# Using Latent Profile Analysis to Examine Associations Between Gestational Chemical Mixtures and Child Neurodevelopment

**by**

**Amanda Yonkman**

B.Sc., University of Northern British Columbia, 2019

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


SIMON FRASER UNIVERSITY
Summer 2021

## Declaration of Committee

| | |
|---|---|
| **Name:** | **Amanda Yonkman** |
| **Degree:** | **Master of Science** |
| **Title:** | **Using Latent Profile Analysis to Examine Associations Between Gestational Chemical Mixtures and Child Neurodevelopment** |
| **Committee:** | **Chair:** **Scott Lear**<br>Professor, Health Sciences |
| | **Lawrence McCandless**<br>Supervisor<br>Professor, Health Sciences |
| | **Angela Kaida**<br>Committee Member<br>Associate Professor, Health Sciences |
| | **Ryan Allen**<br>Committee Member<br>Professor, Health Sciences |
| | **Aimin Chen**<br>Examiner<br>Professor, Epidemiology<br>University of Pennsylvania |

# 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

In this study, we introduced Latent Profile Analysis (LPA) as a novel technique for studying gestational chemical mixtures. Using data from the Maternal-Infant Research on Environmental Chemicals Study, a longitudinal birth cohort study of pregnant Canadian women and their children, we examined the relationship between 30 gestational biomarkers and Verbal IQ, Performance IQ, and Full-Scale IQ. We generated five latent profiles: A Reference profile, a High Level profile, a Low Level profile, a High Organophosphate Pesticides profile, and a Smoking Chemicals profile. Multiple regression analysis showed strong negative associations between the Smoking Chemicals profile and IQ scores. We also found positive associations between the Low Level profile and IQ, and a negative association between the High Level profile and Verbal IQ. However, all 95% confidence intervals spanned the null. After conducting sensitivity analysis comparing LPA with k-means clustering, we concluded that LPA is a promising alternative to other clustering methods.



**Keywords:** Latent Profile Analysis; machine learning; gestational chemical mixtures; maternal-infant health; child neurodevelopment; MIREC

## Acknowledgements

It takes a village to write a thesis. I have way too many people to thank.

First and foremost, I'd like to say a huge thank you to my senior supervisor, Dr. Lawrence McCandless. You have been an incredibly supportive mentor and have taught me so much. It can't have been easy to herd a bunch of graduate students through their theses during a global pandemic, but you've done a fantastic job. It's been a pleasure working with you.

Next, I want to thank my committee members, Dr. Angela Kaida and Dr. Ryan Allen. I've greatly appreciated your insight and support throughout my project. Angela, your GHIRL lab meetings have been the highlight of my week. Ryan, you've always had great, thought-provoking questions that have kept me on my toes.

To the MIREC researchers, my work would not be possible without you. I'd especially like to thank my co-authors Dr. Aimin Chen, Dr. Bruce Lanphear, Dr. Joseph Braun, and Dr. Tye Arbuckle. To the MIREC participants, I saw the length of the questionnaires you had to fill out, and it made me appreciate you all the more.

Thanks to the other students in the McCandless lab who have helped me, directly or indirectly, over the last two years. Harry Zhuang, Jaclyn Parks, Janice Hu, and Shirley Kang, I've learned so much from your work. Josh Alampi, you have been hugely patient in answering all my questions, and I couldn't have done this project without your help.

Last, but not least, a huge shout out to my friends and family. To my parents, I will forever be grateful for your amazing support throughout my education. I love you guys so much. To my brother, David, you don't even like stats that much but you still read my thesis anyway. Thanks, man. Special thanks to my Williams Lake friends who have cheered me on throughout this degree. Kennedi, Kenzie, Gab, and Shroo, I love you all to pieces. And finally, thanks to my Vancouver "household," Maggie, and my "bubble," Chris, Kristyn, Milad, and Steph. It's been a pleasure getting to know you these last two years.

# Table of Contents

# List of Tables

# List of Figures

# List of Acronyms

| | |
|---|---|
| AIC | Akaike Information Criterion |
| BBHC | β-benzene Hexachloride |
| BIC | Bayesian Information Criterion |
| BKMR | Bayesian Kernel Machine Regression |
| BPA | Bisphenol A |
| CHMS | Canadian Health Measures Survey |
| DAG | Directed Acyclic Graph |
| DDE | Dichlorodiphenyldichloroethylene |
| DDT | Dichlorodiphenyltrichloroethane |
| DEHP | Bis-(2-ethylhexyl) Phthalate |
| DEP | Diethylphosphate |
| DETP | Diethylthiophosphate |
| DMDTP | Dimethyldithiophosphate |
| DMP | Dimethylphosphate |
| DMTP | Dimethylthiophosphate |
| FSIQ | Full-Scale Intelligence Quotient |
| LCA | Latent Class Analysis |
| LOD | Limit of Detection |
| LPA | Latent Profile Analysis |
| MBP | Monobutyl Phthalate |
| MBZP | Monobenzyl Phthalate |
| MCPP | Mono-(3-carboxypropyl) Phthalate |
| MEHHP | Mono-(2-ethyl-5-hydroxyhexyl) Phthalate |
| MEHP | Mono-(2-ethylhexyl) Phthalate |
| MEOHP | Mono-(2-ethyl-5-oxohexyl) Phthalate |
| MEP | Mono-ethyl Phthalate |
| MIREC | Maternal-Infant Research on Environmental Chemicals |
| OCP | Organochlorine Pesticides |
| OPP | Organophosphate Pesticides |
| PBDE | Polybrominated Diphenyl Ether |
| PCA | Principal Component Analysis |

| | |
|---|---|
| PCB | Polychlorinated Biphenyl |
| PIQ | Performance Intelligence Quotient |
| VIQ | Verbal Intelligence Quotient |
| WPPSI-III | Wechsler Preschool and Primary Scale of Intelligence - Third Edition |
| WQSR | Weighted Quantile Sum Regression |

# Chapter 1.

# Introduction

## 1.1. Background

### *1.1.1. Gestational Chemical Exposures and Child Neurodevelopment*

Exposure to toxins during the prenatal period can severely impact health outcomes later in life[1–4]. The central nervous system is especially vulnerable to the effects of chemical exposures during this time, as the fetus does not have a fully-formed blood-brain barrier or detoxifying enzymes to protect itself[5–7]. As such, many studies have been conducted on the associations between gestational chemicals and child neurodevelopment[8–18]. Researchers have discovered dozens of gestational neurotoxins; a non-comprehensive summary of these chemicals can be found in the literature review below. However, environmental epidemiologists have primarily focused on the effects of single chemicals or chemical groups in their studies. By failing to study chemical mixtures, researchers may underestimate the impact of aggregate exposures[6,8,9,19]. Low-dose exposures that do not meet the threshold of concern individually may still have severe cumulative effects[20]. It is therefore important for epidemiologists to shift their focus towards gestational chemical mixtures.

Many challenges arise when studying chemical mixtures. First, it is important to be pragmatic when choosing which chemical mixtures to study[20]. Researchers must focus on more prevalent or dangerous mixtures in order to ensure public health relevance and minimize financial cost. Second, if several correlated exposure variables are included in the same multivariate model, collinearity will lead to instability and large standard errors for model parameters[6,21,22]. Finally, epidemiologists do not yet have a standard method for evaluating mixtures[20]. Multivariate regression analysis is commonly used to study smaller, less complex mixtures, but this method loses power when too many variables are included in the model[6,23]. These challenges make it difficult to choose a statistical technique that can accurately estimate the effects and interactions of many chemicals at once[20,22]. For a more in-depth discussion of the challenges of studying chemical mixtures, see Braun et al.'s 2016 review on the topic[20].

### *1.1.2. Machine Learning Techniques*

To mitigate the problems listed above, researchers have turned to unsupervised machine learning techniques for studying chemical mixtures. Machine learning is a branch of

statistics in which a computer system uses algorithms to analyse complex data and detect patterns[24]. Supervised machine learning techniques, in which the researcher specifies the patterns of interest, are a much more common form of machine learning[25]. A popular example of supervised machine learning is regression analysis. Unsupervised machine learning techniques, in which the computer is responsible for detecting patterns without predetermination, are a promising alternative for studying chemical mixtures[24,25].

Several unsupervised methods have been used to study chemical mixtures. Examples include k-means clustering, principal component analysis (PCA), and Bayesian kernel machine regression (BKMR)[9,26–29]. For the purpose of sensitivity analysis, this MSc thesis concerns itself with k-means clustering. In this technique, an algorithm uses Euclidean geometry to separate points in a dataset into distinct, non-overlapping subgroups called clusters[25,26]. This method's relative simplicity makes it a popular choice for epidemiologists. However, it has two key disadvantages. First, the number of k-means clusters must be predetermined by the researcher, and there is not a standard way to make this decision[25,30]. Second, this method does not provide information about classification accuracy at the level of individual data points, meaning researchers cannot see how well each point represents its respective cluster[31]. For these reasons, the results of k-means clustering can be inconsistent and highly dependent on decisions made by the researcher.

### *1.1.3. Latent Profile Analysis*

In this MSc thesis, I introduced a novel technique to the field of environmental epidemiology called Latent Profile Analysis (LPA). LPA is an unsupervised machine learning technique in which an algorithm uses patterns in continuous independent variables to generate homogeneous subgroups called profiles[31,32]. This technique has several advantages over other machine learning techniques. Unlike multiple linear regression analysis, LPA maintains statistical power when working with a large number of variables[32]. Also, because this method allows one to study the effects of mixtures as a whole, researchers do not have to be concerned about collinearity; in fact, models are easier to interpret when the independent variables are highly correlated. Unlike k-means clustering, LPA does not require the researcher to predetermine the number of profiles[33,34]. Quality measures such as the Bayesian Information Criterion (BIC) can be used to choose a model after statistical analysis has begun[33]. Finally, LPA is probabilistic, meaning that it generates the posterior probabilities that each data point will match each profile[31]. This allows researchers to assess classification accuracy. It is for these

reasons that I was interested in using LPA for my thesis project. Its ability to handle models with a large number of highly correlated continuous variables while providing more information than other common techniques makes it extremely promising for studying chemical mixtures.

## 1.2. Literature Review of Gestational Neurotoxins

### *1.2.1. Overview*

In this section, I will be discussing a number of known gestational chemical groups and their associations with child neurodevelopment. It should be noted that due to the methodological nature of this project, this is not a comprehensive review. I did not include every known neurotoxin in my model, as this would be well outside the scope of a MSc thesis. Likewise, I will not be discussing every single study published about the neurotoxins I included in my study. The purpose of this review is to act as a brief introduction to the chemical groups that were used in my models.

### *1.2.2. Heavy Metals: Arsenic, Cadmium, Lead, Manganese, and Mercury*

*Arsenic*

Arsenic is used in manufacturing and can leach into groundwater from contaminated soil[35]. This toxin is a possible predictor of child neurodevelopment[36]. In a 2004 study in Bangladesh, increased arsenic exposure in children was found to be associated with lower IQ scores[37]. However, the effects of prenatal arsenic on child neurodevelopment are under-researched. A study by Wang et al. found that higher prenatal arsenic levels in cord blood were associated with lower neurodevelopment in newborns, but we have little information about the effects on later cognitive development[38].

*Cadmium*

Cadmium is a rare metal that is found as an impurity in underground pipes, which results in trace amounts leaching into drinking water[39]. Although it has not been thoroughly researched, prenatal cadmium exposure is believed to affect neurodevelopment[40]. A systematic review conducted in 2019 found that six of the nine available studies showed negative associations between prenatal cadmium exposure and neurodevelopment[41]. However, more study is needed to understand the exact effects of prenatal cadmium in terms of timing and dose-response.

*Lead*

Lead is a well-known neurotoxin that was once used in food cans, gasoline, and paints[42]. Although it was phased out of commercial use in Canada in the 1970's, Canadian mothers are

still exposed to trace amounts of this metal in food, water, and household products. Gestational lead levels have been linked to lower IQ scores in preschool-aged children, especially boys, as well as an increased likelihood of cognitive disorders later in life[14,36,43]. Researchers have also found a synergistic interaction between prenatal lead and cadmium, which were negatively associated with neurodevelopment[40]. Researchers have not found a threshold of concern below which effects do not occur[44,45].

*Manganese*

Manganese is a naturally-occurring element found in air, water, and soil[46]. Although it is an essential nutrient, studies have found that when maternal blood cadmium levels are too high or too low, it can negatively impact neurodevelopment in young children[47–49]. However, this chemical is also under-researched. A 2019 systematic review by Leonhard et al. found that they could not establish causal effects of prenatal manganese on child neurodevelopment, and concluded that more research is needed[50].

*Mercury*

Mercury is a neurotoxic metal used in various industrial processes[51]. It is persistent in the environment and bioaccumulates in animals. Commonly found in rice and seafood, this toxin enters the body by ingestion and can cross the placenta to the fetus[52]. The effects of gestational mercury exposure have been well-documented; high doses have been found to cause severe neurological issues such as Minamata disease[17,36,53]. However, more research is needed to understand the effects of low mercury levels[52,54].

### *1.2.3. Polychlorinated Biphenyls*

Polychlorinated biphenyls (PCBs) are dioxin-related compounds that were once used in products such as paints, rubber, and hydraulic equipment, but were phased out of use in many countries in the late 1970's[55,56]. Because these chemicals are persistent and bioaccumulative, pregnant mothers are still exposed to trace amounts of them in their everyday lives[18]. PCBs were first studied as neurotoxins in the 1980's, but the effects of low level exposures have been under-researched[57]. Few studies focus on prenatal exposure to PCBs, and those that do have found inconsistent results. A systematic review by Dzwilewski and Schantz showed that prenatal PCBs are negatively associated with linguistic intelligence in older children but not in

toddlers[18]. More research is needed to understand the associations between this chemical group and neurodevelopment.

### *1.2.4. Organochlorine and Organophosphate Pesticides*

Organochlorine pesticides (OCPs) have been banned in many Western countries but are still regularly used in other parts of the world[58]. They are persistent and bioaccumulative, and they can spread long distances easily. Research has shown that prenatal OCP exposure can impact neurodevelopment. A review by Saravi and Dehpour in 2016 found that many types of OCPs are associated with cognitive decline and memory loss[59]. When OCPs were first banned in the 1970's, organophosphate pesticides (OPPs) rose in popularity[60]. Although they are less persistent in the environment than OCPs, OPPs are still considered to be dangerous and are heavily regulated in Western countries. Prenatal exposure to OPPs has been linked to negative effects on neurodevelopment in children[61–63].

### *1.2.5. Polybrominated Diphenyl Ethers*

Polybrominated diphenyl ethers (PBDEs) are persistent, bioaccumulative chemicals that are used as flame retardants in many industrial and commercial products[64]. Although the most dangerous PBDE mixtures have been banned in many countries, pregnant mothers are still regularly exposed to mixtures of PBDEs[65]. A review by Gibson et al. found that prenatal PBDE exposure is negatively associated with cognitive development, although the strength of the association is still not understood[66]. It is possible that these relationships are sex-dependent; a recent study year found associations between PBDEs and neurodevelopment in boys but not girls[67]. Further research is necessary to understand these inconsistencies.

### *1.2.6. Phthalates*

Phthalates are used as plasticizers in many household and industrial products[68]. Studies have found sex-dependent associations between gestational phthalate exposure and child neurodevelopment, although results have been inconsistent. For example, a study by Kim et al. found negative associations between phthalates and neurodevelopment in 6-month-old boys only, while another by Téllez-Rojo found negative associations between mental development indexes and phthalates in girls only[69,70]. More research is needed to clarify these results.

### *1.2.7. Smoking Metabolites*

Exposure to cigarette smoke during the gestational period is linked to many child health outcomes, including neurodevelopment. Cotinine, a metabolite of nicotine, is negatively associated with Verbal IQ and psychomotor function[71–73]. Prenatal smoking has also been linked to increased conduct disorders and schizophrenia[10,74]. However, little study has been done on how smoking chemicals interact with other chemicals. Smoking metabolites should also be studied as part of chemical mixtures to determine whether they act synergistically with other exposures.

### *1.2.8. Gestational Chemical Mixtures and Their Effects on Neurodevelopment*

Historically, most studies on prenatal neurotoxins have failed to consider the effects of complex chemical mixtures[26]. Most research that focuses on mixtures only includes chemicals within the same group; this approach has been used to study the effects of gestational phthalates, pesticides, PBDEs, and metals on cognitive development[9,11,15,16,40,75]. Due to their relative simplicity, these smaller mixtures can be studied using multiple linear regression. However, a few studies have assessed the impacts of more complicated mixtures on IQ. Most of these rely on unsupervised machine learning techniques.

In 2018, Kalloo et al. used two unsupervised machine learning techniques, k-means clustering and PCA, to create clusters and components based on 43 gestational chemical exposures[26]. Their results were later used to study the effects of chemical mixtures on child IQ in the same cohort[8]. In 2020, Tanner et al. used weighted quantile sum regression (WQSR) to study the effects of a mixture of 26 gestational chemicals on IQ[12]. The same year, Guo et al. used BKMR to study the effects of a mixture of gestational heavy metals, pesticides, and phenols on IQ[29].

To my knowledge, LPA has not yet been used to study gestational chemical mixtures. A study by Carrol et al. in 2019 used a similar method called Latent Class Analysis (LCA) to find groups of mothers based on gestational chemical mixtures[76]. Unlike LPA, this method can only create classes using categorical independent variables. The researchers chose to dichotomize their exposure biomarker concentrations, which resulted in a loss of information. Another recent study by Khorrami et al. used LPA to determine associations between air pollution and lung cancer in the general population of Tehran, Iran[77]. The study used chemical exposure variables to create latent profiles showing risk stratification of different air pollutant combinations. While it

is similar to my MSc thesis, this study only used 12 chemical exposures, which were determined by geographical location instead of using biomarkers. Also, the study did not focus on chemical exposures during the gestational period.

## 1.3. Thesis Study Design

### *1.3.1. Data Example: The MIREC Study*

To examine the associations between chemical mixtures and neurodevelopment, I used data from the Maternal-Infant Research on Environmental Chemicals (MIREC) Study. This is a prospective Canadian birth cohort study that began recruitment in 2008 with the aim of finding relationships between gestational exposures and maternal and child health outcomes[78]. The study population includes 2001 pregnant women from 10 Canadian cities. Hundreds of chemical biomarkers were measured in maternal hair, urine, and blood. Questionnaires were also administered to the mothers throughout the pregnancy and in the years following birth. Since 2008, several follow-up studies have been conducted, including the MIREC-CD3 Neurodevelopment Study[79,80]. In this follow-up, MIREC researchers travelled to a subset of the participants' houses and administered questionnaires to the parents and cognitive development tests to the three-year-old children. This thesis concerns itself with the scores from one of these tests, the Wechsler's Preschool and Primary Scale of Intelligence (WPPSI-III) test, which will be described below.

### *1.3.2. The Wechsler's Preschool and Primary Scale of Intelligence (WPPSI-III)*

For this MSc thesis, I used data from the WPPSI-III test, an IQ test that measures various aspects of intelligence through a series of core and supplemental subtests (Table A.1.)[81–83]. In WPPSI-III, subtests include the Receptive Vocabulary test, in which the child points to pictures that correspond with different words; the Information test, in which they answer general questions; the Block Design test, in which they copy three-dimensional designs; and the Object Assembly test, in which they assemble a simple puzzle. Subtest scores are generated based on time and accuracy of task completion.

The subtest results are used to calculate three main IQ scores. First, the Receptive Vocabulary and Information subtests are used to calculate Verbal IQ (VIQ), a measure of linguistic intelligence. Second, the Block Design and Object Assembly subtests are used to calculate Performance IQ (PIQ), a measure of visuospatial intelligence. Finally, the composite Full-Scale IQ (FSIQ) is a combination of VIQ and PIQ and acts as the overall IQ test result. These scores are all scaled to a standardized sample based on the results of the total Canadian population. They have mean levels of 100 and standard deviations of 15[81]. Scoring procedures

can be found in the WPPSI-III Administration and Scoring Manual, and require the use of WPPSI-III scoring and interpretation software[84].

### *1.3.3. Study Objectives*

For this MSc thesis I had three main objectives. The first was to use LPA to create latent profiles based on gestational biomarkers in participants of the MIREC Study. My study population included 517 children who had complete WPPSI-III test scores and biomarker measurements. In my model, I chose to include 30 neurotoxins from the seven chemical groups described in the literature review above. These chemicals were measured in maternal blood and urine during the first trimester of pregnancy[78]. Once I had generated the profiles, my second objective was to determine associations between profile membership and child IQ using regression analysis. I used the posterior probabilities of profile membership as independent variables, and VIQ, PIQ, and FSIQ as dependent variables. Finally, my last objective was to conduct sensitivity analysis using k-means clustering in order to verify my profile generation and regression results. I chose k-means clustering as a comparison method because it has a similar output to LPA, meaning the results are directly comparable[25,26]***.***

This is a manuscript-based thesis. Chapter 2 will be the manuscript that I intend to submit for publication to an epidemiology or biostatistics journal in Fall 2021. Chapter 3 will include my overall conclusions, direction for future study, and implications for researchers and policymakers.

## 1.4. References


1. 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(5-6):373-517.

2. Russ K, Howard S. Developmental Exposure to Environmental Chemicals and Metabolic Changes in Children. *Curr Probl Pediatr Adolesc Health Care*. 2016;46(8):255-285.

3. Rappazzo KM, Coffman E, Hines EP. Exposure to Perfluorinated Alkyl Substances and Health Outcomes in Children: A Systematic Review of the Epidemiologic Literature. *Int J Environ Res Public Health*. 2017;14(7). doi:10.3390/ijerph14070691

4. Boekelheide K, Blumberg B, Chapin RE, et al. Predicting later-life outcomes of early-life exposures. *Environ Health Perspect*. 2012;120(10):1353-1361.

5. Lanphear BP. The impact of toxins on the developing brain. *Annu Rev Public Health*. 2015;36:211-230.

6. 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(2):26001.

7. Sutton P, Woodruff TJ, Perron J, et al. Toxic environmental chemicals: the role of reproductive health professionals in preventing harmful exposures. *Am J Obstet Gynecol*. 2012;207(3):164-173.

8. Kalloo G, Wellenius GA, McCandless L, et al. Chemical mixture exposures during pregnancy and cognitive abilities in school-aged children. *Environ Res*. 2021;197:111027.

9. 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(6):067015.

10. Braun JM, Froehlich TE, Daniels JL, et al. Association of environmental toxicants and conduct disorder in U.S. children: NHANES 2001-2004. *Environ Health Perspect*. 2008;116(7):956-962.

11. Herbstman JB, Sjödin A, Kurzon M, et al. Prenatal exposure to PBDEs and neurodevelopment. *Environ Health Perspect*. 2010;118(5):712-719.

12. Tanner EM, Hallerbäck MU, Wikström S, et al. *Early Prenatal Exposure to Suspected Endocrine Disruptor Mixtures Is Associated with Lower IQ at Age Seven*. Vol 134.; 2020.

13. Kim, W, Lim, Y, Kim, B, et al. Associations between prenatal and concurrent lead concentrations and IQ in preschool children in Korea. *Environmental Epidemiology*. 2019;3:201.

14. Desrochers-Couture M, Oulhote Y, Arbuckle TE, et al. Prenatal, concurrent, and sex-specific associations between blood lead concentrations and IQ in preschool Canadian children. *Environ Int*. 2018;121(Pt 2):1235-1242.

15. 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(8):856-862.

16. Loftus CT, Bush NR, Day DB, et al. Exposure to prenatal phthalate mixtures and neurodevelopment in the Conditions Affecting Neurocognitive Development and Learning in Early childhood (CANDLE) study. *Environ Int*. 2021;150:106409.

17. Sakamoto M, Tatsuta N, Izumo K, et al. Health Impacts and Biomarkers of Prenatal Exposure to Methylmercury: Lessons from Minamata, Japan. *Toxics*. 2018;6(3). doi:10.3390/toxics6030045

18. Dzwilewski KLC, Schantz SL. Prenatal chemical exposures and child language development. *J Commun Disord*. 2015;57:41-65.

19. Monosson E. Chemical mixtures: considering the evolution of toxicology and chemical assessment. *Environ Health Perspect*. 2005;113(4):383-390.

20. 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(1):A6-A9.

21. Mason CH, Perreault WD. Collinearity, Power, and Interpretation of Multiple Regression Analysis. *J Mark Res*. 1991;28(3):268-280.

22. 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(3):284-292.

23. Olvera Astivia OL, Gadermann A, Guhn M. The relationship between statistical power and predictor distribution in multilevel logistic regression: a simulation-based approach. *BMC Med Res Methodol*. 2019;19(1):97.

24. Mclean A. Machine Learning. *Canadian Journal of Nursing Informatics*. 2018;13(3/4). http://search.proquest.com/docview/2308256108/

25. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: With Applications in R*. Vol 103. Springer New York; 2013.

26. Kalloo G, Wellenius GA, McCandless L, et al. Profiles and Predictors of Environmental Chemical Mixture Exposure among Pregnant Women: The Health Outcomes and Measures of the Environment Study. *Environ Sci Technol*. 2018;52(17):10104-10113.

27. Chen H, Zhang W, Zhou Y, et al. Characteristics of exposure to multiple environmental chemicals among pregnant women in Wuhan, China. *Sci Total Environ*. 2021;754:142167.

28. 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.

29. Guo J, Wu C, Zhang J, et al. Prenatal exposure to mixture of heavy metals, pesticides and phenols and IQ in children at 7 years of age: The SMBCS study. *Environ Int*. 2020;139:105692.

30. Syakur MA, Khotimah BK, Rochman EMS, Satoto BD. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conf Ser: Mater Sci Eng*. 2018;336(1):012017.

31. Olivera-Aguilar M, Rikoon SH, Robbins SB. Using Latent Profile Analysis to Identify Noncognitive Skill Profiles Among College Students. *J Higher Educ*. 2017;88(2):234-257.

32. Bonadio FT, Tompsett C. Who Benefits from Community Mental Health Care? Using Latent Profile Analysis to Identify Differential Treatment Outcomes for Youth. *J Youth Adolesc*. 2018;47(11):2320-2336.

