Equation Modeling. In: Teo T, editor. Handbook of Quantitative Methods for Educational Research. Rotterdam: SensePublishers; 2013. pp. 289–314.

63. Hubbard AE. Causal inference and the curse of dimensionality. 41st Annual Meeting of the Society-for-Epidemiologic-Research. 2008. pp. S45–S45.

64. Joliffe IT, Morgan BJ. Principal component analysis and exploratory factor analysis. Stat Methods Med Res. 1992;1: 69–95.

65. Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJS, et al. Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. Struct Equ Modeling. Routledge; 2009;16: 439–476.

66. Sánchez BN, Budtz-Jørgensen E, Ryan LM, Hu H. Structural Equation Models: A Review with Applications to Environmental Epidemiology. J Am Stat Assoc. [American Statistical Association, Taylor & Francis, Ltd.]; 2005;100: 1443–1455.

67. Bryant FB, Yarnold PR. Principal-components analysis and exploratory and confirmatory factor analysis. American Psychological Association; 1995; Available:

https://psycnet.apa.org/record/1995-97110-004

68. Zou H, Hastie T, Tibshirani R. Sparse Principal Component Analysis. J Comput Graph Stat. [American Statistical Association, Taylor & Francis, Ltd., Institute of Mathematical Statistics, Interface Foundation of America]; 2006;15: 265–286.

69. Buja A, Hastie T, Tibshirani R. Linear Smoothers and Additive Models. Ann Stat. Institute of Mathematical Statistics; 1989;17: 453–510.

70. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. Springer, New York, NY; 2013.

71. Lazarevic N, Barnett AG, Sly PD, Knibbs LD. Statistical Methodology in Studies of Prenatal Exposure to Mixtures of Endocrine-Disrupting Chemicals: A Review of Existing Approaches and New Alternatives. Environ Health Perspect. ehp.niehs.nih.gov; 2019;127: 26001.

72. Zanobetti A, Wand MP, Schwartz J, Ryan LM. Generalized additive distributed lag models: quantifying mortality displacement. Biostatistics. academic.oup.com; 2000;1: 279–292.

73. Ramsay TO, Burnett RT, Krewski D. The effect of concurvity in generalized additive models linking mortality to ambient particulate matter. Epidemiology. journals.lww.com; 2003;14: 18–23.

74. Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol. academic.oup.com; 2002;156: 193–203.

75. Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. Characterization of Weighted Quantile Sum Regression for Highly Correlated Data in a Risk Analysis Setting. JABES. Springer US; 2015;20: 100–120.

76. Koenker R, Hallock K. Quantile regression: An introduction. J Econ Perspect. econ.uiuc.edu; 2001;15: 43–56.

77. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Springer series in statistics New York; 2001.

78. Stel VS, Pluijm SMF, Deeg DJH, Smit JH, Bouter LM, Lips P. A classification tree for predicting recurrent falling in community-dwelling older persons. J Am Geriatr Soc. Wiley Online Library; 2003;51: 1356–1364.

79. Mueser KT, Yarnold PR, Rosenberg SD, Swett C Jr, Miles KM, Hill D. Substance use disorder in hospitalized severely mentally ill psychiatric patients: prevalence, correlates, and subgroups. Schizophr Bull. academic.oup.com; 2000;26: 179–192.

80. Nishida N, Tanaka M, Hayashi N, Nagata H, Takeshita T, Nakayama K, et al. Determination of smoking and obesity as periodontitis risks using the classification and regression tree method. J Periodontol. Wiley Online Library; 2005;76: 923–928.

81. York TP, Eaves LJ, van den Oord EJCG. Multivariate adaptive regression splines: a powerful method for detecting disease-risk relationship differences among subgroups. Stat Med. 2006;25: 1355–1367.

82. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat. Institute of Mathematical Statistics; 2001;29: 1189–1232.

83. Rietbergen C, Debray TPA, Klugkist I, Janssen KJM, Moons KGM. Reporting of Bayesian analysis in epidemiologic research should become more transparent. J Clin Epidemiol. Elsevier; 2017;86: 51–58.e2.

84. Papathomas M, Molitor J, Richardson S, Riboli E, Vineis P. Examining the joint effect of multiple risk factors using exposure risk profiles: lung cancer in nonsmokers. Environ Health Perspect. 2011;119: 84–91.

85. Valeri L, Mazumdar MM, Bobb JF, Claus Henn B, Rodrigues E, Sharif OIA, et al. The Joint Effect of Prenatal Exposure to Metal Mixtures on Neurodevelopmental Outcomes at 20-40 Months of Age: Evidence from Rural Bangladesh. Environ Health Perspect. 2017;125: 067015.

86. Bobb JF, Claus Henn B, Valeri L, Coull BA. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. Environ Health. ehjournal.biomedcentral.com; 2018;17: 67.

87. Buckley JP, Hamra GB, Braun JM. Statistical Approaches for Investigating Periods of Susceptibility in Children's Environmental Health Research. Curr Environ Health Rep. Springer; 2019;6: 1–7.

88. Harley KG, Berger K, Rauch S, Kogut K, Claus Henn B, Calafat AM, et al. Association of prenatal urinary phthalate metabolite concentrations and childhood BMI and obesity. Pediatr Res. nature.com; 2017;82: 405–415.

89. Woods MM, Lanphear BP, Braun JM, McCandless LC. Gestational exposure to endocrine disrupting chemicals in relation to infant birth weight: a Bayesian analysis of the HOME Study. Environ Health. 2017;16: 115.

90. Braun JM, Kalkbrenner AE, Just AC, Yolton K, Calafat AM, Sjödin A, et al. Gestational exposure to endocrine-disrupting chemicals and reciprocal social, repetitive, and stereotypic behaviors in 4- and 5-year-old children: the HOME study. Environ Health Perspect. ehp.niehs.nih.gov; 2014;122: 513–520.

91. Hamra GB, Lyall K, Windham GC, Calafat AM, Sjödin A, Volk H, et al. Prenatal Exposure to Endocrine-disrupting Chemicals in Relation to Autism Spectrum Disorder and Intellectual Disability. Epidemiology. journals.lww.com; 2019;30: 418–426.

92. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. Environ Health Perspect. 1994;102 Suppl 8: 33–39.

93. Chiu Y-H, Bellavia A, James-Todd T, Correia KF, Valeri L, Messerlian C, et al. Evaluating effects of prenatal exposure to phthalate mixtures on birth weight: A comparison of three statistical approaches. Environ Int. 2018;113: 231–239.

94. Lenters V, Portengen L, Rignell-Hydbom A, Jönsson BAG, Lindh CH, Piersma AH, et al. Prenatal Phthalate, Perfluoroalkyl Acid, and Organochlorine Exposures and Term Birth Weight in Three Birth Cohorts: Multi-Pollutant Models Based on Elastic Net Regression. Environ Health Perspect. 2016;124: 365–372.

95. Govarts E, Remy S, Bruckers L, Den Hond E, Sioen I, Nelen V, et al. Combined Effects of Prenatal Exposures to Environmental Chemicals on Birth Weight. Int J Environ Res Public Health. 2016;13. doi:10.3390/ijerph13050495

96. Chen Y-H, Ferguson KK, Meeker JD, McElrath TF, Mukherjee B. Statistical methods for modeling repeated measures of maternal environmental exposure biomarkers during pregnancy in association with preterm birth. Environ Health. 2015;14: 9.

97. Kalloo G, Wellenius GA, McCandless L, Calafat A, Sjodin A, Karagas MR, et al. Chemical Mixture Exposures during Pregnancy and Birth Outcomes. ISEE Conference Abstracts. ehp.niehs.nih.gov; 2018. Available: https://ehp.niehs.nih.gov/doi/abs/10.1289/isesisee.2018.O01.04.22

98. Fortmann-Roe S. Understanding the bias-variance tradeoff. citeulike.org; 2012. p. Essays.

99. Kaufman JS. Commentary: Why are we biased against bias? Int J Epidemiol. academic.oup.com; 2008;37: 624–626.

Chapter 3. Effects of gestational exposures to chemical mixtures on birth weight using Bayesian Factor Analysis in the Health Outcome and Measures of Environment (HOME) Study

Published: Zhuang LH, Chen A, Braun JM, Lanphear BP, Hu JMY, Yolton K, McCandless LC. Effects of gestational exposures to chemical mixtures on birth weight using Bayesian factor analysis in the Health Outcome and Measures of Environment (HOME) Study. Environ Epidemiology. 2021 Jun 8;5(3):e159. doi: 10.1097/EE9.0000000000000159. PMID: 34131620; PMCID: PMC8196215.

3.1. Abstract

Background*:* Studying the effects of gestational exposures to chemical mixtures on infant birth weight is inconclusive due to several challenges. One of the challenges is which statistical methods to rely on. Bayesian Factor Analysis (BFA), which has not been utilized for chemical mixtures, has advantages in variance reduction and model interpretation.

Methods*:* We analyzed data from a cohort of 384 pregnant women and their newborns using urinary biomarkers of phthalates, phenols, and organophosphate pesticides (OPs) and serum biomarkers of polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs), perfluoroalkyl substances (PFAS) and organochlorine pesticides (OCPs). We examined the association between exposure to chemical mixtures and birth weight using BFA and compared with multiple linear regression (MLR), and Bayesian kernel regression models (BKMR).

Results*:* For BFA, a 10-fold increase in the concentrations of PCB and PFAS mixtures was associated with an 81g (95%CI -132g, -31g) and 57 g (95%CI -105g, -10g) reduction in birth weight, respectively. BKMR results confirmed the direction of effect. However, the 95% credible intervals all contained the null. For single-pollutant MLR, a 10-fold increases in the concentrations of multiple chemicals were associated with reduced birth weight, yet the 95% CI all contained the null. Variance inflation from MLR was apparent for models that adjusted for co-pollutants, resulting in less precise confidence intervals.

Conclusion*:* We demonstrated the merits of BFA on mixture analysis in terms of precision and interpretation compared to MLR and BKMR. We also identified the association between exposure to PCBs and PFAS and lower birth weight.

3.2. Introduction

Exposure to chemicals mixtures during pregnancy has been associated with perinatal complications and adverse fetal development, such as preterm birth and low birth weight.[1-14] Most epidemiological studies, however, are informed by single-pollutant statistical models, particularly linear and logistic regression models, that do not capture the complex exposure profiles in real-life scenarios among pregnant mothers.[2,15] As researchers move beyond the "one chemical at a time" analysis to evaluate mixture effects,[2,16-18] several challenges related to collinearity among individual chemicals and providing easily interpretable analysis results have arisen.[16,19,20]

To combat the challenge of collinearity, several frequentist approaches such as least absolute shrinkage and selection operator and principal component analysis were developed to reduce collinearity by discarding correlated variables that are less impactful.[16,19] In the context of environmental epidemiology, it is difficult to justify the discarding of chemical variables because variables within a class of chemical mixture often share similar biological pathways. Bayesian methods[21], on the other hand, are gaining attention in the field of environmental epidemiology as an approach to address the challenges without discarding variables. They provide a more explicit quantification of uncertainty than conventional measures, such as p-values by modelling parameters as probability distributions.[22] Moreover, Bayesian methods have the ability to improve the precision of parameter estimates in the presence of collinearity among variables in mixtures compared to traditional methods.[10,15,23,24] This is typically achieved by combining Bayesian techniques with regularization, shrinkage and prior information about model parameters. However, Bayesian procedures tend to increase the computation time and complexity of the analysis.

Factor analysis modelling, also known as latent variable modelling, is widely used in the field of psychology to manage collinearity for characteristics that are difficult to directly measure[25-27] and can be extended to a Bayesian framework. Using latent constructs to linearly quantify the combined effects of an unmeasured variable, like a chemical mixture, is an appealing way to address collinearity challenge while providing interpretable estimates. However, apart from a publication by Ferrari and Dunson, Bayesian factor analysis has not been used in this context[57]. Accordingly, we aimed to illustrate the potential benefits of Bayesian factor analysis (BFA) to estimate the association between chemical mixtures and birth weight using data from the Health Outcomes and Measure of the Environment (HOME) Study, a birth cohort from Cincinnati, Ohio established to study the health impact of various chemical and their mixtures.[28] We also compared our BFA results with two established

methods, multiple linear regression (MLR) and Bayesian Kernel Machine Regression (BKMR) to assess collinearity reductions and interpretability.

3.3. Methods

3.3.1. Health Outcomes and Measures of the Environment (HOME) study

The HOME Study is a prospective birth cohort of pregnant mothers and their infants established in 2003 at the Cincinnati Children's Environmental Health Center, Ohio.[28] The primary goal of the HOME Study is to examine the impact of environmental toxicants on child health. Pregnant mothers who were >18 years old and at 16±3 weeks of gestation and living in a residence built before 1978 were recruited from seven prenatal clinics and hospitals.[28] Out of the 468 women initially enrolled in the study, we excluded 67 women who dropped out before delivery, three stillbirths, nine sets of twins, and five participants missing covariate data. Therefore, 384 mothers who delivered singleton live births, provided biological samples, and had complete sociodemographic information were included in our analysis.

3.3.2. Biomarkers of environmental chemical mixtures

We collected blood and urine samples from participants at approximately 16- and 26-weeks gestation.[28] The Centers for Disease Control and Prevention Environmental Health Laboratories used gas and liquid chromatography-mass spectrometry to measure the concentrations of environmental chemical biomarkers in serum and urine samples as previously described.[28]

With the specific goal of estimating the effect of exposure to environmental chemical mixtures on infant birth weight, we consulted existing literature on birth outcomes to identify potential environmental chemical mixtures to investigate.[10,13,14,29-31] A total of seven classes of chemical mixtures were identified: polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs), phthalates, organochlorine pesticides (OCPs), organophosphate pesticides (OPs), phenols, and Perfluoroalkyl and Polyfluoroalkyl Substances (PFAS). PCBs, PBDEs, and OCPs are lipophilic and were lipid standardized. Phthalate metabolites, phenols, and OPs were creatinine standardized to account for urine dilution. In addition, to preserve the sample size of our analysis, we further restricted our analysis to biomarkers that are widely detected in the population (>80% detected above the limit of detection). Furthermore, to keep the modelling approaches consistent between all chemical classes and mitigate issues with the excessive dimensionality[32] in regression analysis, we selected a total of 35 biomarkers to be

included in our final analysis (Table 3.1). For PCBs, PBDEs, OCPs and PFAS, we used samples measured at 16 weeks to maintain consistency across measures. For phthalates, OPs, and phenol biomarkers, we averaged concentrations in samples collected at 16 weeks and 26 weeks to represent the overall concentrations. For all biomarkers, measurements below the limit of detection were replaced using single imputation according to Lubin et al.[33] The imputed values were sampled from a truncated lognormal distribution with the mean and standard deviation of the concentration of the chemical variables. The detection limit of each specific chemical was set as the upper bound value for imputation. The concentrations of these biomarkers were log10 transformed to reduce the effects of right skewness in the distribution and to assist with the interpretation of the results. The regression coefficients are interpreted as the change in birth weight for every ten-fold increase in the chemical concentrations.

3.3.3. Outcome variable

Infant birth weight, measured in grams (g), was abstracted from the medical records and examined as a continuous variable. To examine fetal growth, we adjusted for gestational age, measured in weeks.14 An alternate way would be to use gestational age-specific birth weight z-scores. However, its interpretation is not straightforward as it reports effect measures in the unit of standard deviation, which results in different absolute amounts of weight across the gestational age spectrum.34

3.3.4. Covariates

A direct acyclic graph was drawn to select confounders based on the relationship among potential covariates, the selected five classes of environmental chemical mixtures and birth weight (Supplementary Figure 1). Exposures to lead and tobacco smoke have been documented to have effects on infant birth weight.[29,35] Therefore, we included the biomarker measurements of lead and cotinine as covariates. Additional covariates in the statistical models included maternal age at delivery, infant sex, race, marital status, maternal education, maternal BMI and annual household income. We excluded maternal BMI from the covariates in the analysis of lipophilic chemicals to avoid duplicate adjustment since the concentrations were already lipid-adjusted, which is directly related to BMI.[36] The effect of gestational duration on birth weight has been documented to be nonlinear,[14] therefore, we used the cubic spline approach for the adjustment of gestational age using the “splines” package in R.[37]

3.3.5. Analytic approach

The primary analytic approach for this study was BFA to estimate the mixture effect of environmental chemicals on birth weight. We also compared our results with BKMR, and additionally, two bridging methods that can be viewed as the intermediate step bridging MLR and BFA. Individual pollutant MLR model was used as a sensitivity analysis because it is the most commonly used method in the literature of environmental epidemiology for continuous outcomes. It also serves as a benchmark for comparison with the more advanced methods.[38] BKMR is a combinatorial method of Bayesian approach and non-linear approximation methods that is gaining attention in the field of environmental epidemiology.[23,39,40]

3.3.6.Approach 1 - Bayesian factor analysis

We used BFA to assess the association between each class of mixture and birth weight while adjusting for covariates. A total of seven BFA models for the seven specific chemical classes were examined. We did not include a BFA model with all chemical classes simultaneously because dimensionality increases dramatically when more variables are included in the model and result in low statistical power with small sample size. We employed the confirmatory factor analysis approach for the purpose of generating regression coefficients that can be interpreted with respect to a specific class of chemical mixture.[41,42] Bayesian techniques were used for regularization and easier interpretation of the parameter estimates. Each class of mixture is represented by a latent variable illustrated by the following set of equations:

$$\mathrm{Y} = \beta_0 + \beta_z \mathrm{Z} + \beta_A \mathrm{A} + \beta_B \mathrm{B} + \epsilon_y \text{ [Equation 1]}$$

$$\mathrm{X_i} = \beta_i + \gamma_i \mathrm{Z} + \epsilon_{xi} \text{ [Equation 2]}$$

for i=1, ..., k where k is the number of chemicals in the mixture, and where β_0, β_i are the Y-intercepts, β_z is the regression coefficient for the latent variable for each mixture class of chemicals, Z is the latent variable representing each mixture class of chemicals, β_c is the vector of regression coefficients of covariates, C is the vector of confounders such as age and household income, $\mathrm{X_i}$ is the ith individual chemical within the class of mixture Z. γ_i is the factor loading score of the ith chemical on mixture Z and ϵ_y and ϵ_{xi} are the normally distributed random errors. For parameters β_z and β_c, we assigned uninformative normal priors with variance equal to 1,000. For y-intercepts β_0, β_i, and random errors ϵ_y ϵ_{xi}, we used the default priors in R package "blavaan".[43] The variance of Z was set to be 1.0 for identifiability.[43]

Markov Chain Monte Carlo (MCMC) sampling was accomplished with the R package

"blavaan"[43] and "rstan"[44] to generate samples from a posterior to estimate parameters of the interest. For each BFA model, the number of iterations were determined experimentally to achieve convergence assessed by the measure of the potential scale reduction factor. As a result, a total of 40,000 iterations were run for samples with 2,000 burn-in iterations.

3.3.7. Approach 2 – Bridging methods between BFA and MLR

To examine the mathematical relationship between MLR and BFA, we included two additional bridging methods to obtain regression estimates that can be viewed as the intermediate steps bridging MLR and BFA. As previously shown in equations 1 and 2, BFA can be conceptually broken down into three hierarchical steps. The first step estimates the latent variable Z for each study participant, which is denoted by the symbol $\hat{\mathrm{Z}}$. The second step computes the parameter estimates $\beta_{\hat{z}}$ for the effect of the latent variable. The third step applies Bayesian prior distributions on the parameter estimates. Therefore, the first bridging method was factor analysis (FA) using the R package[43] "lavaan". The second bridging method was "MLR with extracted factor score", which is MLR incorporating the estimated latent variable Z obtained from the FA model in the absence of individual chemicals, while adjusted for covariates as the following equation:

$$\mathrm{Y} = \beta_0 + \beta_{\hat{z}}\hat{\mathrm{Z}} + \beta_{\mathrm{A}}\mathrm{A} + \beta_{\mathrm{B}}\mathrm{B} \text{ [Equation 3]}$$

where β_0 is the Y-intercept, $\hat{\mathrm{Z}}$ is the estimated factor score, $\beta_{\hat{z}}$ is the regression coefficient for the estimated factor score, β_{c} is the vector of regression coefficients of covariates, C is the vector of confounders and ϵ_{y} is the random error. By comparing the different regression estimates from the sensitivity analyses, we can observe the gradual differences in the precision of estimates.