33. Dziak JJ, Coffman DL, Lanza ST, Li R, Jermiin LS. Sensitivity and specificity of information criteria. *Brief Bioinform*. 2020;21(2):553-565.

34. Oberski D. Mixture Models: Latent Profile and Latent Class Analysis. In: Robertson J, Kaptein M, eds. *Modern Statistical Methods for HCI*. Springer International Publishing; 2016:275-287.

35. Health Canada. Arsenic - Canada.ca. Published March 14, 2008. Accessed April 14, 2020. https://www.canada.ca/en/health-canada/services/food-nutrition/food-safety/chemical-contaminants/environmental-contaminants/arsenic.html

36. Bellinger DC. Prenatal Exposures to Environmental Chemicals and Children's Neurodevelopment: An Update. *Saf Health Work*. 2013;4(1):1-11.

37. Wasserman GA, Liu X, Parvez F, et al. Water arsenic exposure and children's intellectual function in Araihazar, Bangladesh. *Environ Health Perspect*. 2004;112(13):1329-1333.

38. Wang B, Liu J, Liu B, Liu X, Yu X. Prenatal exposure to arsenic and neurobehavioral development of newborns in China. *Environ Int*. 2018;121(Pt 1):421-427.

39. Health Canada. Guidelines for Canadian Drinking Water Quality: Guideline Technical Document – Cadmium - Canada.ca. Published September 1, 1986. Accessed April 14, 2020. https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidelines-canadian-drinking-water-quality-guideline-technical-document-cadmium.html

40. Kim Y, Ha E-H, Park H, et al. Prenatal lead and cadmium co-exposure and infant neurodevelopment at 6 months of age: the Mothers and Children's Environmental Health (MOCEH) study. *Neurotoxicology*. 2013;35:15-22.

41. Liu Z, Cai L, Liu Y, Chen W, Wang Q. Association between prenatal cadmium exposure and cognitive development of offspring: A systematic review. *Environ Pollut*. 2019;254(Pt B):113081.

42. Health Canada. Lead - Canada.ca. Published October 27, 2006. Accessed April 14, 2020. https://www.canada.ca/en/health-canada/services/chemical-substances/fact-sheets/chemicals-glance/lead.html

43. Hu H, Téllez-Rojo MM, Bellinger D, et al. Fetal lead exposure at each stage of pregnancy as a predictor of infant mental development. *Environ Health Perspect*. 2006;114(11):1730-1735.

44. Lanphear BP, Hornung R, Khoury J, et al. Low-level environmental lead exposure and children's intellectual function: an international pooled analysis. *Environ Health Perspect*. 2005;113(7):894-899.

45. Blood Lead Levels in Children | Lead | CDC. Published December 26, 2019. Accessed April 11, 2020. https://www.cdc.gov/nceh/lead/prevention/blood-lead-levels.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fnceh%2Flead%2Facclpp%2Fblood_lead_levels.htm

46. Health Canada. Guidelines for Canadian Drinking Water Quality - Manganese.; 2019. https://www.canada.ca/content/dam/hc-sc/documents/services/publications/healthy-living/guidelines-canadian-drinking-water-quality-guideline-technical-document-manganese/pub-manganese-0212-2019-eng.pdf

47. Mora AM, Arora M, Harley KG, et al. Prenatal and postnatal manganese teeth levels and neurodevelopment at 7, 9, and 10.5 years in the CHAMACOS cohort. *Environ Int*. 2015;84:39-54.

48. Chung SE, Cheong H-K, Ha E-H, et al. Maternal Blood Manganese and Early Neurodevelopment: The Mothers and Children's Environmental Health (MOCEH) Study. *Environ Health Perspect*. 2015;123(7):717-722.

49. Lin C-C, Chen Y-C, Su F-C, et al. In utero exposure to environmental lead and manganese and neurodevelopment at 2 years of age. *Environ Res*. 2013;123:52-57.

50. Leonhard MJ, Chang ET, Loccisano AE, Garry MR. A systematic literature review of epidemiologic studies of developmental manganese exposure and neurodevelopmental outcomes. *Toxicology*. 2019;420:46-65.

51. Health Canada. Mercury and Human Health - Canada.ca. Published November 3, 2004. Accessed April 14, 2020. https://www.canada.ca/en/health-canada/services/healthy-living/your-health/environment/mercury-human-health.html

52. Llop S, Guxens M, Murcia M, et al. Prenatal exposure to mercury and infant neurodevelopment in a multicenter cohort in Spain: study of potential modifiers. *Am J Epidemiol*. 2012;175(5):451-465.

53. Yorifuji T, Kato T, Kado Y, et al. Intrauterine Exposure to Methylmercury and Neurocognitive Functions: Minamata Disease. *Arch Environ Occup Health*. 2015;70(5):297-302.

54. Snoj Tratnik J, Falnoga I, Trdin A, et al. Prenatal mercury exposure, neurodevelopment and apolipoprotein E genetic polymorphism. *Environ Res*. 2017;152:375-385.

55. von Stackelberg K. PCBs. In: Nriagu JO, ed. *Encyclopedia of Environmental Health*. Elsevier; 2011:346-356.

56. Dioxins, Furans and Dioxin-Like Polychlorinated Biphenyls Factsheet. Published May 7, 2021. Accessed May 25, 2021. https://www.cdc.gov/biomonitoring/DioxinLikeChemicals_FactSheet.html

57. Goodman M, Squibb K, Youngstrom E, et al. Using systematic reviews and meta-analyses to support regulatory decision making for neurotoxicants: lessons learned from a case study of PCBs. *Cien Saude Colet*. 2011;16(7):3207-3220.

58. Jayaraj R, Megha P, Sreedev P. Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. *Interdiscip Toxicol*. 2016;9(3-4):90-100.

59. Saeedi Saravi SS, Dehpour AR. Potential role of organochlorine pesticides in the pathogenesis of neurodevelopmental, neurodegenerative, and neurobehavioral disorders: A review. *Life Sci*. 2016;145:255-264.

60. Health Canada. *Report on Human Biomonitoring of Environmental Chemicals in Canada: Results of the Canadian Health Measures Survey Cycle 1 (2007–2009)*.; 2010. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/hecs-sesc/pdf/pubs/contaminants/chms-ecms/report-rapport-eng.pdf

61. Liu P, Wu C, Chang X, Qi X, Zheng M, Zhou Z. Adverse Associations of both Prenatal and Postnatal Exposure to Organophosphorous Pesticides with Infant Neurodevelopment in an Agricultural Area of Jiangsu Province, China. *Environ Health Perspect*. 2016;124(10):1637-1643.

62. Bouchard MF, Chevrier J, Harley KG, et al. Prenatal exposure to organophosphate pesticides and IQ in 7-year-old children. *Environ Health Perspect*. 2011;119(8):1189-1195.

63. Eskenazi B, Kogut K, Huen K, et al. Organophosphate pesticide exposure, PON1, and neurodevelopment in school-age children from the CHAMACOS study. *Environ Res*. 2014;134:149-157.

64. Siddiqi MA, Laessig RH, Reed KD. Polybrominated diphenyl ethers (PBDEs): new pollutants-old diseases. *Clin Med Res*. 2003;1(4):281-290.

65. Health Canada. Polybrominated diphenyl ethers (PBDEs) - information sheet - Canada.ca. Published March 12, 2017. Accessed April 13, 2020. https://www.canada.ca/en/health-canada/services/chemical-substances/fact-sheets/chemicals-glance/polybrominated-diphenyl-ethers-public-summary.html

66. Gibson EA, Siegel EL, Eniola F, Herbstman JB, Factor-Litvak P. Effects of Polybrominated Diphenyl Ethers on Child Cognitive, Behavioral, and Motor Development. *Int J Environ Res Public Health*. 2018;15(8). doi:10.3390/ijerph15081636

67. Azar N, Booij L, Muckle G, et al. Prenatal exposure to polybrominated diphenyl ethers (PBDEs) and cognitive ability in early childhood. *Environ Int*. 2021;146:106296.

68. Health Canada. Phthalates - Canada.ca. Published October 6, 2017. Accessed April 13, 2020. https://www.canada.ca/en/health-canada/services/chemicals-product-safety/phthalates.html

69. Kim Y, Ha E-H, Kim E-J, et al. Prenatal exposure to phthalates and infant development at 6 months: prospective Mothers and Children's Environmental Health (MOCEH) study. *Environ Health Perspect*. 2011;119(10):1495-1500.

70. Téllez-Rojo MM, Cantoral A, Cantonwine DE, et al. Prenatal urinary phthalate metabolites levels and neurodevelopment in children at two and three years of age. *Sci Total Environ*. 2013;461-462:386-390.

71. Polanska K, Krol A, Merecz-Kot D, et al. Environmental Tobacco Smoke Exposure during Pregnancy and Child Neurodevelopment. *Int J Environ Res Public Health*. 2017;14(7). doi:10.3390/ijerph14070796

72. Wehby GL, Prater K, McCarthy AM, Castilla EE, Murray JC. The Impact of Maternal Smoking during Pregnancy on Early Child Neurodevelopment. *J Hum Cap*. 2011;5(2):207-254.

73. Park S, Cho S-C, Hong Y-C, et al. Environmental tobacco smoke exposure and children's intelligence at 8-11 years of age. *Environ Health Perspect*. 2014;122(10):1123-1128.

74. Niemelä S, Sourander A, Surcel H-M, et al. Prenatal Nicotine Exposure and Risk of Schizophrenia Among Offspring in a National Birth Cohort. *Am J Psychiatry*. 2016;173(8):799-806.

75. Coker E, Gunier R, Bradman A, et al. Association between Pesticide Profiles Used on Agricultural Fields near Maternal Residences during Pregnancy and IQ at Age 7 Years. *Int J Environ Res Public Health*. 2017;14(5). doi:10.3390/ijerph14050506

76. Carroll R, White AJ, Keil AP, et al. Latent classes for chemical mixtures analyses in epidemiology: an example using phthalate and phenol exposure biomarkers in pregnant women. *J Expo Sci Environ Epidemiol*. Published online October 21, 2019. doi:10.1038/s41370-019-0181-y

77. Khorrami Z, Pourkhosravani M, Rezapour M, et al. Multiple air pollutant exposure and lung cancer in Tehran, Iran. *Sci Rep*. 2021;11(1):9239.

78. Arbuckle TE, Fraser WD, Fisher M, et al. Cohort profile: the maternal-infant research on environmental chemicals research platform. *Paediatr Perinat Epidemiol*. 2013;27(4):415-425.

79. Follow-up Studies. Accessed January 17, 2021. https://www.mirec-canada.ca/en/about/follow-up-studies/

80. Braun JM, Muckle G, Arbuckle T, et al. Associations of Prenatal Urinary Bisphenol A Concentrations with Child Behaviors and Cognitive Abilities. *Environ Health Perspect*. 2017;125(6):067008.

81. Gordon B. Test Review: Wechsler, D. (2002). The Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III). San Antonio, TX: The Psychological Corporation. *Canadian Journal of School Psychology*. 2004;19(1-2):205-220.

82. Flanagan DP, Alfonso VC. A Critical Review of the Technical Characteristics of New and Recently Revised Intelligence Tests for Preschool Children. *J Psychoeduc Assess*. 1995;13(1):66-90.

83. Wechsler Preschool and Primary Scale of Intelligence. Accessed January 17, 2021. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Preschool-and-Primary-Scale-of-Intelligence-%7C-Third-Edition/p/100000422.html?tab=product-details

84. Early Childhood Measurement and Evaluation. https://www.ualberta.ca/community-university-partnership/media-library/community-university-partnership/resources/tools---assessment/wppsi-iiimay-2012.pdf

# Chapter 2.

# Manuscript

# Using Latent Profile Analysis to Examine Associations Between Gestational Chemical Mixtures and Child Neurodevelopment

## Authorship

Amanda M Yonkman[a], Josh D Alampi[a], Angela Kaida[a], Ryan W Allen[a], Aimin Chen[b], Bruce P Lanphear[a], Joseph M Braun[c], Tye E Arbuckle[d], and Lawrence C McCandless[a]

## Affiliations

[a]Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada
[b]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[c]Department of Epidemiology, Brown University, Providence, Rhode Island, USA
[d]Environmental Health Science and Research Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, Canada

## 2.1. Introduction

The gestational period is a crucial time during which neurotoxic exposures can severely impact cognitive development[1–4]. Although much research has been done on the effects of gestational neurotoxins, most studies have only focused on one chemical at a time[1,3,5–9]. Researchers have used simple regression analysis to study many individual exposures, including lead, bisphenol A (BPA), and dichlorodiphenyldichloroethylene (DDE)[1,3,10–13]. However, comparatively little study has been done on the effects of chemical mixtures[4]. The few studies that do focus on mixtures usually restrict their models to chemicals in the same groups, such as heavy metals or polybrominated diphenyl ethers (PBDEs)[14,15]. This leads to several problems. The first is that these studies are not reflective of reality; we are not exposed to only one chemical at a time, but instead encounter a complex mixture of chemicals from different classes every day[16]. Second, if there are high levels of correlation between exposure variables, studying these chemicals individually can result in collinearity and large standard errors[4,17,18]. Third, researchers may underestimate the collective impact of a group of chemicals if each exposure only has a modest effect on neurodevelopment. Finally, the impact of mixtures may be different from the additive exposure effects, and these interactions are difficult to estimate if chemicals are studied individually[19].

The gap in research on chemical mixtures is due in part to a lack of appropriate statistical methods[2,5,9,18]. While simple statistical techniques such as regression analysis can show the effects of a few chemicals at a time, these methods are poorly equipped to handle a large number of correlated continuous variables[20]. A few unsupervised machine learning techniques have been used to study mixtures. For example, k-means clustering is commonly used to study many variables at once. However, this is a non-probabilistic method that requires the researcher to predetermine the number of clusters, and it is difficult to assess classification accuracy[21]. For this reason, we propose that an unsupervised machine learning technique called Latent Profile Analysis (LPA) be used to study chemical mixtures.

LPA is a model-based technique that detects patterns in continuous independent variables[22]. Although this method is popular in psychology and behavioural sciences, it is much less common in the field of environmental epidemiology[22–26]. The purpose of LPA is to use patterns detected by machine learning to create homogenous, probabilistic subgroups called profiles[22]. Computer algorithms generate a variable called the posterior probability, which is the likelihood that a data point will fall into each profile[20]. LPA has several advantages over other

machine learning techniques. It can handle dozens of variables without losing power, making it more effective than linear regression for studying mixtures[20,22]. Its probabilistic nature allows researchers to assess classification accuracy, providing more accurate and nuanced results than non-probabilistic clustering methods such as k-means clustering[22–24]. Finally, LPA does not require the researcher to predetermine the number of profiles and has more rigorous methods of choosing a model than other unsupervised machine learning techniques[20,24,25].

The aim of this study is to introduce LPA to the field of environmental epidemiology as a tool for studying the effects of complex chemical mixtures on child neurodevelopment. We believe that the advantages of LPA over other more common statistical techniques make it well-suited for this task. In this study, we use LPA to create profiles of pregnant Canadian women based on 30 gestational chemical biomarkers, and then examine associations between profile membership and child neurodevelopment. We also conduct sensitivity analysis, comparing the results found using LPA to those found using k-means clustering.

## 2.2. Methods

### *2.2.1. Study Population*

We used data from the Maternal-Infant Research on Environmental Chemicals (MIREC) Study, an ongoing Canadian birth cohort study that began in 2008[27–29]. The primary objective of the MIREC Study is to examine the role of gestational exposures, measured using biomarkers in maternal blood, urine, and hair, in maternal and child health outcomes. Details about eligibility and exclusion criteria are outlined by Arbuckle et al.[28] Briefly, participants were recruited as follows: Researchers approached 8716 adult pregnant women in 10 Canadian cities, of whom 2001 were eligible and gave their consent in either English or French (Figure B.1.). Participants had to be at least 18 years old, <14 weeks of gestation, and willing to provide cord blood samples and deliver at a local hospital. Exclusion criteria included known fetal abnormalities, a history of medical complications, or illicit drug use. At age 36-48 months, a subset of 610 participating children were included in a follow-up study called the MIREC-CD3 Neurodevelopment visit, in which researchers assessed neurodevelopment through a series of tests and questionnaires[3,30–32]. We restricted our study to the 517 participants with available measures for all first trimester chemicals and neurodevelopment scores of interest[29,31,32].

The MIREC Study received ethics approval from Health Canada and the Institutional Review Board of CHU Sainte-Justine Research Centre. For this project, we also received ethics approval from the Simon Fraser University review board. All participants gave informed consent to take part in this study.

### *2.2.2. Neurodevelopmental Outcomes*

We measured neurodevelopment using the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) test, which was administered during the MIREC-CD3 Neurodevelopment visit when the participants were 36-48 months old[1,31,32]. The WPPSI-III test is designed to assess various aspects of child intelligence, and includes four subtests. Subtest scores are calculated based on time and accuracy and then combined to generate the three main IQ scores, as described in the WPPSI-III Administration and Scoring Manual[33–35]. Scores from the Receptive Vocabulary and Information subtests are combined to generate the Verbal IQ (VIQ) score, which measures linguistic intelligence[34]. Scores from the Block Design and Object Assembly subtests are combined to generate the Performance IQ (PIQ) score, which measures visuospatial intelligence. The overall summary measure of cognitive performance is

called the Full-Scale IQ (FSIQ), which is a composite score calculated from the VIQ and PIQ. All three IQ scores are scaled to a standardized Canadian sample with a mean of 100 and a standard deviation of 15.

### *2.2.3. Biomarkers of Gestational Toxicant Exposure*

In our models, we included 30 potential neurotoxins measured in the first trimester: Five heavy metals (arsenic, cadmium, lead, manganese, and mercury), four organochlorine pesticides [OCPs; β-benzene hexachloride (BBHC), dichlorodiphenyldichloroethylene (DDE), oxychlordane, and trans-nonachlor], five organophosphate pesticides [OPPs; diethylphosphate (DEP), diethylthiophosphate, (DETP), dimethyldithiophosphate (DMDTP), dimethylphosphate (DMP), and dimethylthiophosphate (DMTP)], seven phthalates [monobutyl phthalate (MBP), monobenzyl phthalate (MBZP), mono-(3-carboxypropyl) phthalate (MCPP), mono-(2-ethyl-5-hydroxyhexyl) phthalate (MEHHP), mono-(2-ethylhexyl) phthalate (MEHP), mono-(2-ethyl-5-oxohexyl) phthalate (MEOHP), and mono-ethyl phthalate (MEP)], seven polychlorinated biphenyls (PCBs; Aroclor, PCB118, PCB138, PCB153, PCB170, PCB180, and PCB187), one polybrominated diphenyl ether (PBDE47), and one smoking metabolite (cotinine). These chemicals were chosen based on biomarker availability in the first trimester, as well as previous research on their neurotoxicity[1,36–44]. The metals, OCPs, PCBs, PBDE, and cotinine were measured in the maternal blood, and the OPPs and phthalates were measured in maternal urine[28]. Samples were stored at -20°C until analysis in the Toxicology Laboratory at the Institut national de santé publique du Québec, where they were quantified using gas chromatography/mass spectrometry.

To account for the effects of plasma-lipid levels, we adjusted the OCPs, PCBs, and PBDE for participants' plasma-lipid concentrations[18,45]. We accounted for variation in urine dilution in the OPPs and phthalates by adjusting for specific gravity using the equation:

$$P_c = P[\frac{1.015 - 1}{SG} - 1]$$

where $P_c$ is the standardized chemical concentration, $P$ is the unstandardized chemical concentration, $SG$ is the specific gravity, and 1.015 is the median specific gravity among the study participants[46].

For chemicals below the limit of detection (<LOD), we employed a single imputation "fill-in" method[47]. We first log2-transformed our chemicals to reduce skewness, and then temporarily replaced the values <LOD with LOD/√2. Using these new values and the observed chemical concentrations, we determined the mean and standard deviation of a truncated lognormal distribution. The values <LOD were then replaced with values randomly sampled from this distribution.

### *2.2.4. Covariates*

Trained researchers assessed potential confounders by administering standardized interviews and questionnaires to the mothers during the first and third trimesters[3,30,48]. To determine the appropriate covariates, we constructed a Directed Acyclic Graph (DAG) using information from previous studies (Figure B.2.)[10,49–51]. We adjusted for maternal age, race, education, marital status, household income, and prenatal alcohol consumption. We did not adjust for self-reported prenatal smoking because cotinine was already a factor in profile generation. Although child sex cannot be considered a confounder, we stratified results by sex because studies have found differing effects of gestational exposures on neurodevelopment in boys and girls[1,3,8]. Gestational age and birth weight were excluded from the model because they act as mediators, not confounders[52].

Measures for several of our covariates were missing in a small number of participants. Prenatal alcohol was missing for 4.3% of mothers, household income for 3.3%, and maternal education for 0.5%, which would have resulted in a total of 40 mothers (7.7%) being removed from the analysis. To avoid this, we used a single imputation approach to fill in these missing values.

### *2.2.5. Statistical Analysis*

*Descriptive Statistics*

We assessed the central tendency and distribution of the gestational biomarker concentrations and the children's VIQ, PIQ, and FSIQ scores. We stratified the WPPSI-III scores by our chosen covariates to compare between demographics. To calibrate our inferences about associations between demographic characteristics, chemical mixtures, and neurodevelopment, we used linear regression analyses. We examined the relation between demographic characteristics and WPPSI-III scores, as well as the relation between the individual log2-transformed chemical concentrations and WPPSI-III scores.

*Latent Profile Analysis*

We used LPA to create chemical mixture profiles based on 30 chemical biomarkers. To ensure that the biomarkers had adequate levels of correlation for profile interpretation, we conducted pairwise correlation analysis on the 30 individual chemical biomarkers. We then performed LPA with the RStudio packages tidyLPA and mclust[53]. To choose the number of profiles, we calculated the Bayesian Information Criterion (BIC) of 35 different models, manipulating the number of profiles from one to twelve, as well as assumptions about the variance within profiles and the covariance between chemical exposures. We chose the BIC as a quality measure because it penalizes model complexity more than other measures such as the Akaike Information Criterion (AIC)[54]. In addition to the BIC, we also considered interpretability when choosing a model.

Once we had chosen the number of profiles, we generated the mothers' posterior probabilities of profile membership. We then examined the demographic characteristics of participants in each of the five profiles. Additionally, we calculated the mean chemical concentration in each profile using the formula:

$$W = \frac{\sum P_i W_i}{\sum P_i}$$

where $X_i$ is the log2-transformed chemical concentration in mother *i* and $P_i$ is the posterior probability of profile membership in mother *i*. We converted biomarker concentrations to z-scores to compare the profiles' overall chemical compositions and determine the patterns detected by LPA.

*Regression Analysis of the Latent Profiles*

We used covariate-adjusted multiple linear regression analysis to measure the association between profile membership and WPPSI-III scores, running separate regression models for VIQ, PIQ, and FSIQ. Each model included the posterior probabilities of profile membership for every profile except the reference, as shown in the following equation:

$$Y = \beta_0 + \beta_2 Z_2 + \cdots + \beta_k Z_k + \beta_{c1} C_1 + \cdots + \beta_{cp} C_p \qquad \text{[Equation 1]}$$

where $Y$ is the WPPSI-III score, $Z_2 \ldots Z_k$ are the posterior probabilities that a mother will fall into each of the $k$ profiles, $Z_1$ is the reference profile that is excluded from the model, and $C_1 \ldots C_p$

are *p* confounders. For example, the quantity $\beta_3$ would be the change in mean IQ score as the posterior probability of membership in Profile 3 increases from 0 to 1, implying a 100% probability of inclusion in Profile 3 compared to Profile 1, adjusted for *p* confounders.

*Sensitivity Analysis Comparing LPA with K-means Clustering*

We conducted sensitivity analysis to compare LPA with k-means clustering, a more established method in the field of environmental epidemiology. Unlike LPA, k-means clustering requires that the researcher predetermine the number of clusters[55,56]. We first ran several models and used the "elbow method" to help us make this decision. For interpretation, we generated a heat map of mean biomarker concentration z-scores and compared them with the LPA results.

Additionally, we used covariate-adjusted multiple linear regression analysis to measure the associations between class membership and WPPSI-III scores, once again running separate models for VIQ, PIQ, and FSIQ. We created dummy variables for each of the clusters (with 1 denoting membership to a cluster and 0 denoting non-membership). Each model included all dummy variables except for the reference cluster, as shown in the following equation:

$$Y = \beta_0 + \beta_2 D_2 + \cdots + \beta_k D_k + \beta_{c1} C_1 + \cdots + \beta_{cp} C_p \qquad \text{[Equation 2]}$$

where *y* is the WPPSI-III score, $D_2 \ldots D_k$ are the dummy variables for each of the k clusters, $D_1$ is the reference cluster that is excluded from the model, and $C_1 \ldots C_p$ are *p* confounders. For example, the quantity $\beta_3$ would be the difference in the mean IQ scores of a participant in Cluster 3 compared to a participant in Cluster 1, adjusted for *p* confounders.

## 2.3. Results

### *2.3.1. Study Population and Descriptive Statistics*

The MIREC-CD3 Neurodevelopment visit included 808 participants. 608 children took the WPPSI-III test, and 599 of those finished all subtests and received complete scores. After excluding those with missing chemical information, we included 517 MIREC participants in our study (Figure B.1.). Compared to the average Canadian mother in 2011, participants' mothers were on average older (41% in the 35+ group), predominantly white (86%), with higher education levels (68% completed an undergraduate degree or higher) and household incomes (41% over $100, 000), and lower self-reported prenatal smoking (9%) and alcohol (17%) rates[57]. Parity was similar to the Canadian average (44% on their first child).

In covariate-adjusted regression models of participant characteristics and WPPSI-III scores, we found higher average scores in female children, as well as children with white, more educated, or low-parity mothers (Table B.1.). Prenatal smoking was negatively associated with all three IQ scores, but prenatal alcohol was positively associated with IQ. Maternal age and marital status were linked to higher VIQ but lower PIQ; however, the associations were imprecise with wide 95% confidence intervals.

We calculated the geometric means of each chemical and found them to be similar to those found by Canadian Health Measures Survey (CHMS) at the time (Table 2.1.)[58]. The %>LOD was higher (>80%) for the heavy metals, phthalates, and most of the OCPs and PCBs, and lower for the OPPs, PBDE, and cotinine. We excluded all chemicals with less than 40% >LOD.

Table 2.1. Distribution of gestational chemical biomarkers in participating mothers in the MIREC study during their first trimester of pregnancy, as compared to the geometric mean concentrations found in the average Canadian mother in 2008, measured in the Canadian Health Measures Survey (n = 517).

| | % >LOD | Mean (SD) | GM (GSD) | Percentiles | | | | | CHMS GM[58] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 25th | 50th | 75th | 95th | Max | |
| **Heavy metals - Whole Blood (ug/L)** | | | | | | | | | |
| Arsenic | 96.5 | 1.1 (1.8) | 0.8 (2.0) | 0.6 | 0.8 | 1.2 | 2.3 | 34.5 | 0.9 |
| Cadmium | 97.5 | 0.3 (0.4) | 0.2 (2.1) | 0.1 | 0.2 | 0.3 | 0.7 | 5.1 | 0.4 |
| Lead | 100.0 | 7.2 (4.0) | 6.4 (1.6) | 4.6 | 6.2 | 8.5 | 14.1 | 41.4 | 8.9 |
| Manganese | 100.0 | 9.1 (2.8) | 8.7 (1.4) | 7.1 | 8.8 | 10.4 | 13.7 | 26.9 | 9.8 |
| Mercury | 91.5 | 1.0 (0.9) | 0.6 (2.7) | 0.4 | 0.7 | 1.3 | 2.8 | 7.8 | 0.7 |
| **OCPs - Plasma (ng/L)** | | | | | | | | | |
| BBHC | 64.0 | 6.3 (30.1) | 2.3 (2.7) | < LOD | 2.1 | 3.4 | 9.0 | 500.0 | 4.8 |
| DDE | 99.2 | 90.9 (212.6) | 55.23 (2.2) | 35.7 | 49.1 | 77.1 | 214.6 | 2,656.3 | 102.2 |
| Oxychlor | 92.8 | 2.3 (1.2) | 2.0 (1.8) | 1.5 | 2.2 | 3.0 | 4.5 | 8.4 | 2.3 |
| Transnona | 85.7 | 3.4 (2.3) | 2.9 (1.8) | 2.0 | 3.0 | 4.3 | 7.4 | 18.3 | 3.1 |
| OCP Sum | NA | 5.8 (3.3) | 5.0 (1.7) | 3.5 | 5.1 | 7.2 | 11.7 | 23.9 | NA |
| **OPPs - Urine (ug/L)** | | | | | | | | | |
| DEP | 73.1 | 7.5 (92.5) | 2.6 (2.3) | < LOD | 2.5 | 4.2 | 9.8 | 2,104.8 | 2.0 |
| DETP | 49.1 | 1.1 (1.3) | 0.7 (2.5) | < LOD | < LOD | 1.2 | 3.1 | 15.6 | NA |
| DMDTP | 53.0 | 1.2 (2.3) | 0.5 (3.6) | < LOD | 0.5 | 1.1 | 4.8 | 22.5 | NA |
| DMP | 77.4 | 5.1 (6.5) | 3.2 (2.7) | 1.8 | 3.3 | 6.2 | 14.7 | 71.5 | 2.6 |
| DMTP | 80.7 | 8.1 (11.9) | 3.6 (4.0) | 1.4 | 3.9 | 8.9 | 30.6 | 96.2 | 1.8 |
| **Phthalates - Urine (ug/L)** | | | | | | | | | |
| MBP | 99.6 | 19.8 (37.5) | 12.6 (2.3) | 7.8 | 12.1 | 19.2 | 47.1 | 525.9 | 18.0 |
| MBZP | 99.4 | 8.9 (14.9) | 5.4 (2.5) | 3.1 | 4.8 | 9.1 | 25.3 | 182.0 | 9.3 |
| MCPP | 80.3 | 2.0 (4.7) | 0.9 (3.3) | 0.5 | 0.9 | 1.7 | 7.2 | 72.0 | 1.1 |
| MEHHP | 99.4 | 13.9 (24.2) | 9.5 (2.2) | 6.1 | 9.1 | 14.1 | 34.7 | 355.8 | 20.0 |
| MEHP | 98.1 | 3.3 (4.6) | 2.3 (2.2) | 1.4 | 2.2 | 3.7 | 9.1 | 53.0 | 3.4 |
| MEOHP | 99.4 | 9.2 (13.1) | 6.7 (2.1) | 4.4 | 6.5 | 9.5 | 22.7 | 171.1 | 13.0 |
| MEP | 100.0 | 149.6 (969.3) | 33.2 (4.0) | 12.4 | 26.0 | 71.5 | 416.0 | 20,800.0 | 50.0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| DEHP Sum | NA | 26.1 (40.6) | 18.6 (2.1) | 12.5 | 18.1 | 26.4 | 63.8 | 550.6 | NA |
| **PCBs - Plasma (ng/L)** | | | | | | | | | |
| Aroclor | 98.7 | 84.6 (79.7) | 65.0 (2.0) | 41.1 | 64.2 | 94.6 | 225.1 | 659.6 | NA |
| PCB118 | 77.0 | 3.0 (2.5) | 2.4 (1.9) | 1.7 | 2.5 | 3.4 | 6.6 | 30.2 | 3.1 |
| PCB138 | 94.2 | 5.8 (5.4) | 4.5 (2.0) | 2.9 | 4.5 | 6.6 | 15.1 | 46.8 | 5.5 |
| PCB153 | 99.8 | 10.5 (10.1) | 8.0 (2.0) | 5.0 | 7.6 | 11.8 | 27.9 | 80.9 | 8.2 |
| PCB170 | 58.8 | 2.9 (3.7) | 2.0 (2.2) | < LOD | 2.0 | 3.1 | 8.1 | 40.3 | NA |
| PCB180 | 96.7 | 7.8 (10.1) | 5.5 (2.2) | 3.2 | 5.3 | 8.4 | 21.2 | 114.9 | 5.8 |
| PCB187 | 47.2 | 2.3 (2.4) | 1.7 (2.0) | < LOD | < LOD | 2.5 | 5.7 | 26.9 | NA |
| PCB Sum | NA | 30.0 (29.3) | 23.2 (2.0) | 14.3 | 22.0 | 34.3 | 78.3 | 262.8 | NA |
| **PBDEs - Plasma (ng/L)** | | | | | | | | | |
| BDE47 | 65.0 | 15.0 (45.2) | 7.3 (2.7) | < LOD | 6.8 | 11.7 | 38.1 | 727.3 | 10.8 |
| **Tobacco Metabolites - Plasma (ng/L)** | | | | | | | | | |
| Cotinine | 54.4 | 2871.4 (17325.1) | 7.7 (11.8) | < LOD | 6.5 | 20.0 | 270.0 | 180,000.0 | NA |

### *2.3.2. Latent Profile Analysis*

We created a heat map showing the correlation between 30 log2-transformed biomarkers (Figure 2.1.). We found high levels of correlation within the OCPs, the OPPs, the phthalates (except for MEP), and the PCBs. There was also correlation between several chemical groups. The PCBs, OCPs, and most of the metals were correlated, as were cotinine and cadmium, a component of cigarette smoke[59].

We generated 35 models, changing the number of profiles and the assumptions for variance and covariance to attain the lowest BIC. The highest quality models included a model with five profiles and equal variance and covariance (BIC = 40697) and a model with two profiles and varying variance and covariance (BIC = 40405). Accounting for interpretability, we chose the model with five profiles.

Figure 2.2. shows the LPA results. We found one profile with mean biomarker concentrations similar to those of the total group (the Reference profile), one with high biomarker concentrations (the High Level profile), one with low biomarker concentrations (the Low Level profile), one with high OPPs and low levels of every other chemical (the High OPPs profile), and one with high levels of cotinine and cadmium (the Smoking Chemicals profile). Certain chemicals had greater levels of variation between the profiles than others (Figure 2.2.). Specifically, there was high variance of PCB and smoking chemical concentrations between profiles.

Table 2.2. shows the demographic characteristics of participants in each of the five profiles. For this table, participants were assigned to whichever profile they matched most closely based on their posterior probabilities. We found that the Reference group was comparatively larger (n = 365) and the High Level, Low Level, High OPPs, and Smoking Chemicals profiles were comparatively smaller (n = 33, 20, 79, and 20 respectively). We also examined the spread of the posterior probabilities between profiles to assess classification accuracy. The mothers in the three smaller profiles tended to have high posterior probabilities (mean >0.99), as did those in the Reference profile (mean = 0.97), whereas mothers in the High OPPs profile had lower posterior probabilities (mean = 0.88). The demographic characteristics of the Reference profile were similar to those of the total sample, although mothers were more often white with higher socioeconomic status. The High Level profile had children with much older, non-white mothers. Both the High OPPs profile and the Low Level profile had younger, unmarried mothers with lower socioeconomic status, and the Low Level profile contained mostly girls (80%). The Smoking Chemicals profile tended to have younger, non-white, unmarried

mothers with much lower socioeconomic status and high levels of prenatal smoking, although it should be noted that not every mother in this profile reported prenatal smoking.

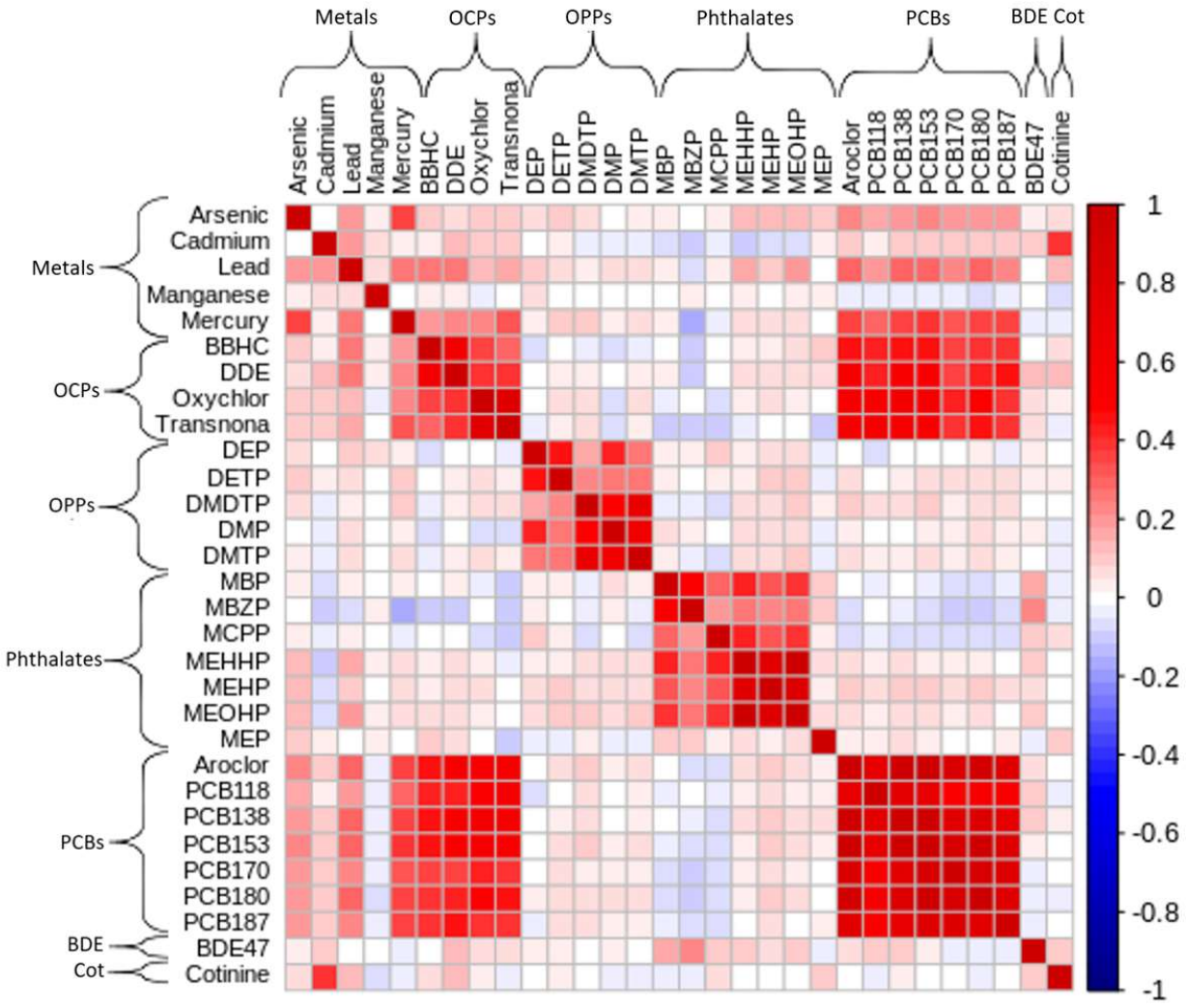


Figure 2.1. Correlation heat map of 30 log2-transformed chemicals measured in participating mothers in the MIREC study during their first trimester of pregnancy (n = 517).

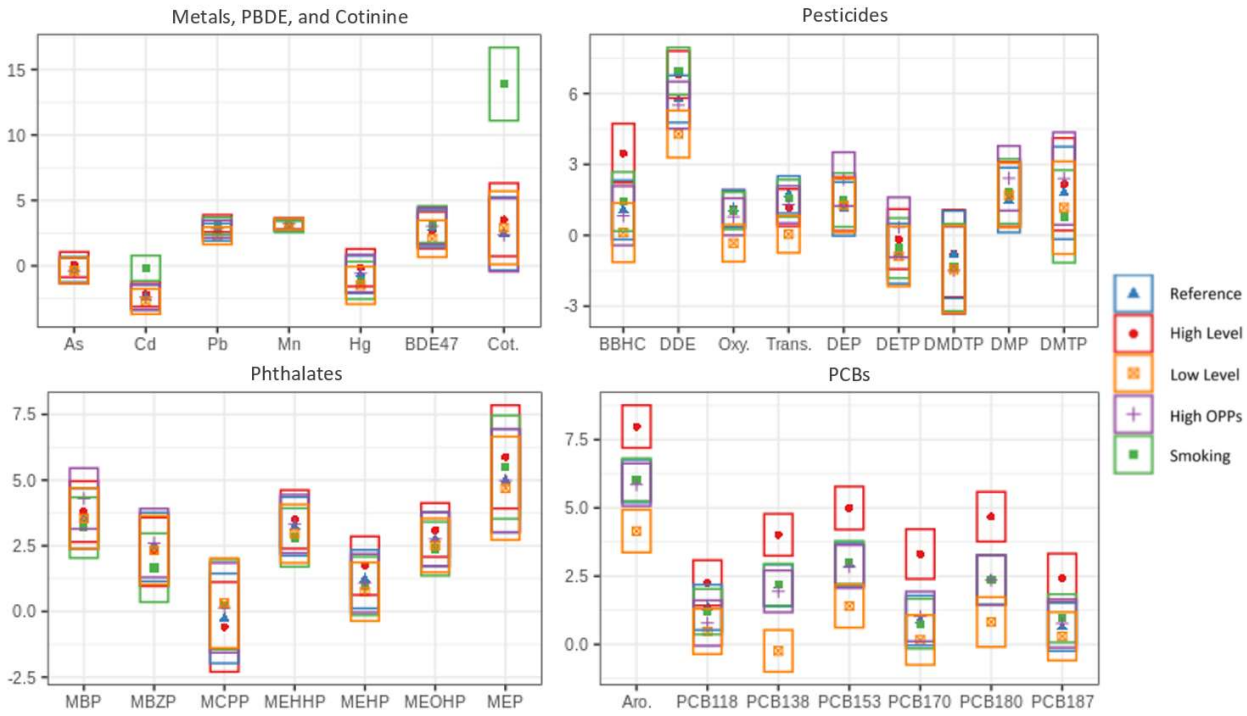


Figure 2.2. Mean log2-transformed chemical compositions of the five profiles generated by Latent Profile Analysis, with boxes showing standard deviation (n = 517).

Table 2.2. Demographic characteristics for total study population and for participants in each latent profile. Participants were assigned in whichever profile they matched most closely using the highest posterior probability.

| | Total | Reference | High Level | Low Level | High OPPs | Smoking Chemicals |
|---|---|---|---|---|---|---|
| | *n = 517* | *n = 365* | *n = 33* | *n = 20* | *n = 79* | *n = 20* |
| **Child Sex** | | | | | | |
| Male | 49 | 49 | 48 | 20 | 54 | 50 |
| Female | 51 | 51 | 52 | 80 | 46 | 50 |
| **Maternal Age** | | | | | | |
| 19-30 | 21 | 17 | 12 | 50 | 29 | 40 |
| 30-35 | 39 | 41 | 30 | 30 | 34 | 30 |
| 35+ | 41 | 42 | 58 | 20 | 37 | 30 |
| **Maternal Race** | | | | | | |
| White | 86 | 90 | 73 | 85 | 78 | 70 |
| Other | 14 | 10 | 27 | 15 | 22 | 30 |
| **Maternal Education** | | | | | | |
| Highschool | 5 | 3 | 6 | 15 | 10 | 20 |
| College | 27 | 25 | 24 | 25 | 37 | 45 |
| Undergrad | 39 | 40 | 42 | 60 | 29 | 30 |
| Grad | 29 | 33 | 27 | 0 | 24 | 5 |
| **Marital Status** | | | | | | |
| Married | 72 | 75 | 70 | 60 | 68 | 55 |
| Unmarried | 28 | 25 | 30 | 40 | 32 | 45 |
| **Household Income** | | | | | | |
| < 40 000 | 10 | 6 | 9 | 20 | 17 | 45 |
| 40 000 - 80 000 | 29 | 27 | 33 | 30 | 34 | 25 |
| 80 000 - 100 000 | 21 | 22 | 12 | 30 | 20 | 10 |
| > 100 000 | 41 | 45 | 46 | 20 | 29 | 20 |
| **Parity** | | | | | | |
| 0 | 44 | 45 | 42 | 45 | 37 | 50 |
| 1 | 41 | 41 | 46 | 20 | 47 | 35 |
| 2 | 12 | 11 | 12 | 15 | 13 | 15 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3+ | 3 | 3 | 0 | 20 | 4 | 0 |
| **Prenatal Smoking** | | | | | | |
| No | 91 | 94 | 91 | 100 | 95 | 30 |
| Yes | 9 | 6 | 9 | 0 | 5 | 70 |
| **Prenatal Alcohol** | | | | | | |
| No | 83 | 82 | 82 | 95 | 87 | 75 |
| Yes | 17 | 18 | 18 | 5 | 13 | 25 |

### *2.3.3. Regression Analysis of the Latent Profiles*

We conducted covariate-adjusted regression analysis (as shown in Equation 1) to measure the association between the posterior probabilities and the three WPPSI-III scores (Table 2.3.). We used the Reference Profile as the reference in our analysis, as this profile was the most representative of the average mother. The Smoking Chemicals profile had a strong negative association with VIQ, PIQ, and FSIQ in both boys and girls, although there was uncertainty in the effect estimates with wide 95% confidence intervals. In the High Level profile, we found a negative association with VIQ in both boys and girls; however, the results for PIQ and FSIQ were inconsistent. In the Low Level and High OPPs profiles, we found a positive association with all three IQ scores in girls but not boys. For all regression coefficients the 95% confidence intervals covered the null value.

Notably, the beta coefficients found using group membership were higher than those of the individual chemical biomarkers or survey questions. For example, in the Smoking Chemicals Profile, the mean FSIQ for all children was 5.7 points [95% CI (11.0, -0.4)] lower than that of the Reference Profile (Table 2.3.). In contrast, the self-reported prenatal smoking variable only showed a 3.8 point [95% CI (-0.2, 7.9)] decrease of FSIQ in all children compared to those with non-smoking mothers (Table B.1.). Furthermore, doubling the cotinine biomarker resulted in only a 0.2 point [95% CI (-0.2, 0.5)] decrease for FSIQ in all children (Table B.2.). Since the log2-cotinine levels in the Smoking Chemicals profile are roughly 12 greater than those of the Reference profile, this means that cotinine by itself would only account for a 2.4 point decrease in FSIQ (Table B.2.; Figure 2.2.). Effect estimates also differed from the additive effects of each of the chemicals. For example, the coefficients in the High Level profile were higher than those of the individual PCBs, but lower than the combined effects of all the PCBs.

Table 2.3. Covariate-adjusted linear regression coefficients showing the associations between latent profile membership and WPPSI-III scores, adjusted for maternal age, race, education, and marital status, household income, parity, and prenatal alcohol, compared to medium level Reference Profile, with 95% confidence intervals. Results are shown for all children, and then stratified by sex (n = 517).

| | VIQ (95% CI) | PIQ (95% CI) | FSIQ (95% CI) |
|---|---|---|---|
| **Intercept (Reference Level Profile)** | | | |
| All | 105.2 | 97.2 | 101.4 |
| Boys | 104.6 | 99.7 | 102.7 |
| Girls | 106.6 | 94.7 | 100.4 |
| **High Level Profile** | | | |
| All | -3.1 (-7.6, 1.3) | 2.2 (-3.1, 7.5) | -0.5 (-5.1, 4.2) |
| Boys | -4.4 (-11.3, 2.6) | -0.7 (-8.7, 7.4) | -2.8 (-10.2, 4.5) |
| Girls | -2.5 (-8.2, 3.1) | 2.9 (-4.3, 10.0) | 0.3 (-5.7, 6.2) |
| **Low Level Profile** | | | |
| All | 0.5 (-5.3, 6.4) | 4.7 (-2.2, 11.7) | 3.0 (-3.1, 9.1) |
| Boys | -3.2 (-17.0, 10.5) | -0.2 (-16.1, 15.7) | -2.2 (-16.8, 12.3) |
| Girls | 0.3 (-5.9, 6.6) | 6.1 (-1.9, 14.0) | 3.6 (-2.9, 10.2) |
| **High OPPs Profile** | | | |
| All | 1.1 (-2.4, 4.5) | -0.3 (-4.4, 3.8) | 0.6 (-3.1, 4.2) |
| Boys | -2.3 (-7.6, 3.0) | -2.0 (-8.2, 4.1) | -2.4 (-8.0, 3.2) |
| Girls | 3.4 (-1.1, 8.0) | 0.3 (-5.5, 6.1) | 2.3 (-2.5, 7.1) |
| **Smoking Chemicals Profile** | | | |
| All | -3.7 (-9.5, 2.2) | -6.4 (-13.3, 0.6) | -5.7 (-11.9, 0.4) |
| Boys | -4.7 (-13.8, 4.3) | -7.8 (-18.3, 2.7) | -7.1 (-16.7, 2.5) |
| Girls | -2.7 (-10.2, 4.8) | -6.6 (-16.2, 2.9) | -5.3 (-13.2, 2.6) |

### *2.3.4. Sensitivity Analysis Comparing LPA and K-means Clustering*

The “elbow method” of determining the number of clusters showed that 5-8 clusters were appropriate for this study population. We chose a model with five clusters generated by k-means clustering to compare with the LPA results. The clusters were similar to the profiles generated by LPA; Figure 2.3. shows a cluster with low to middling biomarker concentrations, one with high biomarker concentrations, one with low biomarker concentrations, one with high OPPs, and one with high concentrations of smoking chemicals. While these patterns were similar to the ones detected by LPA, there were some key differences between the clusters and the profiles. First, the Reference, High Level, Low Level, and High OPPs clusters were all of similar sizes (n = 119, 111, 131, 138 respectively), although the Smoking Chemicals cluster was comparatively small (n = 18; Table B.3.). Second, the range of mean z-scores was smaller in some of the clusters, showing weaker patterns in the High Level and Low Level clusters (Figure 2.3.). Finally, the Reference cluster had lower mean biomarker concentrations and was less representative of the overall population than the Reference profile.

When we conducted covariate-adjusted linear regression analysis shown in Equation 2, we found that the Smoking Chemicals cluster had a consistent negative association with all WPPSI-III scores, although 95% confidence intervals consistently covered the null (Table B.4.). The High Level cluster had negative associations with VIQ and FSIQ, but not PIQ. The Low Level cluster was positively associated with VIQ but negatively associated with PIQ and FSIQ. We found no consistent trends in the High OPPs cluster. No noteworthy patterns were detected when we stratified by child sex.