3.3.8. Approach 3 - Bayesian kernel machine regression

The BKMR model can be described using the following equation:

$$\mathrm{Y}_{\mathrm{bkmr}} = \beta_{\mathrm{bkmr}} + \mathrm{h}(\mathrm{Z}) + \beta_{\mathrm{A}}\mathrm{A} + \beta_{\mathrm{B}}\mathrm{B} + \epsilon_{\mathrm{bkmr}} \text{ [Equation 4]}$$

where β_{bkmr} is the Y-intercept, $\mathrm{h}(\mathrm{X})$ is the vector of exposure-response functions for each individual chemical within the specific class of mixture, β_{c} is the vector of regression coefficients of covariates, C is the vector of confounders, and ϵ_{bkmr} is the random error in the model. The exposure-response functions for each individual chemical were determined by

employing Gaussian kernel functions non-parametrically based on the available data structure.[23] Since the exposure-response functions were determined based on the data, the priors for the parameters of each individual chemical were also specified differently according to the exposure-response functions with details explained by Bobb et al.[45] Similar to BFA, Markov Chain Monte Carlo (MCMC) sampling was also used in BKMR to generate samples from posterior distributions for the estimation of the parameters as well as the dose-response curves illustrated by the cross-section views of the exposure-surface functions. The posterior samples were sampled from a total of 20,000 iterations, which is determined experimentally to achieve convergence assessed by the measure of the potential scale reduction factor. The variable selection feature of "bkmr" R package[45] was not used in our primary analysis since we intended to retain all chemicals in the class in the BKMR model to compare with other methods. A separate analysis of BKMR using the variable selection feature was also used to assess the impact of such feature on the analysis results.

3.3.9. Sensitivity analyses – MLR

We used MLR as our sensitivity analyses to assess the association between each individual chemical and birth weight while adjusted for covariates. We also used MLR to assess the association between each individual chemical and birth weight while adjusted for both covariates and co-pollutants within that class of chemicals.

3.4. Results

3.4.1. Descriptive Statistics

The study participants consisted of 384 mother-singleton newborn pairs. Due to various degrees of the missingness that could not be imputed (missingness due to incomplete biospecimen collection or insufficient volumes for chemical assays instead of measurements below limit of detection), the sample sizes for our analysis were 360 for OPs mixture, 366 for phthalates mixture, 284 for PBDEs mixtures, 237 for OCPs mixtures, 310 for PCBs mixtures, 296 for phenols mixtures and 307 for PFAS mixtures. Mothers who participated in the study were mostly white (62.5%), married (65.6%) and had at least a bachelor's degree (60.1%). The mean infant birth weight was 3,352 grams with a standard deviation of 632 grams. The infant sex ratio was roughly 1.18 to 1 (54.2% female to 45.8% male). Sociodemographic characteristics that were associated with birth weight included maternal age, household income and maternal BMI (Table 3.2). Infant birth weight tended to decrease with increasing maternal age and increased with increasing household income and maternal BMI

(Supplementary Table 3.1).

A high degree of correlation was detected among chemicals within the same class (Figure 3.1). For example, all PCB congeners displayed correlation coefficients in the range of 0.51 (PCB 118 & PCB 180) to 0.99 (PCB 170 & PCB 180) with each other.

3.4.2. BFA analysis results

We ran seven BFA models for the seven different classes of chemical mixtures. The class-specific regression coefficients of each class of the mixture and loading of the individual congeners on the mixture were evaluated by BFA. PCBs and PFAS displayed associations with birth weight reduction and every ten-fold increase in the concentration of the mixture (Figure 3.2). Specifically, the regression coefficients were -81g (95%CI: -132g, -31g) for PCBs and -57g (95%CI: -105g, -10g) for PFAS.

In addition to the mixture-specific effect estimates, we also estimated the loading coefficient of individual chemicals within the class of chemical mixtures, denoted by the quantity from Equation 2 (Figure 3.2). This provides a relative measure of importance for these individual chemicals because it measures how much influence each individual chemical variable contributes to the overall latent mixture variable.42 For the PCB mixture, we observed that PCB 170 and PCB 180 had a stronger impact on the overall latent mixture than PCB 153, PCB 118 and PCB 138. For the PFAS mixture, we observed that PFHXS and PFOS had a stronger impact on the overall latent variable compared to PFNA and PFOA. The BFA results of the other chemicals are represented in Supplementary Figure 2.

We also observed a slight decrease in birth weight with every ten-fold increase in the concentration of OCPs mixture at -16g (95%CI: -66g, 34g) and phenols mixture at -21g (95%CI: -71g, 28g). A slight increase in birth weight was associated with every ten-fold increase in the concentration of PBDEs mixture at 25g (-18g, 69g), phthalates mixture at 49g (95%CI: -3g, 99g), and OPs mixture at 8g (95%CI: -40g, 54g). However, all the credible intervals of OCPs, phenols, PBDEs, phthalates and OPs were imprecise and contained the null value of zero.

3.4.3. Bridging methods results comparing BFA with MLR

We provide detailed comparisons of regression estimates for PCBs and PFAS and their 95% confidence interval across different methods (Figure 3.3). We observed that when

collinearity between the individual chemicals is present, co-pollutant adjustments in MLR results in less reliable parameter estimates and poorer precision for both PCBs and PFAS. Small sample size also played an important role. For PCB 153, for example, we observed that for single pollutant MLR, the precision interval of the estimators was -99g (95%CI: -143g,147g). After co-pollutant adjustments, the precision interval of the estimators inflated to -418g (95%CI: -1645g, 808g). When using the bridging methods for PCBs, the precision interval obtained from MLR with extracted factor score was -24g (95%CI: -89g, 42g), which represents a range of 131g for 95%CI. The precision interval obtained from FA was -43g (95%CI: -110g, 10g), which represents an absolute range of 120g for 95%CI. And finally, the precision interval obtained from BFA was -81g (95%CI: -132g, -31g), which represents an absolute range of 101g. For PFAS, the precision interval obtained from MLR with extracted factor score was -58g (95%CI: -117g, -13g), which represents a range of 104g for 95%CI. The precision interval obtained from FA was -57g (95%CI: -115g, -16g), which represents an absolute range of 99g for 95%CI. And finally, the precision interval obtained from BFA was -57g (95%CI: -105g, -10g), which represents an absolute range of 95g.

Therefore, Figure 3.3 illustrates that the precision of the regression estimates increased when more conceptual steps of BFA were performed. This demonstrates BFA can provide a more precise measure of mixture effects when multiple correlated co-pollutants are in the model.

3.4.4. BKMR analysis results

The dose-response function between individual PCBs and PFAS chemicals and the change in birth weight using BKMR are shown (Figure 3.4), as is the overall association between the chemical mixtures and birth weight (Figure 3.5). Exposure to PCB congeners and PFAS congeners both displayed inverse associations with birth weight. It is apparent that as the concentration quantile of PCB congeners and PFAS congeners increases, then the mean estimate of birth weight decreases. The confidence intervals of the estimates at the extreme ends were the widest due to the smaller sample sizes. The BKMR results for the rest of the chemicals are in Supplementary Figures 3.3-4. All regression estimates obtained from BKMR had 95% interval estimates that were imprecise and contained the null. A separate BKMR analysis using the variable selection feature was also conducted to examine the influence of different variables in the model (Supplementary Table 3.2). According to the results, all individual variables from the seven different classes of chemicals were included in each BKMR model, giving the same results as the BKMR when variable selection is not used. This is reasonable because each class of chemical mixture is selected based on its similar

chemical structures. All individual chemicals within the mixtures have high correlation with one other. Therefore, the model selection did not drop any variables to improve the parameter estimates. The posterior inclusion probability for each chemicals are given in Supplementary Table 3.2.

3.4.5. Sensitivity analysis results

When the biomarkers were analyzed one at a time in MLR while controlling for covariates, then PCB 170, PCB 180, and PCB 153 from the PCBs mixture, PBDE 153 from the PBDE mixture, DDE from the OCPs mixture, MEP from the phthalate mixture, DEP, DMP and DEDTP from the OPs mixture, BPA, MPB and TCS from the phenol mixture and all biomarkers from the PFAS mixture displayed negative associations with birth weight (Table 3.3). All the associations, however, contained the null value of zero.

The regression coefficients and the variances associated with the regression coefficients were both inflated in magnitude after adjusting for co-pollutants within the same mixture class (Table 3.3). Some individual chemicals even showed a reversal in the direction of effect estimates. For example, in the single pollutant model, a ten-fold increase in the concentrations of PCB 170 was associated with a change in birth weight of -118g (95%CI: -348g, 111g). In the model adjusted for co-pollutants, a ten-fold in the concentrations of PCB 170 is associated with a change in birth weight of 1267g (95%CI: 97g, 2437g). These results show that with the presence of collinearity, MLR is inadequate for mixture analysis because variances associated with parameter estimates would inflate to extreme values, resulting in unreliable and imprecise estimates.

3.5. Discussion

Most of the previous mixture analysis of the HOME Study on perinatal outcomes focused on reducing collinearity among individual chemicals. For example, Woods et al. used a hierarchical Bayesian approach to reduce collinearity in the data and generated comprehensive estimates of multiple individual chemical congeners.[10] Kalloo et al. employed both nonparametric (k-means clustering) and parametric approaches (principal component analysis) to generate effect estimates associated with mixtures in conjunction with collinearity reduction.[31] Additionally, numerous papers on the effects of mixtures on other childhood outcomes utilized innovative statistical methods to reduce collinearity.[40,46-52] Yet, few paid attentions to the challenge of interpretability.

We evaluated whether BFA improves precision and interpretability when estimating the health effects of prenatal exposure to chemical mixtures, compared to established methods MLR and BKMR. Among the three methods, BFA produced the most precise effect estimates for the mixture models (Figure 3). The improvement in the precision of the estimate were apparent for both PCBs mixture and PFAS mixtures. Furthermore, the magnitude of the precision improvement was directly related to the degree of correlation among the chemicals. PCBs had higher correlation among each other compared to PFAS and hence had more improvement in precision of the estimates. The improvement of precision is achieved by a combination of latent variable modelling and Bayesian techniques.[26,27] Latent variable modelling alone decreases variance greatly while Bayesian procedures and prior distributions further stabilize the parameter estimates. Meanwhile, BKMR uses non-linear smoothing techniques,[23] which resulted in less precise effect estimates compared with BFA. This is due to the additional amount of variance introduced by allowing non-linearity in the kernel approximating functions. However, both BFA and BKMR performed better than the co-pollutant adjusted MLR models.

MLR showed poor estimate precision and is, therefore, inadequate for mixture analysis.[15-17] The precision of the MLR regression coefficients is drastically reduced after co-pollutant adjustment is made (Table 3). This inflation of variances in the regression estimates is directly related to the degree of collinearity present among the exposures.[20] While the single pollutant regression models could generate more precise estimation, it provided biased estimates because it assumed the absence of the co-pollutant confounding. Additionally, the effect estimates generated in parallel by single pollutant model cannot be simply added arithmetically for mixture effects.[20]

Among the three methods, BFA had the clearest interpretation of the mixture effect estimates. It achieved this by simultaneously modelling the parameter estimates and error terms as depicted in Equations 1 and 2. The regression coefficients generated by BFA can be directly interpreted as the mixture-specific regression coefficients. For example, a change in birth weight for every ten-fold increase in PCB mixture concentration (consists of PCB 118, 138, 153, 170 and 180) is associated with a birth weight change of -81g (95%CI: -132g, -31g), While a change in birth weight for every ten-fold increase in PFAS mixture concentration (consists of PFHXS, PFNA, PFOA and PFOS) is associated with a birth weight change of -57g (95%CI -105g, -10g). Additionally, the BFA modelling framework can be explicitly defined by researchers, making replications and comparisons of results across studies possible. BFA can also be used flexibly for different types of research questions including source-specific mixture studies where the latent construct is defined as the exposure source instead of the

chemical structure, as was in this study.[53]

The latent variable in BFA is a form of dimensional reduction method that can capture information of all the individual chemicals and produce a single index representing the overall exposure of the mixtures.[39,53] The Bayesian framework applied further restrictions to the parameter estimates to reduce variance. In this study, an uninformative gaussian prior with mean zero and variance 1000 were employed for the parameter estimates in the BKMR and BFA. This means that minimal prior information was introduced in our model to influence the final estimates and the information of our data dictate our analysis results. Therefore, the assumptions made in our Bayesian analysis was the same as the non-Bayesian analysis, with the exception that BKMR assumes non-linear relationship while BFA assumes linear relationship. Mathematically, the amount of variance is reduced because only one exposure is modelled instead of all five within the mixture. This simplifies the interpretation and provides a more explicit and specific definition of mixture exposure by capturing all information of the individual chemicals instead of other dimensional reduction methods where a subset of the chemicals is selected.[15,20,54,55]

BKMR provided graphical outputs clearly depicting the dose-response curve. However, the interpretation is challenging because BKMR is an extremely flexible model. During the modelling procedure, different non-linear transformations were used for each exposure. This is advantageous for detection and characterization of non-linear effect in the dose-response curves of the chemical mixtures. However, when the actual relationship of the dose-response is linear, it could introduce additional complexities in the model through the flexible approximation functions. Furthermore, these approximation functions are dependent on the specific data structures of the cohort. This makes direct comparisons of analysis results from different cohorts challenging unless the characteristics of the cohorts are generally comparable, and the same assumption of the biological mechanism are made.

In terms of the interpretation of health effects in our analysis, our BFA results were coherent with our BKMR results. Both methods found that prenatal exposure to PCBs and PFAS were associated with reduced birth weight. Our findings are consistent with previous birth weight studies using the HOME Study data, although the magnitude of the effects may vary slightly due to a combination of reasons such as different data transformation or standardization techniques.[30,31] For example, in a single pollutant study, Rauch et al. found that every 10-fold increase in the concentration of selected OPs (∑DAP, ∑DEP, ∑DMP) was associated with slight negative, but less precise associations with birth weight.[30] In a mixture study, Kalloo et al. used principal component and clustering techniques and found that the

principal component and cluster with mostly OCPs and phenol compounds were associated with a reduction of birth weight, although the 95% CI contained the null.[30] Although the effect estimate computed by Kalloo et al. can be attributed to certain mixtures such as principal components and clusters,[30] it is difficult to interpret the results since the mixture generated were dependent solely on the available data. The mixtures identified can be very different given different make-up of the study population, making direct comparisons of the results difficult.

While the strength of BFA in estimation precision and interpretability is apparent, BFA has several limitations. BFA assumes a linear relationship among the latent mixture which is informed by specific hypothetical causal diagrams and prior distributions of the parameters. If the actual relationship among these variables deviates significantly from the assumptions, BFA results may be biased. Another disadvantage of BFA is the relatively longer computation time needed for achieving model convergences when more variables and more sample sizes are supplied. It should also be noted that depending on the internal variance-covariance relationship present in the data, researchers may need to try different priors and tuning parameters such as the number of burn-in iterations and the ratio of adaptive and posterior sample size for the model to converge successfully. Additionally, our study contained limitations that could not be addressed by simply employing different methods. For example, some of the chemicals analyzed, such as phthalates, may vary during pregnancy. Measurement errors may exist in these non-persistent chemicals because of their short half-lives and that measurements taken at a specific time may not reflect the actual amount of exposure.[56]

In terms of the generalizability for our analysis, the sample size of the HOME Study is relatively modest to examine multiple classes of environmental chemical mixtures. However, the concentrations of environmental chemicals in the HOME Study are similar to those in pregnant women in the U.S. National Health and Nutrition Examination Survey at similar time of enrollment.[58] The estimation of mixture-specific regression coefficients without considering accompanying classes of chemical mixture is a limitation of the study. Ideally, a mixture analysis method includes all classes of chemical exposures, but computational burden is a hinderance. In the future, it is possible to combine different variable selection methods in multiple stages to enhance the estimation of regression coefficients for chemicals from different classes with an increasing sample size.

3.6. Conclusion

We examined three different statistical approaches to characterize and quantify the association between birth weight and prenatal exposures to seven classes of environmental chemical mixtures. We found that PCBs and PFAS displayed strong associations with reduced birth weight. We demonstrated the advantages of BFA in estimate precision and interpretability, while BKMR excels at visualizing dose-response relationships. Therefore, BFA and BKMR can complement each other to provide a more comprehensive interpretation of the mixture-specific effect. We also demonstrated the inadequacy of MLR for mixture assessment, especially in the presence of collinearity.

3.7. References

1. Bergman A, Heindel JJ, Kasten T, et al. The impact of endocrine disruption: a consensus statement on the state of the science. Environ Health Perspect. 2013;121: A104–6.

2. Carlin DJ, Rider CV, Woychik R, Birnbaum LS. Unraveling the Health Effects of Environmental Mixtures: An NIEHS Priority. Environ Health Perspect. 2013;121: a6–a8.

3. Burns JS, Williams PL, Sergeyev O, et al. Serum dioxins and polychlorinated biphenyls are associated with growth among Russian boys. Pediatrics. 2011;127: e59–68.

4. Windham G, Fenster L. Environmental contaminants and pregnancy outcomes. Fertil Steril. 2008;89: e111–6; discussion e117.

5. Wigle DT, Arbuckle TE, Turner MC, et al. Epidemiologic evidence of relationships between reproductive and child health outcomes and environmental chemical contaminants. J Toxicol Environ Health B Crit Rev. 2008;11: 373–517.

6. Schieve LA, Tian LH, Rankin K, et al. Population impact of preterm birth and low birth weight on developmental disabilities in US children. Ann Epidemiol. 2016;26: 267–274.

7. Liu L, Oza S, Hogan D, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet. 2015;385: 430–440.

8. Tanne JH. Preterm and low weight births rise again in the US. BMJ. 2017;358: j3311.

9. Kingsley SL, Eliot MN, Glazer K, et al. Maternal ambient air pollution, preterm birth and markers of fetal growth in Rhode Island: results of a hospital-based linkage study. J Epidemiol Community Health. 2017;71: 1131–1136.

10. Woods MM, Lanphear BP, Braun JM, McCandless LC. Gestational exposure to endocrine disrupting chemicals in relation to infant birth weight: a Bayesian analysis of the HOME Study. Environ Health. 2017;16: 115.

11. Lenters V, Portengen L, Rignell-Hydbom A, et al. Prenatal Phthalate, Perfluoroalkyl Acid, and Organochlorine Exposures and Term Birth Weight in Three Birth Cohorts: Multi-Pollutant Models Based on Elastic Net Regression. Environ Health Perspect. 2016;124: 365–372.

12. Etzel TM, Calafat AM, Ye X, et al. Urinary triclosan concentrations during pregnancy and birth outcomes. Environ Res. 2017;156: 505–511.

13. Shoaff JR, Romano ME, Yolton K, Lanphear BP, Calafat AM, Braun JM. Prenatal phthalate exposure and infant size at birth and gestational duration. Environ Res. 2016;150: 52–58.

14. Harley KG, Engel SM, Vedar MG, et al. Prenatal Exposure to Organophosphorous Pesticides and Fetal Growth: Pooled Results from Four Longitudinal Birth Cohort Studies. Environ Health Perspect. 2016;124: 1084–1092.

15. Lazarevic N, Barnett AG, Sly PD, Knibbs LD. Statistical Methodology in Studies of Prenatal Exposure to Mixtures of Endocrine-Disrupting Chemicals: A Review of Existing Approaches and New Alternatives. Environ Health Perspect. 2019;127: 26001.

16. Taylor KW, Joubert BR, Braun JM, et al. Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology: Lessons from an Innovative Workshop. Environ Health Perspect. 2016;124: A227–A229.

17. Braun JM, Gray K. Challenges to studying the health effects of early life environmental chemical exposures on children’s health. PLoS Biol. 2017;15: e2002800.

18. Braun JM, Gennings C, Hauser R, Webster TF. What Can Epidemiological Studies Tell Us about the Impact of Chemical Mixtures on Human Health? Environ Health Perspect. 2016;124: A6–9.

19. Billionnet C, Sherrill D, Annesi-Maesano I, GERIE study. Estimating the health effects of exposure to multi-pollutant mixture. Ann Epidemiol. 2012;22: 126–141.

20. Hamra GB, Buckley JP. Environmental exposure mixtures: questions and methods to address them. Curr Epidemiol Rep. 2018;5: 160–165.

21. Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics. Br J Math Stat Psychol. 2013;66: 8–38.

22. Eddy SR. What is Bayesian statistics? Nat Biotechnol. 2004;22: 1177.

23. Bobb JF, Claus Henn B, Valeri L, Coull BA. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. Environ Health. 2018;17: 67.

24. Dunson DB. Commentary: practical advantages of Bayesian analysis of epidemiologic data. Am J Epidemiol. 2001;153: 1222–1226.

25. Keith TZ, Reynolds MR. Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. 2018. Available: https://psycnet.apa.org/record/2018-36604-031

26. Vandekerckhove J, Rouder JN, Kruschke JK. Editorial: Bayesian methods for advancing psychological science. Psychon Bull Rev. 2018;25: 1–4.

27. Merkle EC, Wang T. Bayesian latent variable models for the analysis of experimental psychology data. Psychon Bull Rev. 2018;25: 256–270.

28. Braun JM, Kalloo G, Chen A, et al. Cohort Profile: The Health Outcomes and Measures of the Environment (HOME) study. Int J Epidemiol. 2017;46: 24.

29. Braun JM, Daniels JL, Poole C, et al. A prospective cohort study of biomarkers of prenatal tobacco smoke exposure: the correlation between serum and meconium and their association with infant birth weight. Environ Health. 2010;9: 53.

30. Rauch SA, Braun JM, Barr DB, et al. Associations of prenatal exposure to organophosphate pesticide metabolites with gestational age and birth weight. Environ Health Perspect. 2012;120: 1055–1060.

31. Kalloo G, Wellenius GA, McCandless L, et al. Exposures to chemical mixtures during pregnancy and neonatal outcomes: The HOME study. Environ Int. 2020;134: 105219.

32. Hubbard AE. Causal inference and the curse of dimensionality. 41st Annual Meeting of the Society-for-Epidemiologic-Research. 2008. pp. S45–S45.

33. Lubin JH, Colt JS, Camann D, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ Health Perspect. 2004;112: 1691–1696.

34. Wang Y, Chen H-J. Use of Percentiles and Z-Scores in Anthropometry. In: Preedy VR, editor. Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease. New York, NY: Springer New York; 2012. pp. 29–48.

35. Govarts E, Remy S, Bruckers L, et al. Combined Effects of Prenatal Exposures to Environmental Chemicals on Birth Weight. Int J Environ Res Public Health. 2016;13. doi:10.3390/ijerph13050495

36. Keevil VL, Khaw K-T. Overadjustment in regression analyses: considerations when evaluating relationships between body mass index, muscle strength, and body size. The journals of gerontology. Series A, Biological sciences and medical sciences. 2014. pp. 616–617.

37. Walia AS. Cubic and Smoothing Splines in R. In: DataScience+ [Internet]. 30 Jun 2017 [cited 5 Feb 2020]. Available: https://datascienceplus.com/cubic-and-smoothing-splines-in-r/

38. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. Springer Science & Business Media; 2011.

39. Chiu Y-H, Bellavia A, James-Todd T, et al. Evaluating effects of prenatal exposure to phthalate mixtures on birth weight: A comparison of three statistical approaches. Environ Int. 2018;113: 231–239.

40. Valeri L, Mazumdar MM, Bobb JF, et al. The Joint Effect of Prenatal Exposure to Metal Mixtures on Neurodevelopmental Outcomes at 20-40 Months of Age: Evidence from Rural Bangladesh. Environ Health Perspect. 2017;125: 067015.

41. Bryant FB, Yarnold PR. Principal-components analysis and exploratory and confirmatory factor analysis. 1995. Available: https://psycnet.apa.org/record/1995-97110-004

42. Sánchez BN, Budtz-Jørgensen E, Ryan LM, Hu H. Structural Equation Models: A Review with Applications to Environmental Epidemiology. J Am Stat Assoc. 2005;100: 1443–1455.

43. Merkle E, Rosseel Y. blavaan: Bayesian Structural Equation Models via Parameter Expansion. Journal of Statistical Software, Articles. 2018;85: 1–30.

44. Gelman A, Lee D, Guo J. Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization. J Educ Behav Stat. 2015;40: 530–543.

45. Bobb JF, Valeri L, Claus Henn B, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics. 2015;16: 493–508.

46. Liang H, Vuong AM, Xie C, et al. Childhood polybrominated diphenyl ether (PBDE) serum concentration and reading ability at ages 5 and 8 years: The HOME Study. Environ Int. 2019;122: 330–339.

47. Chen A, Yolton K, Rauch SA, et al. Prenatal polybrominated diphenyl ether exposures and neurodevelopment in U.S. children through 5 years of age: the HOME study. Environ Health Perspect. 2014;122: 856–862.

48. Zhang H, Yolton K, Webster GM, et al. Prenatal PBDE and PCB Exposures and Reading, Cognition, and Externalizing Behavior in Children. Environ Health Perspect. 2017;125: 746–752.

49. Braun JM, Yolton K, Stacy SL, et al. Prenatal environmental chemical exposures and longitudinal patterns of child neurobehavior. Neurotoxicology. 2017;62: 192–199.

50. Yolton K, Xu Y, Sucharew H, et al. Impact of low-level gestational exposure to organophosphate pesticides on neurobehavior in early infancy: a prospective study. Environ Health. 2013;12: 79.

51. Yolton K, Xu Y, Strauss D, Altaye M, Calafat AM, Khoury J. Prenatal exposure to bisphenol A and phthalates and infant neurobehavior. Neurotoxicol Teratol. 2011;33: 558–566.

52. Bello GA, Arora M, Austin C, Horton MK, Wright RO, Gennings C. Extending the Distributed Lag Model framework to handle chemical mixtures. Environ Res. 2017;156: 253–264.

53. Petit C, Blangiardo M, Richardson S, Coquet F, Chevrier C, Cordier S. Association of environmental insecticide exposure and fetal growth with a Bayesian model including multiple exposure sources: the PELAGIE mother-child cohort. Am J Epidemiol. 2012;175: 1182–1190.

54. Keil AP, Daza EJ, Engel SM, Buckley JP, Edwards JK. A Bayesian approach to the g-formula. Stat Methods Med Res. 2018;27: 3183–3204.

55. Savitz DA, Wellenius GA. Invited Commentary: Exposure Biomarkers Indicate More Than Just Exposure. Am J Epidemiol. 2018;187: 803–805.

56. Chen X, Xu S, Tan T, Lee ST, et al. Toxicity and estrogenic endocrine disrupting activity of phthalates and their mixtures. Int J Environ Res Public Health. 2014;11: 3156–3168.

57. Ferrari F, Dunson DB. Bayesian Factor Analysis for Inference on Interactions. arXiv [stat.ME]. 2019. Available: http://arxiv.org/abs/1904.11603

58. Woodruff TJ, Zota AR, Schwartz JM. Environmental chemicals in pregnant women in the United States: NHANES 2003-2004. Environ Health Perspect. 2011;119: 878–885.

3.8. Table and figures

Table 3.1 Names and abbreviation of environmental chemical mixtures and the associated individual chemical biomarkers from pregnant women for HOME study, 2003-2006, Cincinnati, OH, n=384

Mixture group	Individual chemical biomarkers
Polychlorinated Biphenyls (PCBs)	PCB 118 PCB 138 PCB 153 PCB 170 PCB 180
Polybrominated Diphenyl Ethers (PBDEs)	PBDE 28 PBDE 47 PBDE 99 PBDE 100 PBDE 153
Organochlorine Pesticides (OCPs)	Dichlorodiphenyldichloroethylene (DDE) Dichlorodiphenyltrichloroethane (DDT) Trans-nonachlor (T_NONA) Oxychlordane (OXYCHLOR) Hexachlorobenzene (HCB)
Organophosphate Pesticides (OPs)	Dimethyldithiophosphate (DMDTP) Diethylthiophosphate (DETP) Diethyl phosphate (DEP)

	Dimethyl thiophosphate (DMTP) Dimethyl phosphate (DMP) Diethyldithiophosphate (DEDTP)
Phthalates	Molar sum of di-2-ethylhexyl phthalate (ΣDEHP)* Mono-benzyl phthalate (MBzP) Mono-n-butyl phthalate (MnBP) Mono-iso-butyl phthalate (MiBP) Mono-ethyl phthalate (MEP)
Phenols	Bisphenol A (BPA) Methyl Paraben (MPB) Benzophenone-3 (BP3) Propyl Paraben (PPB) Triclosan (TCS)
Perfluoroalkyl and Polyfluoroalkyl Substances (PFAS)	Perfluorohexanesulfonic acid (PFHXS) Perfluorononanoic acid (PFNA) Perfluorooctanoic acid (PFOA) Perfluorooctanesulfonic acid (PFOS)

*Weighted molar sum of the DEHP metabolites calculated from:

Mono-(2-ethylhexyl) phthalate (MEHP), Mono(2-ethyl-5-oxohexyl) phthalate (MEHHP)

Mono(2-ethyl-5-oxohexyl) phthalate (MEOHP),Mono-(2-ethyl-5-carboxypentyl) phthalate (MECPP),

expressed in units of ng/mL of MECPP (308 g/mol).

Table 3.2 Distribution of birth weight in relation to participant characteristics among women in the HOME study, 2003-2006, Cincinnati, OH.		
	n (%)	***Birth weight (g) mean ± SD***
All Participants	384 (100%)	3352 ± 632
Maternal Age		
<25	89 (23.2%)	3066 ± 608
25-29	109 (28.0%)	3447 ± 623
30-34	123 (32.0%)	3454 ± 579
35+	63 (16.4%)	3391 ± 674
Education		
Bachelor's Degree or higher	233 (60.1%)	3403 ± 664
Some college or 2-y degree	94 (24.5%)	3272 ± 573
High school diploma or Some high school	57 (14.8%)	3270 ± 577
Race		
White	240 (62.5%)	3484 ± 635

Black	117 (30.5%)	3128 ± 538
Other	27 (7.0%)	3148 ± 680
Marital Status		
Married, Living with partner	252 (65.6%)	3454 ± 637
Not Married, Living with partner	53 (13.8%)	3200 ± 551
Not Living with Partner	79 (20.6%)	3128 ± 596
Household Income		
<$25,000	98 (25.5%)	3123 ± 515
>$25,000 & <$50,000	83 (21.6%)	3392 ± 685
>$50,000 & <$100,000	139 (36.2%)	3472 ± 670
>$100,000	64 (16.7%)	3390 ± 559
Infant Sex		
Male	176 (45.8%)	3473 ± 686
Female	208 (54.2%)	3249 ± 565
Maternal BMI		

Underweight or Normal	161 (41.9%)	3309 ± 582
Overweight	130 (33.9%)	3381 ± 629
Obese	93 (24.2%)	3385 ± 718

Table 3.3 Regression coefficients for the association between individual environmental chemical biomarkers (10-fold increases) and mean birth weight among women in the HOME study, 2003-2006, Cincinnati, OH, using MLR.

	ß adjusted for covariates (95% CI)	***ß adjusted for covariates and other chemicals within the mixture class (95% CI)***
PCBs (n= 310)		
PCB 118	78g (-143g, 286g)	3g (-330g, 336g)
PCB 138	37g (-202g, 263g)	505g (-235g, 1245g)
PCB 153	-99g (-344g, 147g)	-418g (-1645g, 808g)
PCB 170	-118g (-348g, 111g)	1267g (97g, 2437g)
PCB 180	-194g (-423g, 35g)	-1461g (-2696g, -227g)
PBDEs (n= 284)		
PBDE 28	45g (-101g, 191g)	-11g (-381g, 360g)
PBDE 47	65g (-66g, 195g)	-32g (-703g, 638g)
PBDE 99	86g (-38g, 210g)	180g (-186g, 547g)
PBDE 100	17g (-105g, 138g)	24g (-497g, 545g)

PBDE 153	-74g (-189g, 42g)	-153g (-392g, 86g)
OCPs (n= 237)		
DDE	-111g (-358g, 137g)	-312g (-627g, 3g)
DDT	102g (-66g, 271g)	179g (-20g, 377g)
OXYCHLOR	64g (-233g, 361g)	183g (-522g, 889g)
HCB	101g (-284g, 487g)	132g (-358g, 622g)
T_NONA	37g (-208g, 282g)	-96g (-640g, 448g)
Phthalates (n=366)		
ΣDEHP	65g (-26g, 157g)	48g (-55g, 150g)
MEP	-4g (-94g, 86g)	-36g (-134g, 62g)
MiBP	66g (-47g, 181g)	22g (-140g, 184g)
MnBP	73g (-41g, 187g)	29g (-134g, 192g)
MBZP	62g (-38g, 162g)	29g (-107g, 165g)

OPs (n=360)		
DMDTP	20g (-26g, 66g)	34g (-25g, 94g)
DETP	38g (-23g, 98g)	34g (-34g, 102g)
DEP	-29g (-81g, 24g)	-19g (-81g, 43g)
DMTP	17g (-58g, 89g)	35g (-69g, 140g)
DMP	-50g (-106g, 7g)	-81g (-157g, -5g)
DEDTP	-5g (-54g, 44g)	8g (-43g, 59g)
Phenols (n=296)		
BPA	-35g (-177g, 108g)	-32g (-177g, 113g)
MPB	-29g (-136g, 78g)	-69g (-222g, 84g)
BP3	40g (-32g, 111g)	49g (-26g, 123g)
PPB	0g (-86g, 86g)	36g (-86g, 156g)
TCS	-11g (-103g, 82g)	-14g (-110g, 83g)

PFAS (n=307)		
PFHXS	-109g (-282g, 63g)	-41g (-261g, 179g)
PFNA	-251g (-564g, 63g)	-160g (-557g, 237g)
PFOA	-114g (-339g, 112g)	22g (-265g, 310g)
PFOS	-194g (-429g, 42g)	-103g (-469g, 264g)

[1] Total sample size for this analysis was reduced to exclude samples with missing values in one or more of the chemical concentrations after the imputation process. The regression coefficients refer to the association with every two-fold increase in the chemical concentration.

[2] Adjusted for all covariates including cubic-spline gestational age, maternal age, maternal education, race, marital status, household income, infant sex)

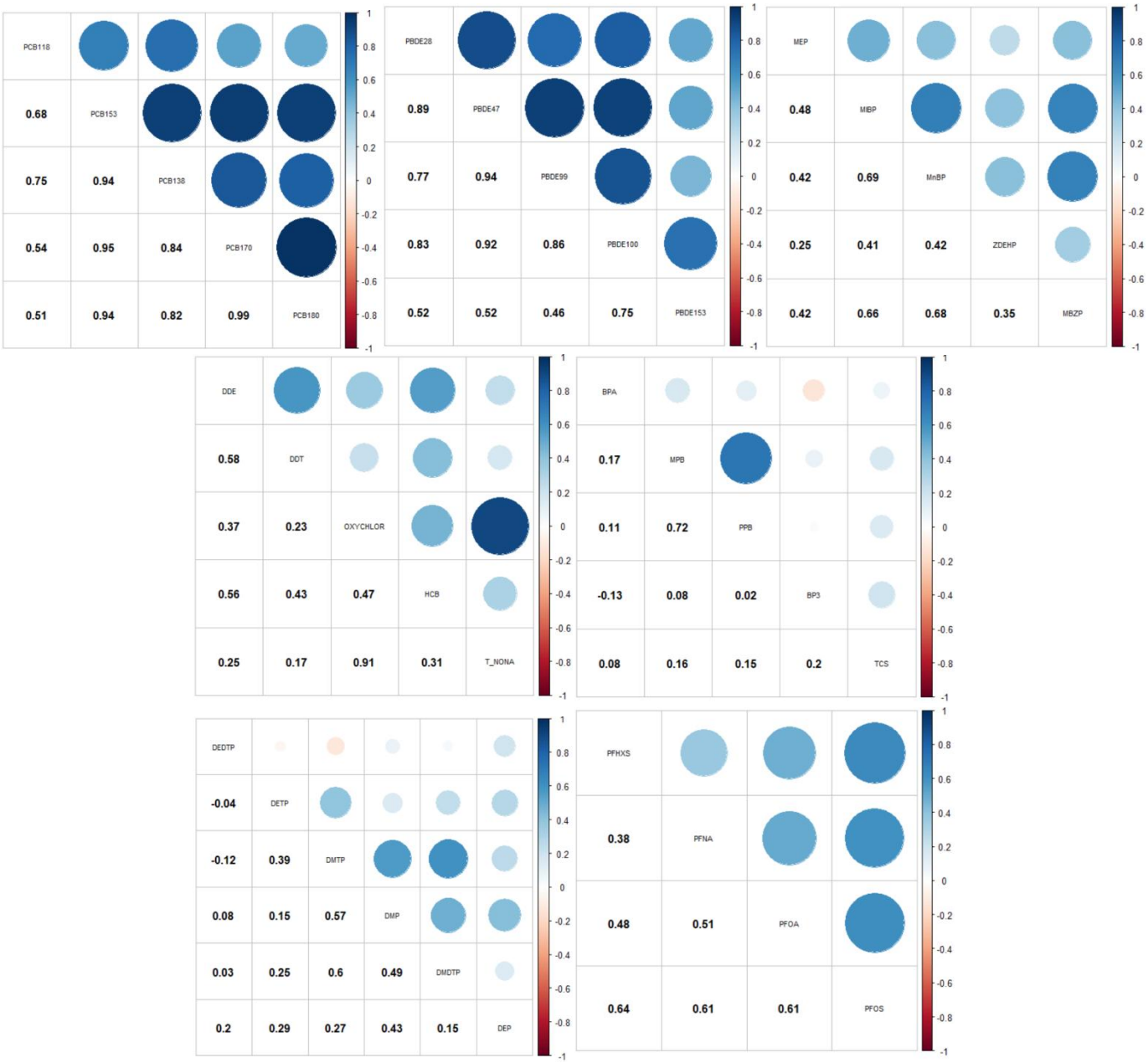
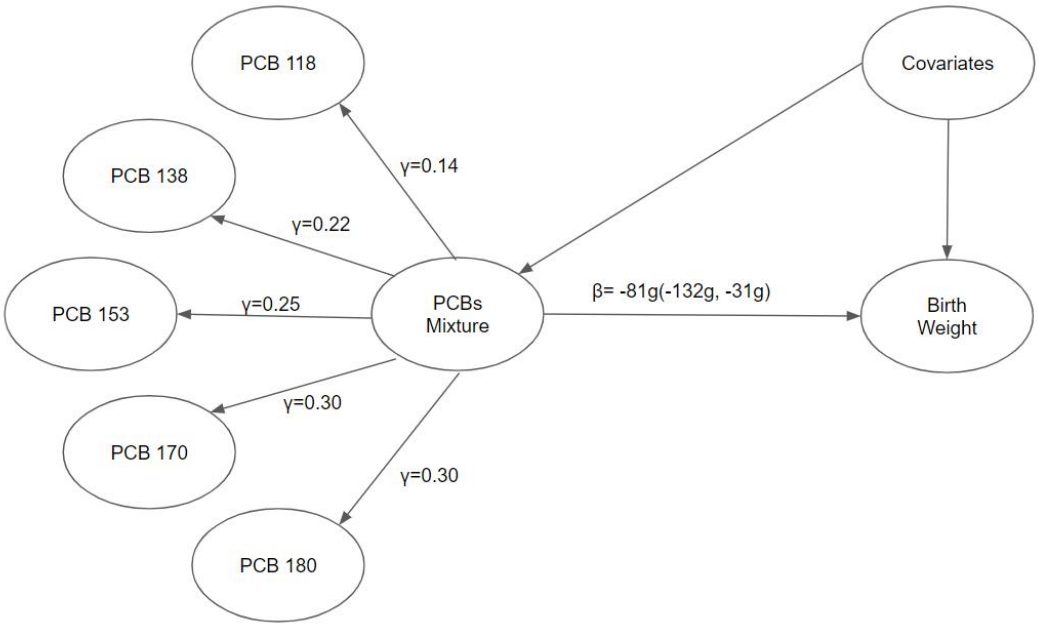


Figure 3.1: Pearson correlation coefficients between environmental chemical biomarkers. The color intensity of shaded circles indicates the magnitude of the correlation. Blue indicates a positive correlation while red indicates a negative correlation.

A.



B.