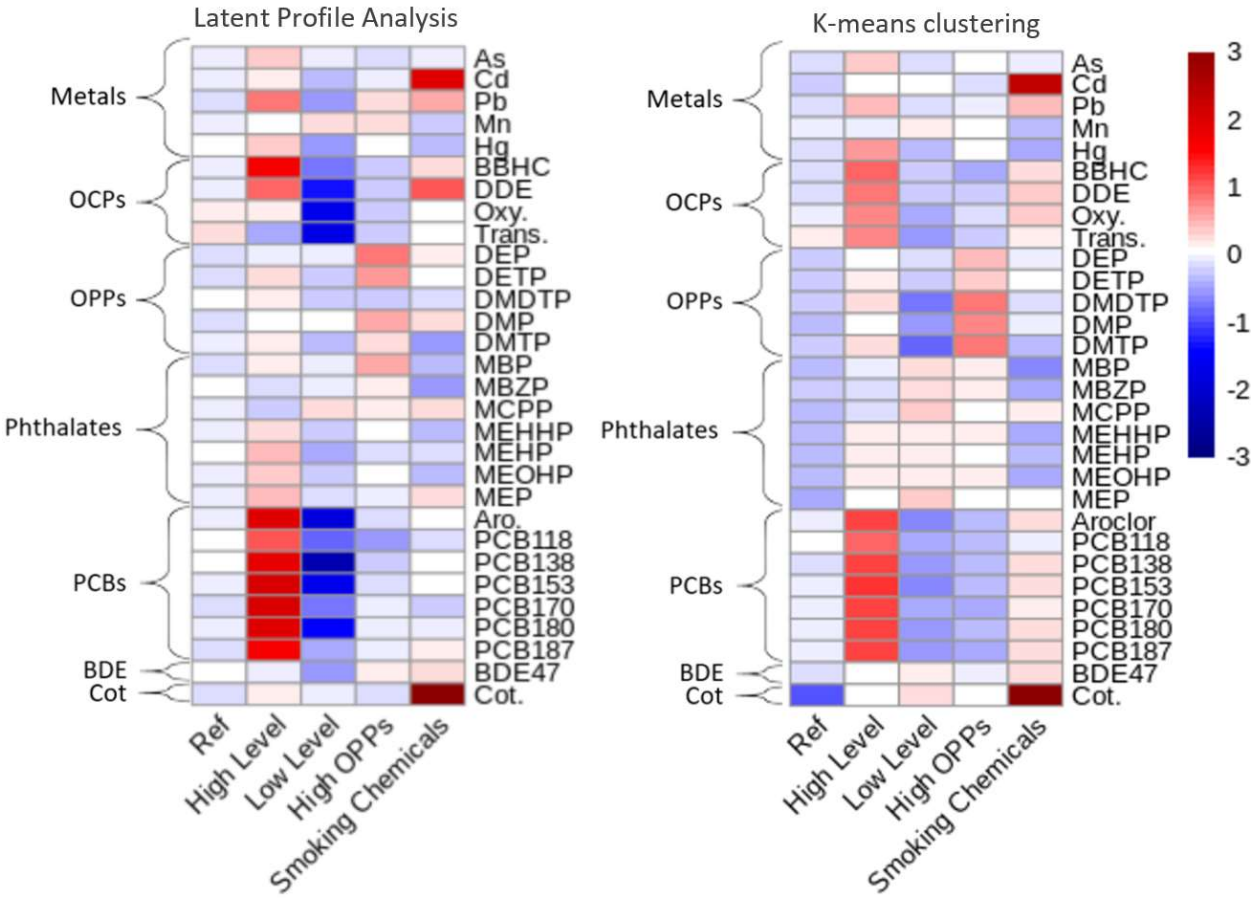


Figure 2.3. Sensitivity analysis comparing LPA results with k-means clustering results, showing z-scores for the mean biomarker concentrations in each latent profile (left) and k-means cluster (right; n = 517).

## 2.4. Discussion

Using LPA, we generated five clearly defined latent profiles which showed risk stratification of gestational chemical mixtures. We found a High Level profile, a Low Level profile, a High OPPs profile, a Smoking Chemicals profile, and a medium level Reference profile. We verified the accuracy of the profiles in a sensitivity analysis comparing LPA with k-means clustering. Unlike k-means clustering, LPA does not have a bias towards creating profiles of equal size, and this allows for the generation of very small profiles with high classification accuracy. The large size of the Reference profile in our study indicates that this profile acted as a catch-all group for mothers who did not fit any of the other patterns.

We found a strong negative association between the Smoking Chemicals profile and VIQ, PIQ, and FSIQ when compared to the Reference profile, although the 95% confidence intervals spanned the null. Furthermore, the magnitude of the regression coefficients for the Smoking Chemicals profile in relation to WPPSI-III scores were larger than the corresponding regression coefficients for self-reported prenatal smoking and the cotinine biomarker by itself (Table 2.3.; Table B.1.; Table B.2.). We did not expect the effect estimates of the Smoking Chemicals profile to be so much greater than the cotinine biomarker, which is a measure that is less prone to error than self-reported smoking. Although more research is needed in this area, our results suggest that the latent profile detected by LPA may be a better tool for studying the health effects of smoking than these measures.

In both boys and girls, the High Level profile had a negative association with VIQ but not PIQ, although the 95% confidence intervals spanned the null. As shown in Figure 2.2., the High Level profile had a higher concentration of PCBs than any of the other chemicals, and can therefore be interpreted as a highly concentrated PCB mixture. We found that the associations in Table 2.3. were reflective of trends in the individual PCBs, which were all negatively associated with VIQ, but had inconsistent associations with PIQ. See Table B.3. for comparison. Our findings are consistent with the literature; studies have shown significant negative relationships between PCBs and VIQ, but not PIQ[60,61]. Studies have been done on VIQ-PIQ discrepancies and their impact on child development and developmental disorders, but more research is needed to determine why these differences occur[62–65].

Both the Low Level profile and the High OPPs profile were positively associated with IQ scores in girls but not boys. This result is consistent with previous studies that have shown that some chemicals may have differing effects on boys and girls[1,3,8]. However, the results in the

Low Level profile may be partly due to the unequal sex distribution; it may be that the larger group of girls are more indicative of the actual patterns for this profile. It should also be noted that the exposure concentrations of the OPP profile were similar to those of the Reference profile, as the OPPs did not show high levels of variance between profiles. This profile also had the lowest classification accuracy of the five. This may explain the inconsistent effect of membership in the profile on IQ.

When we conducted sensitivity analysis using k-means clustering, we found that the five clusters had similar patterns to the five profiles, albeit less pronounced. We found that k-means clustering had three main disadvantages: The unclear method of choosing a model, the tendency towards choosing clusters of the same size, and the non-probabilistic nature of cluster assignment. We chose to use a model with five clusters because that was the number of profiles LPA generated, but this would have been a more difficult decision had we not done LPA first. The clusters were mostly the same size (except for the smaller Smoking Chemicals cluster). This may have resulted in lower classification accuracy, which would explain the lower biomarker variance between clusters. However, because this method does not generate posterior probabilities, we could not assess classification accuracy. Although the small sizes of some of the latent profiles may make them less applicable to large groups of the general population, they are still indicative of patterns within the population, and we found them to be more effective at detecting harmful chemical mixtures than the more generalized clusters. We therefore conclude that LPA is a more useful method for studying the effects of gestational chemical mixtures.

This study builds on other work that has used unsupervised machine learning techniques to estimate the health effects of chemical mixtures. Several similar studies have been done using k-means clustering, which is why we chose this method for sensitivity analysis[2,66–68]. PCA has also been used to determine new variables for chemical mixtures[36,67,69]. One recent study by Carroll et al. used Latent Class Analysis (LCA) to study phthalate and phenols[70]. LCA is similar to LPA, but it only works with categorical independent variables[71]. Therefore, chemical exposures were dichotomized in the study, reducing accuracy[70]. Had we chosen to use this method, we could not have differentiated the profiles nearly as well, and would have missed information about the spread of the PCBs and the existence of the reference group, among other results. Finally, a recent study by Khorrami et al. used LPA to find associations between mixtures of air pollutants and lung cancer[72]. In this study, however,

pollutant concentration was ascertained using geographical location, not chemical biomarkers. Also, that study did not focus on gestational chemical exposures.

Our study has several limitations. The first is that mothers in the MIREC cohort tend to be older, wealthier, more educated, more likely to be white, and less likely to report prenatal smoking and drinking than the Canadian average[28]. Therefore, we may not be accurately reflecting the exposure patterns found in more vulnerable populations. Secondly, there were several chemicals with relatively low %>LODs, and the imputation method we used may have underestimated the variance for these chemicals[47]. Third, due to high levels of correlation, it can be difficult to distinguish between the effects of gestational exposures and chemicals in the postnatal environment[60,73]. Fourth, while much work has been done to ensure the validity of the WPPSI-III test, IQ tests for children often have problems with accuracy[34,74]. Regardless of the training of test administrators, scores can vary based on the children's levels of cooperation, motivation, or fatigue, especially in children with intellectual disabilities[75]. Fifth, exposure misclassification for non-persistent chemicals may have affected the analysis results[18]. For example, the poorly defined profiles for phthalates in Figure 2.2. may be due to the fact that phthalate exposure is measured with error. Finally, the complexity of LPA results can make them difficult to interpret, leading some researchers to prefer simpler methods such as LCA or k-means clustering[26,70]. In certain situations, the simpler methods are more useful. For example, when all the profiles have high classification accuracy, the posterior probabilities are all so close to 0 or 1 that a probabilistic method becomes unnecessary[26]. However, despite these limitations, we believe that LPA is a promising technique that is worthy of more study.

In conclusion, we recommend the use of LPA as a technique for studying chemical mixtures. Although further research is needed to understand LPA's capabilities, we believe that this is an effective alternative to other clustering methods. This technique can find patterns in large, complex datasets while avoiding many of the disadvantages of other machine learning techniques. It generates a helpful new variable that can be used to study the effects of chemical mixtures on other health outcomes.

## 2.5. Declarations

### *2.5.1. Competing Interests*

None declared.


### *2.5.2. Funding*

This work was supported by a Discovery Grant from the Natural Sciences and Engineering Council of Canada (RGPIN-2015-05155). The MIREC Study was funded by Health Canada's Chemicals Management Plan, the Ontario Ministry of the Environment, and a research grant from the Canadian Institute for Health Research (MOP-81285).


### *2.5.3. Disclosure*

JMB was financially compensated for his services as an expert witness for plaintiffs in litigation related to second-hand tobacco smoke exposures.

## 2.6. References


1. Desrochers-Couture M, Oulhote Y, Arbuckle TE, et al. Prenatal, concurrent, and sex-specific associations between blood lead concentrations and IQ in preschool Canadian children. *Environ Int*. 2018;121(Pt 2):1235-1242.

2. Kalloo G, Wellenius GA, McCandless L, et al. Profiles and Predictors of Environmental Chemical Mixture Exposure among Pregnant Women: The Health Outcomes and Measures of the Environment Study. *Environ Sci Technol*. 2018;52(17):10104-10113.

3. Braun JM, Muckle G, Arbuckle T, et al. Associations of Prenatal Urinary Bisphenol A Concentrations with Child Behaviors and Cognitive Abilities. *Environ Health Perspect*. 2017;125(6):067008.

4. Sutton P, Woodruff TJ, Perron J, et al. Toxic environmental chemicals: the role of reproductive health professionals in preventing harmful exposures. *Am J Obstet Gynecol*. 2012;207(3):164-173.

5. Lee W-C, Fisher M, Davis K, Arbuckle TE, Sinha SK. Identification of chemical mixtures to which Canadian pregnant women are exposed: The MIREC Study. *Environ Int*. 2017;99:321-330.

6. 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(2):26001.

7. Li N, Arbuckle TE, Muckle G, et al. Associations of cord blood leptin and adiponectin with children's cognitive abilities. *Psychoneuroendocrinology*. 2019;99:257-264.

8. Green R, Lanphear B, Hornung R, et al. Association Between Maternal Fluoride Exposure During Pregnancy and IQ Scores in Offspring in Canada. *JAMA Pediatr*. Published online August 19, 2019. doi:10.1001/jamapediatrics.2019.1729

9. Lanphear BP. The impact of toxins on the developing brain. *Annu Rev Public Health*. 2015;36:211-230.

10. Taylor CM, Kordas K, Golding J, Emond AM. Effects of low-level prenatal lead exposure on child IQ at 4 and 8 years in a UK birth cohort study. *Neurotoxicology*. 2017;62:162-169.

11. Kim, W, Lim, Y, Kim, B, et al. Associations between prenatal and concurrent lead concentrations and IQ in preschool children in Korea. *Environmental Epidemiology*. 2019;3:201.

12. Lin C-C, Chien C-J, Tsai M-S, Hsieh C-J, Hsieh W-S, Chen P-C. Prenatal phenolic compounds exposure and neurobehavioral development at 2 and 7 years of age. *Sci Total Environ*. 2017;605-606:801-810.

13. Torres-Sánchez L, Schnaas L, Rothenberg SJ, et al. Prenatal p,p´-DDE exposure and neurodevelopment among children 3.5-5 years of age. *Environ Health Perspect*. 2013;121(2):263-268.

14. 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(6):067015.

15. Herbstman JB, Sjödin A, Kurzon M, et al. Prenatal exposure to PBDEs and neurodevelopment. *Environ Health Perspect*. 2010;118(5):712-719.

16. 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(1):A6-A9.

17. 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.

18. 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(3):284-292.

19. Park SK, Tao Y, Meeker JD, Harlow SD, Mukherjee B. Environmental risk score as a new tool to examine multi-pollutants in epidemiologic research: an example from the NHANES study using serum lipid levels. *PLoS One*. 2014;9(6):e98632.

20. Stanley L, Kellermanns FW, Zellweger TM. Latent Profile Analysis: Understanding Family Firm Profiles. *Family Business Review*. 2017;30(1):84-102.

21. James G. *An Introduction to Statistical Learning : With Applications in R*. [Uncorrected edition]. New York : Springer; 2013.

22. Bonadio FT, Tompsett C. Who Benefits from Community Mental Health Care? Using Latent Profile Analysis to Identify Differential Treatment Outcomes for Youth. *J Youth Adolesc*. 2018;47(11):2320-2336.

23. Olivera-Aguilar M, Rikoon SH, Robbins SB. Using Latent Profile Analysis to Identify Noncognitive Skill Profiles Among College Students. *J Higher Educ*. 2017;88(2):234-257.

24. Gabriel AS, Daniels MA, Diefendorff JM, Greguras GJ. Emotional labor actors: a latent profile analysis of emotional labor strategies. *J Appl Psychol*. 2015;100(3):863-879.

25. Contractor AA, Caldas S, Weiss NH, Armour C. Examination of the heterogeneity in PTSD and impulsivity facets: A latent profile analysis. *Pers Individ Dif*. 2018;125:1-9.

26. Pastor DA, Barron KE, Miller BJ, Davis SL. A latent profile analysis of college students' achievement goal orientation. *Contemp Educ Psychol*. 2007;32(1):8-47.

27. Haines DA, Arbuckle TE, Lye E, et al. Reporting results of human biomonitoring of environmental chemicals to study participants: a comparison of approaches followed in two Canadian studies. *J Epidemiol Community Health*. 2011;65(3):191-198.

28. Arbuckle TE, Fraser WD, Fisher M, et al. Cohort profile: the maternal-infant research on environmental chemicals research platform. *Paediatr Perinat Epidemiol*. 2013;27(4):415-425.

29. Krzeczkowski JE, Boylan K, Arbuckle TE, et al. Neurodevelopment in 3-4 year old children exposed to maternal hyperglycemia or adiposity in utero. *Early Hum Dev*. 2018;125:8-16.

30. Oulhote Y, Tremblay É, Arbuckle TE, et al. Prenatal exposure to polybrominated diphenyl ethers and predisposition to frustration at 7 months: Results from the MIREC study. *Environ Int*. 2018;119:79-88.

31. Till C, Green R, Flora D, et al. Fluoride exposure from infant formula and child IQ in a Canadian birth cohort. *Environ Int*. 2020;134:105315.

32. Follow-up Studies. Accessed January 17, 2021. https://www.mirec-canada.ca/en/about/follow-up-studies/

33. Early Childhood Measurement and Evaluation. https://www.ualberta.ca/community-university-partnership/media-library/community-university-partnership/resources/tools---assessment/wppsi-iiimay-2012.pdf

34. Gordon B. Test Review: Wechsler, D. (2002). The Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III). San Antonio, TX: The Psychological Corporation. *Canadian Journal of School Psychology*. 2004;19(1-2):205-220.

35. Wechsler Preschool and Primary Scale of Intelligence. Accessed January 17, 2021. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Preschool-and-Primary-Scale-of-Intelligence-%7C-Third-Edition/p/100000422.html?tab=product-details

36. Kim Y, Ha E-H, Park H, et al. Prenatal lead and cadmium co-exposure and infant neurodevelopment at 6 months of age: the Mothers and Children's Environmental Health (MOCEH) study. *Neurotoxicology*. 2013;35:15-22.

37. Snoj Tratnik J, Falnoga I, Trdin A, et al. Prenatal mercury exposure, neurodevelopment and apolipoprotein E genetic polymorphism. *Environ Res*. 2017;152:375-385.

38. Rodríguez-Barranco M, Lacasaña M, Aguilar-Garduño C, et al. Association of arsenic, cadmium and manganese exposure with neurodevelopment and behavioural disorders in children: a systematic review and meta-analysis. *Sci Total Environ*. 2013;454-455:562-577.

39. Saeedi Saravi SS, Dehpour AR. Potential role of organochlorine pesticides in the pathogenesis of neurodevelopmental, neurodegenerative, and neurobehavioral disorders: A review. *Life Sci*. 2016;145:255-264.

40. Dzwilewski KLC, Schantz SL. Prenatal chemical exposures and child language development. *J Commun Disord*. 2015;57:41-65.

41. Hyland C, Mora AM, Kogut K, et al. Prenatal Exposure to Phthalates and Neurodevelopment in the CHAMACOS Cohort. *Environ Health Perspect*. 2019;127(10):107010.

42. Gibson EA, Siegel EL, Eniola F, Herbstman JB, Factor-Litvak P. Effects of Polybrominated Diphenyl Ethers on Child Cognitive, Behavioral, and Motor Development. *Int J Environ Res Public Health*. 2018;15(8). doi:10.3390/ijerph15081636

43. Bellinger DC. Prenatal Exposures to Environmental Chemicals and Children's Neurodevelopment: An Update. *Saf Health Work*. 2013;4(1):1-11.

44. Liu P, Wu C, Chang X, Qi X, Zheng M, Zhou Z. Adverse Associations of both Prenatal and Postnatal Exposure to Organophosphorous Pesticides with Infant Neurodevelopment in an Agricultural Area of Jiangsu Province, China. *Environ Health Perspect*. 2016;124(10):1637-1643.

45. O'Brien KM, Upson K, Cook NR, Weinberg CR. Environmental Chemicals in Urine and Blood: Improving Methods for Creatinine and Lipid Adjustment. *Environ Health Perspect*. 2016;124(2):220-227.

46. Hauser R, Gaskins AJ, Souter I, et al. Urinary Phthalate Metabolite Concentrations and Reproductive Outcomes among Women Undergoing in Vitro Fertilization: Results from the EARTH Study. *Environ Health Perspect*. 2016;124(6):831-839.

47. Lubin JH, Colt JS, Camann D, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect*. 2004;112(17):1691-1696.

48. Etzel T, Muckle G, Arbuckle TE, et al. Prenatal urinary triclosan concentrations and child neurobehavior. *Environ Int*. 2018;114:152-159.

49. Choi W-J, Kwon H-J, Lim MH, Lim J-A, Ha M. Blood lead, parental marital status and the risk of attention-deficit/hyperactivity disorder in elementary school children: A longitudinal study. *Psychiatry Res*. 2016;236:42-46.

50. Hanscombe KB, Trzaskowski M, Haworth CMA, Davis OSP, Dale PS, Plomin R. Socioeconomic status (SES) and children's intelligence (IQ): in a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS One*. 2012;7(2):e30320.

51. Huang J, Zhu T, Qu Y, Mu D. Prenatal, Perinatal and Neonatal Risk Factors for Intellectual Disability: A Systemic Review and Meta-Analysis. *PLoS One*. 2016;11(4):e0153655.

52. Lederer DJ, Bell SC, Branson RD, et al. Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. *Ann Am Thorac Soc*. 2019;16(1):22-28.

53. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J*. 2016;8(1):289-317.

54. Brewer MJ, Butler A, Cooksley SL, Freckleton R. The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity. *Methods Ecol Evol*. 2016;7(6):679-692.

55. Alade T. Tutorial: How to determine the optimal number of clusters for k-means clustering. Cambridge Spark. Published May 27, 2018. Accessed December 30, 2020. https://blog.cambridgespark.com/how-to-determine-the-optimal-number-of-clusters-for-k-means-clustering-14f27070048f

56. Syakur MA, Khotimah BK, Rochman EMS, Satoto BD. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conf Ser: Mater Sci Eng*. 2018;336(1):012017.

57. Milan A. *Report on the Demographic Situation in Canada*. Statistics Canada; 2011. https://www150.statcan.gc.ca/n1/en/pub/91-209-x/2011001/article/11513-eng.pdf?st=HTao0rh9

58. Health Canada. *Report on Human Biomonitoring of Environmental Chemicals in Canada: Results of the Canadian Health Measures Survey Cycle 1 (2007–2009)*.; 2010. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/hecs-sesc/pdf/pubs/contaminants/chms-ecms/report-rapport-eng.pdf

59. Satarug S, Moore MR. Adverse health effects of chronic exposure to low-level cadmium in foodstuffs and cigarette smoke. *Environ Health Perspect*. 2004;112(10):1099-1103.

60. Jacobson SW. Assessing the impact of maternal drinking during and after pregnancy. *Alcohol Health Res World*. 1997;21(3):199-203.

61. Stewart PW, Lonky E, Reihman J, Pagano J, Gump BB, Darvill T. The relationship between prenatal PCB exposure and intelligence (IQ) in 9-year-old children. *Environ Health Perspect*. 2008;116(10):1416-1422.

62. Yu T-Y, Chen K-L, Chou W, et al. Intelligence quotient discrepancy indicates levels of motor competence in preschool children at risk for developmental delays. *Neuropsychiatr Dis Treat*. 2016;12:501-510.

63. Margolis A, Bansal R, Hao X, et al. Using IQ discrepancy scores to examine the neural correlates of specific cognitive abilities. *J Neurosci*. 2013;33(35):14135-14145.

64. Margolis AE, Davis KS, Pao LS, et al. Verbal-spatial IQ discrepancies impact brain activation associated with the resolution of cognitive conflict in children and adolescents. *Dev Sci*. 2018;21(2). doi:10.1111/desc.12550

65. Peijnenborgh JCAW, van Abeelen SAM, Hurks PPM, et al. Can IQ predict parent-reported behavioral and emotional problems in children with neurological deficiencies? *Eur J Paediatr Neurol*. 2017;21(2):336-343.

66. Caron A, Redon N, Coddeville P, Hanoune B. Identification of indoor air quality events using a K-means clustering analysis of gas sensors data. *Sens Actuators B Chem*. 2019;297:126709.

67. Chen H, Zhang W, Zhou Y, et al. Characteristics of exposure to multiple environmental chemicals among pregnant women in Wuhan, China. *Sci Total Environ*. 2021;754:142167.

68. 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.

69. Veyhe AS, Hofoss D, Hansen S, et al. The Northern Norway Mother-and-Child Contaminant Cohort (MISA) Study: PCA analyses of environmental contaminants in maternal sera and dietary intake in early pregnancy. *Int J Hyg Environ Health*. 2015;218(2):254-264.

70. Carroll R, White AJ, Keil AP, et al. Latent classes for chemical mixtures analyses in epidemiology: an example using phthalate and phenol exposure biomarkers in pregnant women. *J Expo Sci Environ Epidemiol*. Published online October 21, 2019. doi:10.1038/s41370-019-0181-y

71. Oberski D. Mixture Models: Latent Profile and Latent Class Analysis. In: Robertson J, Kaptein M, eds. *Modern Statistical Methods for HCI*. Springer International Publishing; 2016:275-287.

72. Khorrami Z, Pourkhosravani M, Rezapour M, et al. Multiple air pollutant exposure and lung cancer in Tehran, Iran. *Sci Rep*. 2021;11(1):9239.

73. Hofhuis W, de Jongste JC, Merkus PJFM. Adverse health effects of prenatal and postnatal tobacco smoke exposure on children. *Arch Dis Child*. 2003;88(12):1086-1090.

74. Flanagan DP, Alfonso VC. A Critical Review of the Technical Characteristics of New and Recently Revised Intelligence Tests for Preschool Children. *J Psychoeduc Assess*. 1995;13(1):66-90.

75. Jenni OG, Fintelmann S, Caflisch J, Latal B, Rousson V, Chaouch A. Stability of cognitive performance in children with mild intellectual disability. *Dev Med Child Neurol*. 2015;57(5):463-469.

# Chapter 3.
# Conclusion

## 3.1. Summary

In summary, this MSc thesis showed that LPA is a promising technique for studying chemical mixtures, and that it compares favourably to k-means clustering. This study contributes to the growing field on appropriate statistical measures for the analysis of chemical mixtures[1–3]. For my first objective, I used LPA to generate five latent profiles of MIREC participants based on 30 gestational biomarkers. These five profiles were found using patterns in exposure mixtures in different subgroups of the population, highlighting common mixtures in pregnant Canadian women. They identified subgroups with very high and low levels of PCBs respectively, as well as one with high OPPs, and one with high smoking chemicals.

This thesis also contributes to the field of gestational chemical mixtures and their effects on neurodevelopment[4–8]. For my second objective, I not only showed the risk stratification of different chemical mixtures and how they impact IQ scores, but also confirmed that PIQ and VIQ can be affected differently. I found that the Smoking Chemicals profile was associated with a large decrease in all three IQ scores in both sexes, that the High Levels profile was negatively associated with VIQ, and that the Low Level and High OPPs profile were positively associated with IQ in girls but not boys. However, these associations were imprecise and had wide 95% confidence intervals which covered the null values, perhaps due to a small sample size.

For my third and final objective, I conducted sensitivity analysis using k-means clustering. I repeated my first two objectives using this method, first creating five clusters, and then regressing cluster membership against IQ. The clusters I generated had similar patterns to the latent profiles. However, this method's bias toward clusters of the same size led to less chemical variation between the clusters and weaker regression results. I found that, while k-means clustering was faster and easier to use than LPA, the difficulty choosing a model and lack of information about classification accuracy resulted in less useful results than those of LPA. From my experiment I concluded that LPA shows potential as a method for studying chemical mixtures and that it compares favourably against k-means clustering. However, more research is needed to understand LPA's capabilities in terms of number of predictor variables for profile generation and statistical power for determining effects. Future directions of work are discussed in Section 3.3.

## 3.2. Relevance for Policymakers and Environmental Health Researchers

There are two main reasons that stakeholders would find my research relevant to their work; some will be more interested in the methodology I employed, and some will be more interested in my results. From a methodological perspective, this study is relevant to epidemiologists because it introduces a new machine learning technique that can be used to study complex chemical mixtures. Although most neurotoxins are still studied one at a time, exposure mixtures are becoming an increasingly common topic of study for environmental epidemiology[6,7,9,10]. However, researchers are still exploring new techniques for tackling this issue[2]. Many problems arise when studying complex chemical mixtures; researchers must choose the appropriate mixtures to study, and contend with collinearity in highly correlated datasets, the loss of power when a large number of variables are included in the model, and interactions between individual chemicals[1]. LPA is a method that helps avoid these issues. It finds the strongest patterns in a population, allowing researchers to prioritize more common mixtures. It does not lose power when more variables are added to the model. Finally, it allows one to study the effects of mixtures as a whole, avoiding problems with interactions or collinearity. I believe that epidemiologists would do well to consider this method for studies beyond the scope of my thesis.