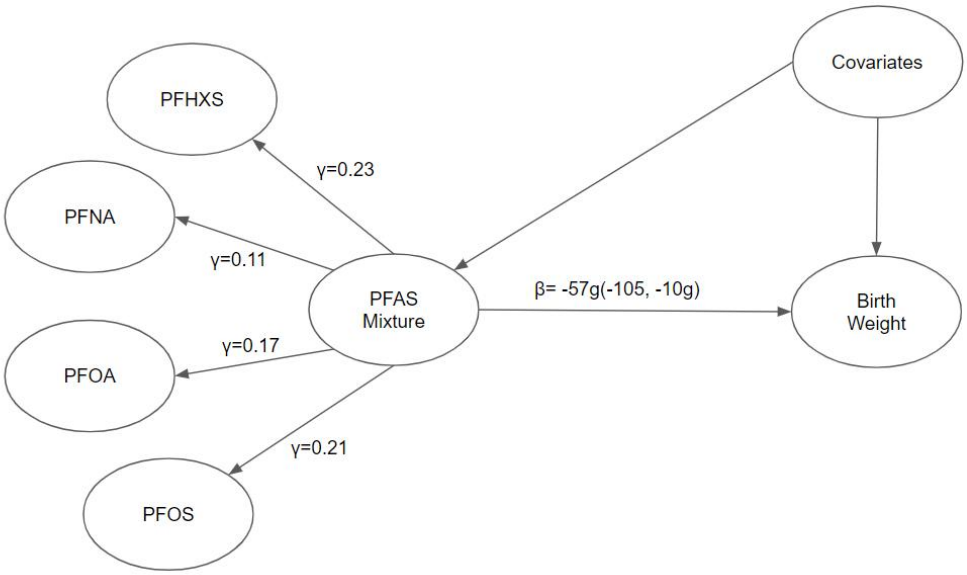
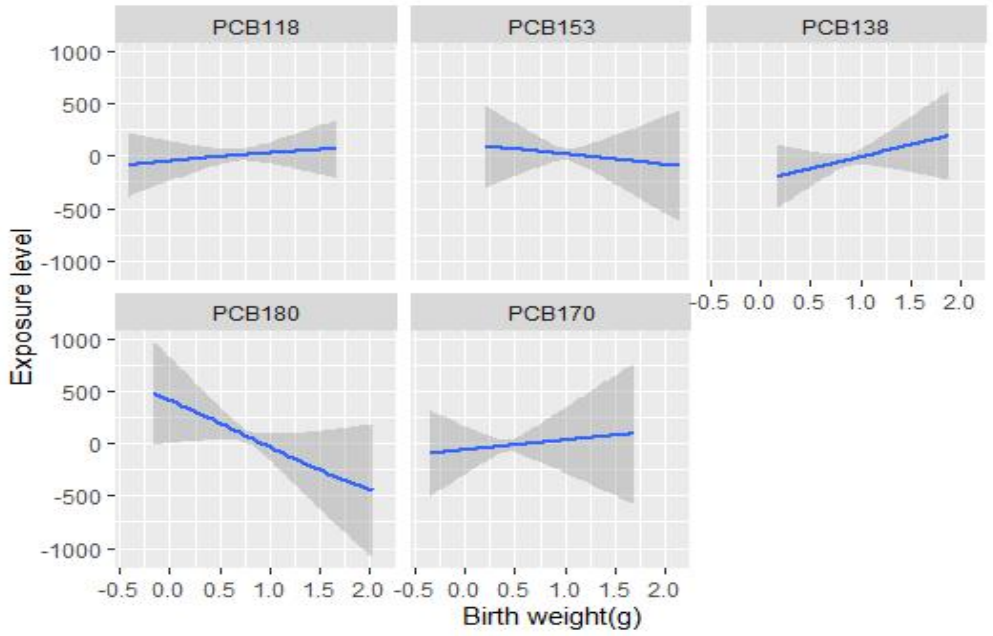


Figure 3.2: The associations between every ten-fold increase of the latent mixture of PCBs (A), the latent mixture of PFAS (B) and birth weight (represented by coefficient **β**) and the factor loadings of the individual congeners onto the latent mixture (represented by coefficient **γ**) among mother-child birth pairs in the HOME Study estimated by Bayesian factor analysis (BFA) adjusted for covariates including cubic-spline gestational age, maternal age, maternal education, race, marital status, household income, infant sex

A.



B.

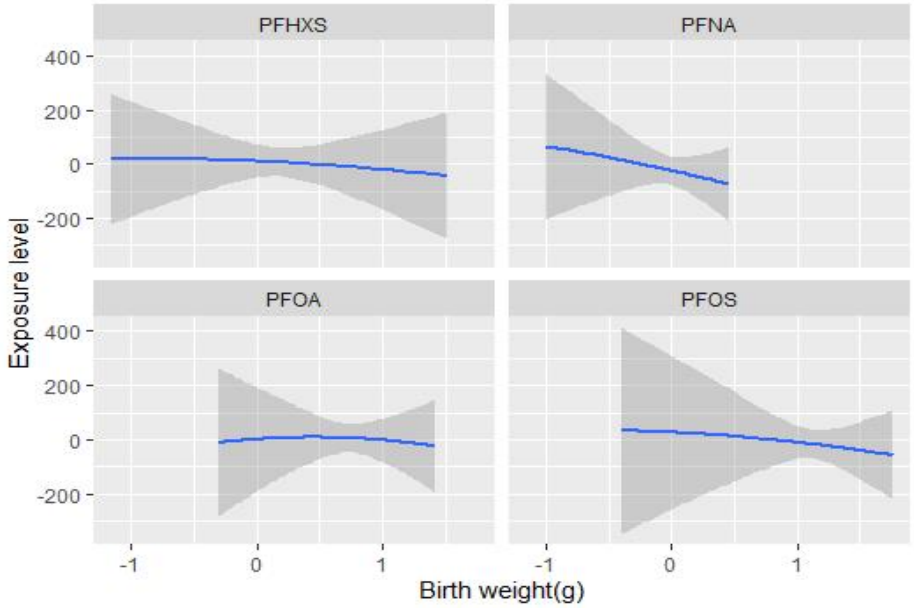
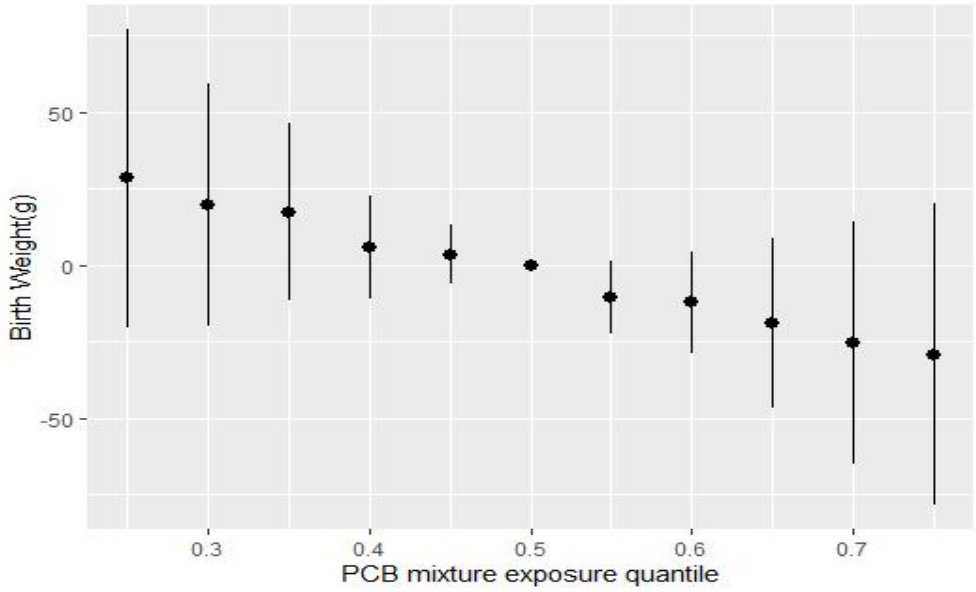


Figure 3.3: Comparisons of the association between PCBs (A), PFAS (B) and change in birth weight (g) according to different methods: The red bars represent the regression estimates **β** with 95% CI for the single pollutant MLR model adjusted for covariates and co-pollutants. The green bars represent the regression estimates **β** with 95% CI for the single pollutant model adjusted for covariates, but not co-pollutants. The blue bars represent regression estimates **β** with 95% CI for three different mixture-specific models related to the factor analysis model outlined (1. MLR with the extracted latent variable. 2. FA. 3. BFA.)

A.



B.

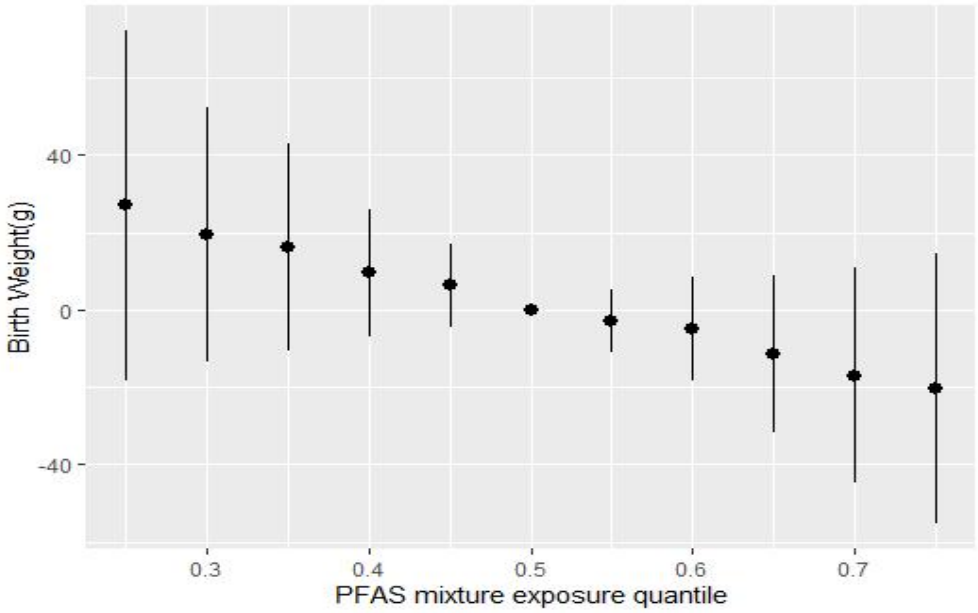


Figure 3.4: Dose-response function (95% credible intervals) between every ten-fold increase in concentrations of selected PCB congeners(A) and birth weight while fixing other PCB congener concentrations at median values and PFAS congeners (B) and birth weight while fixing other PFAS congener concentrations at median values estimated by Bayesian Kernel Machine Regression (BKMR) adjusted for covariates including cubic-spline gestational age, maternal age, maternal education, race, marital status, household income, infant sex.

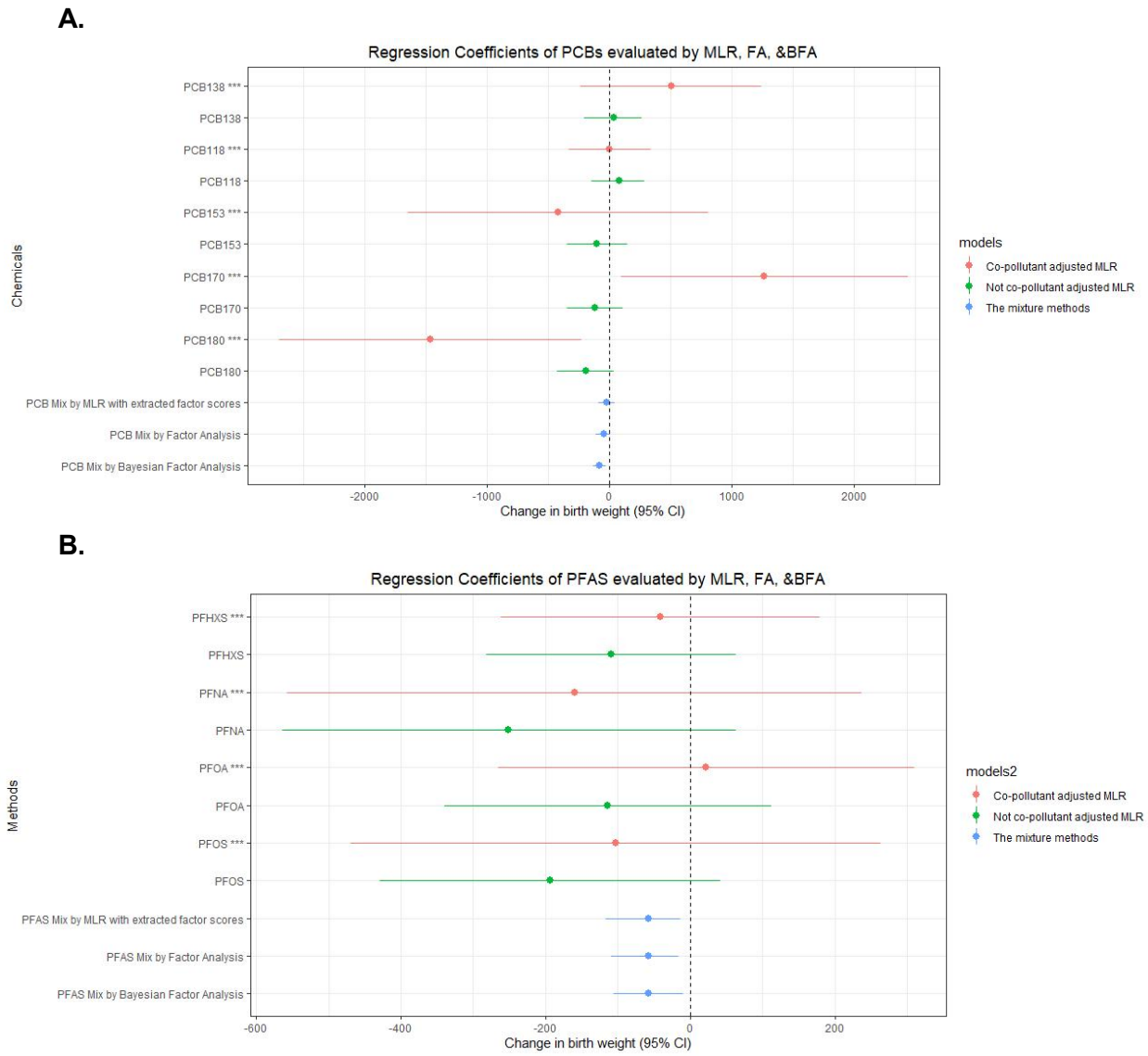


Figure 3.5: Difference in birth weight (95% credible intervals) for different percentiles of the concentrations of all PCB congeners (A) and all PFAS congeners (B) while centering the effect at median concentrations at zero estimated by Bayesian Kernel Machine Regression (BKMR) adjusted for covariates including cubic-spline gestational age, maternal age, maternal education, race, marital status, household income, infant sex.

3.9. Supplementary materials

Supplementary Table 3.1. Regression coefficients for the relation between participant covariates and mean birth weight among women in the HOME study, 2003-2006, Cincinnati, OH. using Multiple Linear Regression (n=384).

	n (%)	Birth Weight Unadjusted mean (95% CI)	Birth Weight Adjusted* mean (95% CI)
Maternal Age			
Intercept		3079g (2955g, 3204g)	759g (-110g, 1629g)
<25	89 (23.2%)	0.0(referent)	0.0(referent)
25-29	109 (28.0%)	369g (201g, 538g)	131g (-14g, 275g)
30-34	123 (32.0%)	377g (212g, 542)	77g (-84g, 239g)
35+	63 (16.4%)	311g (115g, 507g)	74g (-98g, 247g)
Education			
Intercept		3403g (3323g, 3484g)	759g (-110g, 1629g)
Bachelor's Degree or higher	233 (60.1%)	0.0(referent)	0.0(referent)
Some college or 2 year degree	94 (24.5%)	-111g (-259g, 36g)	-39g (-152g, 73g)
High school diploma or Some high school	57 (14.8%)	-114g (-300g, 72g)	-41g (-193g, 112g)
Race			
Intercept		3485g (3408g, 3561g)	759g (-110g, 1629g)
White	240 (62.5%)	0.0(referent)	0.0(referent)

Black	117 (30.5%)	-347g (-480g, -214g)	-124g (-264g, 15g)
Other	27 (7.0%)	-336g (-579g, -95g)	-164g (-354g, 18g)
Marital Status			
Intercept		3455g (3380g, 3531g)	759g (-110g, 1629g)
Married, Living with partner	252 (65.6%)	0.0(referent)	0.0(referent)
Not Married, Living with partner	53 (13.8%)	-254g (-433g, -76g)	-51g (-223g, 120g)
Not Living with Partner	79 (20.6%)	-315g (-470g, -161g)	19g (-159g, 198g)
Household Income			
Intercept		3134g (3014g, 3254g)	759g (-110g, 1629g)
<$25,000	98 (25.5%)	0.0(referent)	0.0(referent)
>$25,000 & <$50,000	83 (21.6%)	297g (158g, 436g)	168g (23g, 313g)
>$50,000 & <$100,000	139 (36.2%)	320g (171g, 470g)	175g (33g, 283g)
>$100,000	64 (16.7%)	290g (151g, 426g)	162g (20g, 305g)
Infant Sex			
Intercept		3473g (3382g, 3563g)	759g (-110g, 1629g)
Male	176 (45.8%)	0.0(referent)	0.0(referent)
Female	208 (54.2%)	-218g (-342g, -95g)	-194g (-284g, -105g)
Maternal BMI			

Intercept		3310g (3213g, 3406g)	759g (-110g, 1629g)
Underweight or Normal	161 (41.9%)	0.0(referent)	0.0(referent)
Overweight	130 (33.9%)	70g (-75g, 214g)	125g (22g, 228g)
Obese	93 (24.2%)	88g (-70g, 247g)	273g (154g, 393g)

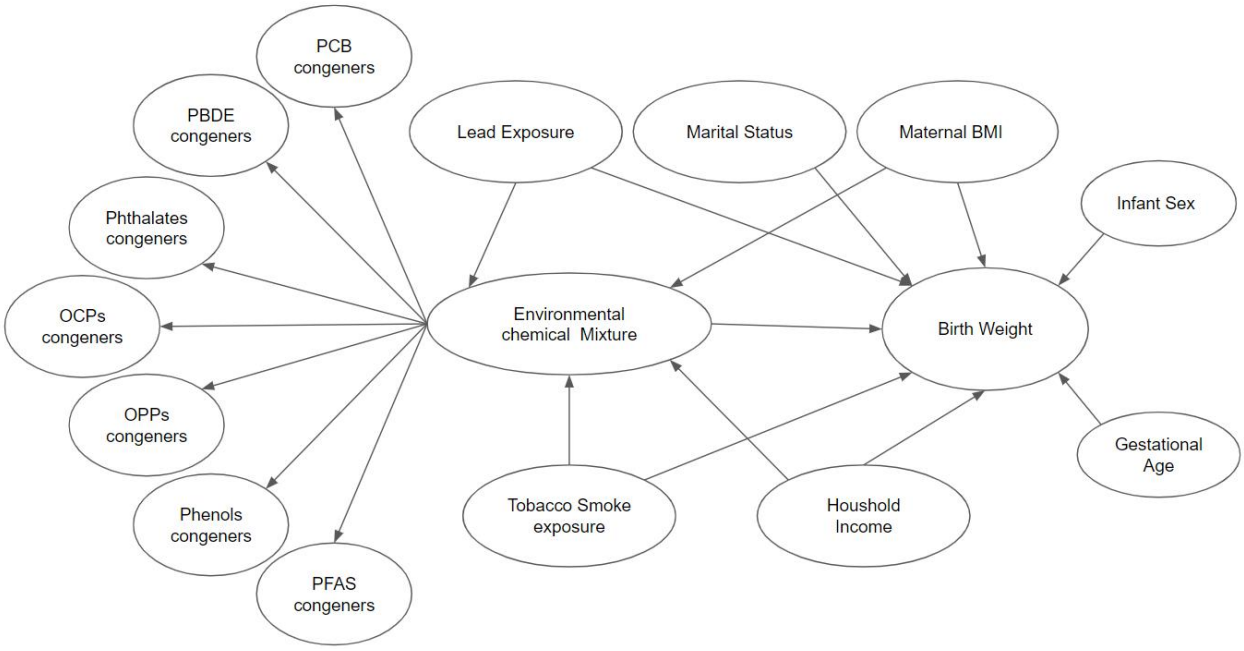
*Adjusted for all other covariates except the covariate being analyzed (cubic-spline gestational age, maternal age, maternal education, race, marital status, household income, infant sex, and maternal BMI)