This work is also relevant to several groups who may be interested in the profiles I generated. Researchers studying gestational exposures need to shift their focus away from single chemical exposures and towards chemical mixtures[7,9,11]. While excellent work has been done in this field on single exposures, studies on mixtures can more accurately reflect the harm of chemicals during pregnancy. That said, it is important to focus on the mixtures that are most relevant to society[1]. My work highlights several specific chemical mixtures that deserve more attention, as they are found in specific subgroups of the population. For example, the High and Low Level profiles show that we need to be paying attention to PCB exposure during pregnancy, because this biomarker varies wildly between mothers and is negatively associated with VIQ. Also, the subgroup of young mothers with higher levels of OPP exposure may be the result of our shift away from OCPs[12]. It will be important to keep an eye on this group in the future.

Environmental epidemiologists need to convey the importance of chemical mixtures to policymakers. Oftentimes, policymakers will focus on single chemical exposures, and will

emphasize “safe” doses extrapolated from studies on high level exposures[13–15]. This is a point of contention in environmental epidemiology. Studies have shown that many chemicals, such as lead, phthalates, and air pollution, have no safe threshold of exposure[11,13,15]. We are beginning to understand that low-dose chemical exposures, especially during vulnerable times such as the gestational period, are far from benign. Despite this, policymakers still ascribe some chemicals a threshold below which neurotoxic effects are negligible, and use these thresholds to make policy decisions[15]. However, any messaging about “safe” doses ignores the fact that exposure to many simultaneous low-dose chemicals can be just as dangerous as exposure to high levels of a single chemical[14]. Government officials and policymakers need to reframe their understanding of this complex issue.

Finally, this study is relevant to one very important group of stakeholders: Pregnant mothers. Information regarding gestational exposures is large in volume and contradictory[16]. This can be overwhelming for parents trying to make decisions about their children’s health. The onus should not be on pregnant mothers to avoid every potential neurotoxin; this would be impossible given these chemicals’ pervasiveness in our environment. However, it is important that people understand that gestational chemical exposures can severely impact child neurodevelopment. Effective knowledge translation about the neurotoxicity of chemical groups such as smoking metabolites, pesticides, and PCBs may help mothers make informed decisions[11]. Obstetricians, gynecologists, and pediatricians can provide guidance and early intervention in this area. Using patient-centred actions and communication strategies, these health workers can help mitigate the impacts of gestational neurotoxins.

## 3.3. Future Topics of Study

My research demonstrates the promising nature of LPA for studying gestational chemical mixtures, which is something that has not been done before in the field of environmental epidemiology. Most applications of LPA in the literature have been in the fields of psychology and behavioural sciences; LPA has been used to study mental health treatment outcomes, noncognitive skill, emotional labour strategies, and PTSD symptoms[17–20]. However, further research is necessary to understand LPA's capabilities, including the number of predictor variables that can be used in a model, the level of correlation the variables can have, and the applicability of profiles between populations. My study opens the doors for a wide variety of new research focusing on LPA and the effects of chemical mixtures.

First, the profiles I generated could be used in future research to study other child health outcomes, such as birth weight or behavioural disorders. This would allow us to understand the effects of these specific chemical mixtures more fully. These studies would have the advantage of being fast and easy to conduct, as profile generation was the most time-consuming part of my project.

This research could also be repeated with a different cohort to further assess these profiles' applicability in other populations. It would be beneficial to test LPA with a larger sample size of Canadian women to see if similar profiles are generated and to try to get adequate power for significant results. It would also be ideal to study populations with higher biomarker concentrations, which tend to be quite low in Canadian women[12]. Finally, repeating this study with a different cohort could confirm the effect sizes of these chemical mixtures on IQ, which were much higher than those of our individual chemical exposures.

If future researchers use LPA to study neurodevelopment, they may choose to use different chemical mixtures to generate the latent profiles. For this project, I used the 30 known neurotoxins that were available in a large portion of my study population[21–28]. Future studies could be conducted to test the limits of how many chemical exposures LPA can handle at once. I would also be interested in conducting similar research without using cotinine as a variable, as it was a very strong driver of chemical profile membership and I would like to see how the profiles would be generated without it.

Finally, environmental epidemiologists may conduct studies using other measures to form profiles. While this is, to my knowledge, the first study to use LPA on gestational chemical mixtures, many studies have been conducted in other fields using different variables that are

relevant to human health[17–20]. I am interested in seeing LPA used on other variables in the field of environmental epidemiology. In particular, I would like to try creating profiles based on both chemical exposures and other demographic variables. Although we did not choose to do this for this project, it is possible to adjust for covariates directly in the model[29]. Overall, I am excited to explore LPA's capabilities further in future projects.

## 3.4. References


1. 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(1):A6-A9.

2. 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(2):26001.

3. Oberski D. Mixture Models: Latent Profile and Latent Class Analysis. In: Robertson J, Kaptein M, eds. *Modern Statistical Methods for HCI*. Springer International Publishing; 2016:275-287.

4. Kalloo G, Wellenius GA, McCandless L, et al. Chemical mixture exposures during pregnancy and cognitive abilities in school-aged children. *Environ Res*. 2021;197:111027.

5. Carroll R, White AJ, Keil AP, et al. Latent classes for chemical mixtures analyses in epidemiology: an example using phthalate and phenol exposure biomarkers in pregnant women. *J Expo Sci Environ Epidemiol*. Published online October 21, 2019. doi:10.1038/s41370-019-0181-y

6. Guo J, Wu C, Zhang J, et al. Prenatal exposure to mixture of heavy metals, pesticides and phenols and IQ in children at 7 years of age: The SMBCS study. *Environ Int*. 2020;139:105692.

7. Tanner EM, Hallerbäck MU, Wikström S, et al. *Early Prenatal Exposure to Suspected Endocrine Disruptor Mixtures Is Associated with Lower IQ at Age Seven*. Vol 134.; 2020.

8. Lanphear BP, Hornung R, Khoury J, et al. Low-level environmental lead exposure and children's intellectual function: an international pooled analysis. *Environ Health Perspect*. 2005;113(7):894-899.

9. Kalloo G, Wellenius GA, McCandless L, et al. Profiles and Predictors of Environmental Chemical Mixture Exposure among Pregnant Women: The Health Outcomes and Measures of the Environment Study. *Environ Sci Technol*. 2018;52(17):10104-10113.

10. Kim, W, Lim, Y, Kim, B, et al. Associations between prenatal and concurrent lead concentrations and IQ in preschool children in Korea. *Environmental Epidemiology*. 2019;3:201.

11. Sutton P, Woodruff TJ, Perron J, et al. Toxic environmental chemicals: the role of reproductive health professionals in preventing harmful exposures. *Am J Obstet Gynecol*. 2012;207(3):164-173.

12. Health Canada. *Report on Human Biomonitoring of Environmental Chemicals in Canada: Results of the Canadian Health Measures Survey Cycle 1 (2007–2009)*.; 2010. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/ewh-semt/alt_formats/hecs-sesc/pdf/pubs/contaminants/chms-ecms/report-rapport-eng.pdf

13. Lanphear BP. Low-level toxicity of chemicals: No acceptable levels? *PLoS Biol*. 2017;15(12):e2003066.

14. Monosson E. Chemical mixtures: considering the evolution of toxicology and chemical assessment. *Environ Health Perspect*. 2005;113(4):383-390.

15. Hill CE, Myers JP, Vandenberg LN. Nonmonotonic Dose-Response Curves Occur in Dose Ranges That Are Relevant to Regulatory Decision-Making. *Dose Response*. 2018;16(3):1559325818798282.

16. Toxins and Pregnancy. Accessed May 12, 2021. https://www.webmd.com/baby/features/pregnancy-and-toxins

17. Bonadio FT, Tompsett C. Who Benefits from Community Mental Health Care? Using Latent Profile Analysis to Identify Differential Treatment Outcomes for Youth. *J Youth Adolesc*. 2018;47(11):2320-2336.

18. Olivera-Aguilar M, Rikoon SH, Robbins SB. Using Latent Profile Analysis to Identify Noncognitive Skill Profiles Among College Students. *J Higher Educ*. 2017;88(2):234-257.

19. Gabriel AS, Daniels MA, Diefendorff JM, Greguras GJ. Emotional labor actors: a latent profile analysis of emotional labor strategies. *J Appl Psychol*. 2015;100(3):863-879.

20. Contractor AA, Caldas S, Weiss NH, Armour C. Examination of the heterogeneity in PTSD and impulsivity facets: A latent profile analysis. *Pers Individ Dif*. 2018;125:1-9.

21. 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(6):067015.

22. Satarug S, Moore MR. Adverse health effects of chronic exposure to low-level cadmium in foodstuffs and cigarette smoke. *Environ Health Perspect*. 2004;112(10):1099-1103.

23. Sakamoto M, Tatsuta N, Izumo K, et al. Health Impacts and Biomarkers of Prenatal Exposure to Methylmercury: Lessons from Minamata, Japan. *Toxics*. 2018;6(3). doi:10.3390/toxics6030045

24. Stewart PW, Lonky E, Reihman J, Pagano J, Gump BB, Darvill T. The relationship between prenatal PCB exposure and intelligence (IQ) in 9-year-old children. *Environ Health Perspect*. 2008;116(10):1416-1422.

25. Azar N, Booij L, Muckle G, et al. Prenatal exposure to polybrominated diphenyl ethers (PBDEs) and cognitive ability in early childhood. *Environ Int*. 2021;146:106296.

26. Loftus CT, Bush NR, Day DB, et al. Exposure to prenatal phthalate mixtures and neurodevelopment in the Conditions Affecting Neurocognitive Development and Learning in Early childhood (CANDLE) study. *Environ Int*. 2021;150:106409.

27. Bouchard MF, Chevrier J, Harley KG, et al. Prenatal exposure to organophosphate pesticides and IQ in 7-year-old children. *Environ Health Perspect*. 2011;119(8):1189-1195.

28. Saeedi Saravi SS, Dehpour AR. Potential role of organochlorine pesticides in the pathogenesis of neurodevelopmental, neurodegenerative, and neurobehavioral disorders: A review. *Life Sci*. 2016;145:255-264.

29. Pastor DA, Barron KE, Miller BJ, Davis SL. A latent profile analysis of college students' achievement goal orientation. *Contemp Educ Psychol*. 2007;32(1):8-47.

# Appendices

## Appendix A. Supplementary Materials for Chapter 1

***Table A.1.***

Descriptions of WPPSI-III subtests for younger age band (aged two years and six months to three years and eleven months). *

| Subtest | Type | Composites | What the child does |
|---|---|---|---|
| **Receptive Vocabulary** | Core | VIQ, FSIQ, and GLC | Points to the picture that corresponds to certain words |
| **Information** | Core | VIQ and FSIQ | Answers general information questions (verbally or by pointing) |
| **Block Design** | Core | PIQ and FSIQ | Uses blocks to copy a design |
| **Object Assembly** | Core | PIQ and FSIQ | Completes puzzles |
| **Picture Naming** | Supplemental | GLC and VIQ (substitute) | Names objects from picture |

*From Gordon B. Test Review: Wechsler, D. (2002). The Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III). San Antonio, TX: The Psychological Corporation. Canadian Journal of School Psychology. 2004;19(1-2):205-220.

***Table A.2.***

Unadjusted linear regression coefficients showing the associations between MIREC participant demographic characteristics and unadjusted mean WPPSI-III scores, with 95% confidence intervals (n = 517).

| | VIQ (95% CI) | PIQ (95% CI) | FSIQ (95% CI) |
|---|---|---|---|
| **Child Sex** | | | |
| Intercept | 106.7 (105.2, 108.3) | 100.9 (99.1, 102.7) | 104.2 (102.6, 105.8) |
| Male | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Female | 5.3 (3.2, 7.5) | 4.2 (1.6, 6.7) | 5.5 (3.2, 7.8) |
| **Maternal Age** | | | |
| Intercept | 107.8 (105.4, 110.3) | 104.0 (101.2, 106.9) | 106.7 (104.1, 109.2) |
| 19-29 | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| 30-34 | 1.3 (-1.8, 4.3) | -0.2 (-3.7, 3.4) | 0.6 (-2.6, 3.7) |
| 35+ | 2.8 (-0.2, 5.8) | -2.3 (-5.7, 1.2) | 0.3 (-2.8, 3.5) |
| **Maternal Race** | | | |
| Intercept | 110.0 (108.8, 111.2) | 103.5 (102.1, 104.9) | 107.6 (106.3, 108.8) |
| White | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Other | -3.6 (-6.8, -0.4) | -3.1 (-6.8, 0.6) | -3.8 (-7.1, -0.5) |
| **Maternal Education** | | | |
| Intercept | 102.5 (97.9, 107.1) | 94.9 (89.4, 100.4) | 98.5 (93.7, 103.4) |
| Highschool | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| College | 2.7 (-2.4, 7.7) | 6.0 (0.0, 12.0) | 4.6 (-0.7, 9.9) |
| Undergrad | 7.7 (2.8, 12.6) | 9.6 (3.7, 15.4) | 9.8 (4.7, 14.9) |
| Grad | 11.3 (6.3, 16.3) | 9.9 (3.9, 15.8) | 12.0 (6.8, 17.3) |
| **Marital Status** | | | |
| Intercept | 110.5 (109.2, 111.8) | 102.2 (100.7, 103.8) | 107.2 (105.8, 108.5) |
| Married | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Unmarried | -3.6 (-6.0, -1.1) | 2.9 (0.0, 5.8) | -0.5 (-3.1, 2.0) |
| **Household Income** | | | |
| Intercept | 105.6 (102.0, 109.1) | 103.9 (99.7, 108.1) | 105.1 (101.4, 108.8) |
| <40 000 | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| 40 000 - 80 000 | 1.6 (-2.5, 5.7) | -2.0 (-6.8, 2.8) | -0.2 (-4.5, 4.1) |

| | | | |
|---|---|---|---|
| 80 000 - 100 000 | 4.2 (-0.1, 8.5) | -1.4 (-6.4, 3.6) | 1.9 (-2.6, 6.3) |
| >100 000 | 6.3 (2.4, 10.2) | 0.1 (-4.5, 4.7) | 3.9 (-0.3, 8.0) |
| **Parity** | | | |
| Intercept | 111.2 (109.5, 112.9) | 104.6 (102.6, 106.5) | 108.9 (107.1, 110.6) |
| 0 | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| 1 | -2.1 (-4.5, 0.3) | -2.3 (-5.1, 0.5) | -2.5 (-5.0, 0.0) |
| 2 | -5.0 (-8.6, -1.3) | -3.3 (-7.6, 0.9) | -4.8 (-8.6, -1.1) |
| 3+ | -8.6 (-14.9, -2.2) | -4.6 (-11.9, 2.8) | -7.7 (-14.3, -1.1) |
| **Prenatal Smoking** | | | |
| Intercept | 109.8 (108.6, 111) | 103.4 (102.0, 104.7) | 107.4 (106.2, 108.6) |
| No | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Yes | -4.0 (-8.0, 0.0) | -3.7 (-8.3, 0.9) | -4.2 (-8.4, -0.1) |
| **Prenatal Alcohol** | | | |
| Intercept | 109.3 (108.1, 110.5) | 102.3 (100.9, 103.7) | 106.5 (105.2, 107.8) |
| No | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Yes | 1.0 (-2.0, 4.0) | 4.6 (1.2, 8.0) | 3.0 (-0.1, 6.1) |

***Table A.3.***

Unadjusted linear regression coefficients showing the associations between prenatal chemical exposures (two-fold increase in chemical exposure) and WPPSI-III scores in participants from the MIREC Study, with 95% confidence intervals (n = 517).

| | VIQ (95% CI) | PIQ (95% CI) | FSIQ (95% CI) |
|---|---|---|---|
| **Heavy Metals** | | | |
| Arsenic | -1.2 (-2.3, 0.0) | -0.4 (-1.7, 1.0) | -1.0 (-2.2, 0.2) |
| Cadmium | -1.0 (-2.1, 0.0) | -1.0 (-2.2, 0.2) | -1.2 (-2.3, -0.1) |
| Lead | -0.8 (-2.5, 0.8) | 1.2 (-0.7, 3.1) | 0.2 (-1.5, 1.9) |
| Manganese | -0.8 (-3.4, 1.7) | -1.4 (-4.4, 1.6) | -1.3 (-4.0, 1.4) |
| Mercury | 0.4 (-0.4, 1.2) | 0.7 (-0.2, 1.6) | 0.6 (-0.2, 1.4) |
| **OCPs** | | | |
| BBHC | 0.1 (-0.7, 0.9) | 0.0 (-0.9, 0.9) | 0.1 (-0.7, 0.9) |
| DDE | -0.2 (-1.3, 0.8) | -0.4 (-1.5, 0.8) | -0.3 (-1.4, 0.7) |
| Oxychlor | 1.5 (0.2, 2.8) | 1.3 (-0.2, 2.9) | 1.6 (0.2, 3.0) |
| Transnona | 1.2 (-0.1, 2.5) | 0.9 (-0.6, 2.4) | 1.1 (-0.2, 2.5) |
| Sum of OCP | 1.4 (0.0, 2.8) | 1.2 (-0.4, 2.8) | 1.4 (0.0, 2.9) |
| **OPPs** | | | |
| DEP | -0.5 (-1.4, 0.4) | -0.6 (-1.6, 0.5) | -0.6 (-1.6, 0.3) |
| DETP | 0.1 (-0.7, 1.0) | 0.5 (-0.5, 1.4) | 0.4 (-0.5, 1.2) |
| DMDTP | 0.7 (0.1, 1.3) | 0.3 (-0.4, 1.0) | 0.6 (0.0, 1.2) |
| DMP | 0.2 (-0.6, 1.0) | 0.1 (-0.8, 1.0) | 0.2 (-0.6, 1.0) |
| DMTP | 0.5 (-0.1, 1.1) | 0.1 (-0.5, 0.8) | 0.4 (-0.2, 1.0) |
| **Phthalates** | | | |
| MBP | -0.5 (-1.5, 0.4) | -1.1 (-2.1, 0.0) | -0.9 (-1.9, 0.1) |
| MBZP | -0.3 (-1.2, 0.5) | 0.1 (-0.9, 1.1) | -0.2 (-1.1, 0.7) |
| MCPP | -0.4 (-1.1, 0.2) | -1.1 (-1.9, -0.4) | -0.9 (-1.5, -0.2) |
| MEHHP | -0.2 (-1.2, 0.8) | -0.7 (-1.9, 0.4) | -0.5 (-1.5, 0.6) |
| MEHP | -0.6 (-1.6, 0.4) | -0.6 (-1.7, 0.5) | -0.6 (-1.7, 0.4) |
| MEOHP | -0.2 (-1.3, 0.9) | -0.8 (-2.1, 0.4) | -0.5 (-1.6, 0.6) |
| MEP | -0.4 (-1.0, 0.2) | -0.1 (-0.7, 0.6) | -0.3 (-0.9, 0.3) |
| Sum of DEHP | -0.3 (-1.3, 0.8) | -0.8 (-2.0, 0.5) | -0.5 (-1.6, 0.6) |

| | | | |
|---|---|---|---|
| **PCBs** | | | |
| Aroclor | 0.5 (-0.6, 1.6) | 0.5 (-0.8, 1.8) | 0.5 (-0.6, 1.7) |
| PCB118 | 0.4 (-0.9, 1.6) | 1.4 (0.0, 2.8) | 1.0 (-0.3, 2.2) |
| PCB138 | 0.4 (-0.7, 1.5) | 0.1 (-1.1, 1.4) | 0.3 (-0.9, 1.4) |
| PCB153 | 0.6 (-0.6, 1.7) | 0.6 (-0.7, 1.9) | 0.6 (-0.5, 1.8) |
| PCB170 | 0.6 (-0.4, 1.6) | 0.9 (-0.3, 2.1) | 0.8 (-0.2, 1.9) |
| PCB180 | 0.9 (-0.1, 1.9) | 0.8 (-0.4, 1.9) | 0.9 (-0.1, 2.0) |
| PCB187 | 1.0 (-0.2, 2.1) | 0.5 (-0.8, 1.8) | 0.8 (-0.3, 2.0) |
| Sum of PCB | 0.6 (-0.6, 1.8) | 0.7 (-0.6, 2.1) | 0.7 (-0.5, 1.9) |
| **PBDEs** | | | |
| BDE47 | -0.8 (-1.6, 0.0) | -0.3 (-1.2, 0.6) | -0.6 (-1.4, 0.2) |
| **Smoking Metabolites** | | | |
| Cotinine | -0.4 (-0.7, -0.1) | -0.3 (-0.6, 0.1) | -0.4 (-0.7, -0.1) |

# Appendix B. Supplementary Materials for Chapter 2

***Table B.1.***

Covariate-adjusted linear regression coefficients showing the relationship between participant demographic characteristics and mean WPPSI-III scores, adjusted for child sex, maternal age, race, and education, marital status, household income, and prenatal smoking and alcohol, with 95% confidence intervals (n = 517).

| | VIQ (95% CI) | PIQ (95% CI) | FSIQ (95% CI) |
|---|---|---|---|
| **Intercept** | 105.4 | 97.1 | 101.3 |
| **Child Sex** | | | |
| Male | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Female | 4.7 (2.6, 6.8) | 4.0 (1.4, 6.5) | 5.0 (2.8, 7.2) |
| **Maternal Age** | | | |
| 19-29 | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| 30-34 | -0.7 (-3.8, 2.4) | -2.2 (-5.9, 1.4) | -1.8 (-5.0, 1.5) |
| 35+ | 1.5 (-1.7, 4.7) | -3.7 (-7.6, 0.1) | -1.3 (-4.7, 2.1) |
| **Maternal Race** | | | |
| White | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Other | -3.4 (-6.6, -0.2) | -1.6 (-5.4, 2.3) | -2.8 (-6.1, 0.5) |
| **Maternal Education** | | | |
| Undergrad | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| College | 0.5 (-4.7, 5.7) | 6.3 (0.0, 12.6) | 3.5 (-1.9, 9.0) |
| Undergrad | 4.6 (-0.6, 9.8) | 9.9 (3.6, 16.1) | 8.2 (2.7, 13.6) |
| Grad | 7.4 (2.0, 12.8) | 10.6 (4.1, 17.1) | 10.1 (4.5, 15.8) |
| **Marital Status** | | | |
| Married | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Unmarried | -2.3 (-4.8, 0.2) | 3.2 (0.2, 6.2) | 0.4 (-2.3, 3.0) |
| **Household Income** | | | |
| <40 000 | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| 40 000 - 80 000 | -1.2 (-5.4, 3.0) | -3.7 (-8.7, 1.3) | -2.7 (-7.1, 1.7) |
| 80 000 - 100 000 | 0.6 (-3.8, 5.0) | -2.6 (-7.9, 2.6) | -0.9 (-5.5, 3.7) |
| >100 000 | 0.6 (-3.7, 4.9) | -2.2 (-7.4, 3.0) | -0.7 (-5.2, 3.8) |
| **Parity** | | | |
| 0 | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |

| | | | |
|---|---|---|---|
| 1 | -2.3 (-4.7, 0.0) | -1.2 (-4.0, 1.7) | -2.0 (-4.4, 0.5) |
| 2 | -4.8 (-8.4, -1.3) | -2.3 (-6.6, 2.0) | -4.1 (-7.9, -0.3) |
| 3+ | -7.9 (-14.1, -1.8) | -1.9 (-9.2, 5.5) | -5.8 (-12.2, 0.6) |
| **Prenatal Smoking** | | | |
| No | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Yes | -2.3 (-6.2, 1.5) | -4.9 (-9.5, -0.2) | -3.8 (-7.9, 0.2) |
| **Prenatal Alcohol** | | | |
| No | 0.0 (ref) | 0.0 (ref) | 0.0 (ref) |
| Yes | 2.1 (-0.8, 4.9) | 5.0 (1.6, 8.4) | 3.9 (0.9, 6.8) |

***Table B.2.***

Covariate-adjusted linear regression coefficients showing the associations between individual chemical biomarkers (2-fold increase in chemical concentration) and WPPSI-III scores in MIREC participants, adjusted for maternal age, race, education, marital status, household income, parity, and prenatal smoking and alcohol, and stratified by child sex, with 95% confidence intervals (n = 517).