Supplementary Table 3.2. Extracted Posterior Inclusion Probability (PIP) for individual chemicals within different chemical mixtures according to BKMR analysis with application of variable selection

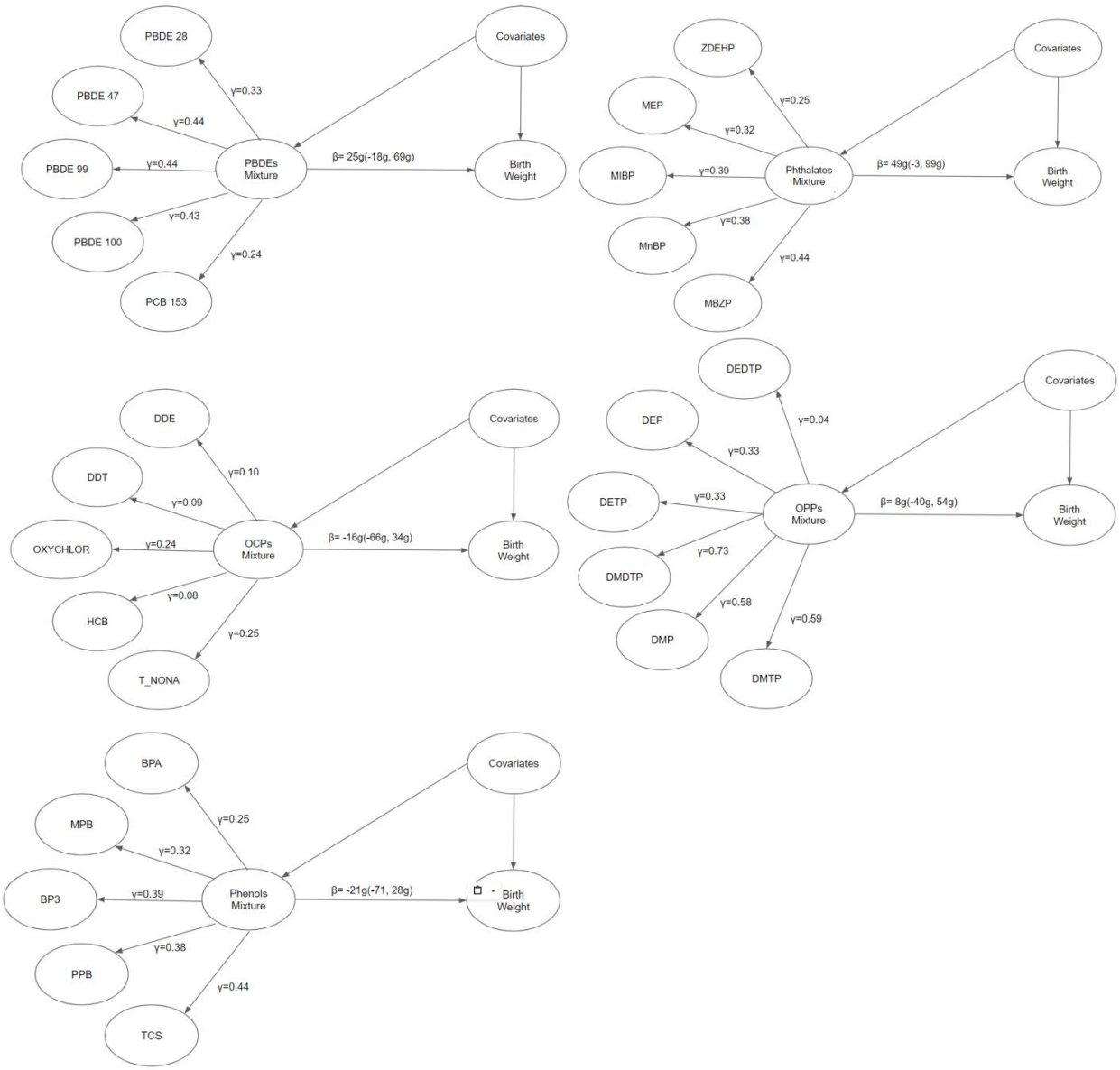
	Extracted Posterior Inclusion Probability (PIP)
PCBs	
PCB 118	0.437
PCB 153	0.447
PCB 138	0.468
PCB 180	0.611
PCB 170	0.540
PBDEs	
PBDE 28	0.496

PBDE 47	0.412
PBDE 99	0.545
PBDE 100	0.393
PBDE 153	0.623
OCPs	
HCB	0.448
OXYCHLOR	0.441
DDE	0.551
DDT	0.484
T_NONA	0.461
OPs	
DEDTP	0.435
DEP	0.533
DETP	0.637
DMDTP	0.756
DMP	0.660
DMTP	0.592
Phthalates	
MEP	0.134

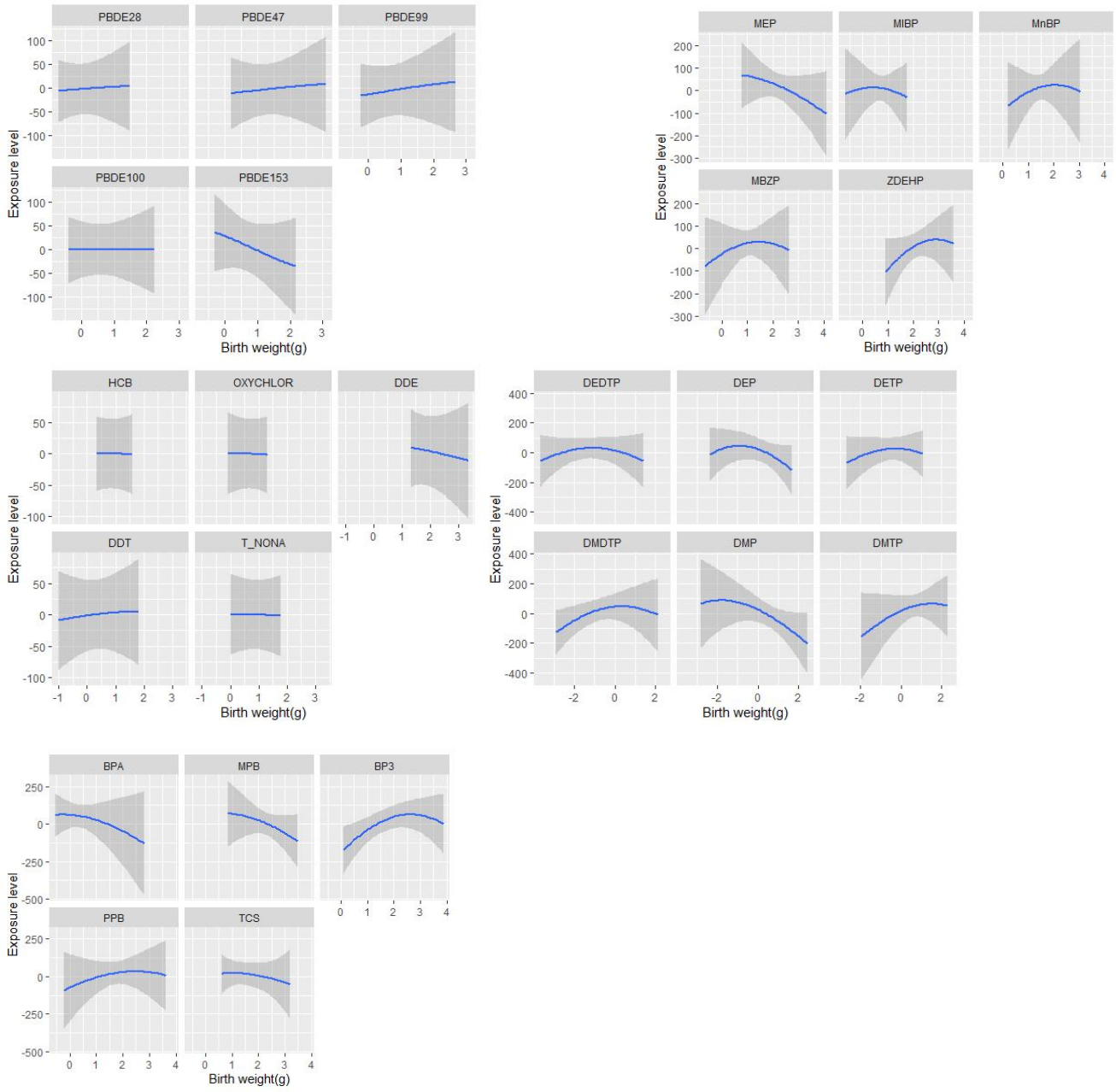
MIBP	0.175
MnBP	0.268
MBZP	0.290
ΣDEHP	0.299
Phenols	
BPA	0.221
MPB	0.207
BP3	0.519
PPB	0.241
TCS	0.190
PFAS	
PFHXS	0.437
PFNA	0.447
PFOA	0.468
PFOS	0.611



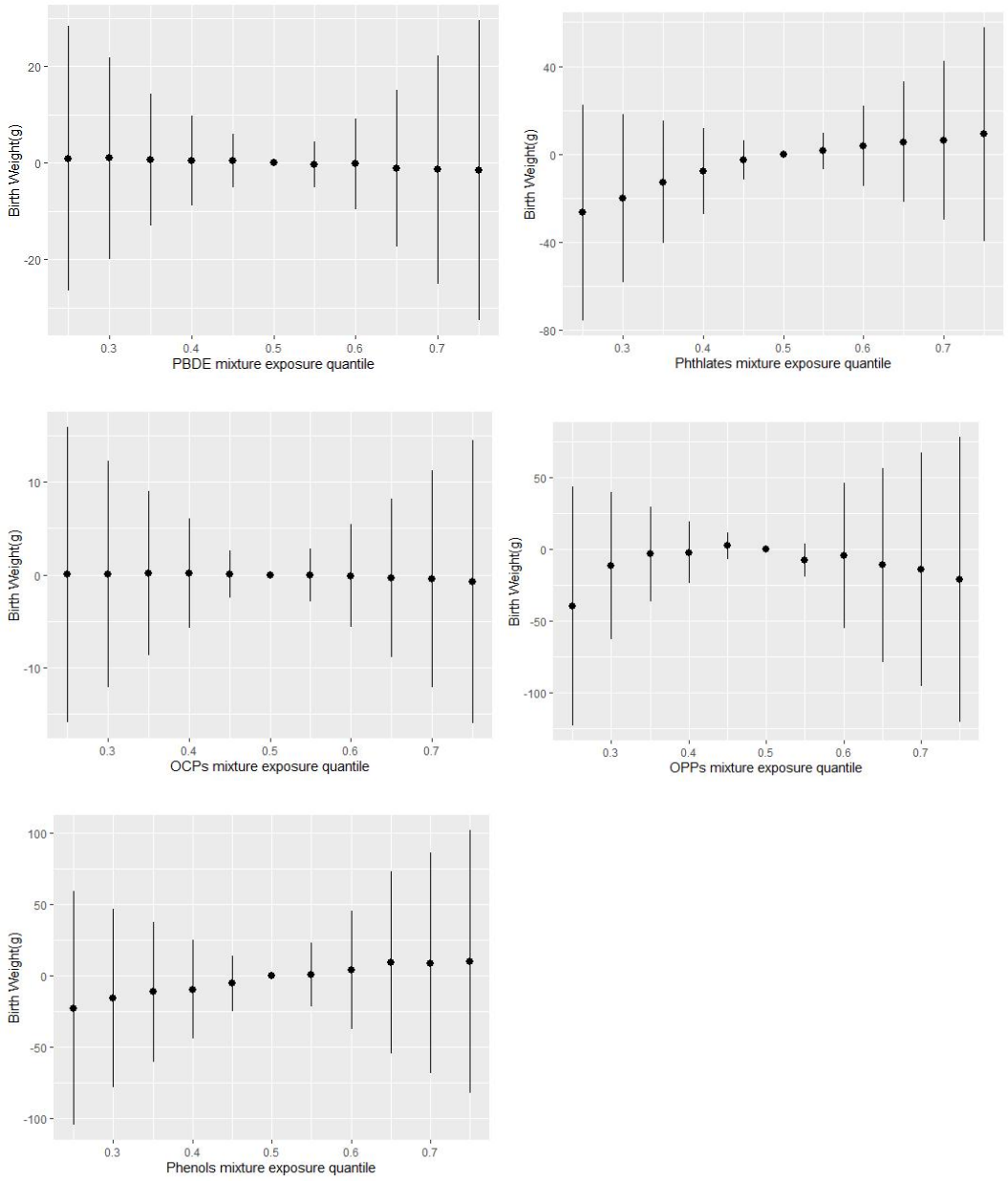
Supplementary Figure 3.1: Directed Acyclic Graph (DAG) for the relationship between exposure to environmental chemical mixtures during pregnancy, birth weight, and covariates.



Supplementary Figure 3.2: The associations between every ten-fold increase of the latent mixture of polybrominated diphenyl ethers (PBDEs), phthalates, organochlorine pesticides (OCPs), organophosphate pesticides (OPs), phenols and birth weight (represented by coefficient **β**) and the factor loadings of the individual PBDE congeners onto the latent mixture (represented by coefficient **γ**) among mother-child birth pairs in the HOME Study estimated by Bayesian factor analysis (BFA) adjusted for covariates including cubic-spline gestational age, maternal age, maternal education, race, marital status, household income, infant sex



Supplementary Figure 3.3: Dose-response function (95% credible intervals) between every ten-fold increase in concentrations of selected PBDE, Phthalates, OCPs, OPs and phenols congeners and birth weight while fixing other congener concentrations at median values estimated by Bayesian Kernel Machine Regression (BKMR) adjusted for covariates including cubic-spline gestational age, maternal age, maternal education, race, marital status, household income, infant sex.



Supplementary Figure 3.4: Differences in birth weight (95% credible intervals) for different percentiles of the concentrations of all PBDE, Phthalates, OCPs, OPs and phenols congeners while centring the effect at median concentrations at zero estimated by Bayesian Kernel Machine Regression (BKMR) adjusted for covariates including cubic-spline gestational age, maternal age, maternal education, race, marital status, household income, infant sex.

Chapter 4. Mediation analysis of thyroid hormones for the associations between environment chemical mixtures and birth weight: The Health Outcome and Measures of Environment (HOME) Study

4.1. Abstract

Background*:* Thyroid hormones are important for fetal development during pregnancy. However, the combined effect of the exposure to chemical mixtures and thyroid hormones on fetal growth have not been investigated. We examined the mediation effect of thyroid hormones on the association between exposure to environmental chemicals and birth weight in order to provide insights on the potential pathway thyroid hormone participates during fetal development.

Methods*:* We analyzed data from a cohort of 214 pregnant women and their newborns using serum biomarkers of polychlorinated biphenyls (PCBs) and perfluoroalkyl substances (PFAS). Serum samples of thyroid hormones (TSH,TT3,FT3,TT4,FT4) were collected during the pregnancy and delivery (maternal: 16 weeks, cord:delivery). We examined the association between exposure to chemical mixtures and birth weight potentially mediated by thyroid hormones using both regression models and latent variable models.

Results*:* We found little evidence of any mediation effect from thyroid hormones on the association between exposure to PCBs, PFAS and infant birth weight. However, we found that the relative impact of mediation by thyroid hormones TT4 and FT4 was greater than the impact of TSH, TT3 and FT3. The use of latent modelling for mediation analysis with chemical mixtures was advantageous in generating more precise confidence intervals compared to regression models.

Conclusion*:* We assessed the mediation effect of thyroid hormones on the association between exposure to PCBs or PFAS and birth weight. Our results demonstrated little evidence of mediation effects while we found that certain thyroid hormones may have a relatively larger impact than other thyroid hormones.

4.2. Introduction

Thyroid hormones are important for fetal development [1]. Specifically, sufficient amounts of thyroid hormone T3, T4 are crucial for central nervous system development [1]. The fetus relies on the maternal supply of thyroid hormones until approximately 18–22 weeks gestation [2]. The fetus continues to depend on maternal inputs for thyroid hormone stabilization even after endogenous fetal production of thyroid hormones begins [3]. The amount of T3, T4 are regulated by a specific pituitary gland hormone TSH[1]. The imbalance of thyroid hormones during pregnancy can lead to growth delays and impairment in neurological functions and lowered IQ scores as shown by previous studies [4]. Low thyroid hormone during pregnancy may also cause neurological impairment [3] and diminished IQ in children.

Laboratory studies have shown that exposure to polychlorinated biphenyls (PCBs) can disrupt thyroid hormone homeostasis in cats and dogs [5]. Prenatal exposure to PCBs were also shown to impact fetal development through thyroid hormone production in rats [33]. Dietary exposures to PCBs were reported to affect circulating levels of thyroid hormones and thyrotropin (TSH) in humans [6]. Exposure to Perfluroalkyl substances (PFAS) can alter circulating levels of thyroid hormones in animal studies [5]. Multiple epidemiological studies have assessed the relationship between various PFAS and thyroid hormones of mothers, neonates, and children, yet the findings were not conclusive [7–9].

The effect of PCBs/PFAS and the effect of thyroid hormones on birth weight have rarely been explored simultaneously. Several studies examined the association between PCBs/PFAS and birth weight [13, 14, 21, 22] while others examined the association between thyroid hormones and birth weight [11,12]. PFAS mixtures were shown to have negligeable association with elevated levels of thyroid hormones [11]. Exposure to environmental chemical mixtures such as PCBs and PFAS were shown to have negative associations with fetal growth indicator such as birth weight [13,14]. Therefore, investigations of the effect of environmental chemicals on fetal growth through the potential mediation of thyroid hormones pathway could be beneficial because it explores the thyroid hormone mechanism involved during fetal growth. The objective of this study was to examine the mediating effect of thyroid hormones on the associations between the exposure to PCBs/PFAS and birth weight. We used multiple linear regression to estimate the effects of each chemical individually, and we used latent variable model to estimate the effects of the mixtures.

4.3. Methods

4.3.1. Health Outcomes and Measures of the Environment (HOME) study

A detailed description of the characteristics of the HOME study is outlined elsewhere [15]. The HOME Study is a prospective birth cohort of pregnant mothers and their infants established in 2003 at the Cincinnati Children's Environmental Health Center, Ohio. To examine the impact of environmental chemicals on child health, pregnant mothers who were >18 years old and at 13-19 weeks of gestation were recruited from seven prenatal clinics and hospitals [15]. The Initial enrollment of the study was 468 women. After adjusting for 67 women dropping out before delivery, nine set of twins and three stillbirths, a total of 384 mothers who delivered singleton live birth were considered for the mediation analysis. We further excluded mothers with missing information on any one of the thyroid hormones, environmental chemicals and sociodemographic variables. The resulted study population consist of 214 mothers for venous cord serum group and 159 mothers for maternal serum group.

4.3.2. Biomarkers of environmental chemical mixtures

We collected blood samples from participants at approximately 16- and 26-weeks gestation [15]. The Centers for Disease Control and Prevention Environmental Health Laboratories used gas and liquid chromatography-mass spectrometry to measure the concentrations of environmental chemical biomarkers in serum samples as previously described [15]. The primary focus of this study is to investigate the impact of mediation from thyroid hormones on the association between environmental mixtures and birth weight, and we built upon the previous study [13] to select two classes of chemicals of interest: polychorinated biphenyls (PCBs) and per/polyfluoroalykyl substances (PFAS). We used samples measured at 16 weeks to maintain consistency across measures. For all biomarkers, measurements below the limit of detection were replaced using single imputation based on a truncated lognormal distribution [16]. The concentrations of these biomarkers were log2 transformed to reduce the effects of right skewness in the distribution and to assist with the interpretation of the results. In regression analysis, the coefficients are interpreted as the change in birth weight for every two-fold increase in the chemical concentrations.

4.3.3 Biomarkers of thyroid hormones

The maternal serum samples were provided by mothers at approximately 16 weeks

gestation and the cord serum samples were collected immediately after delivery [17]. The collected specimens were stored at -70 degree Celsius and later analyzed by the Department of Laboratory Medicine of the University of Washington [11]. Thyroid hormones and antibodies were quantified using a clinical immunoassay analyzer. A total of five thyroid function related biomarkers are included in our analysis: thyroid-stimulating hormone (TSH), free thyroxine (FT4), total thyroxine (TT4), total triiodothyronine (TT3), and free triiodothyronine (FT3) [6]. The concentrations of these hormone biomarkers were also log2 transformed to reduce the effects of right skewness in the distribution and to assist with the interpretation of the results.

4.3.4. Outcome variable

Infant birth weight, measured in grams (g), was abstracted from the medical records and examined as a continuous variable. To examine fetal growth, we adjusted for gestational age, measured in weeks. An alternative measure of the outcome is to use birth weight Z-score standardized by gestational age percentiles [31]. This alternative measure of the outcome statistically avoids the potential complications of modelling the relation between the exposure, the gestational age and the birth weight. However the interpretation of the birth weight Z-score by gestational age percentile is harder to understand, therefore we decided to keep our outcome variable as infant birth weight in grams adjusted by gestational age.

4.3.5. Covariates

A direct acyclic graph was drawn based on the relationship among potential covariates (Figure 4.1). According to Vanderweele [30,32], the potential variables on the exposure-outcome pathway that are associated with the proposed mediator should not be controlled since controlling such variables could induce bias in effect estimates. Therefore, we excluded maternal BMI since it could potentially have an impact on the thyroid hormone pathway [28, 29]. Maternal age at delivery, infant sex, race (white, black, other), marital status (married & living with partner, not married & living with partner, Not living with partner), maternal education (Bachlor's degree or higher, Some college or 2 year degree, High school diploma or some high school), annual household income(x<$25,000, $25,000<x<$50000, $50000<x<$100,000, x>$100,000), tobacco and lead exposure (measured in repeated blood biomarkers concentrations), and gestational age were selected as covariates that should be adjusted in our models. The effect of gestational duration on birth weight has been documented to be nonlinear [18], therefore, we used the cubic spline approach for adjustment for gestational age using the "splines" package in R [19].

4.3.6. Analytic approach: Descriptive statistics

Summary statistics were calculated for two sub-populations of our study. These two sub-populations were distinguished by how the thyroid hormones biomarkers were collected. The cord thyroid hormones biomarkers were measured in cord-serum (n=214) in the first sub-population while the maternal thyroid hormones biomarkers were measured in maternal serum (n=147) in the second sub-population. Both groups had complete information on thyroid hormones, PCBs, PFAS and the associated covariates. We also calculated the median and interquartile range of the thyroid hormones, PCBs, and PFAS among both sub-population of our study.

4.3.7. Approach 1 - Thyroid hormone mediation analysis for individual chemicals

First, we conducted mediation analysis on each of the chemicals individually. We used multivariate linear regression to model the association between each individual PCBs, PFAS and birth weight with mediation by each thyroid hormones, while adjusting for covariates. The model can be represented by the following set of equations:

$$\mathrm{Y} = \beta_0^* + \beta_\mathrm{x}^*\mathrm{X} + \beta_\mathrm{c}^*\mathrm{C} + \epsilon_{\mathrm{y}*} \quad \text{[Equation 1]}$$

$$\mathrm{M} = \gamma_0 + \gamma_\mathrm{x}\mathrm{X} + \gamma_\mathrm{c}\mathrm{C} + \epsilon_\mathrm{M} \quad \text{[Equation 2]}$$

$$\mathrm{Y} = \beta_0 + \beta_\mathrm{x}\mathrm{X} + \beta_\mathrm{m}\mathrm{M} + \beta_\mathrm{c}\mathrm{C} + \epsilon_\mathrm{y} \quad \text{[Equation 3]}$$

where β_0^*, γ_0, β_0 are the Y-intercepts, β_x^* is the regression coefficient between the individual chemicals and birth weight, X is the selected individual chemical, β_c^* is the vector of regression coefficients of covariates, C is the vector of confounders such as age and household income, M is the selected individual thyroid hormone, γ_x is the regression coefficient between the mediator and the individual chemicals, $\beta_\mathrm{x}, \beta_\mathrm{m}$ are the regression coefficients between the individual chemicals, mediator and birth weight in the combined model, and ϵ_y are the normally distributed random errors.

The mediation approaches can be generally characterized as either statistical or causal according to Lee et al. [31]. The statistical mediation analysis assumes linear relationship between exposure, mediator and the outcomes and require normal distribution for these variables. To generate reliable confidence interval estimates, we employed the product method in statistical mediation analysis with bootstrapped sampling. The effect measure of the

mediation was characterized by the direct effect, indirect effect and total effect of the exposure on the outcome. In our study, β_x is the measure of direct effect of individual chemical on birth weight. The measure of the indirect effect is computed by multiplying γ_x and β_m. The measure of the total effect β_x^* is the sum of β_x and $\gamma_x\beta_m$. We employed the R package "mediation" to compute the uncertainty measure of the effect estimate by quasi-Bayesian Monte Carlo method based on normal approximation. For each model, 1,000 simulations were run, and the output of the simulations were bootstrapped to calculate the confidence intervals.

4.3.8. Approach 2 - Thyroid hormone mediation analysis for chemical mixtures

Multivariate regression model can only assess the mediation effect of thyroid hormones between individual chemicals and birth weight. Therefore, building on the work of Zhuang et al.[13], we used latent variable modelling to extract the latent factor score of the respective chemical mixtures. Additionally, this approach provides clear interpretations of the mixture estimates. The latent model is constructed as the following:

$$\mathrm{Y} = \beta_0 + \beta_{\hat{z}}\hat{\mathrm{Z}} + \beta_c\mathrm{C} + \epsilon_{y1} \text{ [equation 4]}$$

$$\mathrm{X_i} = \beta_i + \gamma_i\mathrm{Z} + \epsilon_{xi} \text{ [equation 5]}$$

for i=1, ..., k where k is the number of chemicals in the mixture, β_0, β_i are the Y-intercepts, $\hat{\mathrm{Z}}$ is the estimated factor score, $\beta_{\hat{z}}$ is the regression coefficient for the estimated factor score, $\mathrm{X_i}$ are the individual chemicals within the mixture, Z is the vector of thyroid hormones, γ_i is the the regression coefficients between the individual chemicals and thyroid hormones, β_c is the vector of regression coefficients of covariates, C is the vector of confounders and ϵ_y and ϵ_{xi} are the normally distributed random errors.

After the extraction of the predicted latent factor score $\hat{\mathrm{Z}}$, we reappled it to our mediation analysis model using the R package "lavaan" as the following:

$$\mathrm{Y} = \beta_0^* + \beta_{\hat{z}}\hat{\mathrm{Z}} + \beta_c^*\mathrm{C} + \epsilon_{y*} \text{ [equation 6]}$$

$$\mathrm{M} = \gamma_0 + \gamma_Z\mathrm{Z} + \gamma_c\mathrm{C} + \epsilon_M \text{ [equation 7]}$$

$$\mathrm{Y} = \beta_0 + \beta_{\hat{z}}\hat{\mathrm{Z}} + \beta_m\mathrm{M} + \beta_c\mathrm{C} + \epsilon_{y2} \text{ [equation 8]}$$

where β_0^*, γ_0, β_0 are the Y-intercepts, $\beta_{\hat{z}}$ is the regression coefficient between the latent

mixture and birth weight, $\hat{Z}$ is the selected class of mixture, β_c^* is the vector of regression coefficients of covariates, C is the vector of confounders such as age and household income, M is the selected individual thyroid hormone, γ_Z is the regression coefficient between the mediator and the selected class of mixture, $\beta_{\hat{z}}, \beta_m$ are the regression coefficients between the selected chemical mixture, mediator and birth weight in the combined model, and ϵ_y are the normally distributed random errors. We also conducted exploratory correlation analysis on thyroid hormones, PCBs and PFAS to explore the correlations among these biomarkers to assess whether the latent modelling approaches can perform similarly despite of different degrees of correlations.

4.4. Results

4.4.1. Descriptive Statistics

The demographic characteristics were similar between the cord-TH group and maternal-TH group (Table 4.1). Most of the mother participants were between 25 and 35 years of age(60% in Cord-serum group, 62% in Maternal serum group), non-Hispanic white (63% in both group), had an education of bachelor's degree or higher (60% in Cord-serum group, 56% in Maternal serum group), and were married and living with partners (66% in Cord-serum group, 69% in Maternal serum group). There were more female infants than male infants by a slight margin for both the cord serum group and maternal serum group (45% male, 55% female).

For the biomarker concentrations (Table 4.2), both sub-populations in the cord serum group and the maternal serum group had similar levels for the environmental chemicals (PCBs and PFAS). PCB 153 and PFOS were identified to be the most predominant chemical which had a median of 14.4ng/g lipid (IQR=3.7ng/g lipid) and 14.4 ng/mL (IQR=8.5ng/mL) respectively. For the thyroid hormones, both sub-populations in the cord serum group and the maternal serum group had similar levels of TT4 and FT4. However, for the TSH, TT3 and FT3, the cord serum group had a higher TSH concentration compared to maternal serum group (8.9uIU/mL in Cord-serum group, 1.6uIU/mL in Maternal-serum group) while the maternal serum group had a higher TT3 (50.8ng/dL in Cord-seum group, 161.7ng/dL in Maternal serum group) and FT3 (1.7pg/mL in Cord-serum group, 3.2pg/mL in Maternal serum group).

4.4.2. Individual chemical analysis results

We found limited evidence of the mediation effect from thyroid hormones on the effect between individual PCBs, PFAS and infant birth weight among both the cord serum group and

the maternal serum group (Table 4.3-6). The confidence interval of the regression coefficients of the total effect between individual PCBs, PFAS and birthweight all contained the null except PCB170 and PCB180 in the maternal serum group (Table 4.5). The regression coefficient of the total effect of PCB170 on birth weight was -108g (95%CI: -321g, -14g) while the regression coefficient of the total effect of PCB180 on birth weight was -120g (95%CI: -306g, -53g). The confidence interval of the regression coefficients of the indirect effect of individual PCBs and PFAS mediated by different thyroid hormones on birth weight all contained the null value for both the cord serum group and the maternal serum group.

Although all of the indirect effects estimated by our model contained the null, we were able to compare the magnitude of the mediation relatively between each individual PCBs, PFAS and thyroid hormones by the mean percentage of mediation. For the cord serum group (Table 4.3), the relative impact of mediation by the thyroid hormones on the effect between PCBs and birth weight are slightly different. The indirect effects for FT4 and TT4 were generally greater in magnitude than the effects in other thyroid hormones. Furthermore, this larger FT4 and TT4 effect were apparent on the effect between PFAS and birth weight as well. For the maternal serum group (Table 4.5), the relative impact of mediation by the thyroid hormones of the effect between PCBs and birth weight also displayed greater relative impact among FT4 and TT4 compared to the others. In summary, by the measure of the mediation impact on a relative scale, TT4 and FT4 displayed higher degrees of mediation compared to TSH, TT3 and FT3 for both individual PCBs and PFAS.

4.4.3. Mixture analysis results

Our latent variable model (Table 4.4, Table 4.6) involving PCBs and PFAs mixture demonstrated similar results in both the cord serum group and maternal serum group compared to our individual chemical analysis. While the effect estimate obtained by the latent model still contained the null, the confidence interval was reduced by using latent variables to represent the whole mixture compared to the individual chemical approach. For example, the individual model of PFOA with FT4 as the mediator gave a total effect estimate of -169g (95%CI: -494g,130g) with an indirect effect estimate of -4g (95%CI: -40g, 31g) while the latent model of PFAs mixture with FT4 as the mediator gave an total estimate of -62g (-141g, 18g) with an indirect effect estimate of 3g (95%CI: -5g, 10g). The relative mediation impact of the thyroid hormones agreed with our individual chemical analysis where TT4, FT4 show greater impact compared to TT3, FT3 and TSH in both cord serum and maternal serum group. The maternal serum group had higher percentage of mediation (FT4: -27.8%, TT4: -15.8%) compared with the cord serum group (FT4:-4.8%, TT4: -3.1%). This may be because the maternal serum group have a different thyroid hormone concentration distribution compared to

cord serum group as well as the fact that maternal serum group had a smaller sample size compared to the cord serum group.

4.4.4. Correlation analysis on thyroid hormones, PCBs & PFAS

FT3 and TT3 had a correlation coefficient of 0.72 while TT4 and FT4 had a correlation coefficient of 0.54 (Table 4.7). The correlations among PCBs were higher than PFAs (Table 4.8). Specifically, PCB153, PCB170 and PCB180 had extreme correlation with coefficients of 0.9 while the highest correlation between PFAs was 0.63 between PFHXS and PFOS. This demonstrate that the latent modelling approaches can handle multiple degrees of correlations when doing mixture analysis.

4.5. Discussion

We have found limited evidence for the mediation of thyroid hormones on the associations between exposure to PCBs/PFAs and infant birth weight. By selecting individuals with complete thyroid hormone data, however, the sample size of the study population decreased from 468 to 214 (cord-serum) and 147 (maternal-serum). Another indicator of loss of statistical power was that most of the chemical effect on birth weight also contained the null when previous studies have shown more precise estimate that didn't contain the null with a larger sample size [13]. Although the indirect effects measured all contained the null value, we were nonetheless able to extract some insights by comparing the magnitude of the mean mediation effect estimates across different models. We have found that TT4 and FT4 contributed more to the mediation impact of change in birth weight than TT3, FT3 and TSH for both PCBs and PFAs exposure. These findings were consistent between the individual chemical models and the mixture model.

Thyroid hormones have been documented to be associated with levels of multiple environmental chemicals such as PFAs in the general population [33, 34]. However, Few studies have examined the indirect effect of thyroid hormone on the association between chemical exposure and birth weight, but several have examined on the associations along the exposure-mediation pathways[6,7,9,11,20]. Specifically, numerous published studies looked into the association between PCBs/PFAs and birth weight [13,14,21,22], the association between PCBs/PFAs and thyroid hormones [11,12], and the association between thyroid hormones and fetal growth [23,24]. There is a general consensus in publication for the

negative association between elevated level of PCBs/PFAs concentration in blood and birth weight [13,21,22] as well as the negative association between elevated levels of specfic thyroid hormones (FT4, TSH) and fetal growth [23]. Mixed results were documented for the association between PFAs measured in maternal blood and maternal/neonatal thyroid hormones [9,11]. For example, Lebeaux et al. found that elevated level of PFAS had no association with level of thyroid hormones [11] while Xiao et al. found that every doubling of PFAS was postively associated with TSH level [9]. Our findings were mostly consistent with previous studies of the associations identified along the exposure-mediation pathway, although discrepancies can be found due to distinct sample sizes, sample population and different covariates selected in the modelling process.

Some studies found non-linear relationship between environmental chemicals, thyroid hormones, and fetal growth indicator [25,26]. The impact of thyroid hormones on fetal growth is complicated because the level of TSH, T4 and T3 require delicate balance with each other during the pregnancy term [27]. Specifically, hormone levels needed for healthy fetal growth may differ at times of pregnancy [27]. According to Pirahanchi [27], a healthy thyroid hormone cycle typically has low levels of TSH and high level of FT4/FT3 during the first trimester. As pregnancy continues, TSH increases and peak around the third trimester while FT4/FT3 continue to decrease and plateau, resulting in a consistent level of TT4/TT3 [27]. We collected most of the maternal thyroid hormone and chemical biomarker data around 16 weeks (the beginning of the second trimester). Therefore, we may not capture the impact of thyroid hormone on fetal growth as we were only using data at one particular time. To improve characterizations of the thyroid hormones, future studies should aim to record more data at different time points of the pregnancy to capture the flow and fluctuation of thyroid hormone concentrations. Exposure measures such as change in hormone level in specific time-frames with reference to a healthy thyroid hormone cycle could potentially be a better indicator than hormone concentration level at any given time to explore the impact of the hormones on fetal growth as well as how exposure to various chemical compounds disrupt the normal thyroid hormone cycles.

Additionally, there are several other potential challenges with thyroid hormone mediation analysis. Particularly, potential unmeasured effect modifiers and unmeasured confounding variables in either exposure-mediator pathway or mediator-outcome pathways [35] can impact the effect estimates of the mediation analysis. Although in this particular studies, thyroid diseases were in the exclusion criteria of the study participants [15], individual participants could still be taking supplements such as vitamin pills during the pregnancy that could impact the level of thyroid hormones. Future studies should assess the impact of

potential effect modifiers such as thyroid treatment and supplements when such data is available. For potential confounders along the mediation pathway, there are potential variables that were not measured in the study that can impact the effect estimates. One particular example is the paternal factors such as paternal education. Future studies should also include these factors in modelling if such data is available.

There are several strengths and limitations for this study. The first strength is that this is the first study that comprehensively evaluate chemical exposure, thyroid hormone variation and fetal growth indicator altogether for the HOME study. All previous studies either focused on chemical and fetal growth [13,21], chemical and hormone variation [10] or hormone variation and fetal growth [17]. The second strength is that we demonstrated that with limited amount of data, we can effectively improve our precision estimates by the use of the latent variable analysis to combine multiple individual chemical variables into one latent mixture variable to improve statistical power in our analysis as illustrated in our mixture model.

A major limitation of our study is we assumed a linear relationship between the exposure, mediator and the outcome. While the mechanism of thyroid hormone could be potentially non-linear due to the feedback loop of TSH and TT4,TT3 hormones, future mediation analysis could look into the non-linear relationship. Additionally, the modest sample size of our study which only consists of a sub-population of the HOME study, might resulted in limited external validity when extrapolating to the target population. Certain chemicals had relatively low percentages of Limit of Detection (LODs) and the imputation method we employed in our analysis may have underestimated the variance associated with these chemicals. Lastly, it is essential to consider that exposure misclassification may have influenced the analysis results since we were only using measurements taken at specific time point during pregnancy.

4.6. Conclusion

We assessed the mediation effect of thyroid hormones on the association between exposure to PCBs/PFAs and birth weight using both cord serum and maternal serum. We found limited evidence of mediation effect by thyroid hormones on the association between exposure to PCBs/PFAs and birth weight. Among all thyroid hormones, TT4/FT4 were determined to have more impact on the mediation between chemical exposure and birth weight. We also demonstrated by using latent variable modelling, we can improve statistical power in the case of a relatively small sample size.

4.7. References

1. Chung HR. Iodine and thyroid function. Ann Pediatr Endocrinol Metab. 2014;19: 8–12.

2. Mughal BB, Fini J-B, Demeneix BA. Thyroid-disrupting chemicals and brain development: an update. Endocr Connect. 2018;7: R160–R186.

3. Salazar P, Villaseca P, Cisternas P, Inestrosa NC. Neurodevelopmental impact of the offspring by thyroid hormone system-disrupting environmental chemicals during pregnancy. Environ Res. 2021;200: 111345.

4. Lenters V, Iszatt N, Forns J, Čechová E, Kočan A, Legler J, et al. Early-life exposure to persistent organic pollutants (OCPs, PBDEs, PCBs, PFASs) and attention-deficit/hyperactivity disorder: A multi-pollutant analysis of a Norwegian birth cohort. Environ Int. 2019;125: 33–42.

5. Takaguchi K, Nishikawa H, Mizukawa H, Tanoue R, Yokoyama N, Ichii O, et al. Effects of PCB exposure on serum thyroid hormone levels in dogs and cats. Sci Total Environ. 2019;688: 1172–1183.

6. Turyk ME, Anderson HA, Persky VW. Relationships of thyroid hormones with polychlorinated biphenyls, dioxins, furans, and DDE in adults. Environ Health Perspect. 2007;115: 1197–1203.

7. Itoh S, Araki A, Miyashita C, Yamazaki K, Goudarzi H, Minatoya M, et al. Association between perfluoroalkyl substance exposure and thyroid hormone/thyroid antibody levels in maternal and cord blood: The Hokkaido Study. Environ Int. 2019;133: 105139.

8. Reardon AJF, Khodayari Moez E, Dinu I, Goruk S, Field CJ, Kinniburgh DW, et al. Longitudinal analysis reveals early-pregnancy associations between perfluoroalkyl sulfonates and thyroid hormone status in a Canadian prospective birth cohort. Environ Int. 2019;129: 389–399.

9. Xiao C, Grandjean P, Valvi D, Nielsen F, Jensen TK, Weihe P, et al. Associations of Exposure to Perfluoroalkyl Substances With Thyroid Hormone Concentrations and Birth Size. J Clin Endocrinol Metab. 2020;105. doi:10.1210/clinem/dgz147

10. Vuong AM, Webster GM, Romano ME, Braun JM, Zoeller RT, Hoofnagle AN, et al. Maternal Polybrominated Diphenyl Ether (PBDE) Exposure and Thyroid Hormones in Maternal and Cord Sera: The HOME Study, Cincinnati, USA. Environ Health Perspect. 2015;123: 1079–1085.