| | VIQ (95% CI) | PIQ (95% CI) | FSIQ (95% CI) |
|---|---|---|---|
| **Heavy Metals** | | | |
| *Arsenic* | | | |
| All | -1.3 (-2.5, -0.2) | -0.9 (-2.2, 0.5) | -1.3 (-2.5, -0.1) |
| Boys | -1.6 (-3.3, 0.1) | -0.4 (-2.3, 1.5) | -1.2 (-3.0, 0.6) |
| Girls | -0.7 (-2.2, 0.9) | -0.7 (-2.7, 1.3) | -0.8 (-2.4, 0.8) |
| *Cadmium* | | | |
| All | -1.0 (-2.1, 0.2) | -0.8 (-2.2, 0.5) | -1.1 (-2.3, 0.1) |
| Boys | -1.7 (-3.5, 0.2) | -1.6 (-3.8, 0.5) | -1.9 (-3.9, 0.0) |
| Girls | -0.2 (-1.6, 1.2) | -0.1 (-2.0, 1.7) | -0.2 (-1.7, 1.3) |
| *Lead* | | | |
| All | -1.5 (-3.1, 0.2) | 0.4 (-1.6, 2.3) | -0.6 (-2.4, 1.1) |
| Boys | -0.9 (-3.4, 1.6) | 0.8 (-2.1, 3.7) | 0.1 (-2.6, 2.7) |
| Girls | -2.5 (-4.7, -0.3) | -0.2 (-3.0, 2.6) | -1.6 (-3.9, 0.7) |
| *Manganese* | | | |
| All | -0.1 (-2.6, 2.4) | -0.6 (-3.5, 2.4) | -0.4 (-3.0, 2.2) |
| Boys | -0.5 (-4.1, 3.1) | -0.8 (-5.0, 3.4) | -0.8 (-4.6, 3.0) |
| Girls | -0.3 (-3.7, 3.1) | -1.0 (-5.3, 3.3) | -0.8 (-4.4, 2.8) |
| *Mercury* | | | |
| All | -0.1 (-0.9, 0.7) | 0.5 (-0.4, 1.5) | 0.2 (-0.6, 1.1) |
| Boys | 0.2 (-1.1, 1.4) | 0.7 (-0.8, 2.1) | 0.4 (-0.9, 1.8) |
| Girls | -0.1 (-1.1, 0.9) | 0.8 (-0.5, 2.1) | 0.3 (-0.7, 1.4) |
| **OCPs** | | | |
| *BBHC* | | | |
| All | -0.4 (-1.2, 0.4) | 0.1 (-0.9, 1.1) | -0.2 (-1.0, 0.7) |
| Boys | -0.5 (-1.8, 0.8) | 0.1 (-1.4, 1.6) | -0.2 (-1.6, 1.2) |

| | | | |
|---|---|---|---|
| Girls | -0.4 (-1.5, 0.7) | -0.1 (-1.5, 1.3) | -0.3 (-1.4, 0.9) |
| *DDE* | | | |
| All | -1.0 (-2.1, 0.1) | -0.3 (-1.6, 1.1) | -0.7 (-1.9, 0.4) |
| Boys | -1.5 (-3.2, 0.2) | -0.4 (-2.4, 1.6) | -1.2 (-3.0, 0.6) |
| Girls | -0.4 (-1.9, 1.0) | -0.2 (-2.1, 1.7) | -0.3 (-1.9, 1.2) |
| *Oxychlor* | | | |
| All | -0.1 (-1.5, 1.4) | 1.2 (-0.5, 2.9) | 0.6 (-0.9, 2.1) |
| Boys | -0.5 (-3.0, 2.0) | 1.9 (-1.0, 4.7) | 0.7 (-1.9, 3.3) |
| Girls | 0.2 (-1.6, 2.0) | 1.2 (-1.1, 3.5) | 0.7 (-1.2, 2.6) |
| *Transnona* | | | |
| All | -0.7 (-2.1, 0.7) | 0.6 (-1.1, 2.3) | -0.2 (-1.7, 1.3) |
| Boys | -1.6 (-3.8, 0.7) | 0.0 (-2.6, 2.7) | -1.0 (-3.4, 1.4) |
| Girls | -0.4 (-2.2, 1.5) | 1.2 (-1.1, 3.6) | 0.3 (-1.6, 2.3) |
| *Sum of OCP* | | | |
| All | -0.6 (-2.2, 0.9) | 0.9 (-1.0, 2.8) | 0.0 (-1.6, 1.7) |
| Boys | -1.4 (-4.0, 1.1) | 0.7 (-2.2, 3.6) | -0.5 (-3.2, 2.1) |
| Girls | -0.3 (-2.3, 1.8) | 1.5 (-1.1, 4.1) | 0.5 (-1.6, 2.7) |
| **OPPs** | | | |
| *DEP* | | | |
| All | -0.3 (-1.1, 0.6) | -0.6 (-1.7, 0.4) | -0.5 (-1.4, 0.4) |
| Boys | -1.3 (-2.8, 0.1) | -0.7 (-2.4, 1.0) | -1.2 (-2.7, 0.4) |
| Girls | 0.4 (-0.6, 1.5) | -0.6 (-1.9, 0.8) | -0.1 (-1.1, 1.0) |
| *DETP* | | | |
| All | 0.3 (-0.5, 1.1) | 0.5 (-0.5, 1.4) | 0.4 (-0.4, 1.3) |
| Boys | -0.5 (-1.7, 0.7) | -0.2 (-1.6, 1.1) | -0.4 (-1.6, 0.9) |
| Girls | 1.0 (-0.1, 2.1) | 1.3 (-0.1, 2.7) | 1.3 (0.1, 2.4) |
| *DMDTP* | | | |
| All | 0.4 (-0.2, 1.0) | 0.1 (-0.6, 0.7) | 0.3 (-0.3, 0.9) |
| Boys | 0.6 (-0.4, 1.6) | -0.2 (-1.3, 0.9) | 0.3 (-0.7, 1.3) |
| Girls | 0.1 (-0.6, 0.8) | 0.1 (-0.8, 1.0) | 0.1 (-0.6, 0.9) |
| *DMP* | | | |
| All | 0.2 (-0.6, 0.9) | -0.2 (-1.1, 0.7) | 0.0 (-0.8, 0.8) |

| | | | |
|---|---|---|---|
| Boys | -0.2 (-1.4, 1.0) | -0.1 (-1.4, 1.3) | -0.1 (-1.4, 1.1) |
| Girls | 0.4 (-0.6, 1.4) | -0.5 (-1.8, 0.7) | 0.0 (-1.0, 1.1) |
| *DMTP* | | | |
| All | 0.2 (-0.3, 0.8) | -0.2 (-0.9, 0.5) | 0.0 (-0.5, 0.6) |
| Boys | 0.2 (-0.6, 1.1) | -0.1 (-1.2, 0.9) | 0.1 (-0.8, 1.0) |
| Girls | 0.3 (-0.4, 0.9) | -0.2 (-1.1, 0.6) | 0.0 (-0.7, 0.7) |
| **Phthalates** | | | |
| *MBP* | | | |
| All | -0.3 (-1.2, 0.6) | -1.0 (-2.1, 0.0) | -0.8 (-1.7, 0.2) |
| Boys | -0.9 (-2.3, 0.5) | -1.6 (-3.2, -0.1) | -1.4 (-2.9, 0.0) |
| Girls | 0.7 (-0.5, 1.9) | -0.2 (-1.7, 1.3) | 0.3 (-1.0, 1.5) |
| *MBZP* | | | |
| All | 0.0 (-0.8, 0.8) | 0.1 (-0.9, 1.1) | 0.0 (-0.8, 0.9) |
| Boys | 0.1 (-1.2, 1.4) | 0.3 (-1.2, 1.8) | 0.2 (-1.1, 1.6) |
| Girls | 0.0 (-1.0, 1.1) | 0.0 (-1.4, 1.3) | -0.1 (-1.2, 1.0) |
| *MCPP* | | | |
| All | -0.2 (-0.8, 0.4) | -0.9 (-1.7, -0.2) | -0.6 (-1.3, 0.0) |
| Boys | -0.1 (-1.0, 0.8) | -0.7 (-1.7, 0.4) | -0.4 (-1.3, 0.5) |
| Girls | 0.1 (-0.7, 1.0) | -0.9 (-2.0, 0.2) | -0.4 (-1.3, 0.5) |
| *MEHHP* | | | |
| All | -0.3 (-1.3, 0.6) | -0.9 (-2.0, 0.3) | -0.6 (-1.6, 0.4) |
| Boys | -0.5 (-1.9, 1.0) | -1.0 (-2.7, 0.7) | -0.8 (-2.3, 0.8) |
| Girls | 0.4 (-0.8, 1.7) | -0.3 (-1.9, 1.3) | 0.1 (-1.3, 1.4) |
| *MEHP* | | | |
| All | -0.7 (-1.7, 0.3) | -0.8 (-1.9, 0.3) | -0.8 (-1.8, 0.2) |
| Boys | -0.5 (-1.9, 0.9) | -0.6 (-2.3, 1.0) | -0.6 (-2.1, 0.9) |
| Girls | -0.2 (-1.5, 1.1) | -0.6 (-2.2, 1.0) | -0.5 (-1.8, 0.9) |
| *MEOHP* | | | |
| All | -0.4 (-1.4, 0.7) | -1.0 (-2.3, 0.2) | -0.8 (-1.9, 0.3) |
| Boys | -0.3 (-1.9, 1.3) | -1.1 (-2.9, 0.7) | -0.7 (-2.4, 1.0) |
| Girls | 0.3 (-1.1, 1.7) | -0.6 (-2.3, 1.2) | -0.2 (-1.6, 1.3) |
| *MEP* | | | |

| | | | |
|---|---|---|---|
| All | -0.2 (-0.8, 0.3) | 0.0 (-0.6, 0.7) | -0.1 (-0.7, 0.4) |
| Boys | 0.7 (-0.2, 1.6) | 0.4 (-0.7, 1.4) | 0.7 (-0.3, 1.6) |
| Girls | -0.8 (-1.5, -0.2) | -0.2 (-1.1, 0.6) | -0.6 (-1.3, 0.1) |
| *Sum of DEHP* | | | |
| All | -0.4 (-1.5, 0.6) | -1.0 (-2.2, 0.3) | -0.7 (-1.8, 0.3) |
| Boys | -0.5 (-2.0, 1.1) | -1.1 (-2.8, 0.7) | -0.8 (-2.4, 0.8) |
| Girls | 0.3 (-1.0, 1.7) | -0.4 (-2.2, 1.3) | -0.1 (-1.5, 1.4) |
| **PCBs** | | | |
| *Aroclor* | | | |
| All | -0.9 (-2.1, 0.3) | 0.5 (-0.9, 1.9) | -0.3 (-1.5, 1.0) |
| Boys | -1.1 (-2.9, 0.8) | 0.5 (-1.6, 2.7) | -0.4 (-2.3, 1.6) |
| Girls | -1.1 (-2.7, 0.4) | 0.3 (-1.7, 2.2) | -0.5 (-2.2, 1.1) |
| *PCB118* | | | |
| All | -1.1 (-2.4, 0.2) | 1.0 (-0.5, 2.5) | -0.1 (-1.5, 1.2) |
| Boys | -1.7 (-3.6, 0.2) | 0.4 (-1.8, 2.6) | -0.8 (-2.8, 1.2) |
| Girls | -0.6 (-2.2, 1.1) | 1.5 (-0.6, 3.6) | 0.5 (-1.3, 2.2) |
| *PCB138* | | | |
| All | -1.0 (-2.2, 0.2) | 0.0 (-1.4, 1.4) | -0.6 (-1.8, 0.6) |
| Boys | -1.2 (-3.1, 0.8) | 0.2 (-2.0, 2.4) | -0.6 (-2.7, 1.4) |
| Girls | -1.0 (-2.4, 0.5) | -0.3 (-2.1, 1.6) | -0.7 (-2.3, 0.8) |
| *PCB153* | | | |
| All | -0.8 (-2.0, 0.4) | 0.6 (-0.8, 2.1) | -0.2 (-1.4, 1.1) |
| Boys | -0.9 (-2.8, 1.0) | 0.5 (-1.7, 2.7) | -0.3 (-2.3, 1.7) |
| Girls | -1.1 (-2.7, 0.4) | 0.5 (-1.4, 2.5) | -0.4 (-2.0, 1.2) |
| PCB170 | | | |
| All | -0.3 (-1.4, 0.7) | 1.1 (-0.2, 2.3) | 0.4 (-0.7, 1.5) |
| Boys | -0.5 (-2.2, 1.1) | 0.7 (-1.2, 2.5) | 0.0 (-1.7, 1.8) |
| Girls | -0.6 (-2.0, 0.7) | 1.1 (-0.7, 2.8) | 0.2 (-1.2, 1.6) |
| *PCB180* | | | |
| All | -0.2 (-1.3, 0.8) | 0.9 (-0.3, 2.2) | 0.4 (-0.8, 1.5) |
| Boys | -0.3 (-2.0, 1.4) | 1.1 (-0.9, 3.0) | 0.4 (-1.4, 2.2) |
| Girls | -0.7 (-2.0, 0.7) | 0.5 (-1.2, 2.2) | -0.1 (-1.5, 1.3) |

| | | | |
|---|---|---|---|
| *PCB187* | | | |
| All | 0.2 (-1.0, 1.4) | 0.8 (-0.6, 2.2) | 0.6 (-0.7, 1.8) |
| Boys | 0.3 (-1.6, 2.1) | 0.5 (-1.6, 2.7) | 0.4 (-1.6, 2.4) |
| Girls | -0.2 (-1.7, 1.3) | 0.9 (-0.9, 2.8) | 0.4 (-1.2, 1.9) |
| *Sum of PCBs* | | | |
| All | -0.9 (-2.1, 0.4) | 0.8 (-0.7, 2.2) | -0.1 (-1.4, 1.2) |
| Boys | -1.1 (-3.0, 0.9) | 0.6 (-1.6, 2.9) | -0.3 (-2.4, 1.8) |
| Girls | -1.0 (-2.6, 0.5) | 0.6 (-1.4, 2.6) | -0.3 (-2.0, 1.4) |
| **PBDEs** | | | |
| *BDE47* | | | |
| All | -0.6 (-1.3, 0.2) | -0.2 (-1.1, 0.7) | -0.4 (-1.2, 0.4) |
| Boys | -1.0 (-2.1, 0.2) | -0.2 (-1.6, 1.2) | -0.6 (-1.9, 0.6) |
| Girls | 0.1 (-0.8, 1.1) | 0.0 (-1.2, 1.3) | 0.1 (-0.9, 1.2) |
| **Smoking Metabolites** | | | |
| *Cotinine* | | | |
| All | -0.1 (-0.5, 0.2) | -0.1 (-0.6, 0.3) | -0.2 (-0.5, 0.2) |
| Boys | -0.1 (-0.7, 0.4) | 0.0 (-0.7, 0.6) | -0.1 (-0.7, 0.5) |
| Girls | -0.1 (-0.6, 0.4) | -0.1 (-0.7, 0.5) | -0.1 (-0.6, 0.3) |

***Table B.3.***

Demographic characteristics for the total study population and for participants in each cluster generated by k-means clustering.

| | Total | Ref | High Level | Low Level | High OPPs | Smoking Chemicals |
|---|---|---|---|---|---|---|
| | *n = 517* | *n = 119* | *n = 111* | *n = 131* | *n = 138* | *n = 18* |
| **Child Sex** | | | | | | |
| Male | 49 | 46 | 47 | 56 | 46 | 50 |
| Female | 51 | 54 | 53 | 44 | 54 | 50 |
| **Maternal Age** | | | | | | |
| 19-30 | 21 | 13 | 7 | 30 | 27 | 39 |
| 30-35 | 39 | 41 | 30 | 37 | 47 | 28 |
| 35+ | 41 | 46 | 63 | 34 | 26 | 33 |
| **Maternal Race** | | | | | | |
| White | 86 | 93 | 77 | 85 | 86 | 89 |
| Other | 14 | 7 | 23 | 15 | 14 | 11 |
| **Maternal Education** | | | | | | |
| Highschool | 5 | 3 | 2 | 10 | 4 | 22 |
| College | 27 | 19 | 23 | 35 | 29 | 39 |
| Undergrad | 39 | 45 | 36 | 31 | 43 | 39 |
| Grad | 29 | 34 | 40 | 24 | 24 | 0 |
| **Marital Status** | | | | | | |
| Married | 72 | 84 | 72 | 69 | 68 | 50 |
| Unmarried | 28 | 16 | 28 | 31 | 32 | 50 |
| **Household Income** | | | | | | |
| < 40 000 | 10 | 4 | 9 | 13 | 8 | 39 |
| 40 000 - 80 000 | 29 | 24 | 26 | 37 | 30 | 17 |
| 80 000 - 100 000 | 21 | 26 | 17 | 18 | 23 | 11 |
| > 100 000 | 41 | 46 | 48 | 32 | 39 | 33 |
| **Parity** | | | | | | |
| 0 | 44 | 34 | 55 | 38 | 47 | 50 |
| 1 | 41 | 48 | 37 | 44 | 38 | 33 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | 12 | 14 | 7 | 13 | 12 | 17 |
| 3+ | 3 | 4 | 1 | 5 | 4 | 0 |
| **Prenatal Smoking** | | | | | | |
| No | 91 | 97 | 92 | 93 | 95 | 11 |
| Yes | 9 | 3 | 8 | 7 | 5 | 89 |
| **Prenatal Alcohol** | | | | | | |
| No | 83 | 83 | 78 | 90 | 83 | 56 |
| Yes | 17 | 17 | 22 | 10 | 17 | 44 |

### *Table B.4.*

Covariate-adjusted linear regression coefficients showing the associations between membership in k-means clusters and WPPSI-III scores, adjusted for maternal age, race, education, marital status, household income, parity, and prenatal alcohol, compared to medium level Reference class membership. Results shown for all children, and then stratified by sex, with 95% confidence intervals (n = 517).

| | VIQ (95% CI) | PIQ (95% CI) | FSIQ (95% CI) |
|---|---|---|---|
| **Intercept** | | | |
| All | 105.0 | 99.1 | 102.4 |
| Boys | 103.2 | 99.3 | 101.5 |
| Girls | 106.6 | 98.0 | 102.4 |
| **High Level** | | | |
| All | -1.3 (-4.6, 2.0) | 0.1 (-3.8, 4.0) | -0.8 (-4.2, 2.7) |
| Boys | -2.2 (-7.3, 3.0) | 0.9 (-5.1, 6.8) | -0.7 (-6.2, 4.7) |
| Girls | -0.3 (-4.5, 3.9) | -1.1 (-6.4, 4.2) | -0.8 (-5.4, 3.4) |
| **Low Level** | | | |
| All | 0.3 (-2.9, 3.4) | -2.2 (-6.0, 1.5) | -1.1 (-4.4, 2.3) |
| Boys | 0.6 (-4.2, 5.4) | -1.3 (-6.9, 4.2) | -0.3 (-5.4, 4.8) |
| Girls | 0.7 (-3.5, 4.9) | -2.6 (-7.9, 2.8) | -1.1 (-5.6, 3.3) |
| **High OPPs** | | | |
| All | 0.4 (-2.7, 3.5) | -1.9 (-5.6, 1.7) | -0.7 (-3.9, 2.6) |
| Boys | -1.9 (-6.9, 3.1) | -3.2 (-9.0, 2.6) | -2.7 (-8.1, 2.6) |
| Girls | 2.4 (-1.4, 6.3) | -0.7 (-5.6, 4.1) | -0.7 (-2.9, 5.1) |
| **Smoking Chemicals** | | | |
| All | -2.6 (-9.0, 3.9) | -7.3 (-14.9, 0.3) | -5.6 (-12.4, 1.1) |
| Boys | -3.5 (-13.4, 6.3) | -5.3 (-16.7, 6.0) | -5.0 (-15.4, 5.4) |
| Girls | -1.6 (-9.9, 6.6) | -9.0 (-19.5, 1.6) | -5.6 (-14.7, 2.7) |

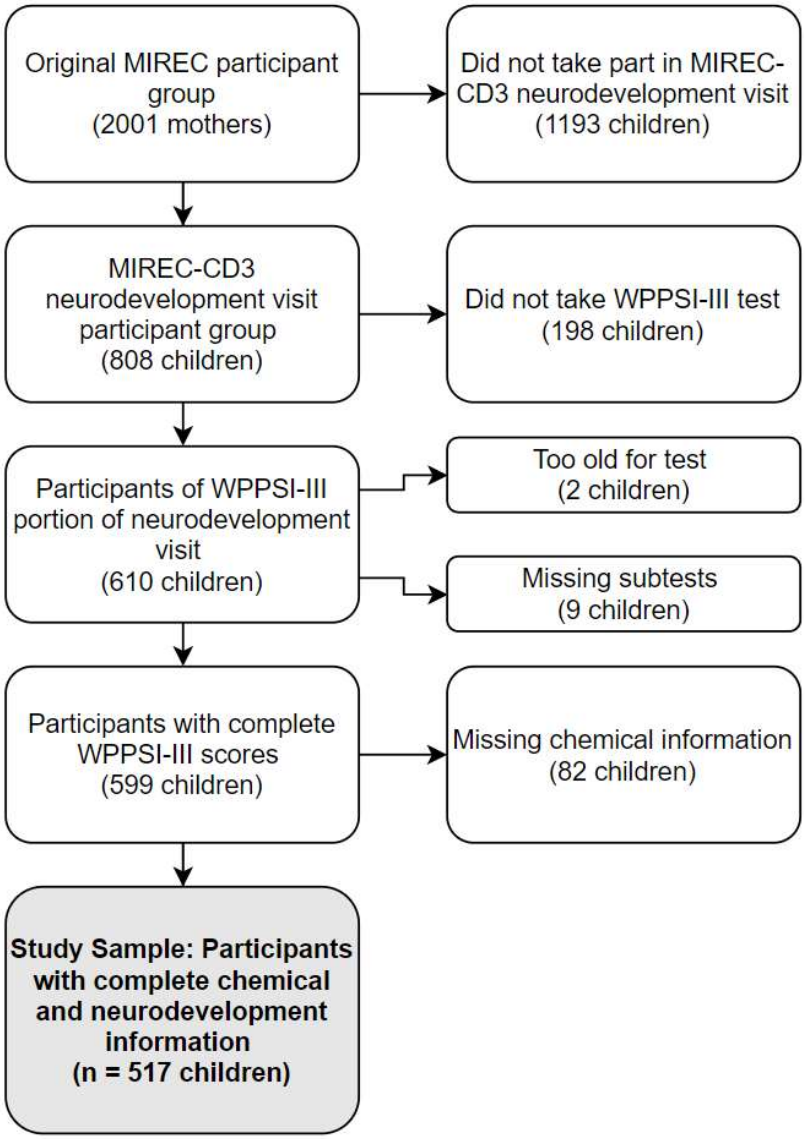


***Figure B.1.***

Study sample flow chart.

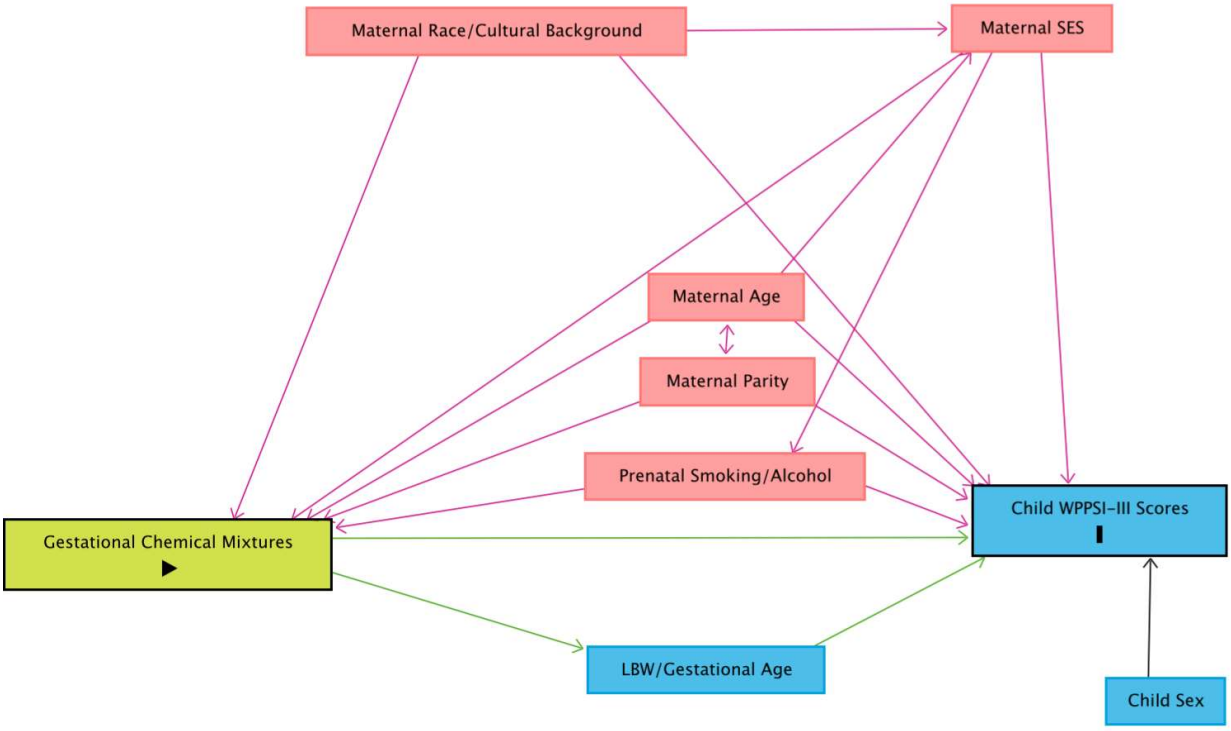


***Figure B.2.***

Directed acyclic graph (DAG) showing confounders and mediators of the effects of gestational chemical mixtures on child WPPSI-III scores.

# Appendix C. LPA Code

```{r basic prep}

# packages
library(mclust)
library(tidyLPA)
library(dplyr)
library(plyr)
library(ggplot2)
library(pheatmap)
library(readr)
library(readxl)
library(Weighted.Desc.Stat)
library(ggpubr)

# preparing the data:

all_data <- read_csv("AY_MIREC_Analysis_July22_2020.csv")
all_data <- filter(all_data, sex2!="NA")
chem_data <- all_data[,c(2:10, 12:23, 25:31, 33, 38)]
chem_data <- rename(chem_data, c("log2.arsenic.t1" = "As", "log2.cadmium.t1"
= "Cd", "log2.lead.t1" =  "Pb", "log2.manganese.t1" =  "Mn",
"log2.mercury.t1" = "Hg", "log2.bbhc.t1" = "BBHC", "log2.dde.t1" = "DDE",
"log2.oxychlor.t1" = "Oxy.", "log2.transnona.t1" = "Trans.", "log2.dep.t1" =
"DEP", "log2.detp.t1" = "DETP", "log2.dmdtp.t1" = "DMDTP", "log2.dmp.t1" =
"DMP","log2.dmtp.t1" =  "DMTP", "log2.mbp.t1" = "MBP", "log2.mbzp.t1" =
"MBZP","log2.mcpp.t1" =  "MCPP","log2.mehhp.t1" =  "MEHHP","log2.mehp.t1" =
"MEHP", "log2.meohp.t1" = "MEOHP","log2.mep.t1" =  "MEP", "log2.aroclor.t1" =
"Aro.", "log2.pcb118.t1" = "PCB118", "log2.pcb138.t1" = "PCB138",
"log2.pcb153.t1" = "PCB153", "log2.pcb170.t1" = "PCB170", "log2.pcb180.t1" =
"PCB180", "log2.pcb187.t1" = "PCB187", "log2.bde47.t1" = "BDE47",
"log2.cot.t1" = "Cot."))
chemicals <- c("As", "Cd", "Pb", "Mn", "Hg", "BBHC", "DDE", "Oxy.", "Trans.",
"DEP", "DETP", "DMDTP", "DMP", "DMTP", "MBP", "MBZP", "MCPP", "MEHHP",
"MEHP", "MEOHP", "MEP", "Aro.", "PCB118", "PCB138", "PCB153", "PCB170",
"PCB180", "PCB187", "BDE47", "Cot.")