11. Lebeaux RM, Doherty BT, Gallagher LG, Zoeller RT, Hoofnagle AN, Calafat AM, et al. Maternal serum perfluoroalkyl substance mixtures and thyroid hormone concentrations in maternal and cord sera: The HOME Study. Environ Res. 2020;185: 109395.

12. Romano ME, Webster GM, Vuong AM, Thomas Zoeller R, Chen A, Hoofnagle AN, et al. Gestational urinary bisphenol A and maternal and newborn thyroid hormone concentrations: the HOME Study. Environ Res. 2015;138: 453–460.

13. Zhuang LH, Chen A, Braun JM, Lanphear BP, Hu JMY, Yolton K, et al. Effects of gestational exposures to chemical mixtures on birth weight using Bayesian factor analysis in the Health Outcome and Measures of Environment (HOME) Study. Environmental Epidemiology. 2021;5: e159.

14. Hu JMY, Arbuckle TE, Janssen P, Lanphear BP, Zhuang LH, Braun JM, et al. Prenatal exposure to endocrine disrupting chemical mixtures and infant birth weight: A Bayesian analysis using kernel machine regression. Environ Res. 2021;195: 110749.

15. Braun JM, Kalloo G, Chen A, Dietrich KN, Liddy-Hicks S, Morgan S, et al. Cohort Profile: The Health Outcomes and Measures of the Environment (HOME) study. Int J Epidemiol. 2017;46: 24.

16. Lubin JH, Colt JS, Camann D, Davis S, Cerhan JR, Severson RK, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ Health Perspect. 2004;112: 1691–1696.

17. Doherty BT, Kosarek N, Hoofnagle AN, Xu Y, Zoeller RT, Yolton K, et al. Maternal, cord, and three-year-old child serum thyroid hormone concentrations in the Health Outcomes and Measures of the Environment study. Clin Endocrinol . 2020;92: 366–372.

18. Snowden JM, Basso O. Causal inference in studies of preterm babies: a simulation study. BJOG. 2018;125: 686–692.
19. Walia AS. Cubic and Smoothing Splines in R. In: DataScience+ [Internet]. 30 Jun 2017 [cited 5 Feb 2020]. Available: https://datascienceplus.com/cubic-and-smoothing-splines-in-r/

20. Berlin M, Barchel D, Brik A, Kohn E, Livne A, Keidar R, et al. Maternal and Newborn Thyroid Hormone, and the Association With Polychlorinated Biphenyls (PCBs) Burden: The EHF (Environmental Health Fund) Birth Cohort. Front Pediatr. 2021;9: 705395.

21. Govarts E, Nieuwenhuijsen M, Schoeters G, Ballester F, Bloemen K, de Boer M, et al. Birth weight and prenatal exposure to polychlorinated biphenyls (PCBs) and dichlorodiphenyldichloroethylene (DDE): a meta-analysis within 12 European Birth Cohorts. Environ Health Perspect. 2012;120: 162–170.

22. Woods MM, Lanphear BP, Braun JM, McCandless LC. Gestational exposure to endocrine disrupting chemicals in relation to infant birth weight: a Bayesian analysis of the HOME Study. Environ Health. 2017;16: 115.

23. Derakhshan A, Peeters RP, Taylor PN, Bliddal S, Carty DM, Meems M, et al. Association of maternal thyroid function with birthweight: a systematic review and individual-participant data meta-analysis. Lancet Diabetes Endocrinol. 2020;8: 501–510.

24. Wu W, Lu J, Ruan X, Ma C, Lu W, Luo Y, et al. Maternal essential metals, thyroid hormones, and fetal growth: Association and mediation analyses in Chinese pregnant women. J Trace Elem Med Biol. 2021;68: 126809.

25. Zdraveska N, Kocova M. Thyroid function and dysfunction in preterm infants-Challenges in evaluation, diagnosis and therapy. Clin Endocrinol . 2021;95: 556–570.

26. Ares Segura S, Casano-Sancho P, Chueca Guindulain M. Assessment of thyroid function in the preterm and/or very low birth weight newborn. Anales de Pediatría (English Edition). 2021;95: 277.e1–277.e8.

27. Pirahanchi Y, Toro F, Jialal I. Physiology, Thyroid Stimulating Hormone. StatPearls. Treasure Island (FL): StatPearls Publishing; 2021.

28. Åsvold BO, Bjøro T, Nilsen TIL, Vatten LJ. Tobacco Smoking and Thyroid Function: A Population-Based Study. Arch Intern Med. 2007;167: 1428–1432.

28. Sun X, Liu W, Zhang B, Shen X, Hu C, Chen X, et al. Maternal Heavy Metal Exposure, Thyroid Hormones, and Birth Outcomes: A Prospective Cohort Study. J Clin Endocrinol Metab. 2019;104: 5043–5052.

29. VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol. 2019;34: 211–219.

30. Lee H, Herbert RD, McAuley JH. Mediation Analysis. JAMA. 2019;321: 697–698.

31. VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. Annu Rev Public Health. 2016;37: 17–32.

32. Gauger KJ, Kato Y, Haraguchi K, Lehmler H-J, Robertson LW, Bansal R, et al. Polychlorinated biphenyls (PCBs) exert thyroid hormone-like effects in the fetal rat brain but do not bind to thyroid hormone receptors. Environ Health Perspect. 2004;112: 516–523.

33. Zhao X, Wang H, Li J, Shan Z, Teng W, Teng X. The correlation between polybrominated diphenyl ethers (PBDEs) and thyroid hormones in the general population: a meta-analysis. PloS one. 2015 May 18;10(5):e0126989.

34. Kim, M. J., Moon, S., Oh, B. C., Jung, D., Ji, K., Choi, K., & Park, Y. J. (2018). Association between perfluoroalkyl substances exposure and thyroid function in adults: A meta-analysis. PloS one, 13(5), e0197244.

35. Valente MJ, Pelham WE, Smyth H, MacKinnon DP. Confounding in statistical mediation analysis: What it is and how to address it. J Couns Psychol. 2017 Nov;64(6):659-671. doi: 10.1037/cou0000242. PMID: 29154577; PMCID: PMC5726285.

4.8. Tables and figures

Table 4.1 Demographic characteristics of study participants with cord and maternal serum data in the HOME study, 2003-2006, Cincinnati, OH.		
	Participants with Cord Serum n (%)	Participants with Maternal Serum n (%)
All Participants	214 (100%)	147 (100%)
Maternal Age		
<25	50 (23%)	29 (20%)
25-35	60 (60%)	91 (62%)
35+	35 (17%)	37 (18%)
Education		
Bachelor's Degree or higher	128 (60%)	82 (56%)
Some college or 2-y degree	54 (25%)	31 (21%)
High school diploma or Some high school	32 (15%)	33 (23%)
Race		

White	135 (63%)	93 (63%)
Black	64 (30%)	41 (28%)
Other	15 (7%)	13 (9%)
Marital Status		
Married, Living with partner	141 (66%)	101 (69%)
Not Married, Living with partner	30 (14%)	16 (11%)
Not Living with Partner	45 (21%)	29 (20%)
Household Income		
<$25,000	56 (26%)	35 (24%)
>$25,000 & <$50,000	47 (22%)	29 (20%)
>$50,000 & <$100,000	75 (35%)	57 (39%)
>$100,000	36 (17%)	25 (17%)
Newborn Sex		

Male	98 (46%)	66 (45%)
Female	116 (54%)	81 (55%)
Maternal BMI		
Underweight or Normal	90 (42%)	65 (44%)
Overweight	73 (34%)	51 (35%)
Obese	51 (24%)	31 (21%)

Table 4.2 Distribution of Thyroid hormones, PCBs, and PFAS across study participants in cord and maternal serum analysis in the HOME study, 2003-2006, Cincinnati, OH.

	Cord Serum Analysis		Maternal Serum Analysis	
Thyroid hormones (mediator)	**n**	**Median(IQR)**	**n**	**Median(IQR)**
TSH (uIU/L)	214	8.9 (6.3)	147	1.6 (1.0)
TT4 (ug/dL)	214	9.6 (1.8)	147	10.2 (1.9)
TT3 (ng/dL)	214	50.8 (16.0)	147	161.7 (23.4)
FT4 (ng/dL)	214	1.0 (0.1)	147	0.7 (0.1)
FT3 (pg/mL)	214	1.7 (0.3)	147	3.2 (0.3)
PCBS (Exposure)	**n**	**Median(IQR)**	**n**	**Median(IQR)**
PCB 118 (ng/g lipid)	214	6.0 (1.7)	147	6.0 (1.6)
PCB 138 (ng/g lipid)	214	10.2 (2.8)	147	10.0 (2.7)
PCB 153 (ng/g lipid)	214	14.4 (3.7)	147	14.2 (3.7)

PCB 170 (ng/g lipid)	214	3.5 (0.3)	147	3.5 (0.3)
PCB 180 (ng/g lipid)	214	8.7 (1.1)	147	8.6 (1.2)
PFAS (Exposure)	**n**	**Median(IQR)**	**n**	**Median(IQR)**
PFHXS (ng/ml)	214	1.6 (1.5)	147	1.6 (1.5)
PFNA (ng/ml)	214	1.0 (0.5)	147	0.9 (0.4)
PFOA (ng/ml)	214	5.6 (4.1)	147	5.6 (4.5)
PFOS (ng/ml)	214	14.5 (8.0)	147	14.3 (9.0)

Table 4.3 Mediation effects of the specific thyroid hormones on birth weight when individual PCBs and PFAS are treated as exposure variables (effect per two-fold increase in biomarker concentration) across study participants in cord serum analysis in the HOME study, 2003-2006, Cincinnati, OH.					
Cord serum analysis	Mediator	Direct effect β (95% CI)	Indirect Effect β (95% CI)	Total Effect β (95% CI)	%Mediated
PCB118	TSH	231g (-9, 544)	2g (-25, 25)	232g (0, 542)	0.7%
PCB138	TSH	118g (-260, 568)	-3g (-41, 20)	116g (-252, 566)	-2.3%
PCB153	TSH	33g (-402, 498)	-1g (-34, 22)	32g (-412, 495)	-3.9%
PCB170	TSH	-20g (-426, 422)	-1g (-40, 25)	-21g (-429, 413)	4.9%
PCB180	TSH	-67g (-450, 393)	-1g (-39, 26)	-68g (-451, 391)	2.0%
PCB118	TT3	229g (-35, 514)	3g (-17, 55)	232g (-10, 552)	1.3%
PCB138	TT3	110g (-263, 577)	6g (-28, 76)	116g (-254, 568)	5.2%
PCB153	TT3	29g (-393, 518)	3g (-34, 48)	32g (-411, 497)	9.4%
PCB170	TT3	-26g (-384, 503)	5g (-46, 104)	-21g (-433, 421)	-23.8%
PCB180	TT3	-78g (-448, 461)	10g (-46, 106)	-68g (-441, 361)	-14.5%
PCB118	FT3	238g (-56, 471)	-6g (-45, 21)	232g (-15, 562)	2.6%
PCB138	FT3	101g (-295, 467)	-5g (-48, 19)	116g (-232, 586)	-4.3%
PCB153	FT3	37g (-452, 453)	-5g (-52, 23)	32g (-400, 498)	-15.6%
PCB170	FT3	-18g (-487, 308)	-3g (-43, 25)	-21g (-435, 420)	14.3%
PCB180	FT3	-64g (-469, 307)	-4g (-49, 22)	-68g (-431, 381)	5.9%
PCB118	TT4	238g (-22, 524)	-6g (-53, 31)	232g (10, 572)	-2.6%
PCB138	TT4	116g (-280, 551)	0g (-4, 27)	116g (-212, 528)	0.2%
PCB153	TT4	-26g (-469, 423)	-6g (-72, 23)	32g (-391, 487)	18.8%
PCB170	TT4	-27g (-424, 430)	6g (-48, 66)	-21g (-439, 423)	-28.6%
PCB180	TT4	-75g (-510, 373)	7g (-47, 69)	-68g (-501, 311)	-10.2%
PCB118	FT4	229g (-113, 427)	3g (-24, 35)	232g (-30, 517)	1.3%
PCB138	FT4	119g (-358, 433)	-3g (-45, 19)	116g (-222, 593)	-2.6%
PCB153	FT4	23g (-421, 515)	9g (-42, 96)	32g (-409, 497)	28.1%
PCB170	FT4	-13g (-463, 412)	-8g (-67, 37)	-21g (-409, 392)	38.1%
PCB180	FT4	-61g (-513, 347)	-7g (-74, 34)	-68g (-482, 367)	10.3%
PFOA	TSH	-178g (-506, 123)	9g (-21, 57)	-169g (-489, 129)	-5.5%
PFNA	TSH	-4g (-439, 413)	0g (-27, 66)	-4g (-415, 414)	-0.2%
PFOS	TSH	-298g (-670, 26)	8g (-34, 73)	-290g (-658, 53)	-2.8%
PFHXS	TSH	-151g (-438, 101)	0g (-3, 21)	-150g (-444, 104)	-0.3%
PFOA	TT3	-167g (-507, 135)	-1g (-45, 33)	-169g (-512, 130)	0.8%
PFNA	TT3	-4g (-450, 420)	0g (-60, 60)	-4g (-439, 419)	-3.8%
PFOS	TT3	-305g (-688, 58)	15g (-39, 74)	-290g (-672, 74)	-5.1%
PFHXS	TT3	-162g (-421, 92)	12g (-30, 52)	-150g(-407, 95)	-7.7%
PFOA	FT3	-170g (-474, 106)	1g (-198, 34)	-169g (-470, 110)	-0.8%
PFNA	FT3	-10g (-453, 413)	0g (-43, 72)	-9g (-457, 415)	-5.2%
PFOS	FT3	-292g (-648, 53)	2g (-3, 39)	-290g (-648, 55)	-0.7%
PFHXS	FT3	-152g (-397, 97)	2g (-16, 33)	-150g (-393, 98)	-1.3%
PFOA	TT4	-166g (-488, 124)	-3g (-31, 22)	-169g (-485, 116)	1.6%
PFNA	TT4	-5g (-450, 421)	0g (-52, 21)	-4g (-455, 417)	-11.1%
PFOS	TT4	-289g (-652, 91)	2g (-38, 17)	-290g (-640, 80)	0.6%

PFHXS	TT4	-147g (-411, 104)	-4g (-45, 30)	-150g (-414, 106)	2.5%
PFOA	FT4	-165g (-496, 134)	-4g (-40, 31)	-169g (-494, 130)	2.4%
PFNA	FT4	-2g (-435, 412)	-2g(-55, 31)	-4g (-443, 390)	53%
PFOS	FT4	-285g (-621, 70)	-5g (-62, 35)	-290g (-652, 74)	1.7%
PFHX	FT4	-145g (-414, 110)	-5g (-58, 41)	-150g (-412, 87)	3.4%

Table 4.4 Mediation effects of the specific thyroid hormones on birth weight when PCBs mixtures, and PFAS mixtures are treated as exposure variables (effect per two-fold increase in latent mixture variable) across study participants in cord serum analysis in the HOME study, 2003-2006, Cincinnati, OH.

Cord serum analysis	Mediator	Direct effect β (95% CI)	Indirect Effect β (95% CI)	Total Effect β (95% CI)	%Mediated
PCBs mixture	TSH	8g (-111, 123)	0g (-9, 7)	8g (-113, 125)	0.1%
	TT3	8g (-103, 132)	0g (-7, 9)	8g (-105, 136)	1.6%
	FT3	8g (-104, 130)	-1g (-13,6)	8g (-103, 127)	-1.0%
	TT4	6g (-107, 132)	2g (-12, 25)	8g (-107, 133)	23.6%
	FT4	8g (-105, 130)	0g (-9, 7)	8g (-105, 130)	-4.0%
PFAS mixture	TSH	-61g (-144, 17)	-1g (-7, 6)	-62g (-146, 15)	1.5%
	TT3	-62g (-143, 19)	0g (-6, 5)	-62g (-145, 20)	0.6%
	FT3	-63g (-147, 18)	1g (-3, 6)	-62g (-141, 18)	-1.3%
	TT4	-65g (-148, 20)	2g (-6, 10)	-64g (-143, 15)	-3.1%
	FT4	-64g (-146, 17)	3g (-5, 10)	-62g (-141, 18)	-4.8%

Table 4.5 Mediation effects of the specific thyroid hormones on birth weight when individual PCBs and PFAS are treated as exposure variables (effect per two-fold increase in biomarker concentration) across study participants in maternal serum analysis in the HOME study, 2003-2006, Cincinnati, OH.					
Cord serum analysis	Mediator	Direct effect β (95% CI)	Indirect Effect β (95% CI)	Total Effect β (95% CI)	%Mediated
PCB118	TSH	56g (-333, 376)	8g (-62, 106)	64g (-325, 395)	12%
PCB138	TSH	-84g (-350, 129)	-4g (-67, 115)	80g (-231, 242)	-5.4%
PCB153	TSH	-71g (-269, 78)	2g (-60, 106)	-69g (-236, 87)	-3.5%
PCB170	TSH	-110g (-326, 22)	2g (-61, 79)	-108g (-321, -14)	-2.2%
PCB180	TSH	-123g (-313, -59)	3g (-57, 81)	-120g (-306, -53)	-2.5%
PCB118	TT3	66g (-326, 414)	-2g (-50, 112)	64g (-306, 431)	-3.1%
PCB138	TT3	78g (-314, 205)	2g (-79, 85)	80g (-201, 312)	3.0%
PCB153	TT3	-68g (-287, 40)	-1g (-65, 70)	-69g (-216, 73)	1.4%
PCB170	TT3	-107g (-276, -20)	-1g (-52, 46)	-108g (-211, -24)	0.7%
PCB180	TT3	-119g (-379, -34)	-1g (-59, 54)	-120g (-386, -45)	0.7%
PCB118	FT3	62g (-337, 461)	1g (-88, 110)	64g (-298, 454)	1.8%
PCB138	FT3	84g (-179, 245)	-4g (-50, 68)	80g (-187, 355)	-4.9%
PCB153	FT3	-71g (-263, 59)	2g (-47, 72)	-69g (-242, 68)	-3.1%
PCB170	FT3	-109g (-382, -22)	1g (-41, 48)	-108g (-383, -22)	-0.9%
PCB180	FT3	-121g (-373, -43)	1g (-34, 47)	-120g (-378, -39)	-0.8%
PCB118	TT4	54g (-312, 432)	10g (-38, 45)	64g (-304, 431)	15.6%
PCB138	TT4	81g (-152, 255)	-1g (-41, 71)	80g (-181, 151)	-1.1%
PCB153	TT4	-70g (-239, 32)	1g (-65, 90)	-69g (-246, 21)	-1.9%
PCB170	TT4	-112g (-384, -43)	4g (-97, 102)	-108g (-391, -22)	-3.6%
PCB180	TT4	-125g (-307, -38)	5g (-99, 103)	-120g (-303, -28)	-4.4%
PCB118	FT4	57g (-368, 408)	6g (-52, 59)	64g (-356, 397)	10.1%
PCB138	FT4	85g (-122, 146)	-5g (-144, 102)	80g (-131, 142)	-6.8%
PCB153	FT4	-74g (-253, 41)	5g (-160, 141)	-69g (-236, 37)	-7.5%
PCB170	FT4	-117g (-342, -24)	9g (-145, 158)	-108g (-351, -14)	-8.0%
PCB180	FT4	-130g (-406, -41)	10g (-150, 150)	-120g (-396, -33)	-8.3%
PFOA	TSH	-132g (-510, 198)	-1g (-57, 30)	-133g (-512, 183)	0.7%
PFNA	TSH	139g (-364, 537)	-5g (-93, 75)	134g (-375, 550)	-3.5%
PFOS	TSH	-93g (-563, 267)	0g (-47, 28)	-94g (-563, 260)	-0.1%
PFHXS	TSH	-57g (-340, 239)	-1g (-37, 26)	-58g (-342, 236)	1.1%
PFOA	TT3	-134g (-549, 234)	1g (-57, 57)	-133g (-543, 204)	-0.7%
PFNA	TT3	135g (-382, 586)	0g (-71, 53)	134g (-381, 578)	-0.3%
PFOS	TT3	-94g (-601, 234)	0g (-47, 27)	-94g (-601, 223)	-0.2%
PFHXS	TT3	-59g (-348, 214)	0g (-25, 43)	-58g (-344, 212)	-0.4%
PFOA	FT3	-131g (-571, 207)	-2g (-49, 35)	-133g (-568, 189)	1.2%
PFNA	FT3	138g (-383, 556)	-3g (-87, 49)	134g (-384, 543)	-2.5%
PFOS	FT3	-95g (-544, 232)	1g (-49, 45)	-94g (-546, 244)	-0.6%
PFHXS	FT3	-59g (-357, 203)	1g (-37, 45)	-58g (-339, 200)	-1.7%
PFOA	TT4	-134g (-552, 172)	1g (-39, 40)	-133g (-485, 116)	-0.9%
PFNA	TT4	122g (-378, 586)	12g (-42, 142)	134g (-382, 605)	9.3%
PFOS	TT4	-99g (-588, 238)	5g (-35, 78)	-94g (-558, 247)	-5.8%

PFHXS	TT4	-65g (-359, 176)	7g (-25, 61)	-58g (-345, 185)	-11.7%
PFOA	FT4	-128g (-555, 190)	-5g (-56, 43)	-133g (-546, 180)	3.4%
PFNA	FT4	139g (-422, 623)	-4g(-71, 94)	134g (-420, 612)	-3.1%
PFOS	FT4	-94g (-635, 270)	0g (-54, 100)	-94g (-603, 260)	0.2%
PFHX	FT4	-64g (-348, 203)	5g (-26, 65)	-58g (-341, 204)	-8.8%

Table 4.6 Mediation effects of the specific thyroid hormones on birth weight when individual PCBs mixtures and PFAS mixtures are treated as exposure variables (effect per two-fold increase in latent mixture variable) across study participants in maternal serum analysis in the HOME study, 2003-2006, Cincinnati, OH.