```

```{r choosing number of profiles and assumptions about variance/covariance}

# Model 1 = equal variance and covariance set to zero
chem_data %>%
  estimate_profiles(models = 1, n_profiles = c(1:12)) %>%
  compare_solutions(statistics = c("BIC", "AIC"))

# Model 2 = varying variance and covariance set to zero
chem_data %>%
  estimate_profiles(models = 2, n_profiles = c(1:10)) %>%
  compare_solutions(statistics = c("BIC", "AIC"))

# Model 3 = equal variance and equal covariance
chem_data %>%
  estimate_profiles(models = 3, n_profiles = c(1:8)) %>%
```

```
  compare_solutions(statistics = c("BIC", "AIC"))

# Model 4 = varying variance and varying covariance
chem_data %>%
  estimate_profiles(models = 4, n_profiles = c(1:5)) %>%
  compare_solutions(statistics = c("BIC", "AIC"))

# The BIC chose Model 3 with 5 profiles and Model 4 with 2 profiles. For
interpretation I'll be using the one with 5 profiles.

```

```{r estimating profiles and making the dataset}

# Generating profiles (~5 minutes):

profile.estimate <- (chem_data %>%
           estimate_profiles(models=3, n_profiles = 5))

# finding the posterior probabilities:
probabilities <- get_data(profile.estimate)
probabilities <- data.frame( "class" = probabilities$Class, "prob_ref" =
probabilities$CPROB2, "prob_high" = probabilities$CPROB1, "prob_low" =
probabilities$CPROB5, "prob_opp" = probabilities$CPROB4, "prob_smoking" =
probabilities$CPROB3, "child_sex" = all_data$sex2)
probabilities <- filter(probabilities, child_sex!="NA")

# making the final dataset:
final_chem <- data.frame("subject_id" = all_data$subject.id, "class" =
probabilities$class, "prob_ref" = probabilities$prob_ref, "prob_high" =
probabilities$prob_high, "prob_low" = probabilities$prob_low, "prob_opp" =
probabilities$prob_opp, "prob_smoking" = probabilities$prob_smoking,
"arsenic" = all_data$log2.arsenic.t1, "cadmium" = all_data$log2.cadmium.t1,
"lead" = all_data$log2.lead.t1,"manganese" = all_data$log2.manganese.t1,
"mercury" = all_data$log2.mercury.t1, "bbhc" = all_data$log2.bbhc.t1, "dde" =
all_data$log2.dde.t1, "oxychlor" = all_data$log2.oxychlor.t1, "transnona" =
all_data$log2.transnona.t1, "dep" = all_data$log2.dep.t1, "detp" =
all_data$log2.detp.t1, "dmdtp" = all_data$log2.dmdtp.t1, "dmp" =
all_data$log2.dmp.t1, "dmtp" = all_data$log2.dmtp.t1, "mbp" =
all_data$log2.mbp.t1, "mbzp" = all_data$log2.mbzp.t1, "mcpp" =
all_data$log2.mcpp.t1, "mehhp" = all_data$log2.mehhp.t1, "mehp" =
all_data$log2.mehp.t1, "meohp" = all_data$log2.meohp.t1, "mep" =
all_data$log2.mep.t1, "aroclor" = all_data$log2.aroclor.t1, "pcb118" =
all_data$log2.pcb118.t1, "pcb138" = all_data$log2.pcb138.t1, "pcb153" =
all_data$log2.pcb153.t1, "pcb170" = all_data$log2.pcb170.t1, "pcb180" =
all_data$log2.pcb180.t1, "pcb187" = all_data$log2.pcb187.t1, "bde47" =
all_data$log2.bde47.t1, "cot" = all_data$log2.cot.t1, "birth_length" =
all_data$birth.length, "birth_weight" = all_data$birth.wt, "gest_age" =
all_data$gest.age, "preterm_birth" = all_data$preterm2, "lbw" =
imputed_data$lbw2, "lga" = all_data$lga2, "live_birth" =
all_data$live.birth2, "child_sex" = all_data$sex2, "small_for_ga" =
all_data$sga2, "alc" = imputed_data$alc2, "city" = all_data$city10, "couple"
= all_data$couple2, "mom_education" = imputed_data$edu4, "household_income" =
imputed_data$income4, "living_status" = all_data$living.status2, "married" =
all_data$married2, "maternal_age" = all_data$mom.age3, "mom_birthplace" =
all_data$mom.birthplace2, "maternal_obesity" = all_data$obese2, "parity" =
all_data$parity4, "prepreg_bmi" = all_data$prepreg.bmi4, "race_aboriginal" =
```

```
all_data$race.aboriginal2, "race_asian2" = all_data$race.asian2,
"race_black2" = all_data$race.black2, "race_latin2" = all_data$race.latin2,
"race_other2" = all_data$race.other2, "race_white2" = all_data$race.white2,
"test_site" = all_data$site11, "smoker" = all_data$smoker2, "wppsi_1" =
all_data$wppsi.1, "wppsi_2" = all_data$wppsi.2, "wppsi_3" = all_data$wppsi.3,
"wppsi_4" = all_data$wppsi.4, "wppsi_5" = all_data$wppsi.5, "viq" =
all_data$viq, "piq" = all_data$piq, "fsiq" = all_data$fsiq,
"general_language" = all_data$general.language)

final_chem <- filter(final_chem, child_sex!="NA")

```

```{r weighted mean values for each chemical in each profile}

# Calculating mean biomarker concentrations of each chemical weighted by
posterior probabilities:
weighted_mean_high <- weighted_mean_ref <- weighted_mean_smoking <-
weighted_mean_opp <- weighted_mean_low <- weighted_sd_high <- weighted_sd_ref
<- weighted_sd_smoking <- weighted_sd_opp <- weighted_sd_low <- rep(0,30)

for(w in 1:30){
  weighted_mean_high[w] <- weighted.mean(final_chem[,(7+w)],
final_chem$prob_high)
  weighted_sd_high[w] <- w.sd(final_chem[,(7+w)], final_chem$prob_high)
  weighted_mean_ref[w] <- weighted.mean(final_chem[,(7+w)],
final_chem$prob_ref)
  weighted_sd_ref[w] <- w.sd(final_chem[,(7+w)], final_chem$prob_ref)
  weighted_mean_smoking[w] <- weighted.mean(final_chem[,(7+w)],
final_chem$prob_smoking)
  weighted_sd_smoking[w] <- w.sd(final_chem[,(7+w)], final_chem$prob_smoking)
  weighted_mean_opp[w] <- weighted.mean(final_chem[,(7+w)],
final_chem$prob_opp)
  weighted_sd_opp[w] <- w.sd(final_chem[,(7+w)], final_chem$prob_opp)
  weighted_mean_low[w] <- weighted.mean(final_chem[,(7+w)],
final_chem$prob_low)
  weighted_sd_low[w] <- w.sd(final_chem[,(7+w)], final_chem$prob_low)
}

weighted_means <- data.frame(chemicals, weighted_mean_ref, weighted_sd_ref,
weighted_mean_high, weighted_sd_high, weighted_mean_low, weighted_sd_low,
weighted_mean_opp, weighted_sd_opp, weighted_mean_smoking,
weighted_sd_smoking)

```

```{r calculating z scores}

# Converting means to z-scores for heat map:
z_arsenic <- z_cadmium <- z_lead <- z_manganese <- z_mercury <- z_bbhc <-
z_dde <- z_oxychlor <- z_transnona <- z_dep <- z_detp <- z_dmdtp <- z_dmp <-
z_dmtp <- z_mbp <- z_mbzp <- z_mcpp <- z_mehhp <- z_mehp <- z_meohp <- z_mep
<- z_aroclor <- z_pcb118 <- z_pcb138 <- z_pcb153 <- z_pcb170 <- z_pcb180 <-
z_pcb187 <- z_bde47 <- z_cotinine <- rep(0,517)

for (z in 1:517){
```

```
  z_arsenic[z] <- (final_chem$arsenic[z]-
mean(final_chem$arsenic))/sd(final_chem$arsenic)
  z_cadmium[z] <- (final_chem$cadmium[z]-
mean(final_chem$cadmium))/sd(final_chem$cadmium)
  z_lead[z] <- (final_chem$lead[z]-mean(final_chem$lead))/sd(final_chem$lead)
  z_manganese[z] <- (final_chem$manganese[z]-
mean(final_chem$manganese))/sd(final_chem$manganese)
  z_mercury[z] <- (final_chem$mercury[z]-
mean(final_chem$mercury))/sd(final_chem$mercury)
  z_bbhc[z] <- (final_chem$bbhc[z]-mean(final_chem$bbhc))/sd(final_chem$bbhc)
  z_dde[z] <- (final_chem$dde[z]-mean(final_chem$dde))/sd(final_chem$dde)
  z_oxychlor[z] <- (final_chem$oxychlor[z]-
mean(final_chem$oxychlor))/sd(final_chem$oxychlor)
  z_transnona[z] <- (final_chem$transnona[z]-
mean(final_chem$transnona))/sd(final_chem$transnona)
  z_dep[z] <- (final_chem$dep[z]-mean(final_chem$dep))/sd(final_chem$dep)
  z_detp[z] <- (final_chem$detp[z]-mean(final_chem$detp))/sd(final_chem$detp)
  z_dmdtp[z] <- (final_chem$dmdtp[z]-
mean(final_chem$dmdtp))/sd(final_chem$dmdtp)
  z_dmp[z] <- (final_chem$dmp[z]-mean(final_chem$dmp))/sd(final_chem$dmp)
  z_dmtp[z] <- (final_chem$dmtp[z]-mean(final_chem$dmtp))/sd(final_chem$dmtp)
  z_mbp[z] <- (final_chem$mbp[z]-mean(final_chem$mbp))/sd(final_chem$mbp)
  z_mbzp[z] <- (final_chem$mbzp[z]-mean(final_chem$mbzp))/sd(final_chem$mbzp)
  z_mcpp[z] <- (final_chem$mcpp[z]-mean(final_chem$mcpp))/sd(final_chem$mcpp)
  z_mehhp[z] <- (final_chem$mehhp[z]-
mean(final_chem$mehhp))/sd(final_chem$mehhp)
  z_mehp[z] <- (final_chem$mehp[z]-mean(final_chem$mehp))/sd(final_chem$mehp)
  z_meohp[z] <- (final_chem$meohp[z]-
mean(final_chem$meohp))/sd(final_chem$meohp)
  z_mep[z] <- (final_chem$mep[z]-mean(final_chem$mep))/sd(final_chem$mep)
  z_aroclor[z] <- (final_chem$aroclor[z]-
mean(final_chem$aroclor))/sd(final_chem$aroclor)
  z_pcb118[z] <- (final_chem$pcb118[z]-
mean(final_chem$pcb118))/sd(final_chem$pcb118)
  z_pcb138[z] <- (final_chem$pcb138[z]-
mean(final_chem$pcb138))/sd(final_chem$pcb138)
  z_pcb153[z] <- (final_chem$pcb153[z]-
mean(final_chem$pcb153))/sd(final_chem$pcb153)
  z_pcb170[z] <- (final_chem$pcb170[z]-
mean(final_chem$pcb170))/sd(final_chem$pcb170)
  z_pcb180[z] <- (final_chem$pcb180[z]-
mean(final_chem$pcb180))/sd(final_chem$pcb180)
  z_pcb187[z] <- (final_chem$pcb187[z]-
mean(final_chem$pcb187))/sd(final_chem$pcb187)
  z_bde47[z] <- (final_chem$bde47[z]-
mean(final_chem$bde47))/sd(final_chem$bde47)
  z_cotinine[z] <- (final_chem$cot[z]-
mean(final_chem$cot))/sd(final_chem$cot)
}

z_scores <- data.frame(final_chem$subject_id, final_chem$class,
final_chem$prob_ref, final_chem$prob_high, final_chem$prob_low,
final_chem$prob_opp, final_chem$prob_smoking, z_arsenic, z_cadmium, z_lead,
z_manganese, z_mercury, z_bbhc, z_dde, z_oxychlor, z_transnona, z_dep,
z_detp, z_dmdtp, z_dmp, z_dmtp, z_mbp, z_mbzp, z_mcpp, z_mehhp, z_mehp,
z_meohp, z_mep, z_aroclor, z_pcb118, z_pcb138, z_pcb153, z_pcb170, z_pcb180,
z_pcb187, z_bde47, z_cotinine)
```

```
# Finding the weighted means of the z_scores
weighted_zmean_high <- weighted_zmean_ref <- weighted_zmean_smoking <-
weighted_zmean_opp <- weighted_zmean_low <- weighted_zsd_high <-
weighted_zsd_ref <- weighted_zsd_smoking <- weighted_zsd_opp <-
weighted_zsd_low <- rep(0,30)
for(w in 1:30){
  weighted_zmean_high[w] <- weighted.mean(z_scores[,(7+w)],
z_scores$final_chem.prob_high)
  weighted_zsd_high[w] <- w.sd(z_scores[,(7+w)],
z_scores$final_chem.prob_high)
  weighted_zmean_ref[w] <- weighted.mean(z_scores[,(7+w)],
z_scores$final_chem.prob_ref)
  weighted_zsd_ref[w] <- w.sd(z_scores[,(7+w)], z_scores$final_chem.prob_ref)
  weighted_zmean_smoking[w] <- weighted.mean(z_scores[,(7+w)],
z_scores$final_chem.prob_smoking)
  weighted_zsd_smoking[w] <- w.sd(z_scores[,(7+w)],
z_scores$final_chem.prob_smoking)
  weighted_zmean_opp[w] <- weighted.mean(z_scores[,(7+w)],
z_scores$final_chem.prob_opp)
  weighted_zsd_opp[w] <- w.sd(z_scores[,(7+w)], z_scores$final_chem.prob_opp)
  weighted_zmean_low[w] <- weighted.mean(z_scores[,(7+w)],
z_scores$final_chem.prob_low)
  weighted_zsd_low[w] <- w.sd(z_scores[,(7+w)], z_scores$final_chem.prob_low)
}

weighted_zmeans <- data.frame(chemicals, weighted_zmean_ref,
weighted_zsd_ref, weighted_zmean_high, weighted_zsd_high, weighted_zmean_low,
weighted_zsd_low, weighted_zmean_opp, weighted_zsd_opp,
weighted_zmean_smoking, weighted_zsd_smoking)

```

```{r heatmap of z scores}

# colour:
breaksList = seq(-3,3, by = 0.1)
col <- colorRampPalette(c("navy", "blue", "white", "red",
"red4"))(length(breaksList))

#labels:
x_lab <- c("Ref", "High Level", "Low Level", "High OPPs", "Smoking
Chemicals")

# using pheatmap because it lets me do a legend and play with the labels,
which heatmap and heatmap.2 don't let me do:

pheatmap(weighted_zmeans[,-c(1,3,5,7,9,11)],
        Colv=NA, Rowv=NA, cluster_rows = FALSE, cluster_cols = FALSE, #
getting rid of the the dendrogram and the clustering
        cellwidth = 20, # changing the size of the cells
        color=col, breaks = breaksList, # picking the colours
        labels_col = x_lab, labels_row = chemicals, angle_col = 45) # row and
column labels

```

```{r plotting the profiles}

# mean and sd biomarker concentrations in each profile

# heavy metals, cotinine, and bde47:

metals_bde_cot <- c("As", "Cd", "Pb", "Mn", "Hg", "BDE47", "Cot.")
metals_plot <- plot_profiles(profile.estimate, rawdata=FALSE, ci=NULL,
variables = metals_bde_cot) + rremove("xlab") + rremove("ylab") +
rremove("legend") + theme(text = element_text(size=9.75))
metals_plot

# OCPs and OPPs:

pesticides <- c( "BBHC", "DDE", "Oxy.", "Trans.", "DEP", "DETP", "DMDTP",
"DMP", "DMTP")
pesticides_plot <- plot_profiles(profile.estimate, rawdata=FALSE, ci=NULL,
variables = pesticides) + rremove("xlab") + rremove("ylab") +
rremove("legend") + theme(text = element_text(size=9.75))
pesticides_plot

# Phthalates:

phthalates <- c("MBP", "MBZP", "MCPP", "MEHHP","MEHP", "MEOHP", "MEP")
phthalates_plot <- plot_profiles(profile.estimate, rawdata=FALSE, ci=NULL,
variables = phthalates) + rremove("xlab") + rremove("ylab") +
rremove("legend") + theme(text = element_text(size=9.75))
phthalates_plot

# PCBs:

pcbs <- c("Aro.", "PCB118", "PCB138", "PCB153", "PCB170", "PCB180", "PCB187")
pcbs_plot <- plot_profiles(profile.estimate, rawdata=FALSE, ci=NULL,
variables = pcbs) + rremove("xlab") + rremove("ylab") + rremove("legend") +
theme(text = element_text(size=9.75))
pcbs_plot


ggarrange(metals_plot, pesticides_plot, phthalates_plot, pcbs_plot, nrow=2,
ncol=2)

```

```{r unadjusted probabilistic multiple regression analysis}

# regression analysis with all the probabilistic profiles in one model,
unadjusted

# VIQ
lm_unadj_all_viq <- lm(final_chem$viq ~ final_chem$prob_high +
final_chem$prob_low + final_chem$prob_opp + final_chem$prob_smoking +
final_chem$prob_ref)
summary(lm_unadj_all_viq)
lm_unadj_all_viq$coefficients

# PIQ
```

```
lm_unadj_all_piq <- lm(final_chem$piq ~ final_chem$prob_high +
final_chem$prob_low + final_chem$prob_opp + final_chem$prob_smoking +
final_chem$prob_ref)
summary(lm_unadj_all_piq)

# FSIQ
lm_unadj_all_fsiq <- lm(final_chem$fsiq ~ final_chem$prob_high +
final_chem$prob_low + final_chem$prob_opp + final_chem$prob_smoking +
final_chem$prob_ref)
summary(lm_unadj_all_fsiq)
```

```{r stratifying by sex}

final_boys <- filter(final_chem, child_sex==0)
final_girls <- filter(final_chem, child_sex==1)

# VIQ for boys:

lm_unadj_all_boys_viq <- lm(final_boys$viq ~ final_boys$prob_high +
final_boys$prob_low  + final_boys$prob_opp  + final_boys$prob_smoking +
final_boys$prob_ref)
summary(lm_unadj_all_boys_viq)

# PIQ for boys:
lm_unadj_all_boys_piq <- lm(final_boys$piq ~ final_boys$prob_high +
final_boys$prob_low + final_boys$prob_opp + final_boys$prob_smoking +
final_boys$prob_ref)
summary(lm_unadj_all_boys_piq)

# FSIQ for boys:
lm_unadj_all_boys_fsiq <- lm(final_boys$fsiq ~ final_boys$prob_high +
final_boys$prob_low + final_boys$prob_opp + final_boys$prob_smoking +
final_boys$prob_ref)
summary(lm_unadj_all_boys_fsiq)

# VIQ for girls:

lm_unadj_all_girls_viq <- lm(final_girls$viq ~ final_girls$prob_high +
final_girls$prob_low + final_girls$prob_opp + final_girls$prob_smoking +
final_girls$prob_ref)
summary(lm_unadj_all_girls_viq)

# PIQ for girls:
lm_unadj_all_girls_piq <- lm(final_girls$piq ~ final_girls$prob_high +
final_girls$prob_low + final_girls$prob_opp + final_girls$prob_smoking +
final_girls$prob_ref)
summary(lm_unadj_all_girls_piq)

# FSIQ for girls:
lm_unadj_all_girls_fsiq <- lm(final_girls$fsiq ~ final_girls$prob_high +
final_girls$prob_low + final_girls$prob_opp + final_girls$prob_smoking +
final_girls$prob_ref)
summary(lm_unadj_all_girls_fsiq)

```

```{r dummy variables for other demographics}

final_chem <- mutate(final_chem, sex_male = ifelse(child_sex == 0, 1, 0),
sex_female = ifelse(child_sex == 1, 1, 0), age_nineteen = ifelse(maternal_age
== 1, 1, 0), age_thirty = ifelse(maternal_age == 2, 1, 0), age_thirtyfive =
ifelse(maternal_age == 3, 1, 0), maternal_race = ifelse(race_aboriginal == 1
| race_asian2 == 1 | race_black2 == 1 |race_latin2 == 1 | race_other2 == 1,
1, 0 ), edu_hs = ifelse(mom_education == 1, 1, 0), edu_college =
ifelse(mom_education == 2, 1, 0), edu_undergrad = ifelse(mom_education == 3,
1, 0), edu_grad = ifelse(mom_education == 4, 1, 0), income_under_forty =
ifelse(household_income == 1, 1, 0), income_forty = ifelse(household_income
== 2, 1, 0), income_eighty = ifelse(household_income == 3, 1, 0),
income_hundred = ifelse(household_income == 4, 1, 0), parity_zero =
ifelse(parity == 1, 1, 0), parity_one = ifelse(parity == 2, 1, 0), parity_two
= ifelse(parity == 3, 1, 0), parity_three_plus = ifelse(parity == 4, 1, 0),
non_smoker = ifelse(smoker == 0, 1, 0), smoker = ifelse(smoker == 1, 1, 0),
no_prenatal_alc = ifelse(alc == 0, 1, 0), prenatal_alc = ifelse(alc == 1, 1,
0))

# Updating the stratified dataset:

final_boys <- filter(final_chem, child_sex==0)
final_girls <- filter(final_chem, child_sex==1)

```


```{r adjusted linear regression}

# all children:

lm_adj_all_viq <- lm(final_chem$viq ~ final_chem$prob_high +
final_chem$prob_low + final_chem$prob_opp + final_chem$prob_smoking +
final_chem$prob_ref + final_chem$age_thirty + final_chem$age_thirtyfive +
final_chem$race_other2 + final_chem$edu_college + final_chem$edu_undergrad +
final_chem$edu_grad + final_chem$married + final_chem$income_forty +
final_chem$income_eighty + final_chem$income_hundred + final_chem$parity_one
+ final_chem$parity_two + final_chem$parity_three_plus +
final_chem$prenatal_alc)
summary(lm_adj_all_viq)

lm_adj_all_piq <- lm(final_chem$piq ~ final_chem$prob_high +
final_chem$prob_low + final_chem$prob_opp + final_chem$prob_smoking +
final_chem$prob_ref + final_chem$age_thirty + final_chem$age_thirtyfive +
final_chem$race_other2 + final_chem$edu_college + final_chem$edu_undergrad +
final_chem$edu_grad + final_chem$married + final_chem$income_forty +
final_chem$income_eighty + final_chem$income_hundred + final_chem$parity_one
+ final_chem$parity_two + final_chem$parity_three_plus +
final_chem$prenatal_alc)
summary(lm_adj_all_piq)

lm_adj_all_fsiq <- lm(final_chem$fsiq ~ final_chem$prob_high +
final_chem$prob_low + final_chem$prob_opp + final_chem$prob_smoking +
final_chem$prob_ref + final_chem$age_thirty + final_chem$age_thirtyfive +
final_chem$race_other2 + final_chem$edu_college + final_chem$edu_undergrad +
final_chem$edu_grad + final_chem$married + final_chem$income_forty +
final_chem$income_eighty + final_chem$income_hundred + final_chem$parity_one
```

```
+ final_chem$parity_two + final_chem$parity_three_plus +
final_chem$prenatal_alc)
summary(lm_adj_all_fsiq)
# boys:

lm_adj_boys_viq <- lm(final_boys$viq ~ final_boys$prob_high +
final_boys$prob_low + final_boys$prob_opp + final_boys$prob_smoking +
final_boys$prob_ref + final_boys$age_thirty + final_boys$age_thirtyfive +
final_boys$race_other2 + final_boys$edu_college + final_boys$edu_undergrad +
final_boys$edu_grad + final_boys$married + final_boys$income_forty +
final_boys$income_eighty + final_boys$income_hundred + final_boys$parity_one
+ final_boys$parity_two + final_boys$parity_three_plus +
final_boys$prenatal_alc)
summary(lm_adj_boys_viq)

lm_adj_boys_piq <- lm(final_boys$piq ~ final_boys$prob_high +
final_boys$prob_low + final_boys$prob_opp + final_boys$prob_smoking +
final_boys$prob_ref + final_boys$age_thirty + final_boys$age_thirtyfive +
final_boys$race_other2 + final_boys$edu_college + final_boys$edu_undergrad +
final_boys$edu_grad + final_boys$married + final_boys$income_forty +
final_boys$income_eighty + final_boys$income_hundred + final_boys$parity_one
+ final_boys$parity_two + final_boys$parity_three_plus +
final_boys$prenatal_alc)
summary(lm_adj_boys_piq)

lm_adj_boys_fsiq <- lm(final_boys$fsiq ~ final_boys$prob_high +
final_boys$prob_low + final_boys$prob_opp + final_boys$prob_smoking +
final_boys$prob_ref + final_boys$age_thirty + final_boys$age_thirtyfive +
final_boys$race_other2 + final_boys$edu_college + final_boys$edu_undergrad +
final_boys$edu_grad + final_boys$married + final_boys$income_forty +
final_boys$income_eighty + final_boys$income_hundred + final_boys$parity_one
+ final_boys$parity_two + final_boys$parity_three_plus +
final_boys$prenatal_alc)
summary(lm_adj_boys_fsiq)

#girls:

lm_adj_girls_viq <- lm(final_girls$viq ~ final_girls$prob_high +
final_girls$prob_low + final_girls$prob_opp + final_girls$prob_smoking +
final_girls$prob_ref + final_girls$age_thirty + final_girls$age_thirtyfive +
final_girls$race_other2 + final_girls$edu_college + final_girls$edu_undergrad
+ final_girls$edu_grad + final_girls$married + final_girls$income_forty +
final_girls$income_eighty + final_girls$income_hundred +
final_girls$parity_one + final_girls$parity_two +
final_girls$parity_three_plus + final_girls$prenatal_alc)
summary(lm_adj_girls_viq)

lm_adj_girls_piq <- lm(final_girls$piq ~ final_girls$prob_high +
final_girls$prob_low + final_girls$prob_opp + final_girls$prob_smoking +
final_girls$prob_ref + final_girls$age_thirty + final_girls$age_thirtyfive +
final_girls$race_other2 + final_girls$edu_college + final_girls$edu_undergrad
+ final_girls$edu_grad + final_girls$married + final_girls$income_forty +
final_girls$income_eighty + final_girls$income_hundred +
final_girls$parity_one + final_girls$parity_two +
final_girls$parity_three_plus + final_girls$prenatal_alc)
summary(lm_adj_girls_piq)
```

```
lm_adj_girls_fsiq <- lm(final_girls$fsiq ~ final_girls$prob_high +
final_girls$prob_low + final_girls$prob_opp + final_girls$prob_smoking +
final_girls$prob_ref + final_girls$age_thirty + final_girls$age_thirtyfive +
final_girls$race_other2 + final_girls$edu_college + final_girls$edu_undergrad
+ final_girls$edu_grad + final_girls$married + final_girls$income_forty +
final_girls$income_eighty + final_girls$income_hundred +
final_girls$parity_one + final_girls$parity_two +
final_girls$parity_three_plus + final_girls$prenatal_alc)
summary(lm_adj_girls_fsiq)
```

## Appendix D. K-means Clustering Code

```{r basic prep}

# loading packages
library(dplyr)
library(pheatmap)
library(readr)
library(readxl)
set.seed(314)

# preparing the data:
all_data <- read_csv("AY_MIREC_Analysis_July22_2020.csv")
all_data <- filter(all_data, sex2!="NA")
imputed_data <- read_csv("AY_MIREC_Analysis_Feb_2021.csv")
chem_data <- all_data[,c(2:10, 12:23, 25:31, 33, 38)]

# renaming the variables:
chem_data <- dplyr::rename(chem_data, Arsenic = log2.arsenic.t1, Cadmium =
log2.cadmium.t1, Lead = log2.lead.t1, Manganese = log2.manganese.t1, Mercury
= log2.mercury.t1, BBHC = log2.bbhc.t1, DDE = log2.dde.t1, Oxychlor =
log2.oxychlor.t1, Transnona = log2.transnona.t1, DEP = log2.dep.t1, DETP =
log2.detp.t1, DMDTP = log2.dmdtp.t1, DMP = log2.dmp.t1, DMTP = log2.dmtp.t1,
MBP = log2.mbp.t1, MBZP = log2.mbzp.t1, MCPP = log2.mcpp.t1, MEHHP =
log2.mehhp.t1, MEHP = log2.mehp.t1, MEOHP = log2.meohp.t1, MEP = log2.mep.t1,
Aroclor = log2.aroclor.t1, PCB118 = log2.pcb118.t1, PCB138 = log2.pcb138.t1,
PCB153 = log2.pcb153.t1, PCB170 = log2.pcb170.t1, PCB180 = log2.pcb180.t1,
PCB187 = log2.pcb187.t1, BDE47 = log2.bde47.t1, Cotinine = log2.cot.t1)

```

```{r determining the number of k-clusters}

# 5 clusters would be best, since I had 5 profiles, but first I want to check
to make sure it's an appropriate number here

# first with the elbow method:

ss_within <- rep(0, 30)
for(t in 1:30){
  ss_within[t] <- (kmeans(as.matrix(chem_data), centers = t, iter.max = 10,
nstart=30))$tot.withinss
}
plot(c(1:30), ss_within)

# It looks like the "crook of the elbow" is between 5 and 8 clusters. This is
not an exact method so I'll try a second one: minimizing the SS within in the
cluters but maximizing the SS between them:

ss_within <- ss_between <- ss_ratio <- rep(0,30)
for(u in 1:30){
  ss_within[u] <- (kmeans(as.matrix(chem_data), centers = u, iter.max = 10,
nstart=30))$tot.withinss
  ss_between[u] <- (kmeans(as.matrix(chem_data), centers = u, iter.max = 10,
nstart=30))$betweenss
  ss_ratio[u] <- ss_between[u]/ss_within[u]
```

```
  }
plot(c(1:30), ss_ratio)

# Still hard to see, but there does seem to be a bit of a hockey stick look
between clusters 5 and 6

# 5 clusters should be fine

```

```{r k-means clustering with 5 clusters}

# k-means clustering results if we set the number for clusters to 5 (to match
LPA):

clusters5 <- kmeans(chem_data, centers=5, iter.max = 10, nstart=30)
clusters5

```

```{r calculating z-scores of k-clusters}

z_arsenic <- z_cadmium <- z_lead <- z_manganese <- z_mercury <- z_bbhc <-
z_dde <- z_oxychlor <- z_transnona <- z_dep <- z_detp <- z_dmdtp <- z_dmp <-
z_dmtp <- z_mbp <- z_mbzp <- z_mcpp <- z_mehhp <- z_mehp <- z_meohp <- z_mep
<- z_aroclor <- z_pcb118 <- z_pcb138 <- z_pcb153 <- z_pcb170 <- z_pcb180 <-
z_pcb187 <- z_bde47 <- z_cotinine <- rep(0,5)

for (z in 1:5){
  z_arsenic[z] <- (clusters5$centers[z,1]-
mean(chem_data$Arsenic))/sd(chem_data$Arsenic)
  z_cadmium[z] <- (clusters5$centers[z,2]-
mean(chem_data$Cadmium))/sd(chem_data$Cadmium)
  z_lead[z] <- (clusters5$centers[z,3]-
mean(chem_data$Lead))/sd(chem_data$Lead)
  z_manganese[z] <- (clusters5$centers[z,4]-
mean(chem_data$Manganese))/sd(chem_data$Manganese)
  z_mercury[z] <- (clusters5$centers[z,5]-
mean(chem_data$Mercury))/sd(chem_data$Mercury)
  z_bbhc[z] <- (clusters5$centers[z,6]-
mean(chem_data$BBHC))/sd(chem_data$BBHC)
  z_dde[z] <- (clusters5$centers[z,7]-mean(chem_data$DDE))/sd(chem_data$DDE)
  z_oxychlor[z] <- (clusters5$centers[z,8]-
mean(chem_data$Oxychlor))/sd(chem_data$Oxychlor)
  z_transnona[z] <- (clusters5$centers[z,9]-
mean(chem_data$Transnona))/sd(chem_data$Transnona)
  z_dep[z] <- (clusters5$centers[z,10]-mean(chem_data$DEP))/sd(chem_data$DEP)
  z_detp[z] <- (clusters5$centers[z,11]-
mean(chem_data$DETP))/sd(chem_data$DETP)
  z_dmdtp[z] <- (clusters5$centers[z,12]-
mean(chem_data$DMDTP))/sd(chem_data$DMDTP)
  z_dmp[z] <- (clusters5$centers[z,13]-mean(chem_data$DMP))/sd(chem_data$DMP)
  z_dmtp[z] <- (clusters5$centers[z,14]-
mean(chem_data$DMTP))/sd(chem_data$DMTP)
  z_mbp[z] <- (clusters5$centers[z,15]-mean(chem_data$MBP))/sd(chem_data$MBP)
  z_mbzp[z] <- (clusters5$centers[z,16]-
mean(chem_data$MBZP))/sd(chem_data$MBZP)
```

```
  z_mcpp[z] <- (clusters5$centers[z,17]-
mean(chem_data$MCPP))/sd(chem_data$MCPP)
  z_mehhp[z] <- (clusters5$centers[z,18]-
mean(chem_data$MEHHP))/sd(chem_data$MEHHP)
  z_mehp[z] <- (clusters5$centers[z,19]-
mean(chem_data$MEHP))/sd(chem_data$MEHP)
  z_meohp[z] <- (clusters5$centers[z,20]-
mean(chem_data$MEOHP))/sd(chem_data$MEOHP)
  z_mep[z] <- (clusters5$centers[z,21]-mean(chem_data$MEP))/sd(chem_data$MEP)
  z_aroclor[z] <- (clusters5$centers[z,22]-
mean(chem_data$Aroclor))/sd(chem_data$Aroclor)
  z_pcb118[z] <- (clusters5$centers[z,23]-
mean(chem_data$PCB118))/sd(chem_data$PCB118)
  z_pcb138[z] <- (clusters5$centers[z,24]-
mean(chem_data$PCB138))/sd(chem_data$PCB138)
  z_pcb153[z] <- (clusters5$centers[z,25]-
mean(chem_data$PCB153))/sd(chem_data$PCB153)
  z_pcb170[z] <- (clusters5$centers[z,26]-
mean(chem_data$PCB170))/sd(chem_data$PCB170)
  z_pcb180[z] <- (clusters5$centers[z,27]-
mean(chem_data$PCB180))/sd(chem_data$PCB180)
  z_pcb187[z] <- (clusters5$centers[z,28]-
mean(chem_data$PCB187))/sd(chem_data$PCB187)
  z_bde47[z] <- (clusters5$centers[z,29]-
mean(chem_data$BDE47))/sd(chem_data$BDE47)
  z_cotinine[z] <- (clusters5$centers[z,30]-
mean(chem_data$Cotinine))/sd(chem_data$Cotinine)
}

z_scores <- data.frame(z_arsenic, z_cadmium, z_lead, z_manganese, z_mercury,
z_bbhc, z_dde, z_oxychlor, z_transnona, z_dep, z_detp, z_dmdtp, z_dmp,
z_dmtp, z_mbp, z_mbzp, z_mcpp, z_mehhp, z_mehp, z_meohp, z_mep, z_aroclor,
z_pcb118, z_pcb138, z_pcb153, z_pcb170, z_pcb180, z_pcb187, z_bde47,
z_cotinine)
```

```{r}

# Making the heat map with z-scores:

# getting the colour ready
breaksList = seq(-3, 3, by=0.1)
col <- colorRampPalette(c("navy", "blue", "white", "red",
"red4"))(length(breaksList))

#labels:
x_lab <- c("Low Level", "Ref", "High OPPs", "High Smoking", "High Level")
chemicals <- c("As", "Cd", "Pb", "Mn", "Hg", "BBHC", "DDE", "Oxy.", "Trans.",
"DEP", "DETP", "DMDTP", "DMP", "DMTP", "MBP", "MBZP", "MCPP", "MEHHP",
"MEHP", "MEOHP", "MEP", "Aroclor", "PCB118", "PCB138", "PCB153", "PCB170",
"PCB180", "PCB187", "BDE47", "Cot.")

# heat map:
pheatmap(t(z_scores), # the t transposes it (switches rows and columns) so
that it faces the right way
        Colv=NA, Rowv=NA, cluster_rows = FALSE, cluster_cols = FALSE, #
getting rid of the dendrogram and the clustering
```

```
        cellwidth = 20, # changing the size of the cells
        color=col, breaks = breaksList, # colouring the chart
        labels_col = x_lab, labels_row = chemicals, angle_col = 45) # picking
the colours

```

```{r}
# changing the order to better compare:

reorder <- z_scores[c(2,5,1,3,4),]

x_lab <- c("Ref", "High Level", "Low Level", "High OPPs", "Smoking
Chemicals")

pheatmap(t(reorder), # the t transposes it (switches rows and columns) so
that it faces the right way
        Colv=NA, Rowv=NA, cluster_rows = FALSE, cluster_cols = FALSE, #
getting rid of the the dendrogram and the clustering
        cellwidth = 20, # changing the size of the cells
        color=col, breaks = breaksList, # colouring the chart
        labels_col = x_lab, labels_row = chemicals, angle_col = 45) # picking
the colours

```

```{r}
# making a final dataset with dummy variables for each cluster

x1 <- ifelse(clusters5$cluster == 1, 1, 0)
x2 <- ifelse(clusters5$cluster == 2, 1, 0)
x3 <- ifelse(clusters5$cluster == 3, 1, 0)
x4 <- ifelse(clusters5$cluster == 4, 1, 0)
x5 <- ifelse(clusters5$cluster == 5, 1, 0)

final_chem <- data.frame("subject_id" = all_data$subject.id, "cluster" =
clusters5$cluster, "high_cluster" = x5, "ref_cluster" = x2, "low_cluster" =
x1, "opp_cluster" = x3, "smoking_cluster" = x4, "Arsenic" =
all_data$log2.arsenic.t1, "Cadmium" = all_data$log2.cadmium.t1, "Lead" =
all_data$log2.lead.t1,"Manganese" = all_data$log2.manganese.t1, "Mercury" =
all_data$log2.mercury.t1, "BBHC" = all_data$log2.bbhc.t1, "DDE" =
all_data$log2.dde.t1, "Oxychlor" = all_data$log2.oxychlor.t1, "Transnona" =
all_data$log2.transnona.t1, "DEP" = all_data$log2.dep.t1, "DETP" =
all_data$log2.detp.t1, "DMDTP" = all_data$log2.dmdtp.t1, "DMP" =
all_data$log2.dmp.t1, "DMTP" = all_data$log2.dmtp.t1, "MBP" =
all_data$log2.mbp.t1, "MBZP" = all_data$log2.mbzp.t1, "MCPP" =
all_data$log2.mcpp.t1, "MEHHP" = all_data$log2.mehhp.t1, "MEHP" =
all_data$log2.mehp.t1, "MEOHP" = all_data$log2.meohp.t1, "MEP" =
all_data$log2.mep.t1, "Aroclor" = all_data$log2.aroclor.t1, "PCB118" =
all_data$log2.pcb118.t1, "PCB138" = all_data$log2.pcb138.t1, "PCB153" =
all_data$log2.pcb153.t1, "PCB170" = all_data$log2.pcb170.t1, "PCB180" =
all_data$log2.pcb180.t1, "PCB187" = all_data$log2.pcb187.t1, "BDE47" =
all_data$log2.bde47.t1, "Cotinine" = all_data$log2.cot.t1, "birth_weight" =
all_data$birth.wt, "gest_age" = all_data$gest.age, "preterm_birth" =
all_data$preterm2, "lbw" = all_data$lbw2, "lga" = all_data$lga2, "child_sex"
= all_data$sex2, "small_for_ga" = all_data$sga2, "alc" = imputed_data$alc2,
"city" = all_data$city10, "mom_education" = imputed_data$edu4,
```

```
"household_income" = imputed_data$income4, "married" = all_data$married2,
"maternal_age" = all_data$mom.age3, "parity" = all_data$parity4,
"race_aboriginal" = all_data$race.aboriginal2, "race_asian2" =
all_data$race.asian2, "race_black2" = all_data$race.black2, "race_latin2" =
all_data$race.latin2, "race_other2" = all_data$race.other2, "race_white2" =
all_data$race.white2, "smoker" = all_data$smoker2, "wppsi_1" =
all_data$wppsi.1, "wppsi_2" = all_data$wppsi.2, "wppsi_3" = all_data$wppsi.3,
"wppsi_4" = all_data$wppsi.4, "wppsi_5" = all_data$wppsi.5, "viq" =
all_data$viq, "piq" = all_data$piq, "fsiq" = all_data$fsiq)
final_chem
```

```{r demographic dummy variables}

final_chem <- mutate(final_chem, sex_male = ifelse(child_sex == 0, 1, 0),
sex_female = ifelse(child_sex == 1, 1, 0), age_nineteen = ifelse(maternal_age
== 1, 1, 0), age_thirty = ifelse(maternal_age == 2, 1, 0), age_thirtyfive =
ifelse(maternal_age == 3, 1, 0), maternal_race = ifelse(race_aboriginal == 1
| race_asian2 == 1 | race_black2 == 1 |race_latin2 == 1 | race_other2 == 1,
1, 0 ), edu_hs = ifelse(mom_education == 1, 1, 0), edu_college =
ifelse(mom_education == 2, 1, 0), edu_undergrad = ifelse(mom_education == 3,
1, 0), edu_grad = ifelse(mom_education == 4, 1, 0), income_under_forty =
ifelse(household_income == 1, 1, 0), income_forty = ifelse(household_income
== 2, 1, 0), income_eighty = ifelse(household_income == 3, 1, 0),
income_hundred = ifelse(household_income == 4, 1, 0), parity_zero =
ifelse(parity == 1, 1, 0), parity_one = ifelse(parity == 2, 1, 0), parity_two
= ifelse(parity == 3, 1, 0), parity_three_plus = ifelse(parity == 4, 1, 0),
non_smoker = ifelse(smoker == 0, 1, 0), smoker = ifelse(smoker == 1, 1, 0),
no_prenatal_alc = ifelse(alc == 0, 1, 0), prenatal_alc = ifelse(alc == 1, 1,
0))
```

```{r unadjusted regression analysis}

# unadjusted regression analysis

# VIQ
klm_unadj_all_viq <- lm(final_chem$viq ~ final_chem$high_cluster +
final_chem$low_cluster + final_chem$opp_cluster + final_chem$smoking_cluster
+ final_chem$ref_cluster)
summary(klm_unadj_all_viq)
klm_unadj_all_viq$coefficients

# PIQ
klm_unadj_all_piq <- lm(final_chem$piq ~ final_chem$high_cluster +
final_chem$low_cluster + final_chem$opp_cluster + final_chem$smoking_cluster
+ final_chem$ref_cluster)
summary(klm_unadj_all_piq)

# FSIQ
klm_unadj_all_fsiq <- lm(final_chem$fsiq ~final_chem$high_cluster +
final_chem$low_cluster + final_chem$opp_cluster + final_chem$smoking_cluster
+ final_chem$ref_cluster)
summary(klm_unadj_all_fsiq)

```

```{r adjusted regression analysis}

# adjusted regression analysis

# VIQ
klm_adj_all_viq <- lm(final_chem$viq ~ final_chem$high_cluster +
final_chem$low_cluster + final_chem$opp_cluster + final_chem$smoking_cluster
+ final_chem$ref_cluster + final_chem$age_thirty + final_chem$age_thirtyfive
+ final_chem$race_other2 + final_chem$edu_college + final_chem$edu_undergrad
+ final_chem$edu_grad + final_chem$married + final_chem$income_forty +
final_chem$income_eighty + final_chem$income_hundred + final_chem$parity_one
+ final_chem$parity_two + final_chem$parity_three_plus +
final_chem$prenatal_alc)
summary(klm_adj_all_viq)
klm_adj_all_viq$coefficients

# PIQ
klm_adj_all_piq <- lm(final_chem$piq ~ final_chem$high_cluster +
final_chem$low_cluster + final_chem$opp_cluster + final_chem$smoking_cluster
+ final_chem$ref_cluster + final_chem$age_thirty + final_chem$age_thirtyfive
+ final_chem$race_other2 + final_chem$edu_college + final_chem$edu_undergrad
+ final_chem$edu_grad + final_chem$married + final_chem$income_forty +
final_chem$income_eighty + final_chem$income_hundred + final_chem$parity_one
+ final_chem$parity_two + final_chem$parity_three_plus +
final_chem$prenatal_alc)
summary(klm_adj_all_piq)

# FSIQ
klm_adj_all_fsiq <- lm(final_chem$fsiq ~ final_chem$high_cluster +
final_chem$low_cluster + final_chem$opp_cluster + final_chem$smoking_cluster
+ final_chem$ref_cluster + final_chem$age_thirty + final_chem$age_thirtyfive
+ final_chem$race_other2 + final_chem$edu_college + final_chem$edu_undergrad
+ final_chem$edu_grad + final_chem$married + final_chem$income_forty +
final_chem$income_eighty + final_chem$income_hundred + final_chem$parity_one
+ final_chem$parity_two + final_chem$parity_three_plus +
final_chem$prenatal_alc)
summary(klm_adj_all_fsiq)

```

```{r unadjusted regression stratified by sex}

# stratifying by sex
final_boys <- filter(final_chem, child_sex == 0)
final_girls <- filter(final_chem, child_sex == 1)

# unadjusted regression analysis - boys
# VIQ
klm_unadj_boys_viq <- lm(final_boys$viq ~ final_boys$high_cluster +
final_boys$low_cluster + final_boys$opp_cluster + final_boys$smoking_cluster
+ final_boys$ref_cluster)
summary(klm_unadj_boys_viq)
klm_unadj_boys_viq$coefficients

# PIQ
```

```
klm_unadj_boys_piq <- lm(final_boys$piq ~ final_boys$high_cluster +
final_boys$low_cluster + final_boys$opp_cluster + final_boys$smoking_cluster
+ final_boys$ref_cluster)
summary(klm_unadj_boys_piq)

# FSIQ
klm_unadj_boys_fsiq <- lm(final_boys$fsiq ~ final_boys$high_cluster +
final_boys$low_cluster + final_boys$opp_cluster + final_boys$smoking_cluster
+ final_boys$ref_cluster)
summary(klm_unadj_boys_fsiq)

# unadjusted regression analysis - girls

# VIQ
klm_unadj_girls_viq <- lm(final_girls$viq ~ final_girls$high_cluster +
final_girls$low_cluster + final_girls$opp_cluster +
final_girls$smoking_cluster + final_girls$ref_cluster)
summary(klm_unadj_girls_viq)
klm_unadj_girls_viq$coefficients

# PIQ
klm_unadj_girls_piq <- lm(final_girls$piq ~ final_girls$high_cluster +
final_girls$low_cluster + final_girls$opp_cluster +
final_girls$smoking_cluster + final_girls$ref_cluster)
summary(klm_unadj_girls_piq)

# FSIQ
klm_unadj_girls_fsiq <- lm(final_girls$fsiq ~ final_girls$high_cluster +
final_girls$low_cluster + final_girls$opp_cluster +
final_girls$smoking_cluster + final_girls$ref_cluster)
summary(klm_unadj_girls_fsiq)

```

```{r}

# adjusted regression analysis - boys

# VIQ
klm_adj_boys_viq <- lm(final_boys$viq ~ final_boys$high_cluster +
final_boys$low_cluster + final_boys$opp_cluster + final_boys$smoking_cluster
+ final_boys$ref_cluster + final_boys$age_thirty + final_boys$age_thirtyfive
+ final_boys$race_other2 + final_boys$edu_college + final_boys$edu_undergrad
+ final_boys$edu_grad + final_boys$married + final_boys$income_forty +
final_boys$income_eighty + final_boys$income_hundred + final_boys$parity_one
+ final_boys$parity_two + final_boys$parity_three_plus + +
final_boys$prenatal_alc)
summary(klm_adj_boys_viq)
klm_unadj_boys_viq$coefficients

# PIQ
klm_adj_boys_piq <- lm(final_boys$piq ~ final_boys$high_cluster +
final_boys$low_cluster + final_boys$opp_cluster + final_boys$smoking_cluster
+ final_boys$ref_cluster + final_boys$age_thirty + final_boys$age_thirtyfive
+ final_boys$race_other2 + final_boys$edu_college + final_boys$edu_undergrad
+ final_boys$edu_grad + final_boys$married + final_boys$income_forty +
final_boys$income_eighty + final_boys$income_hundred + final_boys$parity_one
```

```
+ final_boys$parity_two + final_boys$parity_three_plus +
final_boys$prenatal_alc)
summary(klm_adj_boys_piq)

# FSIQ
klm_adj_boys_fsiq <- lm(final_boys$fsiq ~ final_boys$high_cluster +
final_boys$low_cluster + final_boys$opp_cluster + final_boys$smoking_cluster
+ final_boys$ref_cluster + final_boys$age_thirty + final_boys$age_thirtyfive
+ final_boys$race_other2 + final_boys$edu_college + final_boys$edu_undergrad
+ final_boys$edu_grad + final_boys$married + final_boys$income_forty +
final_boys$income_eighty + final_boys$income_hundred + final_boys$parity_one
+ final_boys$parity_two + final_boys$parity_three_plus +
final_boys$prenatal_alc)
summary(klm_adj_boys_fsiq)

# adjusted regression analysis - girls

# VIQ
klm_adj_girls_viq <- lm(final_girls$viq ~ final_girls$high_cluster +
final_girls$low_cluster + final_girls$opp_cluster +
final_girls$smoking_cluster + final_girls$ref_cluster +
final_girls$age_thirty + final_girls$age_thirtyfive + final_girls$race_other2
+ final_girls$edu_college + final_girls$edu_undergrad + final_girls$edu_grad
+ final_girls$married + final_girls$income_forty + final_girls$income_eighty
+ final_girls$income_hundred + final_girls$parity_one +
final_girls$parity_two + final_girls$parity_three_plus +
final_girls$prenatal_alc)
summary(klm_adj_girls_viq)
klm_unadj_girls_viq$coefficients

# PIQ
klm_adj_girls_piq <- lm(final_girls$piq ~ final_girls$high_cluster +
final_girls$low_cluster + final_girls$opp_cluster +
final_girls$smoking_cluster + final_girls$ref_cluster +
final_girls$age_thirty + final_girls$age_thirtyfive + final_girls$race_other2
+ final_girls$edu_college + final_girls$edu_undergrad + final_girls$edu_grad
+ final_girls$married + final_girls$income_forty + final_girls$income_eighty
+ final_girls$income_hundred + final_girls$parity_one +
final_girls$parity_two + final_girls$parity_three_plus +
final_girls$prenatal_alc)
summary(klm_adj_girls_piq)

# FSIQ
klm_adj_girls_fsiq <- lm(final_girls$fsiq ~ final_girls$high_cluster +
final_girls$low_cluster + final_girls$opp_cluster +
final_girls$smoking_cluster + final_girls$ref_cluster +
final_girls$age_thirty + final_girls$age_thirtyfive + final_girls$race_other2
+ final_girls$edu_college + final_girls$edu_undergrad + final_girls$edu_grad
+ final_girls$married + final_girls$income_forty + final_girls$income_eighty
+ final_girls$income_hundred + final_girls$parity_one +
final_girls$parity_two + final_girls$parity_three_plus +
final_girls$prenatal_alc)
summary(klm_adj_girls_fsiq)
```