Maternal serum analysis	Mediator	Direct effect β (95% CI)	Indirect Effect β (95% CI)	Total Effect β (95% CI)	%Mediated
PCBs mixture	TSH	-101g (-219, 12)	2g(-12,18)	-99g (-215, 15)	-2.2%
	TT4	-101g (-224, 20)	2g (-17, 15)	-99g (-224, 20)	-7.3%
	TT3	-101g (-219, 19)	2g (-18,27)	-99g (-219, 14)	-1.6%
	FT3	-107g (-237,8)	7g (-15,40)	-99g (-227,11)	-2.0%
	FT4	-103g (-213, 5)	4g (-14, 22)	-99g (-211, 6)	-3.6%
	Mediator	Direct effect β (95% CI)	Indirect Effect β (95% CI)	Total Effect β (95% CI)	%Mediated
PFAS mixture	TSH	-21g (-126, 86)	-1g (-4, 2)	-22g (-124, 85)	4.5%
	TT4	-24g (-128, 80)	5g (-9, 21)	-18g (-122, 86)	-27.8%
	TT3	-22g (-126, 83)	0g (-4, 3)	-22g (-125, 84)	1.8%
	FT3	-22g (-127, 80)	1g (-5, 7)	-21g (-121, 84)	-4.8%
	FT4	-22g (-126, 82)	3g (-7, 12)	-19g (-123 ,85)	-15.8%

Table 4.7 Correlation among the specific Thyroid hormones across study participants in cord/maternal serum analysis in the HOME study, 2003-2006, Cincinnati, OH.					
	TSH	TT4	TT3	FT4	FT3
TSH	1	-0.08	0.14	-0.08	0.07
TT4	-0.08	1	0.30	0.54*	0.27
TT3	0.14	0.30	1	0.16	0.72*
FT4	-0.08	0.54*	0.16	1	0.27
FT3	0.07	0.27	0.72*	0.27	1
* : Correlation significantly different at α=0.05					

Table 4.8 Correlation among the specific environmental chemicals within mixture across study participants in cord/maternal serum analysis in the HOME study, 2003-2006, Cincinnati, OH.					
	PFOA	PFNA	PFOS	PFHXS	
PFOA	1	0.48	0.51	0.37	
PFNA	0.48	1	0.58*	0.44	
PFOS	0.51	0.58*	1	0.63*	
PFHXS	0.37	0.44	0.63*	1	
	PCB118	PCB138	PCB153	PCB170	PCB180
PCB118	1	0.73*	0.67*	0.54*	0.50
PCB138	0.73*	1	0.94*	0.84*	0.81*
PCB153	0.67*	0.94*	1	0.95*	0.94*
PCB170	0.54*	0.84*	0.95*	1	0.99*
PCB180	0.50	0.81*	0.94*	0.99*	1
* : Correlation significantly different at α=0.05					

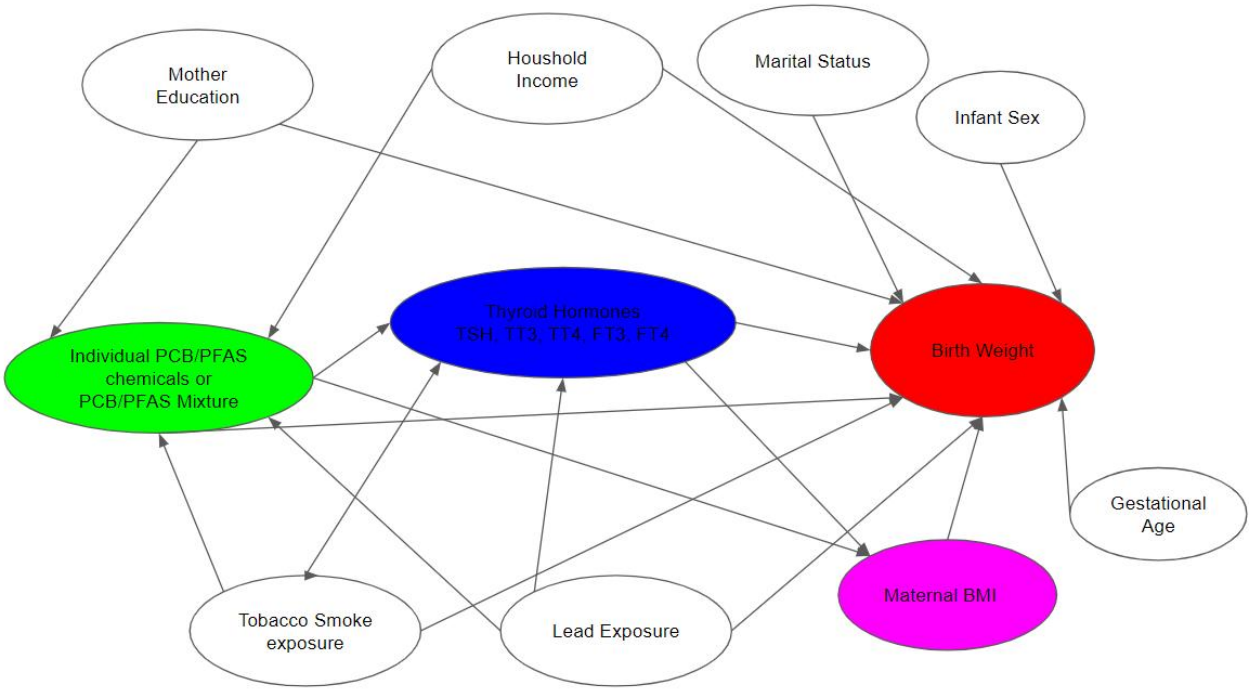


Figure 4.1: Directed Acyclic Graph (DAG) for the relationship between exposure to environmental chemical mixtures during pregnancy, thyroid hormones, birth weight, and covariates. (Green:Exposure of interest, Blue: Mediator, Red: Outcome, Purple: Collider, White: Confounder)

Chapter 5. Conclusion

In summary, this MSc thesis has demonstrated the potential of using novel latent variable approaches for estimating the health effects of chemical mixtures, showcasing their favorable comparison to traditional approaches. We explored the association between environmental chemical exposures during pregnancy and infant birth weight as well as the association between environmental chemical exposures and infant birth weight considering potential mediation of thyroid hormones. With the presence of co-linearity among the different chemical biomarkers and thyroid hormone biomarkers, we examined a novel analysis approach that utilize Bayesian methods and latent variable methods to achieve more reliable estimates compared to traditional approaches.

5.1. Contribution

For our first objective, we utilized three distinct statistical methods to examine the relationship between prenatal exposures to seven classes of environmental chemical mixtures and birth weight. Our findings revealed directional associations between PCBs, PFAS, and reduced birth weight. A 10-fold increase in the concentrations of PCB and PFAS mixtures was associated with an 81g (95%CI -132g, -31g) and 57 g (95%CI -105g, -10g) reduction in birth weight, respectively. For single-pollutant models, a 10-fold increases in the concentrations of multiple chemicals were associated with reduced birth weight, yet the 95% CI all contained the null. Variance inflation from MLR was apparent for models that adjusted for co-pollutants, resulting in less precise confidence intervals.

For our second objective, we investigated the mediation effect of thyroid hormones on the relationship between exposure to PCBs/PFAs and birth weight, considering both cord serum and maternal serum samples. Our results provided limited evidence of a mediation effect by thyroid hormones on the association between PCBs/PFAs exposure and birth weight. Notably, among all the thyroid hormones examined, TT4/FT4 were found to have a greater impact on mediating the relationship between chemical exposure and birth weight. However caution is warrated in interpreting these results because the 95% confidence intervals from the analysis were vary wide and imprecise.

For our third objective, through comparison of other existing methods, we demonstrated the benefits of Bayesian Factor Analysis (BFA) and latent variable modelling in terms of precise estimation and model interpretability. Specifically, for the first project, Bayesian Kernel Machine Regression (BKMR) proved to be useful in visualizing dose-

response relationships. Therefore, combining BFA and BKMR allows for a more comprehensive interpretation of the mixture-specific effect. Furthermore, we demonstrated the limitations of Multiple Linear Regression (MLR) in both of our projects. When assessing complex mixtures with the presence of colinearity either among exposure chemicals or potential mediating thyroid hormones, variance inflation from MLR was apparent, resulting in less precise confidence intervals.

5.2. Relevance for environmental health researchers and/or policy makers

For environmental health researchers and policy makers, this thesis would be relevant in the result findings of our analysis as well as the methodology we employed. Our findings in this thesis helped identify specific chemical mixtures and thyroid hormones that warrant more attentions. In particular, exposure to PCBs and PFAs were found to be associated with reduction in birth weight. Among different thyroid hormones, TT4/FT4 were found to have a greater impact compared to other thyroid hormones for potential mediation on the association between the PCBs, PFAs and infant birth weight. Environmental epidemiologists should communicate the importance of the chemical mixtures to policymakers. Policymakers often focus on single chemical exposures and rely on "safe" dose extrapolations on high-level exposures. However, low-dose mixture exposures of multiple chemicals are significantly more difficult to quantify and extrapolate, especially during vulnerable periods like gestation. It is crucial for policymakers to understand that exposure to multiple low-dose chemicals can be just as hazardous as high-level exposure to a single chemical. A deeper understanding of this complex mixture relation is necessary for informed policy decisions.

From a methodological perspective, this study is relevant to epidemiologists as it introduces new techniques that can be applied to study complex chemical mixtures. The approaches addresses some of the common issues in mixture modelling such as poor precision and difficulty in interpretation. These new approaches allow the examination of mixture effects as a whole, while mitigating problems with interactions and collinearity. To aid with using Bayesian factor analysis, we have supplied our codes (Appendix A), which shows the LAVAAN package in R. Epidemiologists can consider implementing Bayesian techniques and latent variable modelling for studies beyond the scope of this thesis. One of the commonly identified problems in statistical methods is the challenge of providing meaningful parameter estimation within the public health context. The interpretability of results is crucial as they inform policies and practices. Models that offer good interpretability often tend to have less flexibility, as it is difficult to provide clear interpretations when the exposure-response relationship becomes more complex. Our approaches show promises in generating a

framework that can be used for modelling complex mixture modelling and provide easy to follow interpretation of the analysis results.

5.3. **Limitations and Future work**

For both of our applications, we utilized the structures of latent variable method, which assumes the linear relationship between the variables involved. Although this assumption is informed by specific hypothetical causal diagrams, if the true relationship among these variables significantly deviates from these assumptions, then the resulting estimates may be biased [1-3]. Therefore, for future studies, innovation in non-linear modelling can be beneficial for comparison of analysis results and offer different interpretation of the effect of multiple chemicals. Another limitation for our studies was the modest sample size, especially when we are investigating multiple classes of environmental chemical mixtures and thyroid hormones. Future studies should investigate the relationship between the variables with different cohorts, preferably with a larger sample size that is more accommodating. For example, recently Hu et al. [4] applied Bayesian kernel machine regression to model the health effects of chemical mixtures on infant birthweight using data from the Maternal-Infant Research on Environmental Chemicals (MIREC) study. This study uses a sample size of nearly 2000 pregnant women and their infants. Future work could apply the Bayesian factor analysis method to the MIREC data.

Other limitations of our study concern the challenges of measurement of chemical exposure using biomarkers [3]. Certain chemicals had relatively low percentages of Limit of Detection (LODs) and the imputation method we employed in our analysis may have underestimated the variance of parameter estimates in the birthweight models. It is also important to note that we were only using measurements taken at one specific time point during pregnancy for quantification of the chemical exposure and hormone levels during pregnancy, exposure misclassification could potentially influenced the analysis results [5]. Future research should examine different ways of exposure classification such as repeated measurements of chemicals during pregnancy [6-9].

For both of our projects, we picked infant birth weight as the outcome for evaluation. This allows easier interpretation of the results and gives conveniences in modelling as infant birth weight is a continuous measurable variable. However, fetal growth is a complicated phenomenon that should not be characterized just by birth weight. Alternative measure such as birth weight Z-scores that incorporates gestational age in fetal growth characterization can be helpful and future research should examine these alternative approaches. Another major prenatal health outcome is preterm birth. Future studies may investigate further into survival

analysis approaches that incorporate Bayesian and latent variable approaches to generate more insight on the complicated relationship between environmental chemicals and duration of gestation. For example, a Cox proportional hazards model [10] can be used to examine the relationship between the individual chemicals with respect to time to delivery before 37 weeks (preterm birth). For example, Hu et al. modelled preterm birth and chemical exposures in the MIREC data [11]. Hazard ratios (HRs) can be computed to compare the hazard of preterm birth relative to increase in exposure to chemical during pregnancy. The results of such studies could provide a more comprehensive picture of the relationship between duration of gestation, preterm birth and environmental chemical exposures.

5.4. **Reference**

1. Hu JMY, Zhuang LH, Bernardo BA, McCandless LC. Statistical Challenges in the Analysis of Biomarkers of Environmental Chemical Exposures for Perinatal Epidemiology. Current Epidemiology Reports. 2018;5: 284–292.

2. Braun JM, Gray K. Challenges to studying the health effects of early life environmental chemical exposures on children's health. PLoS Biol. 2017;15: e2002800.

3. Sexton K. Cumulative risk assessment: an overview of methodological approaches for evaluating combined health effects from exposure to multiple environmental stressors. Int J Environ Res Public Health. 2012;9: 370–390.

4. Hu, J.M.Y., Arbuckle, T.E., Janssen, P. et al. Associations of prenatal urinary phthalate exposure with preterm birth: the Maternal-Infant Research on Environmental Chemicals (MIREC) Study. Can J Public Health 111, 333–341 (2020). https://doi.org/10.17269/s41997-020-00322-5

5. Savitz DA, Wellenius GA. Invited Commentary: Exposure Biomarkers Indicate More Than Just Exposure. Am J Epidemiol. 2018;187: 803–805.

6. Marateb HR, Mansourian M, Adibi P, Farina D. Manipulating measurement scales in medical statistical analysis and data mining: A review of methodologies. J Res Med Sci. 2014;19: 47–56.

7. Solomon JD, Vallero D, Benson K. Evaluating risk: A revisit of the scales, measurement theory, and statistical analysis controversy. 2017 Annual Reliability and Maintainability

Symposium (RAMS). 2017. pp. 1–6.

8. Youn E, Jeong MK. Class dependent feature scaling method using naive Bayes classifier for text datamining. Pattern Recognit Lett. 2009;30: 477–485.

9. Lubin, J. H., Colt, J. S., Camann, D., Davis, S., Cerhan, J. R., Severson, R. K., et al. (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. Environmental Health Perspectives, 112(17), 1691–1696.

10. Harrell, F. E., & Jr. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer.

11. Hu, J. M. Y., Arbuckle, T. E., Janssen, P., Lanphear, B. P., Zhuang, L. H., Braun, J. M., Chen, A., & McCandless, L. C. (2021). Prenatal exposure to endocrine disrupting chemical mixtures and infant birth weight: A Bayesian analysis using kernel machine regression. Environmental research, 195, 110749, https://doi.org/10.1016/j.envres.2021.110749

Appendix. Example Analysis Code

```
##BFA for single class with confounding/no confounding
library(blavaan)


PCB.model <- '
# latent variable definitions
LAVPCB =~ logPCB118+logPCB153+logPCB170+logPCB180+logPCB74

# regressions
bwt ~ LAVPCB+ga + as.factor(momrace3cat)+as.factor(marital3cat) + momagedeliv+gender+as.factor(momedu4cat)+income.k+wtgain
'
PCB.model



fitpcb <- bsem(PCB.model, data = PCBANA,
               jagcontrol = list(method = "rjparallel"), adapt =100, burnin=1000, sample=5000)

fitpbde <- bsem(PBDE.model, data = PBDEANA,
               jagcontrol = list(method = "rjparallel"), adapt =100, burnin=1000, sample=5000)

fitpht <- bsem(pht.model, data = PhthANA,
                jagcontrol = list(method = "rjparallel"), adapt =100, burnin=1000, sample=5000)

fitocp <- bsem(OCP.model, data = OcpsANA,
                jagcontrol = list(method = "rjparallel"), adapt =100, burnin=1000, sample=5000)

fitopp <- bsem(OPP.model, data = OppsANA,
                jagcontrol = list(method = "rjparallel"), adapt =100, burnin=1000, sample=5000)

summary(fitpcb, standardized=T, fit.measures=T, rsq=T)
```

```
library(lavaan)
library(splines)

##log transformation##
PCBS.Cord$TSH <- log(PCBS.Cord$TSH)/log(10)
PCBS.Cord$TT4 <- log(PCBS.Cord$TT4)/log(10)
PCBS.Cord$FT4 <- log(PCBS.Cord$FT4)/log(10)
PCBS.Cord$FT3 <- log(PCBS.Cord$FT3)/log(10)
PCBS.Cord$TT3 <- log(PCBS.Cord$TT3)/log(10)

PCBs.wk$spline_ga1 <- bs(PCBs.wk$ga)[,1]
PCBs.wk$spline_ga2 <- bs(PCBs.wk$ga)[,2]
PCBs.wk$spline_ga3 <- bs(PCBs.wk$ga)[,3]
PCBs.wk$momrace3cat1 <- ifelse(PCBs.wk$momrace3cat==1, 1, 0)
PCBs.wk$momrace3cat2 <- ifelse(PCBs.wk$momrace3cat==2, 1, 0)

PCBs.wk$marital3cat1 <- ifelse(PCBs.wk$marital3cat==1, 1, 0)
PCBs.wk$marital3cat2 <- ifelse(PCBs.wk$marital3cat==2, 1, 0)

PCBs.wk$momedu4cat1 <- ifelse(PCBs.wk$momedu4cat==1, 1, 0)
PCBs.wk$momedu4cat2 <- ifelse(PCBs.wk$momedu4cat==2, 1, 0)
PCBs.wk$momedu4cat3 <- ifelse(PCBs.wk$momedu4cat==3, 1, 0)

PCBs.wk$bmi3cat1 <- ifelse(PCBs.wk$bmi3cat==1, 1, 0)
PCBs.wk$bmi3cat2 <- ifelse(PCBs.wk$bmi3cat==2, 1, 0)


PCB.Cord.model <- '
# latent variable definitions
LAVPCBs =~ PCB118+PCB138+PCB153+PCB170+PCB180

# direct effect
bwt ~ c*LAVPCBs + momrace3cat+marital3cat + momagedeliv+gender+momedu4cat+income.k
# mediator
FT3 ~ a*LAVPCBs
bwt ~ b*TT3

# indirect effect (a*b)
ab := a*b
# total effect
total := c + (a*b)
'

PCB.cord.fit <- sem(PCB.Cord.model, data = PCBS.Cord, std.lv=TRUE)
summary(PCB.cord.fit)
parameterestimates(PCB.cord.fit)
PCBS.Cord$latent<-as.vector(lavPredict(PCB.cord.fit))
PCBS.Cord<-as.data.frame(PCBS.Cord)
```